الجمهورية العربية السورية
جامعة البعث
كلية الطب البيطري

الإحصاء الحيوي

القسم النظري

الدكتور
ياسر العمر

لطلاب السنة الأولى

مديرية الكتب و المطبوعات

٢٠٠٦-٢٠٠٧

المقدمة

مع أن علم الإحصاء ضروري لكثير من العلوم فإنه وسيلة أساسية لعلوم صحة الحيوان والعلوم البيطرية . ولا سيما تقاريرها ودراساتها المتعلقة بإنتاج الحيوان وجائحات الأمراض ودراسة عوامل خطورة المرض ولتقييم استراتيجيات إنتاجية الحيوان وكذا استراتيجيات التحكم واستئصال الأمراض المعدية .

ولذلك يحتاج الباحثون في علم الحيوان والطب البيطري إلى مثل هذا النمط من العلوم لتنظيم وتوثيق واعتماد نتائج تتعلق بالمسوحات الوبائية.

ومع كثرة المؤلفات في علم الإحصاء الإجتماعي والطبي يندر وجود كتب عربية أو أجنبية تختص بدراسة الإحصاء في صحة الحيوان والعلوم البيطرية . ولذلك لا يرى طالب الطب البيطري وعلم الحيوان ارتباطاً بين المقرر المدروس والمشكلات التي يتناولها الطبيب البيطري والمهتمون بدراسة علم الحيوان وصحته، مما يؤثر على الطالب ويبدي له انفصال المقرر المدروس عن المشاكل التي يتناولها الطبيب البيطري والمهتمون بدراسة علم الحيوان وصحته .

ولذلك وضعنا هذا الكتاب دليلاً للإحصاء المتعلق بصحة الحيوان وأمراضه وإننا نقدم كتابنا هذا " الإحصاء الحيوي " إلى إخوتنا الطلبة في الوطن العربي وإلى

زملائنا الأساتذة المحترمين وإلى الأطباء البيطريين ليكون مرجعاً في العلوم الطبية البيطرية التي تتناول التقييم العام للثروة الحيوانية وتفعيل الاستراتيجيات الخاصة بطرق الوقاية والتحكم بالمرض .

وأخيراً نتمنى أن يكون هذا الجهد المتواضع خطوة متقدمة في دراسة الإحصاء الذ يتناول تقييم إنتاجية الحيوان وصحته وخطوة أخرى في البحث عن الحقائق السليمة لدراسة صحة الحيوان وإنتاجيته لنضع المعالم الأولى في سبيل تطوير الخدمات البيطرية بمعرفتنا بواقعنا ومن ثم وضع النظم الجديدة في سبيل تطوير مهنة الطب البيطري. و الله من وراء القصد .

المؤلفان

المحتويات

الفصل الأول
CHAPTER ONE
المدخل : مفهوم علم الإحصاء
INTRODUCTION : WHAT IS STATISTICS
مفهوم علم الإحصاء What is statistics

يعرف الإحصاء بأنه علم من العلوم الرياضية لأن أسسه وقوانينه مستمدة من القوانين الرياضية البحتة إلا أنه يتميز بكونه علماً تطبيقياً يتناول بالدراسة الناحية الكمية لمختلف العلوم التطبيقية ومنها العلوم الطبية البيطرية .

ومن هنا يمكننا أن نعرف علم الإحصاء بأنه علم يشمل مجموعة الطرائق والأدوات العلمية التي تتبع في الحصول على المعلومات الكمية والنوعية وتبويبها وعرضها وتحليلها وتلخيصها ولما كانت المعلومات تتحول بالمعنى الرياضي إلى أرقام كمية تتمثل بالبيانات عُرِّف علم الإحصاء بأنه دراسة جمع البيانات وتصنيفها وتبويبها وتلخيصها وعرضها وتحليلها لاستنباط الاستنتاجات العلمية الشاملة من هذه البيانات الجزئية .

إن الأعداد الهائلة للنصوص الأولية في موضوع الإحصاء المدرسي تشير إلى أهمية مواضيع الإحصاء المدرسي للعلوم الحيوية طلاب ومَنْ يقومون في مجال البحوث العلمية.

وفي الحقيقة فإن هناك كثير من النصوص تتطلب منا أن نكتشف الطريقة الذكية لتمثيل البيانات المطلوبة .

إن المشكلة التي تواجهنا في الإحصاء الحيوي هي كما يأتي :

عندما نقوم بترتيب وتنسيق المشاهدات المعبرة عنها رقمياً في علوم الأحياء فإنك ستجد القيم المحصول عليها على شكل بيانات مشتتة التوزيع . (وهذا يحتم علينا معرفة) ما إذا كانت ناجمة الاختلافات في هذا التوزيع ناجمة عن أسباب متعلقة بموضوع الدراسة أو أنها تكون نتيجة لفرق طبيعي متعلق بخلفيات غاية مواضيع الدراسة وعلاقة ذلك بالعوامل الأخرى ممكناً . كما أننا نحتاج إلى أن نقيم الوسطَ الحسابي الرقمي فعلياً وإلى أن نمثل هذه الأرقام في طريقة تجعل ارتباط تفسيرها بالعوامل الأخرى ممكناً .

وإذا ما أخذنا كلمة الإحصاء بمعناها الحرفي فإننا نجد أن لها مرادفات عدة فقد ورد ذكرها في القرآن الكريم بمعنى الحصر والعد (وأحصى كل شيء عدداً). وكثيراً ما نصادف العديد من البيانات الإحصائية التي تصدر عن وزارة الزراعة والتي تعبر عن أعداد المواليد والنفوق والإصابة بالأمراض المرتبطة بالثروة الحيوانية وغيرها ... ومن هنا كان لفظ الإحصاء دالاً على الطرائق الإحصائية المختلفة التي تستخدم في تحليل وتفسير البيانات الكمية أو العددية . ويعتمد

الإحصاء على كثير من النظريات الإحصائية والرياضية والمعادلات المختلفة التي يضعها المتخصصون في العلوم الإحصائية الرياضية.

إن علم الإحصاء يركز على دراسة النقاط الآتية :

١- تصميم الدراسة تصميماً تتضح به المعلومات والبيانات المجموعة اتضاحاً فعالاً.

٢- جمع البيانات .

٣- تحليل البيانات .

٤- تمثيل المعلومات وتلخيصها تلخيصاً مناسباً وذلك باستخدام الأشكال أما الجداول المناسبة .

٥- تفسير التحليل للنتائج بطريقة تربط بين موجودات الدراسة بدقة .

إن عرض النقاط السابقة ودراستها في علم الأحياء يُعبر عنه بالقياس الحيوي (biometry) لكننا نؤثر استخدام تعبير أشيع وهو ما يدعى بعبارة الإحصاء Statistics لنغطي كافة نقاط الدراسة .

وبهذا يتبين أن الإحصاء قد أصبح إحدى الوسائل الأساسية في علوم الأحياء الحديثة .

دور الإحصاء في تسلسل خطوات البحث العلمي

The Role of Statistics in the Steps of Scientific Research

من التساؤلات الأساسية الشائعة لدى كلٍّ من طلبة الطب البيطري وعلوم الحيوان هو : لِمَ أحتاج إلى أن أدرس الإحصاء ؟

إن الأساس الرياضي البحث للإجابة على ذلك وتعطي تفسيرات غير مؤكدة غالباً . وفي الحياة المهنية على العموم كثير من الأمثلة التي توضح أهمية الإحصاء في العلوم الحيوية وغيرها من العلوم:

١- المنشورات العلمية المطبوعة The Published – Scientific Literature إن مثل هذه المنشورات مليئة بالطرائق الإحصائية المستخدمة والتي تظهر بوضوح استخدام تعبير (الوسط الحسابي $\mp$ الوسط الحسابي للخطأ للمعياري) أي [mean $\mp$ Standard error of mean (SEM)] للحصول على المعنوية الإحصائية بواسطة قيم الاحتمالية (P) أو باستخدام اختبارات t ستدنت (student t) أو تحليل مربع كاي (χ^2) أو استخدام طرق تحليل التباين . ونجد أن المعلومات الممثلة فيها مختصرة ومبوبة . وبدون إلمام الدراسة بالتحليل

الإحصائي لا نقبل استنتاجات عمل هذا الباحث مطلقاً ، لأننا نكون عندئذٍ غير قادرين على فحص قوة الدعم والربط للبيانات مع الواقع . باختراع الحاسب الذي يقوم بتنسيق وتنظيم البيانات والكثير من حسابات حول المشاهدات التي حصلنا عليها ويقوم أيضاً بتلخيصها وتبويبها وتبيان أهميتها وحتى نقوم بكل ما ذكرناه في جهاز الحاسب ليُحقق الفائدة المرجوة من هذه المعلومات فإننا نحتاج حتماً إلى نظرة شاملة بدراستها إحصائياً .

٢- وفي علوم صحة الحيوان تتزايد أعداد الخدمات التشخيصية المستقلة والتي تحلل العينات من أجل تحقيق الفائدة في رصد صحة الحيوان Health monitoring والمحافظة عليها جيداً . هذا العمل يجرى بواسطة الخدمات المخبرية والتي توجب التركيز دائماً على نوعية التحكم والدقة في القياسات المتبعة للأغراض التشخيصية وكونَنا قادرين على تقديم دلالات واضحة لتفسير النتائج المحصول عليها في المخابر .

٣- الصناعات الدوائية والكيميائية الزراعية Pharmaceutical and Agrichemical Industries : وهي تتطلب إثبات كل من أمان منتجاتها وفعاليتها بطريقة لا تقبل الجدل . مثل هذه البيانات لآلية هذه الصناعات تتطلب طرائق إحصائية لتقييم وتوضيح أساس كل من هاتين النقطتين المدروستين . ففي حالة الصناعات الدوائية نجد أن تطويرها يحتاج إلى فهم أهمية تصميم الدراسة

للتأكد بما يكفي من أعداد الحيوانات المستخدمة في المجموعات المعالجة كي تجرى تجارب مثل هذه الصناعات بما هو جدير بالثقة . إن لجنة ترخيص المنتجات البيطرية ويطلب منها فهم علم الإحصاء لتقييم مدى فعالية المواد العلاجية المصنعة وتأثيرها على الحيوان المريض .

٤- ويزداد التركيز حالياً على أنظمة الأمانة ونوعية الطعام الخاص بالاستهلاك البشري . لأن المنتجات من أصل حيواني ذات أهمية في هذا المجال وتحديد مدى أمانها يقع على عاتق الطبيب البيطري فعلى سبيل المثال : المنتج الدوائي المتعلق بسحب زمني معين قبل الذبح يعتمد على التداخل الدوائي وديناميكية الدواء الخاصة بالمنتج وهذا يتعلق على سبيل المثال بمنتج الحليب بعد عملية معالجة الحيوان وكذا بالنسبة للثمالات النسيجية الناجمة عن المبيدات العشبية والمبيدات الحشرية وإمكانية تلوث الذبائح بالصادات الحيوية المقاومة للجراثيم . وعلى أية حال فإن التوصيات والمقترحات وكذا القوانين الناظمة المناسبة تُقيّم بالدراسات التجريبية وكذا التقييم الإحصائي . وهذا يتطلب من الخبراء أن يكون لديهم خبرة بالطرائق الإحصائية المناسبة كي يؤدوا عملهم في سن القوانين المناسبة والصحيحة اعتماداً على هذه الطرائق .

ويمكن أن نوجز فوائد علم الإحصاء الأساسية في ثلاثة مجالات متعلقة بالطب البيطري وعلم الحيوان :

١- ما لم تكن هناك بيانات إحصائية لقطاع الإنتاج الحيواني والثروة الحيوانية وبصورة مستمرة وعلى نحو سليم فإن عملية التخطيط الاقتصادي تكون فاشلة لتطوير الثروة الحيوانية ومنتجاتها . ناهيك عن الخسائر الجسيمة التي قد تلحق بالاقتصاد الوطني .

٢- تعتبر دراسة الطرائق الإحصائية واستعمالها في التجارب الحيوية والمشاهدات التجريبية أداة مهمة في جمع وتبويب وتحليل البيانات وتفسير النتائج الحيوية .

٣- إن دراسة العلوم الإحصائية ونظرياتها المختلفة تسهل وتساعد على فهم الباحث في الحصول على الحقائق وفق أفضل النتائج وأدها ومطابقتها فيزيولوجيا لوقائع الأحداث المدروسة .

طريقة البحث الإحصائي ومراحله

Method of Statistical Approach and its Steps

تشمل طريقة البحث الإحصائي مراحل مختلفة تختلف تبعاً لإطار الدراسة والهدف المطلوب إنجازه إلا أنها عموماً تشمل المراحل الآتية :

١- جمع البيانات الإحصائية بالطرائق المناسبة والتي سنتطرق إليها في فصول لاحقة .

٢- تصنيف وتبويب البيانات .

٣- عرض البيانات بطريقة أو أكثر من طرائق العرض المناسبة .

٤- دراسة البيانات المجموعة وتحليلها بالطرائق الإحصائية الوصفية أو التحليلية أو بكلتيهما معاً .

٥- استخلاص النتائج المحصول عليها من تحليل بيانات الدراسة .

إلا أنه يجب أن ننوه هنا بأن كل مرحلة من مراحل هذه الدراسة مهمة ويجب إتمامها بدون أخطاء مع اعتبار هذه المراحل مرتبطاً بعضها ببعض وأن كل مرحلة تبدأ من نهاية المرحلة السابقة وتعتمد عليها وعلى دقة نتائجها.

أنواع المتغيرات Types of Variables

المتغير هو عبارة عن قيمة خاصة تتعلق بها صفة أو عضو قابل للتغير في النوع أو الكم من حيوان إلى آخر في نفس المجتمع الحيواني المدروس ، وقد تكون الصفة المتغيرة وصفية أي لا يمكن قياسها مباشرة كاللون أو الحالة الصحية لحيوانات الدراسة وقد تكون الصفة المتغيرة صفة كمية كالصفات والخواص التي يمكن قياسها مباشرة مثل الأطوال والأوزان ... الخ .

وهنا يجب أن نتحدث عن القيمة الإحصائية Variate والتي تعبر عن قيمة إحدى صفات حيوانات الدراسة المتغيرة والتي يمكن أن تأخذ قيماً تختلف في حدودها من القيم الفردية في أحد الحيوانات إلى قيم فردية أخرى أو من قيم شاملة لمجموعة من الحيوانات إلى مجموعة أخرى . فعلى سبيل المثال صفة الطول والوزن وحجم

الحظيرة وتعداد الدم وفعالية الأنزيمات ولون غطاء الجلد والنسب المئوية للنعاج الحاملة في القطيع .. إلخ ويرمز للقيمة الإحصائية بالرمز (X) فإذا كان لدينا مجموعة أطوال لمجموعة من الخيول فيأخذ كل حيوان طولاً فردياً أول مثلاً ويرمز له بـ X_1 والثاني بـ X_2 والثالث بـ X_3 .. وهكذا حتى الطول الأخير Xn ويكون مجموع القيم الإحصائية هو :

$$\sum_{1}^{n} X = X_1 + X_2 + X_3 + ... + \times n$$

حيث يعبر Σ عن المجموع المشتق من الحرف الإغريقي Sigma كما يعبر عن المجموع بالرمز (S) والمشتق من الكلمة الإنكليزية Sum أو Summation إلا أن ذلك غير مرغوب فيه منعاً للخلط بين هذا الرمز ونفس الرمز (S) الذي يعبر عنه أحياناً بالانحراف القياسي كأحد مقاييس التشتت والتي سترد في فصل لاحق . ومن الملاحظ أنه وضع تحت الرمز Σ الرقم وفوقه الحرف n وذلك للتعبير عن أن هذا المجموع هو عبارة عن مجموع الأفراد في حيوانات الدراسة والتي تأخذ القيم من الحيوان الأول (١) حتى الرقم الأخير (n) وعلى نحو واضح فإن بعض هذه المميزات المذكورة الخواص يكون أكثر قابلية لتقديرها كمياً من الأخريات . (ومن أجل بعض المتغيرات يمكننا أن نحول الرقم العددي إلى فئة رقمية وبذلك يمكننا إنشاء نوع من

المستوى العددي إلا أن بعض القيم الأخرى يمكن أن تكون على شكل مستوى رقمي حقيقي تمتد عندها قيمة المتغير) .

كما أن البيانات الرقمية تأخذ أشكالاً مختلفة بما أن الفهم الطبيعي للبيانات وتصنيف المتغيرات هو الخطة الهامة الأولى في اختيار الطريقة الإحصائية المناسبة .

يظهر الملحق C مخططات الخوارزمية (Flow chart) والتي توضح آلية هذه الفكرة وتقوم على الجمع مابين الاختيار المناسب للطريقة الإحصائية وتحليل بيانات معينة .

ويمكن أن يُميَّز الأنماط الرئيسية للمتغير بطريقة نظامية وذلك بتحديد إمكانية أخذ المتغير لقيمة من قيمتين مميزتين ، أو قيمة واحدة من عدة قيم مميزة ، أو أي قيمة ضمن المجال المعطى لهذا المتغير . وعملياُ ً يمكن أن يأخذ أحد الأنواع الآتية :

أ- متغير نوعي/أو فئوي Qualitative/ Categorical variable :

وهو عبارة عن عدد ينتمي إلى فئة واحدة من مجموع فئتين اثنتين أو أكثر من الفئات المميزة لهذا المتغير فالمتغير من نوع ثنائي/بيزي (a binary v.) أو ما يدعى dichotomous وهو نوع خاص من المتغيرات من أنماط المتغيرات الفئوية الاسمية والتي تعرف فقط في حال وجود فئتين ، كأن نقول مثلاً : بقرة حامل أو غير

حامل أو أن نقول هذا الحيوان ذكر أو أنثى . وتتراوح قيم هذا المستوى من المتغيرات ما بين القيمة الصفر إلى القيمة واحد (٠-١) .

ويمكن أن تلخص المستويات الخاصة بالمتغير الفئوي بما يأتي :

١- مستوى / قياسي / اسمي Nominal Scale : وهي عبارة عن فئات مميزة تعرف المتغير بأنها ذات قيم غير مرتبة Unordered وكل منها يمكن أن يشار إليه باسم كأن نقول لون غطاء الجلد (بني – رمادي) .

٢- مستوى ترتيبي Ordinal Scale : وهو يضم فئات متغير تتميز ببعض الترتيبات الداخلية لكن هذا الترتيب لا يتضمن أي معنى بالنسبة للمتغير وتعرف الفواصل بين الفئات المختلفة حسب المتغير : على سبيل المثال قياس حالة الجسم في مصطلح الأمراض الباطنية Body Condition Score وكذا كثافة الومضان الفلوراسيني للخلايا عند استخدام المجهر الومضاني الفلوراسيني ودرجة حركية الحيوانات المنوية في العينة . هذه المستويات تعطي غالباً قيماً عددية تتراوح بين ١ إلى n .

ب- المتغير الكمي Quantitative variable

وهو يتضمن قيماً عددية معرفة جيداً بمستواها وهنا نميز نوعين من المستويات أو المتغيرات الكمية :

١- المستوى غير المستمر (متغيرات منقطعة) :

A discrete (Discontinuous) Scale

نقول عن المتغير الإحصائي X إنه متغير منقطع إذا كان [a , b] ∈X وكانت هناك قيمة واحدة على الأقل تقع ضمن المجال ويستحيل للمتغير X أن يأخذها. ومثال هذا المستوى ، البيانات التي يمكن أن تأخذ قيماً خاصة ذات تعداد نموذجي مثل حجم الحظيرة وكذا أعداد الأفراخ النافقة في حضانة البيض والموسم الإدراري (أو عدد مرات الحمل) . لنأخذ مثلاً عدد الحيوانات المعالجة يومياً في المشفى البيطري في كليتنا ولنفرض أن هذا العدد لا يقل عن ٢ ولا يزيد عن ١٠ ، أي [2 , 10] ∈X وهنا نقول إننا أمام متغير إحصائي منقطع ،أما بالنسبة للعدد ٢ ، ٣ مثلاً فلا يمكن أن يأخذ المتغير X هذه القيمة .

٢- المستوى المستمر أو المتغيرات الإحصائية المستمرة

Variable a Continuous Scale or a Continuous

وتكون هنا جميع القيم مقبولة نظرياً في هذا المستوى وربما تكون القيم ذات حد أعلى وحد أدنى . وبالتالي فإن المتغير X هنا يأخذ أية قيمة من مجموعة الأعداد الحقيقية التي تمثل بمستقيم الأعداد الحقيقية R (real numbers) .

فإذا كان المتغير X متغيراً إحصائياً مستمراً وكان [a,b]∈ X فإن X يمكن أن يأخذ أية قيمة من a إلى b بما فيها الأعداد الصحيحة والنسبية والعشرية غير

النسبية مثل π والقيمة e . ومثال ذلك صفات الطول والوزن والسرعة وتركيز مكون كيميائي في الدم أو البول . وتكون قيم مستوى هذا المتغير ذات مدى غير محدود أي يمكن أن يصل بقيمته إلى اللانهاية . والمثال الرقمي لهذا المتغير أن نأخذ إنتاج الحليب لإحدى مزارع الأبقار الحلوب الحكومية فإذا كان إنتاج الحليب اليومي الأعظمي لإحدى الأبقار في هذه المزرعة ٢٥ كغ وأدنى إنتاج حليب يومي لإحدى الأبقار في هذه المزرعة هو ١٠ كغ فإن X يضم المجال $X \in [10.25]$ فإذا ما اخترنا أي حيوان في هذه المزرعة فإن إنتاجه يمكن أن يكون واقعاً ضمن المجال المذكور أعلاه مهما كانت دقته كإنتاج الحليب الوزني اليومي ١٢.٢ كغ أو ١٥.٥ أو ١٧.٧ كغ على سبيل المثال .

الاختلافات في القياسات Variations in Measurements

من المعروف أنه عندما تتكرر المشاهدة ونريد تقدير ظاهرة حيوية خاصة لا تتطابق القياسات إلا نادراً . جزء من هذا الاختلاف يكون بسبب اختلاف طبيعي في المادة الحيوية المقيسة . فعلى سبيل المثال : نقول إن جميع الأبقار تأكل نفس الكمية من الأعشاب كل يوم وإن كان هناك اختلاف في كل من وزن الجسم ومكونات العلف من المياه مع أخذها بعين الاعتبار . وسنستخدم عبارة الفرق أو الاختلاف الحيوي Biological Variation للتعبير عن هذه الظاهرة مع أن بعض الأخصائيين

يستخدمون عبارة الخطأ الحيوي Biological Error (والحق أن عبارة الخطأ الحيوي تقودنا خطأً إلى تعبير مشوه على حين أن هذا الاختلاف لا يرتبط بنتيجة خطأ متعلق بآلة أو بإنسان وإنما خطأ طبيعي ويرتبط بالاختلاف الحيوي للحيوان المدروس) .

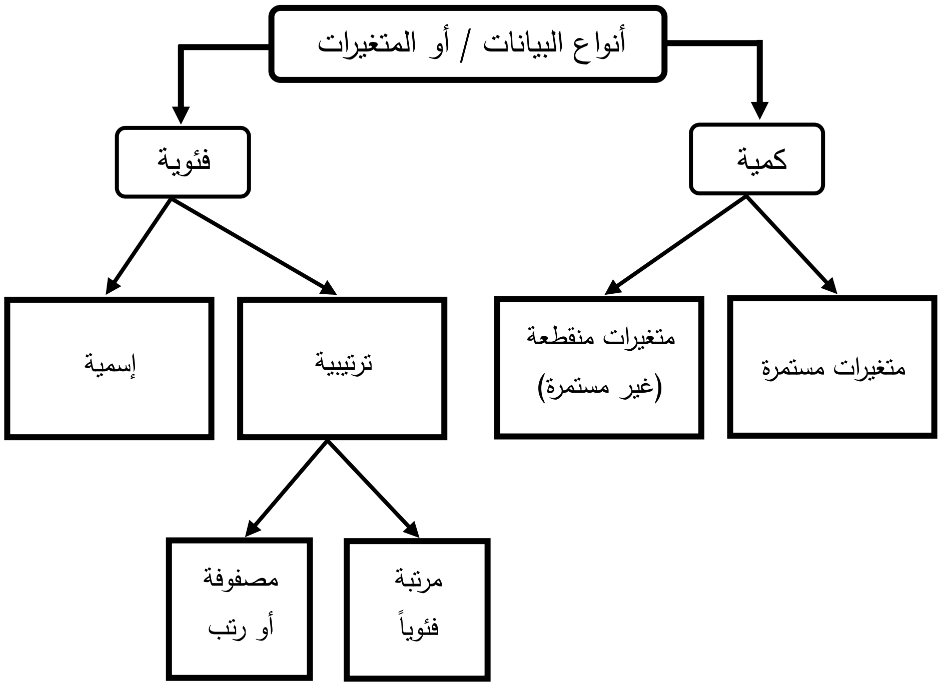
ولكن إذا قمنا باختيار الحيوانات فردياً حسب مواصفات خاصة قبل البدء بجمع البيانات فإننا سنكون قادرين على خفض مدى الاختلاف الحيوي إلا أنه لا يمكننا تحديده . ويعتمد هذا الاختيار غالباً على خواص ذلك الحيوان أو مواصفات (على سبيل المثال الأنواع الحيوانية أو النوع الحيواني والذرية الجرثومية وعمر الحيوان وجنس الحيوان ودرجة النضج ووزن الجسم وقطعان متخصصة بإنتاج الحليب أو اللحم وغيرها ...). ولذلك يعتمد هذا الاختيار على عوامل خاصة ضمن شروط التقصي المستخدم في الدراسة. يوضح الشكل رقم (١) أنماط المتغيرات .

الاختلافات نوعان :

١- الاختلاف الحيوي biological : وهو عائد إلى اختلاف طبيعة المادة الحيوية المقيسة بذاك .

٢- الاختلاف الفني technical : وهو عائد إلى فروق حيوية في القياسات

الشكل رقم (١) أنماط البيانات / المتغيرات في علم الإحصاء



مثال لأنماط المتغيرات الإحصائية :

تحتوي مزرعة على 100 بقرة منها 29 عجلاً و 71 بقرة بالغة كان من بين هاتين المجموعتين عشر عجلات (بِكار) و (55) بقرة بالغة في حالة حمل بالتشخيص الولادي . ومن نفس المجموعتين المذكورتين في البداية كانت لدينا المواصفات الآتية :

58 بقرة في حالة صحية جيد ومنها 48 بقرة حاملاً .

8 بقرات في حالة جسمية متوسطة ومنها 5 بقرات حوامل .

5 أبقار في حالة صحية سيئة ومنها بقرتان حاملان .

وبما أن عدد العجلات (بِكار) الحوامل كانت تسعة وعشرين (٢٩) فقد كان لها بعض المواصفات فـ ١٩ في حالة صحية جيدة منها ٨ عجلات حوامل وتسع عجلات في حالة صحية متوسطة منها عجلتان حاملان وعجلة واحدة في حالة صحية سيئة . كيف يمكن أن نمثل هذه المعلومات في شكل بيانات رقمية ؟ .

الحل :

حالة الجسم الصحي في كل من العجلات من الأبقار في المزرعة المذكورة أعلاه .

	العجلات		الأبقار	
حالة الجسم	عدد المشاهدات	%	عدد المشاهدات	%
جيدة	١٩	٦٦.٥	٥٨	٨١.٧
متوسطة	٩	٣١.٠	٨	١١.٣
سيئة	١	٣.٤	٥	٧.٠
الإجمالي/المجموع	٢٩	١٠٠.٠	٧١	١٠٠.٠

ولذلك فإن الدراسات المطبقة على سبيل المثال على قطعان أبقار اللحم يفترض أن لا تستخدم في قطعان أبقار الحليب والمخصصة لإنتاج الحليب .

وإضافة إلى الاختلاف الحيوي هناك غالباً فروق حيوية في القياسات المتكررة لنفس الموضوع : هذه الاختلافات أو الأخطاء الفنية Technical variation تكون نتيجة أسباب متعلقة بالآلة أو ناجمة عن أخطاء متعلقة بالإنسان المستخدم للآلة مثل

هذه الاختلافات يجب أن نأخذها بعين الاعتبار عند إجراء أي دراسة كي لا تكون القيمة التي حصلنا عليها هي قيمة حقيقية مطلقة .

الفصل الثاني
CHAPTER TWO
تعاريف ومفاهيم إحصائية
STATISTICAL DEFINITIONS AND CONCEPTS

Biological Variations الاختلاف الحيوي

العبارات المرتبطة بنوعية القياس

Terms Relating to Measurement Quality

هناك عبارتان هامتان أساساً في فهم المبادئ الأساسية للقياس الحيوي وهما مصطلح الدقة Precision والتدقيق Accuracy وكذلك يعتبر فهمهما أساسياً عند بدء الدراسة للأخذ بعين الاعتبار طبيعة قياس البيانات .

فالدقة : مصطلح يشير إلى كيفية تكرار المشاهدات متوافقة جيداً مع البيانات الأخرى عند المقارنة .

التدقيق : وهو مصطلح يشير إلى كيفية توافق القيمة المشاهدة مع القيمة الحقيقية توافقاً جيداً .

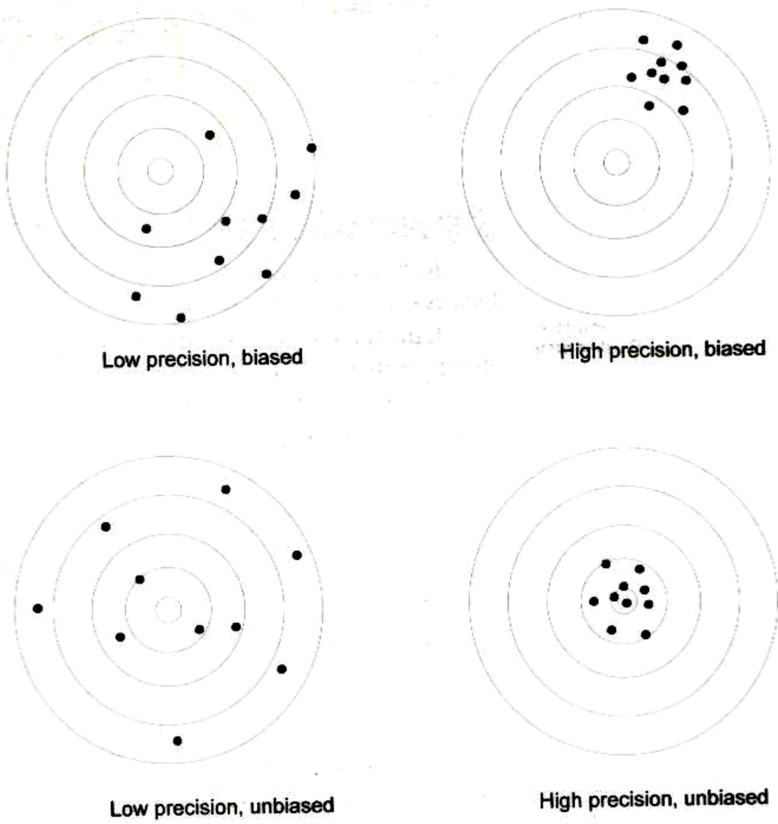
ولنحاول فهم هاتين العبارتين بالشكل رقم (٢) والذي يشير إلى نقاط التصويب في حدقة العين نسبياً والتي تختلف بقربها من القيمة الحقيقية ، ففي هذا الشكل نجد أن

ثمة سواءً في التدقيق والدقة الخاصة في التصويب و كذا يوجد تدقيق سيئ ودقة جيدة . إضافة إلى أنه ثمة دقة وتدقيق جيد من حيث التصويب .

التكرارية Repeatability :

هذا المصطلح يركز على تقدير التشابه للتكرارات والتي غالباً ما تكون مزدوجة كما أن قياسات تقنية معينة بواسطة جهاز معين أو شخص ما أو القائم على القياس تكون في شروط معينة ، على سبيل المثال القياسات التي يجريها الشخص القائم على القياس في نفس المخبر تتم بتقييم الأخطاء الفنية التي يمكن أن تحدث فيه .

الشكل رقم 1 : تقييم الدقة و التدقيق



طريقة التوافق Method Of Agreement :

يركز هذا المصطلح على تحديد كيف يمكن لطريقة قياس واحدة أو اثنتين أو أكثر أن تتوافق في نفس تقديرها مع طريقة أخرى . على سبيل المثال القياسات التي يجريها عدد من الأشخاص القائمين على التشغيل المختلفين والذين يعملون في مخابر مختلفة ويمكن أن يستخدموا طرقاً مختلفة للحصول على نفس قيمة القياس .

الثباتية Stability :

يركـز هـذا المصـطلح علـى القيـاس المتكـرر لفتـرة طويلـة . إذ إن المخـابر التشخيصية يُحتفظ فيها عـادة بمـادة مرجعيـة لفحص الثباتيـة فـي القيـاس مـع مـرور الزمن .

الصلاحية Validity :

وهو مصطلح قياس يركز على تحديد ما إذا كان القياس فعلياً وعلـى تحديد النقاط المراد دراستها والتي يجب إجراء قياس لها . وكذا الفترة التي أُجري القياس فيها أهـي بعيـدة أو قريبـة . ففـي المفهـوم السـريري يُقـارن القيـاس مـع المعيـار الـذهبي Golden Standard الأمثل والذي سيذكر ويشرح لاحقاً .

المجتمعات الحيوانية والعينات Populations and Samples

إن مفهوم المجتمع الحيواني يشمل القياسات التي تجرى على العينة المأخوذة منه اعتماداً على أسس معينة . يتضمن المجتمع الحيواني جميع ممثلات مجموعة محددة وتمثل العينة مجموعة صغيرة من مجموعة كبيرة مأخوذة من المجتمع الحيواني المراد دراسته ونحن نريد هنا اختيارَ عينـة كبيرة كبراً كافياً بهذه الطريقة والتي تمثل المجتمـع الحيـواني ولا بـد لنـا مـن تعريـف أمـرين أساسـيين قبـل تُعـرّف علـى أنمـاط المجتمع الحيواني .

تعريف الواحدة الإحصائية :

تعرف الواحدة (الوحدة) الإحصائية بكل كائن أو ظاهرة أو أي شيء يشترك في صفة أو أكثر وتدور الدراسة حولها . كما أن المجتمع الإحصائي يعرف بجميع وحداته .

أنماط المجتمع الحيواني الإحصائي Types of Population

يعرف المجتمع الحيواني الإحصائي بأنه جملة العناصر التي نريد دراستها وتُعرّف إلى صفاتها والتي هي موضوع الدراسة فإذا أردنا مثلاً أن نبحث في مدى تطبيق التقنيات الحديثة المستخدمة في مختلف العيادات البيطرية أو مراكز الصحة الحيوانية في إحدى المحافظات فسيتحدد المجتمع الإحصائي بجملة العيادات البيطرية أو مراكز الصحة الحيوانية المنتشرة في تلك المحافظة أما إذا أردنا أن ندرس الأمراض في مزارع الأبقار الحلوب فسيتألف المجتمع الإحصائي هنا من مجموع مزارع الأبقار الحلوب في سوريا .

وسوف نستخدم فيما يأتي كلمة / حيوان / على العموم لنقترح ما يسمى بوحدة التقصي Unit of Investigation لكننا أيضاً سوف نستخدم عبارات أخرى ككلمة فرد حيواني / Individual / أو كلمة حالة / Case / لأننا نريد أن نعتاد استعمال مصطلحات مختلفة ضمن منهجنا الجامعي هذا .

إن مجتمع الحيوانات يمكن أن يمثل بما يأتي :

- أفراد الحيوانات The Individuals

مثال ذلك كافة أبقار القطيع ، كافة أبقار اللحم في المزرعة ، كافة حيوانات القطيع (بما فيها عجلات ، ثيران ، أبقار) .

- **قياسات متغير خاص The Measurements of a Particular Variable**

على سبيل المثال قياس أوزان الكبد وأطوال العظم وقياس هرمونات الدم أو مستويات الأنزيم في عضو من أعضاء الجسم .

- **أعداد عناصر معينة Numbers of Items :**

فمثلاً لمنطقة معينة أو حجم معين لشيء ما أو زمن معين) . ومثال ذلك التعداد الخلوي للدم أو تعداد بيوض البراز باستخدام طرق الإشعاعات الجزئية المنبعثة من أجهزة إشعاع معين .

والمجتمع بمفهومه الفلسفي العام إما أن يكون مجموعة حقيقية (وهنا يكون محدوداً finite) أو مجموعة نظرية (وهنا يكون غير محدود infinite) .

مثال :

إذا ما كنا مهتمين بدراسة معدل النمو عند الخنازير في منطقة ما فإن المجتمع هنا جميع خنازير المنطقة . وهذا مجتمع حقيقي محدود . أما إذا أردنا معرفة

تأثير علفة تجريبية على هذه الخنازير فإننا سنجرب تعليف حيوانات هذه الوجبة وهذه الحيوانات تمثل عينة ، والتي يجب أن تكون حيوانات هذه العينة ممثلة لحيوانات هذا المجتمع نظرياً . ونظرياً على الأقل يمكننا أن نقيس فعلياً حالات محددة ضمن هذه العينة المأخوذة من هذا المجتمع الحيواني إلا أن هذه العينة تمثل مجتمعات غير محدودة (لا نهاية لها) والتي تمثلها هذه المجتمعات فقط على شكل ما يدعى العينة Sample .

نتيجة : تعريف العينة الإحصائية : تعرف العينة الإحصائية بأنها مجموعة جزئية من المجتمع الإحصائي المدروس تسحب عناصرها من الوحدات الإحصائية للمجتمع الإحصائي الكلي . بعد سحب العينة نقوم بالدراسة الإحصائية عليها ثم نعمم نتائج دراستها على المجتمع الذي تنتمي إليه والمرغوب دراسته وذلك بحدود معينة ومقبولة من حيث درجة الثقة .

المعاينة العشوائية Random Sampling

إننا نقوم بتفحص مفهوم العينة Sample مع مراجعة العبارات المشكلة لمصطلح المجتمع الحيواني . إن العينة يجب أن تكون ممثلة للمجتمع الحيواني المأخوذة منه لنحصل على نتائج تفيد في التطبيق على المجتمع الحيواني ذي الأعداد الكبيرة . ولكي نحصل على عينة ممثلة للمجتمع الحيواني نختار الأفراد الحيوانية

اختياراً عشوائياً (اعتماداً على طريقة المصادفة By chance أو الاحتمالية Probability) من كافة الأفراد المشكلة للمجتمع الحيواني. هذا المفهوم للاختيار العشوائي Random Selection يعتمد أساساً على طرق معينة ومتعلقة بعملية الاختيار (وهذا ما سنتحدث عنه لاحقاً) .

أساساً يجب أن نستخدم طريقة موضوعية لتحقيق عينة نحصل عليها عشوائياً والتي تعتمد أساساً على سلسلة الأرقام العشوائية طريقةً للاختيار .

ويمكن أن نحصل على سلسلة الأرقام هذه من جداول الأرقام العشوائية . أو يمكن أن نحصل عليها من الأرقام العشوائية في الحاسب الشخصي والتي يمكن إنشاؤها بطريقة سهلة . وكذا يمكن الحصول عليها باستخدام الآلة الحاسبة الإلكترونية العلمية .

نتيجة :

لاحظ من أجل توزيع أفراد الحيوانات الداخلة في المجموعات المعالجة إن المبادئ الأساسية للتوزيع العشوائي (الطريقة العشوائية) يجب أن توظف بالاستفادة من التوزيع العشوائي لنتجنب التأثير الموضعي لدراسة معينة ولنتأكد من أن المجموعات هي في شكل يمكن مقارنته .

ومرة ثانية تكون سلسلة الأرقام العشوائية مطلوبةً لتقدم لنا توزيعاً هدفه الأفراد الحيوانية أو الحيوانات المعالجة ولذلك فإن أسباب وجود أي فروقات متكررة في أثناء إنجاز التحليل بين المجموعات يمكن أن تعرف وتحدد تحديداً مناسباً .

أنواع الطرق الإحصائية Types of Statistical Procedures

يمكن أن تقسم الطرق الإحصائية إلى نمطين اثنين : إحصاء وصفي Descriptive Statistics وإحصاء استنتاجي Inferential Statistics .

١- الإحصاء الوصفي Descriptive Statistics

نستخدم هذه التقنيات لإنقاص حجم البيانات المجموعة وتلخيص اتجاهات البيانات وميولها ضمن الإدارة المدروسة الإضافةً إلى أهميتها في تمثيلها النتائج تمثيلاً واضحاً . من هذه الطرق يمكننا أن نرسم الأشكال والمخططات والجداول مع بيانات وصفية رقمية . ويشمل الوصف الرقمي قياسات كحساب المجال الذي تتأرجح فيه مركزية البيانات مثل الوسط الحسابي /Arithmetic mean/ أو المنوال /mode/ وقياسات التشتت للبيانات Dispersion measures كالفرق variance أو المدى range وستشرح هذه القياسات شرحاً مفصلاً لاحقاً .

تعريف الإحصاء الوصفي Descriptive Statistics :

هو كما تدل تسميته الإحصاء الذي يصف الواقع حقيقة وكأنما نصور صورة فوتوغرافية للمجتمع الإحصائي بكامله . فاللجوء إلى الإحصاء الوصفي يتطلب دراسة للمجتمع الإحصائي بكافة وحداته . فهو يعني عرض حقيقة الظاهرة المدروسة عرضاً ملخصاً بالأساليب الإحصائية المتبعة في جمع وتبويب وعرض البيانات وكذلك تحليلها بإجراء جميع الحسابات اللازمة والتوصل لمعالم وخواص المجتمع الإحصائي التي يجب أن تظهر في الصورة .

٢- الاحصاء الاستنتاجي (أو الاستدلالي)Inferential Statistics

وهو يتضمن الطرق الإحصائية الاستنتاجية وتقدير حدود المجتمع الحيواني مستخدمين بيانات العينة . يشمل الحد parameter α قياسات مثل المعدل mean أو النسب proportion ويصف هذا الإحصاء جوانب معينة لتوزيع المتغيرات . ويتبع هذا التقدير عموماً استخدام النظرية الإحصائية (والتي سنتناولها تفصيلاً في الفصول القادمة) . وفي الفصول القادمة شرح مفصل لهذا النوع من الإحصاء لاستخدام النظرية الإحصائية وكذا مخطط الخوارزمية flow chart المثبت في الملحق والذي يقدم دليلاً سريعاً للاختيار المناسب .

نتيجة :

يعرف الإحصاء الاستنتاجي أو الاستدلالي بأنه ذلك الإحصاء الذي هدفه التوصل إلى صورة للمجتمع الإحصائي وذلك عن طريق دراسة ذلك الجزء من المجتمع الإحصائي والمسمى بالعينة . فالاستدلال هو التوصل من الخاص إلى العام ، أي من الجزء (العينة) إلى العام (المجتمع الكلي) ويمكن تعريف الإحصاء الاستدلالي بمجموعة الطرق المستخدمة في التوصل لنتائج المجتمع بالاعتماد على بيانات العينة المسحوبة من ذلك المجتمع .

استنتاج Conclusion :

قمنا في هذا الفصل بتوضيح الأفكارَ والطرائق التي ستعرض في الفصول اللاحقة كما قدمنا شرحاً وجيزاً للمفاهيم الإحصائية التي سنتناولها لاحقاً إذ أنه من المهم أن تكون لدينا فكرة أولية عن الطرائق والمواد الإحصائية التي ستعرض لاحقاً قبل الدخول فيها على نحو مفصل ومسهب .

وقد هدف هذا الفصل إلى أن اكتساب معرفة عن معنى البيانات التي أنشأناها ويجب أن نتذكر أن البيانات المنظمة تنظيماً سيئاً لا نحفز تحليلها إحصائياً جيداً . فبالإسناد إلى علم الإحصاء الحيوي سنكوّن فكرة عن تلخيص البيانات وتفسير النتائج المترتبة عليها . ويعطينا دلائل لشرح وتفسير ما حولنا . وسنركر في الفصول القادمة على تطوير فهم للطرائق الإحصائية وتحديدها للمساعدة في تفسير النتائج .

الفصل الثالث

CHAPTER THREE

الإحصاء الوصفي (1)

DESCRIPTIVE STATISTICS (1)

جمع البيانات الإحصائية وتبويبها وعرضها

Collected Data , Handling and Presenting

تعتبر عملية جمع البيانات وتنظيمها وعرضها خطوة أساسية قبل البدء بأي دراسة إحصائية لأن العمل الإحصائي يتركز عملياً في خطوات لاحقة لتحليل هذه البيانات . وقد تعرفنا في الفصل الأول أنماطَ البيانات الإحصائية وهي جزء أساسي للبدء بعملية التبويب والعرض .

قد لا يكون مجال في هذا الفصل لشرح تفاصيل السؤالين اللذين قد يتبادران إلى الذهن وهما : ما البيانات التي يجب أن تجمع من أجل الدراسة المطلوبة ؟ ومن أين يمكن جمعها ؟

ذلك أن موضوع الدراسة نفسه يحدد نوع البيانات المطلوبة ويحدد عموماً مصدر هذه البيانات ، فعلى سبيل المثال إذا أردنا معرفة الوسط الحسابي لأوزان المواليد الحديثة الولادة في مبقرة حمص الحكومية فهذا يتطلب منا جمع البيانات المتعلقة بأوزان المواليد الحديثة الولادة ، وهنا نجد أن مصدر البيانات لن يكون بالعملية الصعبة حيث أن إدارة المبقرة أو المحطة فيها كافة البيانات التي نريدها ومن

الغير الممكن أن يخطر بذهننا الذهاب إلى المديرية العامة لمؤسسة المباقر لجمع هذه البيانات .

وعموماً في حالة عدم توفر المصدر الذي يمكن أن نحصل منه على البيانات اللازمة لنا نظراً لخصوصيتها فإننا سنضطر إلى جمع البيانات بأنفسنا ميدانياً .

ولا بد لنا من تحديد وسائل الاتصال الفعالة المتعلقة بالوحدات الإحصائية وهي قد تكون عن طريق الاتصالات الهاتفية أو المقابلات الشخصية أو المراسلة بالبريد هذا إذا كانت الوحدات الإحصائية موضوع الدراسة يمكن الحصول عليها من خلال ذلك .

لكن إذا كان عملنا تجريبياً أي موضوع الدراسة موضوع تجريبي كأن نرغب في الحصول على البيانات الإحصائية المتعلقة بأعداد فئران الحقل الموجودة في إحدى المناطق فهنا لا بد من اللجوء إلى طريقة القياس المباشر . ولا بد من الإشارة إلى أن وسيلة الاتصال الأفضل لجمع البيانات تتوقف على أمور عدة منها التكلفة المادية وسرعة جمع البيانات ودقة هذه البيانات .

وبعد تحديد الوسيلة لعملية جمع البيانات فلا بد لنا من إعداد ما يسمى بالاستمارة أو (الاستبيان) والتي يجب ملؤها بالمعلومات والبيانات المطلوبة وذلك من

أجل إجراء الدراسة الإحصائية عليها لاحقاً فما هي الاستمارة الإحصائية أو الاستبيان الشخصي ؟ .

الاستمارة الإحصائية Statistics Questionnaire

تتألف الاستمارة أو الاستبانة أو (البيان) من صفحة واحدة أو عدة صفحات وتحتوي على الأسئلة التي يريد الباحث إجابة عنها ويترك فراغ بجانب كل سؤال لتسجيل الإجابة عليه .

ويجب أن نعلم أن تصميم الاستبانة أو (البيان) يختلف من باحث لآخر وحسب الغرض والهدف من موضوع الدراسة إلا أن هناك بعض الشروط التي يجب مراعاتها عند تصميم الاستبانة أو (البيان) وأهم هذه الشروط :

١- أن تكون الأسئلة موجزة .

٢- أن تكون مرتبة منطقياً .

٣- أن تكون واضحة ولا تحتاج إلى تفسير .

٤- أن يكون عددها أقل ما يمكن بشرط عدم تأثير ذلك على كمية المعلومات المراد الحصول عليها .

٥- يفضل عموماً الأسئلة ذات الإجابة بنعم أو لا ، أو ذات الإجابات المختصرة .

٦- يفضل عموماً أن تكون إجابات الأسئلة قابلة إلى تحويلها إلى شكل رقمي .

تلخيص البيانات Summarizing Data

نجمع البيانات عادة مع احتساب المعلومات التي يمكن أن نحولها إلى بيانات مكونة من عدة أجزاء . مما يؤدي إلى إشكال بسيط يعرض عندما نحلل بياناتٍ متضمنةً على مشاهداتٍ قليلةً لعدد قليل من الحيوانات . وعلى أية حال فإن كمية البيانات التي نحصل عليها بإزدياد يوماً بعد يوم سوف تصعب علينا تصوّر ما يحدث في هذه المشكلة المدروسة أو تلك ولذلك فإن المرحلة الأولى للعمل على إيجاد تصور لهذه البيانات هي تنظيم هذه البيانات لتقييم الاختلاف في القيم عموماً . وعندئذ فإن الطريقة التي تعطينا تصوراً لهذه البيانات هي إنقاص كميتها وإدارتها مما يؤدي إلى الحصول على تصور معين للمساعدة في فهم وتفسير مثل هذه البيانات ، مثل هذا التصور المبدع يمكن إنجازه بِ :

١- **العرض الجدولي** : وذلك بعرض الملامح العامة للبيانات المجموعة وهذه ستناقش لاحقاً في هذا الفصل .

٢- **العرض بالرسوم** : وذلك باستخدام الأشكال والمخططات التي توضح لنا أنماطاً

٣- **القياسات الرقمية** : التي تلخص البيانات المجموعة (جميع هذه الطرق ستعرض بالتفصيل لاحقاً في هذا الفصل) .

ستعرض بالتفصيل لاحقاً في هذا الفصل) .

توزيعات التكرار التجريبية

Empirical Frequency Distributions

ما هو التوزيع التكراري:

What is a Frequency Distribution

يظهر التوزيع التكراري تكرارات حدوث المشاهدات في قاعدة البيانات ، أي توزيع المشاهدات الموجودة على عدد مناسب من الفئات ومن ثم إيجاد عدد المشاهدات الموزعة على كل فئة . وغالباً ما يدعى توزيع البيانات للمشاهدات بالتوزيع التكراري التجريبي بالمقارنة مع التوزيع الاحتمالي النظري .
والذي يظهر بواسطة نموذج رياضي معين .

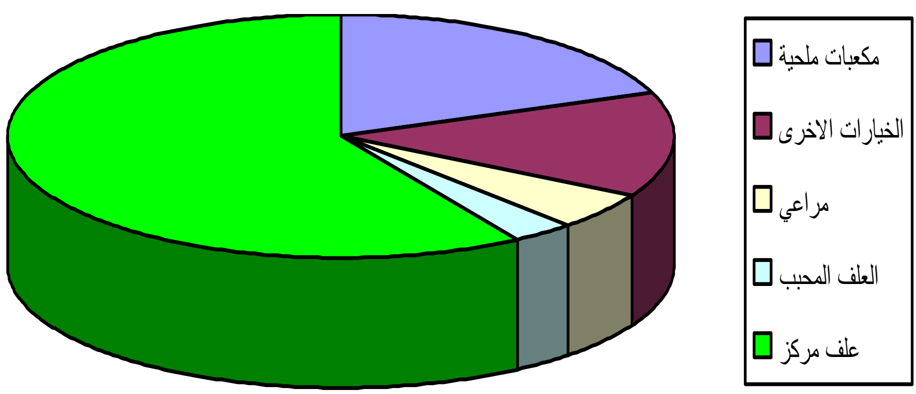
ومن الأهمية الحيوية بمكان أن نفهم بوضوح أن نميز بين المتغيرات النوعية والكمية قبل أن نحاول استخدام التوزيع التكراري بما أن نوع المتغير سوف يقدم لنا الشكل الأمثل تناسباً لتوزيع البيانات فعندما يكون المتغير نوعياً Qualitative variable .

فإن التوزيع التكراري للمشاهدات يتضمن عندئذ تكرار حدوث المشاهدات في كل فئة من هذا المتغير وهنا يمكن أن ننظم هذه البيانات في جدول يمثل كل فئة منها أو يمكن أن نمثلها في مخطط بياني كمخطط مضلعات أو مخطط دائري .

فعلى سبيل المثال إذا كان المتغير يمثل طرائق مختلفة للمعالجات للوقاية من نقص المغنزيوم في الدم Hypomagnesaemia في الأبقار الحلوب فإن أعداد المزارع المشاهدة باستخدام كل طريقة سوف يشمل ما يسمى بالتوزيع التكراري لهذه المزارع عن هذه الحالة وهنا فإن البيانات يمكن أن توضح في شكل مخطط دائري كالشكل رقم (٣) .

فهو مخطط دائري يظهر النسب المئوية للقطعان المطبق عليها طرق التحكم النوعية من نقص المغنزيوم والتي يستخدمها مزارعو الأبقار الحلوب في ٢٧٨ قطيع حلوب (المصدر دراسة ١٩٩٦ Mc Coyetal) .

الشكل رقم ٣: النسب المئوية للقطعان المطبق عليها طرق التحكم النوعية من نقص المغنزيوم على شكل مخطط دائري



وعندما يكون موضوع الدراسة متعلقاً بالمتغيرات الكمية quantitative variables سواء كانت على شكل بيانات غير مستمرة / متقطعة أو بيانات مستمرة يسهُل فهم البيانات واستيعابها بإنشاء ما بين 5 إلى 15 حداً فاصلاً لها . ومن المفضل أن تكون هذه الفواصل غير متداخلة أو على شكل فئات وأن تشمل مدى من القيم يتراوح فيه المتغير المدروس وبهذه الحالة نحدد أعداد المشاهدات المنتمية إلى كل فئة (تكرار الفئات) ولإكمال بيانات تكرارات الفئات نأخذ عندئذ التوزيع التكراري .

لنأخذ مثلاً التوزيع التكراري لأعداد الولادات للأبقار الحلوب الإناث في المزرعة ذات الاسم X في الجدول رقم (١) .

جدول رقم (١) : التوزيع التكراري لأعداد الولادات للأبقار الحلوب الإناث في المزرعة ذات الاسم X

٠	٢	٢	١	٥	١	٣	٠	٠	١
١	٤	٠	١	٠	١	٢	٥	٠	١
١	٣	١	١	١	٠	٣	١	١	٠
٣	٢	٧	١	٣	٠	٢	٤	١	٠
١	٦	٠	٠	٠	٠	٣	١	٢	٠
١	٣	١	٣	٠	١	٠	١	٢	١
١	١	٠	٤	٢	٠	١	٢	٥	٠
٠	٢	٢	١	١	١	٢	١	٠	٣
٥	٠	٠	٠	١	٠	١	٠	٣	٠

		١	١	٧	٢	٢	١	٤	٠

و سيتم حل هذا المثال بعد أن نتعرف باقي أنواع التوزيعات التكرارية .

التوزيع التكراري النسبي (المرتبط)

Relative Frequency Distribution

مع أن التوزيع التكراري هو طريقة مفيدة لوصف المشاهدات الملاحظة فمن الصعب أن نقارن بين اثنين أو أكثر من التوزيعات التكرارية إذا كان إجمالي عدد المشاهدات في كل توزيع مختلفاً . والطريقة الوحيدة لحل هذه المشكلة أن تحسب النسب المئوية للمشاهدات في كل صف أو فئة وهذا ما يدعى بالتكرارات النسبية /المرتبطة/ ويحصل على كل منها بتقسيم التكرار المطلق لكل فئة على العدد الكلي لعدد المشاهدات أو التكرارات . وهكذا فإن مجموع التكرارات النسبية لكل الفئات هو رقم موحد أو ما يدعى بالرقم (١٠٠ %) .

نتيجة :

$$\text{التكرار النسبي} = \frac{\text{التكرار المطلق}}{\text{العدد الكلي للتكرارات}}$$

وغالباً ما تؤخذ الأجزاء المئوية لهذا الاختيار والتي تشكل ٢٥% و٧٥% بعين الاعتبار (وتدعى هذه بالربع المنخفض والربع المرتفع أو الأعلى نسبياً) . فـ ٢٥% من المشاهدات تمتد دون الربع المنخفض و٧٥ % من المشاهدات تمتد علوياً

من الربع المرتفع . ويدعى المجال الذي بين الربعين بالمدى داخل الأرباع (Inter-quartile range) وإن القيمة ٥ % و٩٥ % تشمل القيمة المركزية ٩٠ من المشاهدات ويمكن أن يظهر كيفية تقييم هذه النسب المئوية بإستخدام مضلع التوزيع التكراري التجميعي.

الجدول رقم ٢ : حل مثال الجدول التكراري لأعداد الولادات في إناث الأبقار الحلوب في المزرعة X

عدد الولادات	التكرار المطلق	التكرار النسبي%	التكرار التجميعي	التكرار التجميعي النسبي%
٠	٢٩	٢٩.٦	٢٩	٢٩.٦
١	٣٤	٣٤.٧	٦٣	٦٤.٣
٢	١٤	١٤.٣	٧٧	٧٨.٦
٣	١٠	١٠.٢	٨٧	٨٨.٨
٤	٤	٤.١	٩١	٩٢.٩
٥	٤	٤.١	٩٥	٩٦.٩
٦	١	١.٠	٩٦	٩٨.٠
٧	٢	٢.٠	٩٨	١٠٠.٠

الجداول Tables

الجدول هو ترتيب شقي للبيانات التي تشمل عادة أعداداً أو كلمات ممثلة في صفوف وأعمدة وتفرض جملة من الحقائق لتمييز وفهم البيانات ويعرض تخطيط الجدول بواسطة شكل البيانات واعتماداً على ذلك تختلف نماذج أنماط الجداول حسب اختلاف البيانات . وهي طريقة مفيدة وجيدة لتمثيل البيانات .

وعموماً يجب أن نتذكر أن المبدأ الأكثر أهمية في إدارة وإنشاء الجداول هو أن تكون واضحة جيداً . وفيما يأتي القواعد المستخدمة لبناء جداول واضحة وجيدة التقييم :

١- يتضمن الجدول معلومات واضحة مع وجود عنوان رئيسي وعناوين فرعية واضحة للجداول ومصممة تصميماً متناسقاً .

٢- أن تتضمن عناوين مختصرة لكل صف وعمود .

٣- أن تتضمن وحدات القياس .

٤- عندما يكون الغرض من تشكيل الجدول هو إعطاء ملخص إحصائي (كدراسة الوسط الحسابي) يجب أن يتضمن إلى جانبه ما يدعى بحد الثقة (Confidence Interval) .

٥- أن يتضمن إشارة فقط إلى درجة الدقة عند لزومها كالدرجة المعنوية أو غيرها كدليل للقارئ .

٦- أن لا يتضمن معلومات كثيرة .

٧- تذكر أن يتضمن معلومات يمكن إحصاؤها بسهولة بإستخدام الأعمدة منه من استخدام معلومات متقاطعة في صفوف .

٨- إذا كانت إحدى القيم صفراً فيجب أن تَكْتَب صفراً لا أن تَتْركَ فراغاً أو تضع شحطة بدلاً من الصفر .

٩- يفضل تجنب الخطوط الأفقية والأفضل ترك فراغات بدلاً من الخطوط الأفقية

الأشكال أو الرسوم البيانية Diagrams

إن الغرض من العرض البياني للبيانات هو إعطاء صورة جميلة للبيانات بطريقة سهلة ومرتبة تجذب القارئ وتؤثر عليه . ولما كان جميع القراء يميلون إلى رؤية الأشكال أكثر من مشاهدة الأرقام الجافة في الجداول شاعت طرقُ العرض البياني كثيراً ولا سيما بعد استخدام برامج الحاسوب الخاصة بإنشاء الأشكال البيانية بصورة جميلة وجذابة . وفيما يأتي القواعد الأساسية للرسومات البيانية :

١- يجب أن يفسر الرسم البياني نفسه بنفسه .

٢- يجب أن يكون بسيطاً ولا يتضمن معلومات غير ضرورية .

٣- يجب أن يعرض وحدات القياس المدروسة .

٤- يجب أن يظهر التنقيط بوضوح في المخططات النقطية .

٥- يفضل أن يكون عنوان الرسم البياني بالأسفل حتى لا يتداخل مع الرسم .

٦- وعندما يعرض المخطط البياني أكثر من متغير واحد يجب التمييز بينها بعلامات خاصة .

٧- ويفضل عدم الكتابة على الرسم البياني .

٨- علينا أن ننوه بأن الرسم أو المخطط البياني يجب أن لا يحتوي على متغيرات كثيرة .

٩- يجب التأكد من أن طريقة عرض البيانات تمثل كافة المعلومات المناسبة والضرورية.

طرق الرسم أو العرض البياني

Methods of Presenting of Diagrams

الرسم بالأعمدة Histogram Chart :

وهنا يمثل طول العمود القيمة الرقمية للمتغير الذي يمثله وبمقارنة أطوال الأعمدة الموجودة في الرسم البياني نكون على فكرة عن الاختلاف في المتغير المدروس . ويجب مراعاة القواعد العامة عند الرسم بالأعمدة .

١- عرض قاعدة الأعمدة يجب أن تكون متساوية .

٢- ارتفاع العمود يمثل القيمة الرقمية للمتغير .

٣- المسافة بين الأعمدة يجب أن تكون متساوية ويجب مراعاة التناسق للمسافات بين الأعمدة وعرض الورقة .

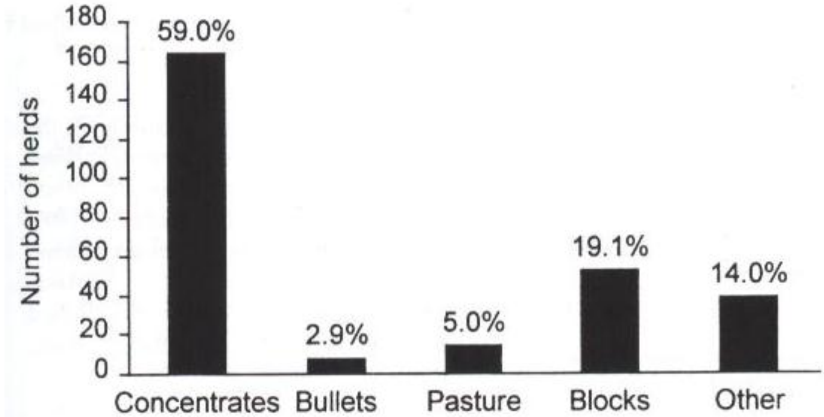
٤- تمثل الأعمدة غالباً بصورة عمودية ويمكن أن تمثل بصورة أفقية ولا سيما إذا كان عدد الأعمدة كثيراً .

٥- يفضل ترتيب الأعمدة تنازلياً أو تصاعدياً .

٦- وإذا كان هناك أكثر من متغير سيمثل بطريقة الأعمدة فيجب التمييز بين أعمدة المتغيرات باستخدام الألوان المختلفة أو التظليل أو التخطيط .

٧- يجب أن تكون تسمية العمود في أسفله لا فوقه .

الشكل (٤) طرق التحكم النوعية لنقص المغنزيوم في بعض القطعان باستخدام المربين لِطرق نوعية (مركزات – حبوب – مرعى – مكعبات مغنزيوم وطرق أخرى).



اللوحة الدائرية Pie Chart

تستعمل هذه الطريقة لتوضيح النسب الجزئية للظاهرة المدروسة ويفضل أن يكون خط البداية عند الساعة (١٢) ولعدها توزع الأجزاء الأخرى باتجاه عقرب الساعة وتبدأ من الجزء الأكبر إلى الجزء الأصغر .

ولتحويل النسب المئوية إلى درجات الزوايا نضرب النسبة المئوية بالرقم (٣.٦) لأن مجموع درجات الدائرة تساوي (٣٦٠) إذ إن ٣٦٠ ÷ ١٠٠ = ٣.٦

وإذا كانت النسب المئوية غير معطاة فإننا نقسم قيمة كل مشاهدة على مجموعها ونضرب بالرقم ٣٦٠ فتعطينا قيمة الزاوية حسب القانون التالي :

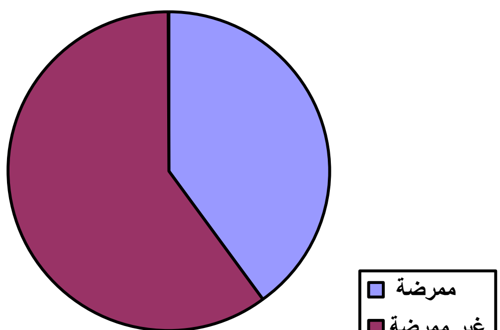
$$\text{درجة الزاوية} = \frac{\text{العدد} \times ٣٦٠}{\text{المجموع الكلي}}$$

ويمكن حسابها بطريقة أخرى وذلك بتقسيم الزاوية المركزية (٣٦٠) على مجموع قيم الظاهرة والناتج يضرب بكل قيمة فنحصل على انقطاع الزاوي المطلوب .

ولنذكر مثالاً لذلك توزيع الذراري المرضية وغير المرضية للمرض الحموي في الجدول رقم(٢).

الجدول رقم ٢: **مثال عن تمثيل مرض الإسهال الحموي عند الأبقار**

نوع المسبب المرضي	العدد	النسب المئوية	درجة الزاوية
ذرية ممرضة	٦٠	٣٠	١٠٨
ذرية غير ممرضة	٨٠	٤٠	١٤٤



الشكل رقم (٥) : نموذج للوحة الدائرية

الخط البياني Diagram :

يرسم الخط البياني على محورين تعاقدياً المحور الأول يدعى بالمحور الأفقي وهو ما يرمز له رياضياً بالمحور السيني (س أو X باللغة الإنكليزية) والمحور الثاني وهو المحور العمودي وهو ما يرمز له رياضياً بالمحور العيني (ع أو Y باللغة

الإنكليزيـة) وعـادة مـا يمثـل المحـور السـيني المتغيرَ المسـتقلَ كـالعمر وكميـة الإنتـاج والقيم الدمويـة وغيرها من المتغيرات المستقلة فـي حين يمثل المحـور العينـي المتغير غير المستقل (التابع) كعدد الحيوانات المريضة أو مستوى حدوث المرض .

وعند الرسم البياني يجدر بنا الأخذ بعين الاعتبار ما يأتي :

١- تحديد مسطرة القياس للمحورين X و Y .

٢- يفضـل أن لا يزيـد عـدد الخطـوط البيانيـة علـى ثلاثـة خطـوط ليسـهل علـى القارئ فهم للرسم البيانـي .

٣- يحسـب المحـور السـيني X بالمقيـاس الحسـابي بينمـا يحسـب المحـور العينـي بالمقيـاس اللوغـاريتمي أو الحسـابي وإذا حسـب المحـور العينـي بالمقيـاس اللوغـاريتمي فيسمى بالرسم البيـاني نصـف اللوغـاريتمي ويسـتخدم هـذا النمط مـن الرسومات البيانية عندما يكون اهتمامنا بتبيان نسب التغيير ولـيس القيمـة المطلقـة للمتغير ، فعلـى سبيل المثـال لإظهـار اتجـاه مرض السـارز أو بعض السرطـانات نهتم بنسب التغيير في حدوث المرض أكثر من الفروقات المطلقة .

الفصل الرابع
CHAPTER FOUR
المقاييس الرقمية

NUMERICAL MEASURES

تشمل قياسات الإحصاء الوصفي جمع البيانات وتبويبها وتلخيصها ومن ثم عرضها . وهذا يعني أن هذا النوع من الإحصاء يهتم بوصف البيانات الإحصائية بدون الدخول في عملية التحليل الإحصائي للوصول إلى الاستنتاجات العلمية وتعتبر أهمَ وسائل الإحصاء الوصفي هي مقاييسُ الميول أو النزعة المركزية Measures of Central Tendency ومقاييس التشتت Measures of Dispersion فعندما ننظر في توزع قيم بيانات إحصائية كإنتاج كمية الحليب اليومية لكل بقرة في إحدى مزارع منطقة من المناطق سوف نجد أن هذه القيم تتوزع حول قيمة متوسطة /مركزية/ إذ يكثر عددها حول هذه القيمة المتوسطة ويقل تدريجياً مع الابتعاد عنها . تسمى المقاييس التي تقيس مثل تلك القيمة المتوسطة مقاييسَ النزعة المركزية . إلا أن دراسة قيم النزعة أو الميول المركزية والتي تلقى الضوء على بعض الجوانب الهامة من الظاهرة تبقي في الخفاء جوانب أخرى على قدرٍ كبيرٍ من الأهمية .

إذ إننا غالباً ما نحتاج إلى دراسة مدى تبعثر أو بعد أو انحراف قيم الظاهرة عن تلك القيمة المتوسطة . تسمى المقاييس التي تقيس انحراف القيم عن القيمة المتوسطة مقاييسَ التشتت . وتعتبر المرحلة الأولى في أي دراسة هي حساب تلك

المقاييس والذي يعتبر بداية أولية في العمل الإحصائي من حيث تحليل البيانات الإحصائية بعد عمليات جمع البيانات وتصنيفها وعرضها بالطرق المناسبة .

المقاييس الموضعية Measures of Location

أو ما يسمى بمقاييس الميول أو النزعة المركزية Measures of central Tendency أو ما يدعى بالمعدل Average إن عبارة المعدل تشير إلى قياس من القياسات العديدة للميول المركزية للبيانات . فعندما نذكر في مثالنا السابق أن معدل إنتاج الحليب للبقرة الواحدة باليوم الواحد في إحدى محطات تربية الأبقار هو ٢٠ ليتراً فمع علمنا أن بعض الأبقار تنتج أكثر من ٢٠ ليتراً وبعضها تنتج أقل من هذه الكمية يمكننا وضع استنتاج أن حجم إنتاج الحليب في هذه المحطة يتمركز حول القيمة ٢٠ ليتراً . وهناك العديد من هذه المقاييس إلا أن أهمها هي : الوسط الحسابي والوسط الهندسي والوسيط والمنوال .

وهناك مقياس غير مركزي يسمى بالمئين ويعبر حقيقة مقياس موضعي يمثل الوسيط في إحدى حالاته الهامة .

١- الوسط الحسابي The Arithmetic Mean

وهذا المقياس أكثر شيوعاً لدى عموم الناس يحسب هذا المقياس بجمع كافة قيم المشاهدات الموجودة في البيانات وتقسيمها على عدد هذه المشاهدات في هذه البيانات .

فإذا رمزنا للمتغير المستمر بالرمز (X) وكان هناك عدد من المشاهدات (n) في العينة فعندئذ يكون الوسط الحسابي في العينة (X) كما يأتي :

$$\overline{X} = \frac{\Sigma \times}{n}$$

وفي جداول التوزيع التكرارية غير المجمعة يؤخذ التكرار بالحسبان فتصبح الصيغة كما يأتي :

$$\overline{X} = \frac{\Sigma fx}{\Sigma f = n}$$

مثال : لتكن لدينا قيم البوتاسيوم في البلاسما الدموية كوحدة قياس (mmol / L) مول / ليتر لأربعة عشر كلباً كما يأتي :

و٤.٥٤ و ٤.٦٨ و ٣.٩٢ و ٤.٣٥ و ٤.٨٧ و ٤.٣٧ و ٤.٨٣ و ٤.٧٣ و ٤.٤٠ و ٤.٦٦ و ٤.٥٩ و ٤.٥٧ و ٥.٢٤ و ٤.٢١ .

فإن الوسط الحسابي هو :

$$\overline{X} = \frac{\Sigma \times}{n}$$

$$\overline{X} = \frac{63.96}{14} = 4.57(mmol / L$$

مثال : لنأخذ بيانات ناتجة من دراسةِ تركيز النطاف في السائل المنوي لواحد وأربعين ثوراً.

٧٥ – ١١٠ – ١٢٣ – ١٣٩ – ١١٠– ٨٦ – ١٢٤ – ١٤٢ – ٩٠ – ١١٠ – ١٢٥ – ١٤٢ – ١١٣ – ١٢٦ – ١٤٢ – ٩٥ – ١١٧ – ١٢٧ – ١٤٣ – ٩٧ – ١١٨ – ١٢٨ – ١٤٣ – ٩٧ – ١٢٠ – ١٢٩ – ١٤٨ – ١٠٢ – ١٢١ – ١٣٣ – ١٥٩ – ١٠٤ – ١٢٢ – ١٣٨ – ١٠٤ – ١٢٢ – ١٣٨ – ١٠٥ – ١٢٣ – ١٣٩ .

الحل:

الوســط الحســـابي = مجمــوع المشـــاهدات / عـــدد المشـــاهدات =٤٨١٣/٤١ = ١١٧.٣٩٠

٢- الوسيط The Median

وهـو عبـارة عـن أحـد المقـاييس الأخـرى للميـول المركزيـة الشـائعة التكـرار باســتخدامها . ويعــرف الوســيط بأنـــه القيمـــة المركزيـــة (الوســيطة) لمجموعــة مــن المشاهدات (n) في عينة ما والتي تكون مرتبة ، فعلى سبيل المثال ترتب المشاهدات تصاعدياً أو تنازلياً بمعنى آخر هو القيمة التي تفصل مجموعة بيانات إحصائية إلى جزئين متساويين فصلاً يجعل قيم أحدهما أقل من قيمة الوسيط بينما تكون قيم الجزء الثاني أكبر منها . وبهذا يكون الوسيط هو القيمة المنصفة لذلك الترتيب التصاعدي والتنازلي لهذه المشاهدات .

نرمز للوسيط بالرمز (Me) ويحسب الوسيط على النحو الآتي :

١- إذا كان عدد المشاهدات (n) وهي فردية فإن الوسيط يساوي :

$$Me = \frac{n+1}{2}$$

إذ إن :

n : تمثل عدد الحدود

وبهذا تكون قيمة الحد ذي الترتيب المقدر هي قيمة الوسيط .

٢- إذا كانت المشاهدات (n) زوجية فإن الوسيط يقع في القيم المنصفة لقسمي المشاهدات.

مثال : إن الأوزان بالغرام لمجموعة من ذكور خنازير غينيا المكونة من ١٩ حيواناً هي الآتي :

٣٥٠ و ٤٩٩ و ٤١٢ و ٥٥٦ و ٧٨٩ و ٩٩١ و ٣١٤ و ٦٤٢ و ٣٨٨ و ٥١٠ و٥٩٨ و ٢٩٧ و ٤٥٥ و ٨٦٣ و ٥٣٦ و ٦٨٥ و ٤٢١ و ٣٣٣ و ٤٧٤

المطلوب : أوجد قيمة الوسيط للمجموعة المذكورة أعلاه .

الحل : لحساب قيمة الوسيط نرتب القيم أولاً وليكن تصاعدياً فنحصل على :

٢٩٧ و٣١٤ و ٣٣٣ و ٣٥٠ و ٣٨٨ و ٤١٢ و ٤٢١ و ٤٥٥ و ٤٧٤ و ٤٩٩ و٥١٠ و ٥٣٦ و ٥٥٦ و ٥٩٨ و ٦٤٢ و ٦٨٥ و ٧٨٩ و ٨٦٣ و ٩٩١ .

فالملاحظ أن عدد المشاهدات هو فردي (١٩) قيمة ولذلك نحسب ترتيب الوسيط :

$$\frac{n+1}{2}=\frac{19+1}{2}=10$$

ولذلك فإن الوسيط هو القيمة المعلمة بصندوق وهي القيمة العاشرة أي تكون قيمة الوسيط ممثلة بالحد العاشر وتكتب Me = 499 وهذه القيمة تتوسط مجموعة بيانات المثال فتقسمها إلى جزئين متساويين عَشْر عن يسارها أصغر منها بالقيمة وتسعة قيم عن يمينها أكبر منها .

وفي الحالة التي يكون فيها جدول n أي عدد قيمها زوجياً فإن ترتيب الوسيط 2 / (n + 1) سيعتبر عدداً كسرياً وسيقع الوسيط عندها بين قيمتين في التوزيع ويكون الوسيط في هذه الحالة عبارة عن الوسط الحسابي لهاتين القيمتين .

مثال : أوجد قيمة الوسيط للمشاهدات الآتية :

١، ٢، ، ٣، ٤

الحل :

بما أن هذه القيم مرتبة تصاعدياً نحسب ترتيب الوسيط مباشرة :

$$\frac{n+1}{2} = \frac{4+1}{2} = 2.5$$

ومن المشاهدات المذكورة أعلاه نجد أن الحد 2.5 غير موجود ويقع بين الترتيبين الثاني والثالث المقابلين للقيمتين ٢ و ٣ فإننا نعتبر أن قيمة الوسيط هي قيمة الوسط الحسابي لهاتين القيمتين :

$$\text{Me} = \frac{2+3}{2} = 2.5$$

ونجد فعلاً أن قيمة الوسيط 2.5 تقسم البيانات الأربعة المعطاة إلى جزئين متساويين .

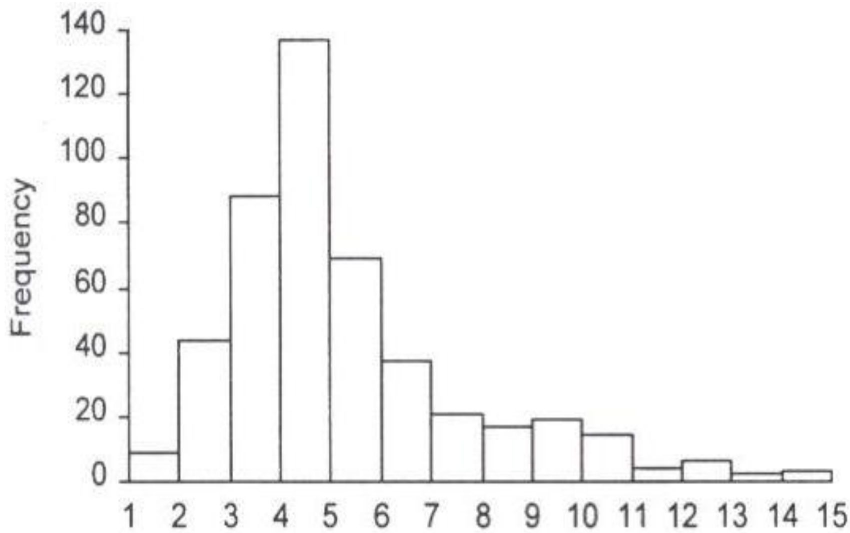
٣- المتوسط الهندسي The Geometric Mean :

إذا أخذ لغاريتم كل قيمة في البيانات (إما للأس y أو للأس e) والتي تكون فيها تكراراتها موزعة بكثرة إلى الناحية اليمنى من المحور السيني فإن هذه البيانات تتحول بتناسق نظامي إلى شقي محور السينات (اليميني واليساري) إذا ما

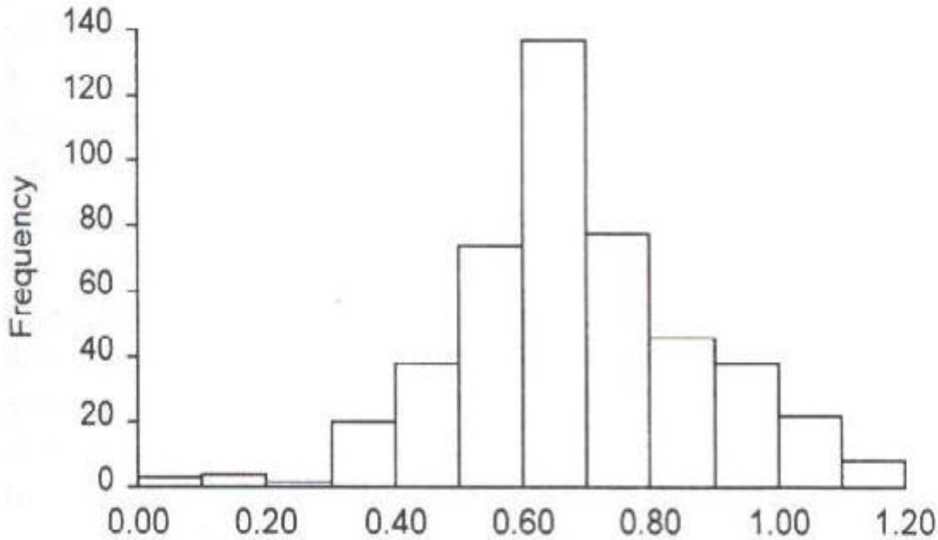
أخـذنا القـيم بشـكلها اللغـاريتمي . و يمثـل الشـكلان رقـم ٦ و ٧ الفـرق بـين التكـرار التوزيعي لقيم الخلايا الجسمية في الحليب و التكرار التوزيعي للقيم اللغاريتميـة فـي إحدى مناطق تربية الأبقار الحلوب.

الشكل (٦) التكرار التوزيعي لقيم الخلايا الجسمية في الحليب

في إحدى مناطق تربية الأبقار الحلوب



الشكل(٧) : التكرار التوزيعي للقيم اللغاريتمية للخلايا الجسمية
في الحليب في إحدى مناطق تربية أبقار الحلوب



وفي هذه الحالة فإن الوسط الحسابي للقيم المحولة لغاريتمياً هو أحد مقاييس الميول المركزية المفيدة . إلا أن مساوئ هذا المقياس هو أننا نتعامل مع مقياس لغاريتمي ولذلك نحول البيانات إلى قياسها الأصلي بحساب الرقم الصحيح المضاد للقيم اللغاريتمية anti logarithm وهذا ما يدعى الوسطَ الهندسي . وفي الدراسات الحيوية عندما تكون البيانات موزعة إلى اليمين أو اليسار من إحدى المحاور البيانية نحسب المتوسط الهندسي لنحصل على معدل القيم المراد قياسها وتصبح أكثر تناسقاً يميناً ويساراً .

يُحسب المتوسط الهندسي رياضياً كما يلي:

$$G = \sqrt[n]{X_1 x X_2 \; x X_3 ......Xn}$$

إذ إن :

$\chi_1 \ldots\ldots \chi_n$: قيم المشاهدات ($\chi > 0$)

n : عدد قيم المشاهدات وهو الجذر النوني لجداءات مجموعة من القيم

وهكذا يُستخدم الوسط الهندسي وسطاً للمعدلات والنسب في كثير من الحالات (وليس في جميعها) كحالة معدل حدوث المرض وهو يمثل الوسط الحسابي في تفسيرها . إذ إن الوسط الحسابي لا يجوز استخدامه في حالات المعدلات والنسب (إلا في حالات خاصة لن نتعرض لها هنا) .

مثال : احسب معدل حدوث مرض داء السكري في عدد من البلدان لأربع سنوات حسب الأرقام المدرجة أدناه (لاحظ أننا عندما نقول معدل حدوث مرض في العلوم الطبية لا تعني المتوسط الحسابي و أنما قيمة عدد الحالات الجديدة للمرض مقسوما على المتوسط الحسابي للحيوانات الواقعة تحت خطر الإصابة خلال فترة زمنية محددة).

0.030 , 0.053 , 0.025 , 0.061

الحل : المتوسط الهندسي لهذه المعدلات المرضية:

$$G = \sqrt[5]{0.030 \times 0.053 \times 0.025 \times 0.06} = 0.25$$

أي معدل حدوث مرض داء السكري للفترة الزمنية المعتبرة هو 0.25 سنوياً .

٤ – المنوال The Mode :

يرمـز للمنـوال بـالرمز Mo وهـو مقيـاس معـروف إلا أن هـذا المقيـاس لا يسـتخدم دائمـاً مقيـاسً للميـول المركزيـة ويعـرف بأنـه هـو القيمـة الأكبـر تكـراراً فـي مشاهدات العينة.

وتختلف قيمة المنوال عن قيمة كل من الوسط الحسابي والوسيط وعادة يكون هناك منوال واحد إذا كان حجم العينة كبيراً ولكن يمكن أن يكون منوالان في مجموعة قـيم عينـة واحـدة إذا تسـاوى تكـرار القيمتـين الأكبـر تكـراراً وعندئـذ يقـال عـن توزيـع مشاهدات العينة بأنه ثنائي المنوال Bimodal .

ن مسوغ اعتبار المنوال من مقاييس النزعة المركزية يرتكز إلى حقيقة كونه التكـرار الأكثـر لإحـدى القيم ، ولـتكن درجتـه 50.000 خليـة / مليمتـر مكعـب لقيم الخلايا الجسمية في الحليب في إحدى المزارع يعني أن قيم هذه الخلايا تدور حول هـذه القيمـة فـي القطـع المـأخوذة منـه العينـة فلتحديـد المنـوال فـي مجموعـة بيانـات محصول عليها نبحث عن القيمة الأكثر تكراراً فيها وتكون هي المنوال .

مثال : لنأخذ قيم كمية الإنتاج لمواسم الإدرار لخمسين بقرة في إحدى مزارع محافظـة حماة / القيم مقدرة باللتر / . أوجد قيمة المنوال لكمية الحليب في الموسم الإدراري من القيم المدرجة أدناه في الجدول رقم (3).

جدول رقم (3): كمية انتاج الحليب في الموسم الادراري (ليتر)

2551	2605	2681	2695	2731
2749	2780	2792	2799	2806
2808	2821	2839	2860	2865
2873	2879	2887	2891	2909
2954	2915	2929	2935	2941
2933	2964	2964	2985	2986
3024	3011	3012	3018	3024
3088	3032	3057	3058	3062
3170	3113	3129	3144	3166
	3231	3240	3381	3303

الحل :

من القيم المدرجة أعلاه نجد أن قيمة 2964 هي القيمة الأكثر تكراراً في هذه البيانات ولذلك فهي قيمة المنوال .

ويجب أن نذكر أن الإحصائيين لا يفضلون استخدام المنوال لتخليص البيانات للأسباب التالية :

١- تعتمد قيمة المنوال على دقة قيم البيانات المقيسة .

٢- بعض التوزيعات ليس فيها قيمة منوال بينما يكون لبعض توزيعات البيانات أكثر من منوال .

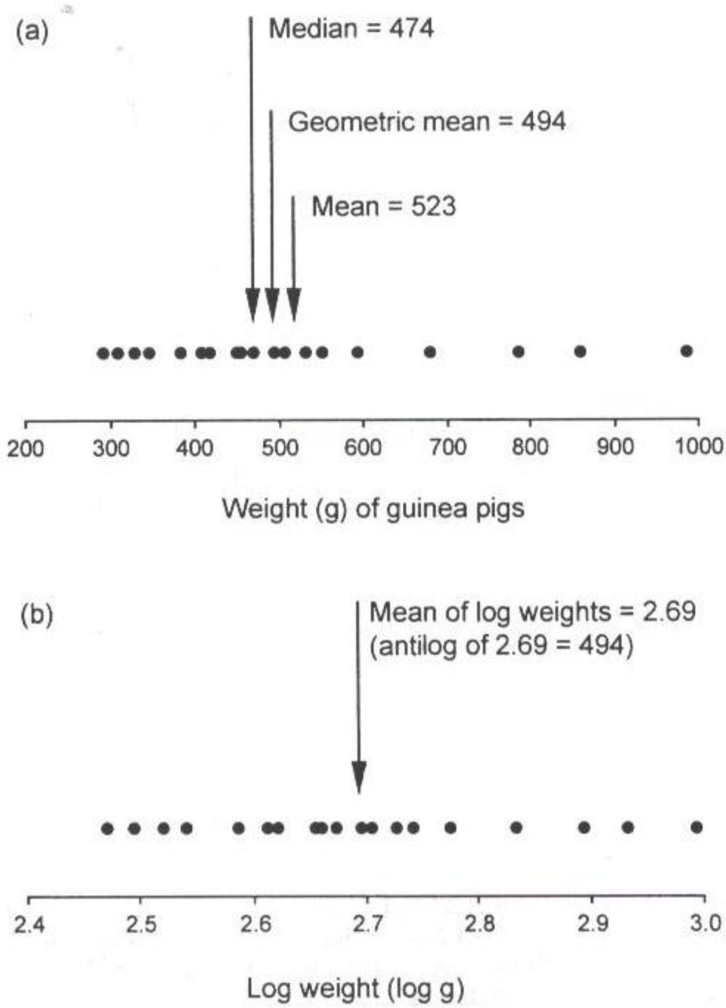
مثال : أوجد قيمة المنوال من مجموعة البيانات الآتية ؟

612 , 8 , 11 , 6 , 8 , 10 , 9 , 7

الحل : نجد أن لدينا منوالين $M_{o1} = 8$, $M_{o2} = 6$ لأن لهما نفس التكرار وهو التكرار الأكثر في التوزيع .

وفي بعض البيانات لا نجد منوالاً مثل 9 , 10 , 11 , 8 , 2 , 1 , 6

الشكل (٨) : أوزان خنازير غينيا المستخدمة في التجربة مع مقارنة الأوزان المستخدمة في التجربة باستخدام القياسات اللوغاريتمية



مقاييس التشتت (الانتشار)

Spread Measures:Measures of Dispersion Spread

وللأسباب التي ذكرت في مستهل هذا الفصل كان من الجدير بنا البحث عن مقاييس أخرى إلى جانب مقاييس النزعة المركزية التي تقيس لنا تشتت قيم الظاهرة عن متوسطاتها وذلك من أجل فهم تفسير ظاهرة الدراسة فهماً جيداً .

هناك العديد من مقاييس انتشار البيانات وكل منها له مفهوم يختلف عن الآخر . ولما كان التشتت لمجموعة من المشاهدات تعني الاختلاف في قيمها . كانت مقاييس التشتت تعطينا فكرة عن درجة الاختلاف والتباعد في قيم مجموعة من المشاهدات .

وأهم مقاييس التشتت هي :

١- المدى The Range :

يعرف المدى بأنه الفرق بين أعلى قيمة وأصغر قيمة في مجموعة المشاهدات .

وطريقة حساب المدى كما هو واضح سهلة . لكن هذا المقياس يعتبر ضعيفاً لقياس درجة التشتت إذ إنه لا يؤخذ بالحسبان سوى قيمتين من قيم مجموعة المشاهدات .

مثال : في المثال السابق الخاص بقيم إنتاج الحليب الموسمي نجد أن أصغر قيمة كانت 2551 لتر وهو يمثل القيمة أو الحد الأدنى من المدى بينما نجد أن أكبر قيمة

إنتاج حليب موسمي كانت 3303 لتر وهو يمثل القيمة أو الحد الأعظمي من المدى .

وبهذا نجد أن المدى في مجموعة المشاهدات في المثال السابق تتراوح بين 2551– إلى 3303 .

٢- التباين The Variance

يحدد الفرق بحساب انحراف أو تباعد كل قيمة من المشاهدات عن الوسط الحسابي . وكلما كانت قيمة المشاهدات قريبة من الوسط الحسابي كانت قيمة التباين أو الفرق صغيرة .

فإذا رمز لعدد المشاهدات بـ n ولكل قيمة من المشاهدات بـ × وللوسط الحسابي بـ $\overline{X}$ وللفرق بـ S^2 فإن التباين أو الفرق يحسب كما يأتي :

التباين = مجموع مربع انحراف القيم عن الوسط الحسابي / عدد المشاهدات – 1

$$S^2 = \frac{\Sigma(\times - \times^-)^2}{n-1}$$

مثال : لنعد إلى مثال قيم البوتاسيوم في بلاسما الدم كما يأتي (مقدرة بالمول /لتر) لأربعة عشر كلباً .

4.37 , 4.87 , 4.35 , 3.92 , 4.68 , 4.54 , 5.24 , 4.27 , 4.59 , 4.86 , 4.40 , 4.73 , 4.83 , 4.21

$$\times^- = \frac{63.96}{14} = 4.57 mmol/L$$

$$S^2 = \frac{(4.37-4.57)^2 + ... + (4.21-4.57)^2}{14-1}$$

$$S^2 = \frac{1.32297}{13} = 0.10177(mmol/L)^2$$

٣- الانحراف المعياري The Standard Deviation :

يرمز للانحراف المعياري بالرمز SD (Stanndard deviation) ويرمز له غالباً بالرمز σ (سيغما) وهو الحرف الإغريقي الصغير لحرف (سيغما) الكبير Σ وهو ببساطة عبارة عن الجذر التربيعية المربعة للتباين حيث نأخذ جذره التربيعي ليعطينا الانحراف المعياري .

ويمكن أن تعتبر الانحراف المعياري معدلاً لانحرافات قيم المشاهدات عن الوسط الحسابي . ويرمز عموماً بالرمز (S) في العينة والرمز σ للانحراف المعياري في كامل المجتمع الحيواني أو البشري .

وهكذا يحسب الانحراف المعياري رياضياً كما يأتي :

$$S = \sqrt{\text{التباين}} \quad = S^2$$

$$S = \sqrt{\frac{\Sigma(\times i - \times^{-})^2}{n}}$$

والجدير بالذكر أن الانحراف المعياري لا يعطي صورة لمتوسط انحرافات القيم عن الوسط الحسابي بل يجذر متوسط مربع الانحرافات عن الوسط الحسابي ويعتبر الانحراف المعياري من أهم مقاييس التشتت لكونه يدخل في معظم عمليات التحليل الإحصائي .

ويمكن أن نحسب الانحراف المعياري بسهولة باستخدام الآلات الحاسبة .

ويجب أن نميز أن هناك مفتاحين في بعض الآلات الحاسبة واحداً يرمز له بالرمز S وهو لحساب الانحراف المعياري ضمن العينة ومفتاحاً آخر يرمز له بالرمز σ وهو لحساب الانحراف المعياري ضمن المجتمع المدروس كاملاً .

وفي بعض الآلات الحاسبة نجد أنه يرمز بالرمز σ_n لحساب الانحراف المعياري في عينة ما.

مثال :

في مثال بيانات موجودات البوتاسيوم في بلازما الدم لأربعة عشر كلباً الذي ذكرناه سابقاً وقدرنا فيه قيمة الوسط الحسابي وقيمة التباين أو الفرق يمكننا أيضاً أن نحسب قيمة الانحراف المعياري .

$$S = \sqrt{0.010177} = 0.319 mmal / L$$

ويمكننا القول إن الانحراف المعياري هو أفضل مقياس لدراسة تشتت البيانات لأن الانحراف المعياري تستخدم فيه كافة قيم المشاهدات المدروسة وهو يستخدم لدراسة حد الثقة(والذي سوف ندرسه في الفصول اللاحقة) لمعرفة دقة النتائج التي حصلنا عليها في اختباراتنا المعملية . فعلى سبيل المثال لحساب مستوى الثقة ٩٥ % لقيم المشاهدات (على سبيل المثال : البوتاسيوم في بلاسما الدم) نجري الآتي :

الوسط الحسابي $\mp$ الانحراف المعياري .

$$CI \quad \%95 = \bar{X} + 2SD$$

حيث أن:

CI= هو حد الثقة (مستوى الثقة) .

$\bar{X}$ = الوسط الحسابي .

S = الانحراف المعياري .

4- معامل الاختلاف Coefficient of Variation :

يعبر عن الانحراف المعياري في بعض الأحيان بنسبة مئوية للوسط الحسابي ولذلك يطلق هذا الاسم على الانحراف المعياري النسبي ويستعمل في التحليل الإحصائي للمقارنة بين تشتت توزيعات مختلفة . وهو ذو أهمية خاصة لمقارنة شدة الإصابة عند استخدام اختبار الاليزا (اختبار المقايسة المناعية الإنظيمية). ويدعى هذا القياس بمعامل الاختلاف ويرمز له بالرمز CV . ويعتبر الانحراف المعياري مقياساً جيداً للتشتت لمجموعة من المشاهدات ولكننا لا نستطيع مقارنة الانحراف المعياري لمجموعتين أو أكثر من المشاهدات لسببين رئيسيين هما :

1- الاختلاف في وحدة القياس .

2- الاختلاف في قيمة الوسط الحسابي .

ولذلك ولمقارنة درجة التشتت في مجموعتين من المشاهدات لا بد من إيجاد مقياس يعتمد على نسبة التشتت بدلاً من القيمة المطلقة للتشتت وهذا المقياس كما ذكر في بداية حديثنا عن معامل الاختلاف الذي يمثل النسبة المئوية للانحراف المعياري من الوسط الحسابي ويحسب رياضياً كما يأتي :

الانحراف المعياري × 100

الوسط الحسابي٧

معامل الاختلاف = ـــ

$$CV = \frac{SD}{\overline{\overline{X}}} X100$$

$$\mathrm{CV} = \frac{\sigma}{\overline{\overline{\mathrm{X}}}} \times 100$$

مع أن :

CV = معامل الاختلاف .

SD أو σ = الانحراف المعياري .

$\overline{X}$: الوسط الحسابي .

٥- المدى المرجعي The Reference Range :

يهتم في بعض الأحيان بوصف قيم المدى الخاصة المتغير والتي تعرف صحة الحيوانات المدروسة وهذا ما يدعى بالمدى المرجعي reference range أو ما يدعى بالفاصل أو الحد المرجعي reference interval ويستخدم هذا المدى لتحديد ما إذا كانت الحيوانات الفردية يمكن وصفها بأنها تنتمي إلى حيوانات سليمة صحياً أو إلى حيوانات مريضة ويدعى المدى المرجعي أحياناً بالمدى الطبيعي normal range ومن الأفضل أن نتجنب استخدام العبارة السابقة لأنها سوف تخلق تشويشاً للقارئ فبينما تشير كلمة طبيعي إلى الحيوانات السليمة تصف كلمة طبيعي في التعبير الإحصائي توزيعاً معيناً للبيانات (والذي سيشرح لاحقاً) . وفي هذا الكتاب سوف نميز هاتين العبارتين باستخدام أحرف أبجدية إنكليزية صغيرة للحالة الصحية وأخرى أبجدية كبيرة بالنسبة للتوزيع الإحصائي .

علـى كـل إن المـدى المرجعـي يمكـن أن يحسـب لحـدود الثقـة 90 % و 95 % و 99 % (ستشرح لاحقاً) إلا أننا غالباً ما نستخدم حد الثقة 95 % وهنا يكون مدى الخطأ 1.96 وبالتالي يحسب المدى المرجعي يحسب كما يأتي :

$$RR = \overline{X} \mp 1.96\ SD$$

مع أن :

$\overline{X}$: الوسط الحسابي ، SD : الانحراف المعياري ويستخدم هنا عندما يكون توزيع البيانات طبيعياً Normal Distribution وعلى كل عندما يكون توزيع البيانات غير طبيعي يمكن أن نحسب المدى المرجعي بفاصل (2.5) بدلاً من 1.96 و 97.5 % بدلاً من 95 % كنسب مئوية من التوزيع التجريبي لقيم المشاهدات .

تمـــارين

١- أجب بكلمة صح أو خطأ للعبارة المدرجة أدناه؟

أ- إن الشكل المناسب لرسم متغير ذي توزيع تكراري مستمر:

١- برسم المستطيلات

٢- اللوحة الدائرية

٣- بوكس و يسكر

ب-الانحراف المعياري هو عبارة عن:

١- هو قياس الانتشار

٢- هو الفرق بين النسب المئوية 5 و 95

٣- هو أكبر من مدى المشاهدات

٤- مقاييس معدل انحراف المشاهدات عن الوسط الحسابي

٥- هو مربع قيمة الفرق

ت-المدى المرجعي (يتضمن مشاهدات 95 %) لمتغير معين:

١- يمكن أن يستخدم لتحديد ما إذا كانت الحيوانات محتملة الإصابة بالمرض عندما تكون قيمتها ممثلة لمتغير معروف

٢- هو يساوي الوسط الحسابي ± الانحراف المعياري إذا كانت البيانات موزعة طبيعياً .

٣- هو يساوي الفرق بين القيم العظمى و الصغرى للمشاهدات في قاعدة البيانات

س٢- احسب ما يأتي:

- المدى و الفرق و الانحراف المعياري لعينة البيانات المدرجة أدناه :

- تركيز فيتامين E (ميكرومول /ليتر) في مصل الدم مأخوذا من 12 عجلة يُظهر أعراضاً سريريةً للمتلازمة العضلية غير الاعتيادية (الشاذة)

٤.٢، ٣.٣ ، ٧.٠ ، ٥.١ ، ٣.٤ ، ٢.٥ ، ٨.٦ ،. ٣.٥ ، ٢.٩ ، ٤.٩ ، ٥.٤،

س٣- احسب الوسط الحسابي و الوسيط لقاعدة البيانات الآتية لأوزان فئران تجارب:
٥٤.١ ، ٤٩.٨ ، ٤٦.٠ ، ٤٤.١ ، ٣٤.٠ ، ٥٢.٦ ، ٥٤.٤ ، ٥٦.١ ، ٥٢.١ ، ٥١.٩ ، ٥٤.٠ ، ٥٨.٠ ، ٣٢.٧ ، ٥٨.٥ .

س٤- احسب الوسط الحسابي و الوسيط لكل من بيانات إنتاج الحليب اليومي خلال شهر كامل لإحدى مزارع إنتاج الحليب في سوريا في البيانات المدرجة أدناه

رقم المزرعة	إنتاج الحليب اليومي (كغ)
١	٢٠
٢	١٣
٣	١٤
٤	١٥
٥	١٦
٦	١٧
٧	١٨
٨	١٩
٩	٢٠
١٠	٢١
١١	١٢
١٢	١٤
١٣	١٤
١٤	١٥
١٥	١٦
١٦	١٧
١٧	٢٠

١٨	٢١
١٩	٢٥
٢٠	٢٣
٢١	١٠
٢٢	١١
٢٣	٢٠
٢٤	٢٠
٢٥	٢٠
٢٦	٢١
٢٧	٢٣
٢٨	٢٤
٢٩	٢٧
٣٠	٢٤
٣١	٢١

س٥- احسب المدى المرجعي لكل من قيم معايير الأضداد لمرض الإسهال الفيروسي في إحدى الدراسات الوبائية في البيانات المرجة أدناه.

٣٣٣٤	٤٥٥٧
٥٥٥	٨٨٩
٥٦٧٨	٦٧٨٩
٣٤٥٦	٨٨٩
٧٧٧	٥٦٧
٨٩٠	٥٦٦
٩٠٠	٥٦٦٧

١٠٠٠	٥٤٦٦
٢٠٠٠	٦٧٧٨
٣٠٠٠	٥٦٧
٢٥٠٠	٤٥٦٦
٤٠٠٠	٦٧٨
٦٠٠٠	٥٦٧
٥٦٧٨	٥٦٧
٧٨٩٦	٨٩٩
٥٦٧٨	٥٦٧
٦٧٨٩	٥٦٧٧
٦٧٨٩	٥٦٧٧
٤٤٥٦	٧٨٩٩
٦٦٦٥	٥٦٧
٥٦٧٨	٣٤٥
٥٦٧٨	٦٧٨
١٠٠٠	٦٧٨
١٢٢٣	٥٦٧
١٢٣٤٥	٥٠٠
٦٧٨	٨٩٠
٧٨٩	٦٧٨
٦٧٨	٧٨٩
٥٦٧	٥٦٧
٦٧٨٩	٦٨٩

الفصل الخامس

CHAPTER FIVE

الاحتمالية والتوزيعات الاحتمالية

PROBABILITY AND PROBABILITY DISTRIBUTIONS

العلاقة الوثيقة للاحتمالية بعلم الإحصاء

The Relevance of Probability to Statistics

بعد أن ناقشنا في الفصول السابقة العمليات التي يرجع إليها تلخيص البيانات وعرض النتائج المحصول عليها من مجموعات من الحيوانات ، وهذه الطرائق معروفة كما ذكرنا بالإحصاء الوصفي Descriptive Statistics وتعتبر خطوة مهمة وتعطى الأولوية في أي تحليل ، وعادة ما يراد الحصول على النتائج من عينة ممثلة لقطعان من حيوانات كبيرة العدد ، وهنا يراد الحصول على استنتاجات من هذه القطعان باستخدام بيانات العينة .

لنفترض على سبيل المثال أن الوسط الحسابي والانحراف المعياري لتركيز الحديد في مصل الدم في عينة عشوائية في قطيع من الأبقار مكونٍ من 59 رأساً كان على التوالي 27.64 ميكرومول/ لتر. و6.36 ميكرومول/لتر (M mol /L) نسبياً فمن غير المحتمل أن تكون هذه النتائج المحصول عليها في هذه العينة هي نتائج تعريفية لجميع المشاهدات في هذا القطيع . وعموماً يراد غالباً استخدام هذه المعلومات للحصول على استنتاج لشيء ما في هذا القطيع . إن هناك شكاً ما ومجال البت فيه مصاحب لهذه الاستنتاجات المنوطة بهذا القطيع ، هذا الشك في الحساب يمكن تقديره بواسطة ما يدعى بالاحتمالية Probability والتي هي أساس

الاستنتاج الإحصائي Fundamental statistical inference والذي يقدم ربطاً بين نتائج العينة والمجتمع الحيواني المأخوذة منه .

وسنناقش مفاهيم الإحصاء الاستنتاجي Inferential statistic في الفصول اللاحقة عندما نعرض طرق أخذ العينات وتوزيعات العينات أما هنا فسيعرض هذا الفصل ويناقش مفاهيم الاحتمالية بشيء من التفصيل .

مدخل عن الاحتمالية Introduction on Probability

تعني كلمة الاحتمال لغوياً إمكانية حدوث حادث معين ، وهناك أمثلة عديدة لفهم معنى الاحتمالية ، على سبيل المثال إذا رمينا قطعة من النقود مرصعاً على إحدى حوافها ذيل حصان وعلى الجانب الآخر صورة لملكة ما تاريخياً فيقال إن هناك احتمالية أو فرصة لظهور صورة الملكة هي ½ أو 0.5 أي 50 % . أو يقال مثلاً هناك احتمال لحدوث فيضان إذا كان مطرٌ غزيرٌ . إلا أن مثل هذا التعبير بمعناه اللغوي لا يعطينا فكرة عن مدى إمكانية حدوث هذا الاحتمال وهذا يشكل أحد الاختلافات الجوهرية بين المفهوم اللغوي للاحتمال والمفهوم الرياضي أو الإحصائي له . فالاحتمال في علم الإحصاء يمثل قيمة رقمية تعبر عن إمكانية حدوث الحادث وتقيس شدة إمكانية هذا الحدث .

إن العرض الأكثر شمولاً لتفسير الاحتمالية في الاستنتاج الإحصائي هو أن تنسب الاحتمالية إلى نسب مئوية لعدد مرات حدوث حدث ما خلال تكرار هذا العدد مرات عديدة في أثناء تجربة أو تجارب عديدة أنجزت بنفس الظروف .

إن نتيجة أي من هذه التجارب أو المحاولات يجب أن تكون مستقلة عن التجربة أو المحاولات الأخرى . على سبيل المثال إذا كان الاهتمام بتقدير احتمالية أعداد حيوانات في أكثر من ثلاث حظائر تحتوي على قطعان من الأغنام فمن المنطقي أن يتم تعداد للحيوانات في هذه الحظائر خلال فترة طويلة من الزمن ولنفترض عاماً كاملاً ومن ثم تقسم هذه الأعداد على العدد الإجمالي لهذه الحظائر. وهذا ما يدعى بالتعريف التكراري للاحتمالية لأنه يعنى بعد تكرارات حادثة ما في عدد كبير من هذه التكرارات لتجارب مشابهة وبهذا تعرف الاحتمالية بأنها تكرار نسبي لحدث ما في تجارب مكررة وضمن ظروف أو شروط متشابهة .

التعريف التقليدي للاحتمال

The Traditional Definition of Probability

نظرياً ، يمكن أن تعرف الاحتمالية P (A) بعدد الحالات النسبي الملائمة في التجربة إلى عدد الحالات الممكنة كما يأتي :

$$P(A) = \frac{m}{n}$$

بإعتبار :

P (A) : الاحتمالية لتحقيق حدث ما .

m : عدد الحالات الملائمة المراد مشاهدتها .

n : عدد الحالات الممكنة الحدوث والتي يرمز لها عادة بالرمز Ω .

مثال :

إذا القي حجر نرد متجانس هو A بإعتبار أن A تمثل أي عدد فردي وهكذا على سبيل المثال m = 4 أربع حالات ملائمة [1 , 4 , 7] n = 8 , A = ثماني حالات ممكنة الحدوث هي (١ و ٢ و ٣ و ٤ و ٥ و ٦ و ٧ و ٨) = Ω وبالتالي احتمالية A هي :

$$P(A) = \frac{m}{n} = \frac{4}{8} = 0.5$$

التعريف الإحصائي للاحتمالية

The Statistical Definition of Probability

فهنا يميل الإحصائيون إلى الاعتماد على المفهوم التطبيقي للتكرار النسبي فإذا رمينا قطعة نقود لا نهتم هنا بكون القطعة متجانسة أو غير متجانسة تكرر ظهور الملكة في قطعة النقود 65 مرة = m من أصل n = 500 رمية فيكون التكرار النسبي لظهور الصورة هو $\frac{65}{500}$ أي $\frac{m}{n}$ وهكذا يعرف الاحتمال إحصائياً لتحقيق حدث A بقيمة نهاية التكرار النسبي عندما ينتهي عدد المحاولات n إلى اللانهاية .

$$P(A) = \frac{Lim \frac{m}{n}}{n \to \infty}$$

خصائص الاحتمالية Properties of a Probability

من الواضح أن الاحتمالية تعرف بأنها تكرار نسبي أو نسبة مئوية معينة لتكرار وبالتالي فإن قيمتها الرقمية يجب أن تتراوح ما بين الصفر (0) والواحد (1) بإعتبار أن قيمة الاحتمالية (0) تعني أن الحدث لا يمكن أن يحدث بينما قيمة الاحتمالية (1) تعني أن الحدث يمكن حدوثه وغالباً ما تحول قيم الاحتمالية إلى نسبة مئوية (وتتراوح ما بين الصفر و 100 %) أو يعبر عنها بتناسب (مثلاً يحدث الحدث مرة واحدة كل ثلاث مرات من تكرار المشاهدات) .

في بعض الأحيان لا نعير الاهتمام لحدوث حدث ما إلا أننا نركز على الحدث الذي لا يحدث عادة . وهذا يتبع خصائص الاحتمالية التي يعبر عنها بأن احتمالية حدث لا يحدث عادة هي (1 – احتمالية حدوث الحدث) ولذلك احتمالية إصابة هرة (قطة صغيرة) بالتهاب الحنجرة والقصبات الفيروسي بعد التحصين بفترة من (9) إلى (13) أسبوعاً من العمر هو 0.04 (في مكان محدد وزمن ما) فتكون الاحتمالية لتحقيق الحماية الكافية من هذا المرض هي 1 – 0.04 = 0.96 .

قوانين الاحتمالية Rules of Probability

هناك قانونان أساسيان يحكمان ويسيطران على الاحتمالات وهما :

١- قانون جمع الإحتمالات Additional rule

عندما يكون هناك حدثان متبادلا التأثير استثنائياً يكون حدوث كلا الحدثين في نفس الوقت أمراً غير منطقي ولا يمكن حدوثه ، وعندئذ تكون الاحتمالية لحدوث واحد من الحدثين هي حاصل مجموع الاحتمالات لكل حدث .

مثال :

لنفترض أنه يوجد في صندوق كرتوني عدد متساو من كلاب ذات ألوان وأشكال مائلة إلى اللون الأسود (5 أشكال مختلفة) فإن احتمالية أخذ أشكال مرصعة باللون الفاتح أو متباينة الشكل من الصندوق الكرتوني هو حاصل مجموع احتمالية الكلاب المرصعة باللون الفاتح (5/1) واحتمالية الأشكال المتباينة اللون (5/1) وهذا يعني :

$$1/5 + 1/5 = \frac{2}{5} = 0.4$$

احتمالية أخذ أحد الشكلين .

٢- قانون ضرب الإحتمالات Multiplication rule

عندما يكون هناك حدثان مستقلان في حدوثهما ، تكون الاحتمالية لحدوث كلا الحدثين هي حاصل حدوث الاحتمالات الفردية في كل منهما .

مثال :

إذا تم التلقيح الاصطناعي لبقرة من سلالة الفريزيان هو لشتاين في يوم محدد أُكد وجود تشخيص إيجابي للحمل فإن احتمالية هذه البقرة لتلد خلال 278 يوم لاحق (وهو الوسط الحسابي لفترة الحمل) هو 0.5 (50 %) فإذا لقحت بقرتان من سلالة الفريزيان اصطناعياً في نفس اليوم (وأُثبت الحمل الإيجابي في كلتيهما) فإن الاحتمالية لولادة كلتيهما خلال فترة 278 يوم لاحق **هي** :

٠.٢٥ = ٠.٥ × ٠.٥ أي جداء احتمال حدوثهما .

التوزيعات الاحتمالية Probability Distribution

المتغيرات العشوائية Random Variables :

يعرف المتغير العشوائي بأنه المتغير الذي يمكنه أن يأخذ أية قيمة من مجموعة مقادير أو قيم من نتائج تجربة لا يمكن أن تعرف نتيجتها قبل إجرائها وحيث يرتبط تحقق تلك القيم باحتمالات معينة .

وبهذا نجد أن العشوائية في مفهوم الاحتمالات تعرف بأنها عدم المعرفة المسبقة لنتيجة المحاولة أو التجربة المجراة وذلك مع أن الحصول على أية نتيجة للتجربة يرتبط ارتباطاً ما باحتمال معين . ومن أمثلة العشوائية كما ذكرنا سابقاً رمي حجر النرد ونتيجة الرقم الذي سيظهر عقب التجربة (رمي الحجر)، لن تُعرف مسبقاً ، فنقول عندئذ إن المتغير هنا هو متغير عشوائي ، وتمثل مجموعة قيمة من خلال

جمـــع النتـــائج التـــي نحصـــل عليهـــا مـــن جـــرّاء محاولـــة رمـــي الحجـــر (الأرقام من 1-6) .

وبهذا يتكون التوزيع الاحتمالي من كافة القيم التي يأخذها المتغير العشوائي مع إمكانية كافة الاحتمالات المرافقة له . وهناك أعداد كبيرة من التوزيعات الاحتمالية يمكن أن نميزها بواسطة ما يسمى بالمتغير العشوائي هذا وهناك نوعان من المتغيرات العشوائية وهما :

١- المتغيرات العشوائية المنقطعة أو المجزاة أو المقسمة (المنفصلة) : Discreet random variables

والتي تأخذ قيماً محددة في قاعدة البيانات أو مجاميع قابلة للعد ففي المثال المذكور أعلاه في رمي حجر نجد أنه في حال تغير المتغير العشوائي هو الأعداد الطبيعية الصحيحة الموجبة من 1 إلى 6 ، نجد أن هذا المتغير منقطع لأنه لا يمكن أن يأخذ جميع القيم الحقيقية المحصورة ضمن حدي المجال مثل القيم 0.25 أو 2.5 أو 4.2 وغيرها .

٢- المتغيرات العشوائية المستمرة (القيم المتصلة) : Continuous variables

وهي التي تأخذ قيماً غير محدودة ضمن قاعدة البيانات في مدى القيم كما فصلنا ذلك في الفصل الأول . إن المتغير العشوائي ذا القيم المستمرة يمكن أن يأخذ أية قيمة منتمية إلى مجال تغير المتحول العشوائي . كما هو الحال في قياس معدل

إنتاج حليب في مزرعة أبقار الحلوب اليومي من ١٨ – ٢١ كغ فالمتغير هنا يمكن أن يأخذ أي قيمة بين القيمتين ١٨ / و / ٢١ كالقيمة ١٦.٥ / ١١٩.٢/ .

التوزيع الاحتمالي Probability Distribution

التوزيعات الاحتمالية المنفصلة / غير المستمرة

Discrete Probability Distributions

يعرف التوزيع الاحتمالي المنفصل بأنه يربط الاحتمالية لكل متغير له علاقة تبادلية مع كل متغير يعرف كمتغير عشوائي منفصل علماً أن مجموع هذه الاحتمالات هو الرقم واحد كوحدة كاملة .

ومثال المتغير العشوائي المنفصل ببساطة نجده في نظرية ماندلين الوراثية Mendelion inheritance حيث لدينا شفعان من المورثات يتمثل بالمورث T وهو المورث المسيطر والمورث t وهو المورث المتلقي . ففي القطط من السلالة العديمة الذيل (Manx) تعطي المورثات المتلقية مواليد تختلف عن الأبوين وهو المورث ذو الرمز t وهو يترافق مع حالة انعدام الذيل لكن اجتماع مورثين متلقيين متجانسين يعطي مواليد من المحتمل أن تنفق لاحقاً ولن يكون لهذه الأجنة فرصة التطور ولنرمز لهذين المورثين بالرمز (t t) . بينما ينتج عن وجود مورثين مختلفين من النوع T t ولادةُ قطط من سلالة منعدمة الذيل ويعطي اجتماع المورثين T T المتجانسين مواليد قطط طبيعية مع وجود ذيل لها .

عندما نزاوج أبوين من القطط من سلالة منعدمة الذيل يكون لدينا أربعة احتمالات وراثية : T T و T t و t T و t t كما هو موضح في الجدول رقم (4) .

الجدول رقم (4) المميزات الوراثية للقطط من أبوين عديمي الذيل (Tt) .

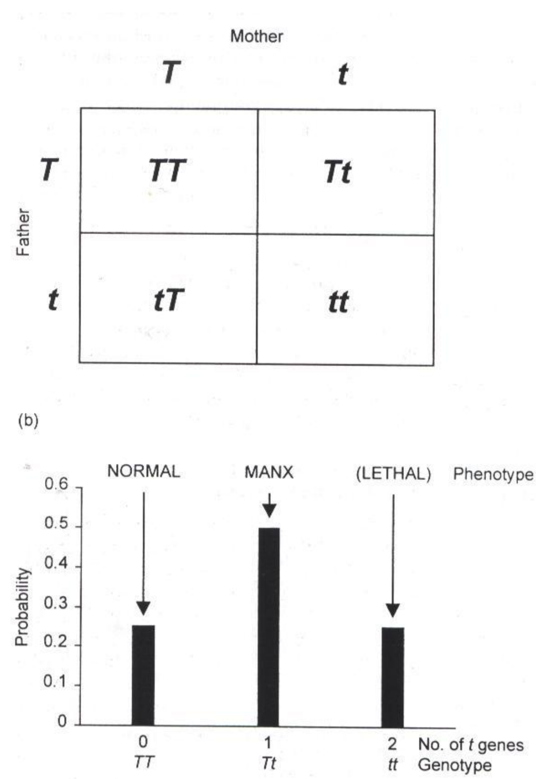
	الأم			
t		T		
T t سلالة منعدمة الذيل		T T سلالة طبيعية	T	
			الأب	
t t السلالة المميتة		t T السلالة المنعدمة الذيل	T	

ولنمثل هذه الحالة في المخطط البياني فيوجد هناك ثلاثة احتمالات مترافقة مع هذه الظاهرة فنجد في محور السينات نتائج ثلاث احتمالات (T T , T t , tt) وهي تمثل الحالات الطبيعية والسلالة المنعدمة الذيل والسلالة المميتة وهذا يعرف بالمتغير العشوائي وكل احتمالية تمثل برسم مستطيل وإن مجموع الاحتمالات الثلاثة المرتبطة بالنتائج تمثل وحدة كاملة أي تساوي الواحد وهي على الشكل الآتي:

١= ٠.٢٥ + ٠.٥٠ + ٠.٢٥

وهذا يوضح في الشكل رقم (٩) .

الشكل رقم (٩): احتمالات الحالات الطبيعية والسلالة المنعدمة الذيل والسلالة المميتة



هناك العديد من أنماط التوزيعات الاحتمالية المنفصلة والمختلفة عن بعضها تستخدم على نحو خاص وتتناسب باستخدامها مع العلوم الحيوية وهي التوزيعات الاسمية الثنائية الحدين The Binomial Distribution وتوزيعات بويسون Poisson Distribution وكما أشرنا سابقاً في هذا الفصل تكون هذه التوزيعات المنفصلة عموماً مقربةً للتوزيعات المستمرة .

التوزيع ثنائي الحدين : The Binomial Distribution

إن التوزيع ثنائي الحدين يرتبط في هذه الحالة بنوع من المتغيرات ذات النوع الثنائي التقسيم للاستجابة الحيوية وهو ما يدعى بالمتغير الثنائي الحدين أو ما يدعى

بمتغير بينري Binary . حيث توجد نتيجتان ممكنتان سنسميها عبارة تجريبية (trial) . على سبيل المثال ، نقول حيوان حامل أو حيوان غير حامل وفي حالة أخرى نقول مثلاً إن هذا الحيوان تظهر عليه الأعراض السريرية لمرض معدٍ أو لا تظهر عليه هذه الأعراض .

ومن هنا يمكننا أن نعرف التوزيع الاسمي الثنائي الحدين بأنه المتغير العشوائي والذي يمثل عدد النجاحات (نرمز في الإحصاء الحيوي بكلمة نجاح للتجربة الممثلة لحالة صحية للحيوانات كالمرض وكذا في حالة حدوث حدث ما لحيوان معين كالحمل) في سلسلة من المشاهدات عددها (n) في تجارب مستقلة عن بعضها وفي كل تجربة يمكن أن تنجح (مع احتمالية π) أو تفشل (وعبارة الفشل ترمز في علم الإحصاء الحيوي ليرمز لحالة الصحة أي عدم وجود المرض أو عدم حدوث حدث ما لحيوان معين كعدم وجود الحمل) (مع احتمالية $\pi - 1$) . ونظرياً هناك سلسلة من المشاهدات عددها (n) من النتائج في هذه الحالة .

لما كانت هناك إمكانية لتشاهد نجاحاً واحداً أو اثنين أو ثلاثة أو حتى عدداً من النجاحات n في تجارب عددها n . فيرتبط التوزيع الاسمي الثنائي الحد باحتمالية كل نتيجة . وهذا يعني أن الفرق أو التباين π n (في حالة النجاح) (أي حدوث

حدث كالمرض أو الحمل) والاحتمالية $(1 - \pi)$ π n (حالة الفشل: أي عدم حدوث الحدث كالصحة (أي عدم وجود المرض وانعدام الحمل) نسبياً .

ويعتبر التوزيع الثنائي الحد هاماً كثيراً في الإحصاء الحيوي لأنه يؤثر يلعب دوراً في تحليل النسب المئوية للأحداث الحادثة . ومثالُ ذلك دراسة مقارنة مستويات الأضداد للحيوانات المصابة بداء الدوران Leptospira والمجهضة والمشخصة مخبرياً إيجابياً لهذا المرض ومستويات الأضداد في الحيوانات الطبيعية وغير المصابة بهذا المرض فنجد مثلاً مستوى الأضداد في تلك الحيوانات المجهضة أعلى معنوياً من تلك الحيوانات الطبيعية .

توزيع بواسون The Poisson Distribution

وهو توزيع احتمالي منفصل آخر يستخدم في علوم الحيوان والطب و الطب البيطري ويعرف توزيع بواسون بأنه يتضمن متغيرات عشوائية يتميز فيها توزيع بواسون بتمثله على شكل حساب مجموع أعداد من الحوادث الحادثة عشوائياً على نحو مستقل في الزمان والمكان وبمعدل ثابت ويكون الوسط الحسابي والتباين في هذا التوزيع مساوياً معدلاً ثابتاً .

وعلى سبيل المثال نستخدم توزيع بواسون لربط احتمالات لمجموع من الأعداد المرمزة لحوادث معينة (كالأمراض مثلاوغيرها من الأعداد الكمية) كأن نقول مثلاً ٥٥٥٠ من الحوادث الصغيرة والمأخوذة عينةً من حوادث الإصابات الإشعاعية

والمعبر عنها حسابياً بتعداد الإصابات الإشعاعية في زمن معين أو قد يكون هذا المجموع الحسابي أعداداً فنقول (٣٥) من الخلايا الدموية لحجم محلول في العينة أو أعداد البيوض الطفيلية أو أعداد النباتات السامة أو أعداد الحالات المرضية لكل حيوان لمرض ما في زمن معين ولنأخذ مثالاً لتوزيع بواسون أعدادَ حالات التهاب الضرع خلال ستة أشهر و الممثلة في الجدول رقم (٥) .

جدول رقم (٥): حالات التهاب الضرع

عدد حالات التهاب الضرع	أرقام الحيوانات
٢	(١)
١	(٢)
٥	(٣)
٢	(٤)
٣	(٥)
٢	(٦)
٣	(٧)
٤	(٨)
٥	(٩)
٦	(١٠)

التوزيعات الاحتمالية المستمرة (التوزيعات الطبيعية)

Continuous Probability Distributions

(Normal Distributions)

العلاقة بين التوزيعات الاحتمالية المنفصلة والمستمرة :

Relationship between Discrete and Continuous Probability Distributions

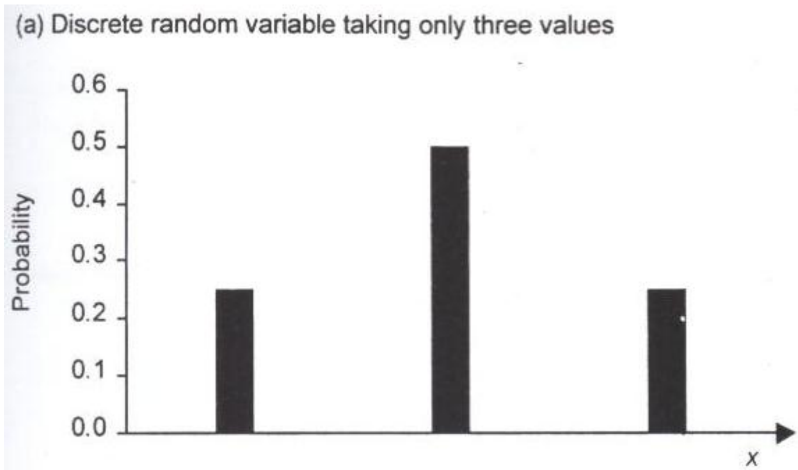
لكي نفهم هذه العلاقة بين التوزيعات الاحتمالية المستمرة والمنفصلة نستعين بهذه الأشكال الموضحة :

يشار إلى الشكل **a – ١٠** كمثال للتوزيعات الاحتمالية المنفصلة .

حيث يأخذ في هذا الشكل a – ١٠ : المتغير العشوائي المنفصل يأخذ ثلاث قيم

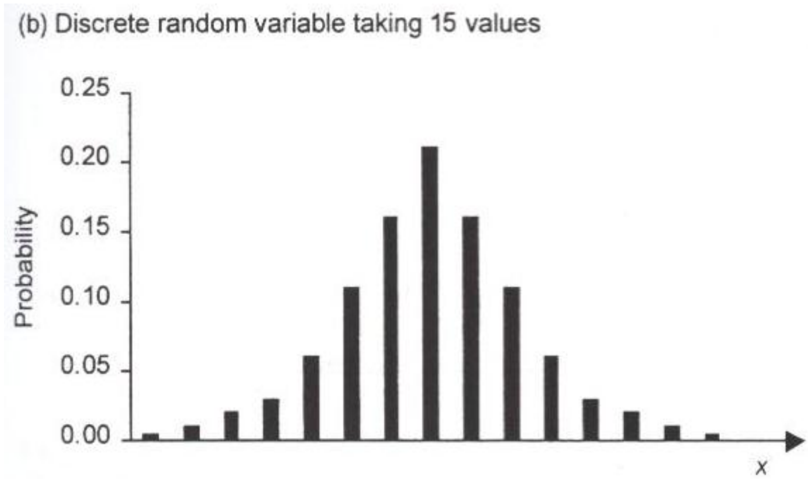
(٠.٢٥ و ٠.٢٥ و ٠.٥٠)

الشكل رقم (١٠ a): المتغير العشوائي غير المستمر يأخذ ثلاث قيم



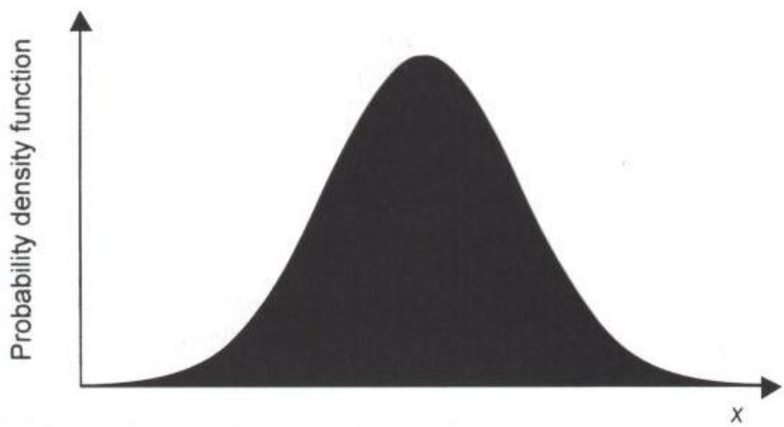
وببساطة شديدة إن كافة الحوادث الممكنة تمثل أفقياً على المحور السيني بينما تتوضع على المحور العمودي العيني الاحتمالية لكل حدث كما أن مجموع الأطوال لمجموع الاحتمالات على هذا المحور يجب أن تساوي الواحد أيضاً ويشير الشكل b . 10 إلى توضيح للتوزيعات الاحتمالية المنفصلة لمجموعة كبيرة من البيانات لكن أعدادها محدودة وهنا أيضاً يمثل قيماً غير مستمرة لأعداد الحوادث الممكنة للمتغيرات العشوائية .

شكل 10.b المتغير العشوائي المنفصل يأخذ عدة قيم



بينما يشير الشكل 10.C إلى تمثيل التوزيع الاحتمالي لمتغيرٍ عشوائي مستمر . وبمقارنة ذلك بالتوزيع الاحتمالي المنفصل فإن المتغير هنا يأخذ قيم أعداد غير محدودة (لا نهائية) ولذلك فمن المستحيل أن تمثل بخطوط منفصلة عن بعضها وتمثل المنطقة المظللة هنا إجمالي مجموع الاحتمالية للوحدة الممثلة ويدعى المنحى الذي يحدد ويعرف المنطقة بقانون الكثافة الاحتمالية probability density Function والتي توصف من خلال هذا القانون الذي سيرد ذكره بعد قليل.

الشكل 10.C : قانون الكثافة الاحتمالية لمتغير عشوائي مستمر



وإن مجموع إجمالي المنطقة المظللة (الشكل 10.C) تحت قانون الكثافة الاحتمالية هو مساوٍ القيمة واحد . وإن احتمالية المتغير العشوائي المستمر تمتد بين نقاط محددة معينة وتساوي تلك المنطقة الواقعة تحت تأثير قانون الكثافة الاحتمالية بين هذه الحدود وتحسب احتمالات هذه التوزيعات بواسطة قوانين الكثافة الاحتمالية والتي ستشرح لاحقاً باستخدام قانون مربع كاي (χ^2) واختبار t ستدنت وتوزيعات F من خلال قوانين محددة لهذه الاختبارات

التوزيع الطبيعي أو توزيع غازيان

The Normal (or Gassian) Distribution

إن التوزيع الطبيعي أو ما يدعى بتوزيع غازيان هو نسبة للباحث C . F . Gauss وهو باحث ألماني في القرن الثامن عشر وهو باحث رياضي يعتبر من أهم الأشخاص الذين وضعوا أسساً للتوزيع المستمر وهو قانونه في نظرية أخذ العينات والتي سوف تشرح في فصول لاحقة

إن عبارة طبيعي (Normal) لا تعني أن نطبق التوزيع الاحتمالي لمتغير عشوائي نموذجياً ومع ذلك فهو يعتبر تقريباً جيداً لهذا التوزيع في العديد من المتغيرات الحادثة طبيعياً أو تلك التي تمثل مجموعات حيوانية فردية غير مريضة ولنميز التوزيع الطبيعي من التفسيرات الطبيعية الأخرى سنستخدم حرفاً كبيراً (N) للإشارة إلى التوزيع الطبيعي في هذا الكتاب .

توضح الأشكال 11.a , 11.b . 11.C العلاقة بين المنطقة تحت قانون الكثافة الاحتمالية .

احتمالية (A = (X o < X <X1 .

احتمالية (B = (X < X 0

احتمالية (C = (X > X 1

يعتبر التوزيع الطبيعي من أهم التوزيعات الاحتمالية المستمرة التي تستخدم في معظم المجالات التطبيقية ومختلف فروع الطب والعلوم الاقتصادية والزراعية

والاجتماعية . وتتساوى في هذا التوزيع قيم الوسط الحسابي والوسيط والمنوال وبالتالي فإن مسقط قمة التوزيع على المحور السيني يقسم السطح المحصور بين منحني التوزيع والمحور السيني إلى قسمين متساويين تماماً ويعتبر التوزيع متناظراً حول مسقط قمته أي متناظراً حول الوسط الحسابي للتوزيع

إن التوزيع الطبيعي هو توزيع نظري حيث في الغالب نجد أن المشاهدات التي نشاهدها من خلال متغير ما في مجموعة من الأفراد لها توزيع تكراري تجريبي وتشابه التوزيع الطبيعي (N) وعندئذ نضع افتراضاً هو أن التوزيع لهذا المتغير في هذه المجتمعات الحيوانية هو توزيع طبيعي . وإذا كان هذا افتراضاً معقولاً فإنه يمكننا أن نستخدم خصائص التوزيع الطبيعي لتقييم الاحتمالات المطلوبة . ويتميز التوزيع الطبيعي بالخصائص التالية :

١- يوصف التوزيع الطبيعي كاملاً من خلال حدين اثنين وهما الوسط الحسابي والانحراف المعياري . يُرمز لهذين الحدين بالحرفين اليونانيين γ (للوسط الحسابي) و σ (الانحراف المعياري) والقانون الرياضي لقانون الكثافة الاحتمالية قد أهمل هنا فقط لتبسيط هذا الوصف .

٢- يشكل التوزيع الطبيعي وحدة نموذج كاملة

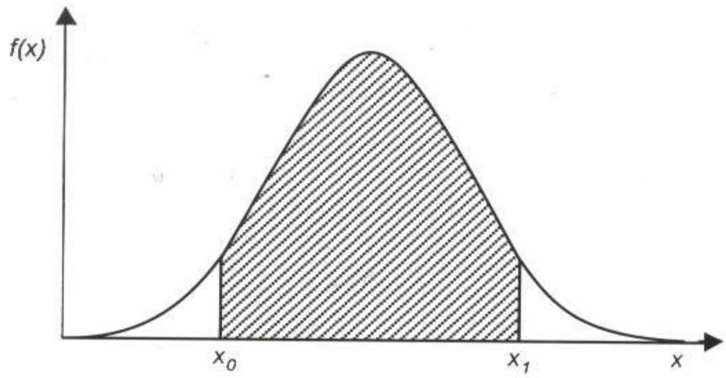
٣- التوزيع الطبيعي متناسق بنظام حول الوسط الحسابي وهذا يطبق في المنحني التوزيعي حيث إن الوسط الحسابي في الجزء اليميني من التوزيع هو خيال مرآة

للمنحني في الجزء اليساري منه وهو يصف هذا التوزيع بشكل الجرس Bell – Shaped .

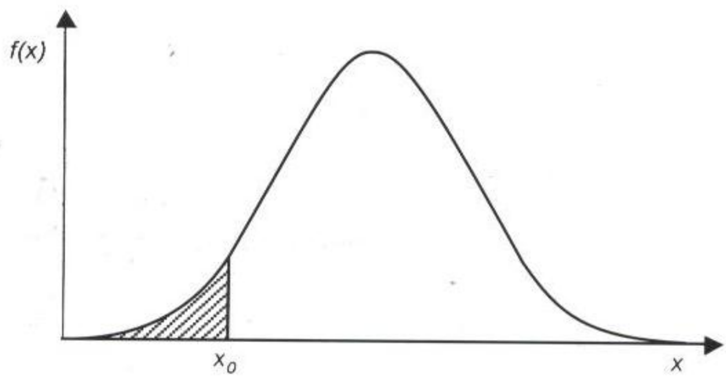
٤- وهذا يعني أنه في التوزيع الطبيعي يتساوى الوسط الحسابي والوسيط والمنوال

٥- إذا بقي الانحراف المعياري بدون أن يتغير فإن هناك زيادة في قيمة الوسط الحسابي تؤثر على المنحني التوزيعي أفقياً من جهة اليمين . وبالمقابل إن أي انخفاض في قيمة الوسط الحسابي تؤثر على المنحني أفقياً من الناحية اليسرى للتوزيع كما هو موضح في الشكل 12.a .

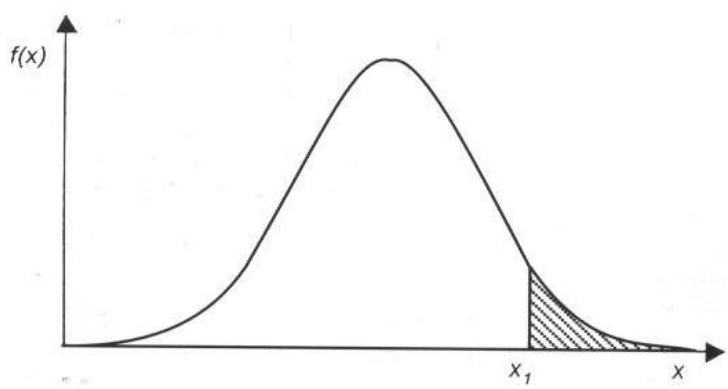
الشكل 12.a



(b)

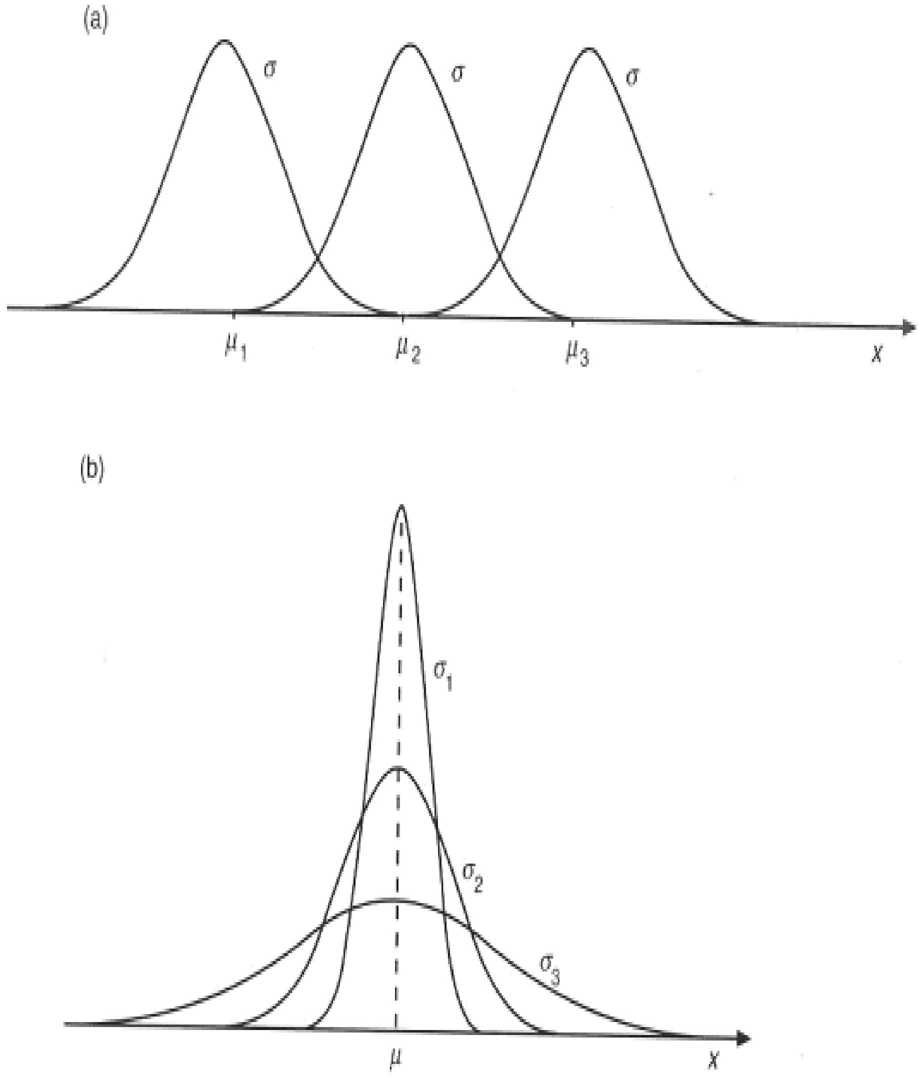


(c)



٦- كما أن انخفاض قيمة الانحراف المعياري للمنحني تجعل المنحني أكثر تضيقاً (ترقيقاً) أي يصبح ذا قمة طويلة وأكثر ارتفاعاً وبالمقابل إن زيادة الانحراف المعياري تجعل المنحني أكثر انفتاحاً (انفراجاً) أي ثخانة وأقصر وأكثر انبساطاً كما هو موضح في الشكل (13) .

الشكل (١٣ و ١٤) التأثير على التوزيع الطبيعي مع تغير في الوسط الحسابي والانحراف المعياري



٧- إن الحدود ما بين (ó −M) و (ó + M) تشمل ٦٨ % من التوزيع كما هو موضح في الشكل (١٤) .

٨- وتتضـمن الحـدود مـا بـين ($M - 1.96$ ó) و ($M + 1.96$ ó) ٩٥ % مـن التوزيع كما هو موضح في الشكل (١٣ و ١٤) وهذه الحقيقة غالباً ما تستخدم في حساب المدى المرجعي كما وضحنا في الفصول السابقة .

٩- كما أن الحدود ما بين ($M - 2.58$ ó) و ($M + 2.58$ ó) تشمل ٩٩ % من التوزيع (كما هو موضح في الشكل 13 و ١٤) .

وهكذا إذا أردنا معرفة عدد أو نسبة القيم المحصورة بين قيمتين X_1 و X_2 من قيم التوزيع يمكنك أن تبحث عن نسبة السطح المحصور بين X_1 و X_2 من السطح الإجمالي وذلك بإجراء تكامل تابع للكثافة الاحتمالي $\int_{x1}^{x2} f(x)dx$ لكن صعوبة هذا الحساب تقودنا إلى البحث عن طريقة أخرى لحساب السطح المحصور بين X_1 و X_2 ولذلك نستعين بتوزيع آخر يسهل التعامل معه يسمى التوزيع الطبيعي المعياري .

التوزيع الطبيعي المعياري

Standardized of Normal Distribution

إذا كان Z متغيراً عشوائياً يخضع للتوزيع الطبيعي المعياري فيكون الوسط الحسابي لهذا التوزيع صفراً $Z = 0$ وعندئذ فإن الانحراف المعياري لهذا يعادل القيمة +١ (ó = 1) وحيث إن الوسط الحسابي والانحراف المعياري لا يتغيران فإن هذا التوزيع يكون ثابتاً . كما أن هذا التوزيع طبيعي كما نوهنا به في التوزيع الطبيعي

بحيث يحصر من السطح تحت المنحنى قيمة من السطح الكلي لمجموع قيم التوزيع ويعطي تابع الكثافة الاحتمالي لهذا التوزيع على الصورة .

$$\int(Z) = \frac{1}{\sqrt{2\pi}} eZ^{2/2}$$

كل قيمة للمتغير العشوائي Z قيمة السطح لما قبل القيمة Z المحصور ما بين منحنى التوزيع والمحور الأفقي .

$$\int_{-\infty}^{Z} F(Z)dz = \int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} e^{-Z^{2/2}} dz$$

ويرفق في ملحق هذا الكتاب جدول للتوزيع الطبيعي المعياري والذي يعطي القيم للسطح ما بين Z الموجب والصفر . يتألف هذا الجدول من عدد من الأعمدة والصفوف خصص العمود الأول فيها للقيم Z والتي تحتوي الرقم الصحيح ورقماً عشرياً واحداً بعد الفاصلة بينما يخصص الصف الأول للرقم العشري الثاني (أجزاء المئة) من قيمة Z والتي تتراوح قيمتها في الجدول من −٣.٠٩ إلى ٣.٠٩ والتي تحصر عملياً كافة قيم التوزيع (جميع أجزاء السطح) .

مثلاً إذا أردنا معرفة قيمة السطح لما قبل Z=1064 فسنجد في المعادلة التالية:

$$\boxed{Z < 1.64 = 0.949\cdot}$$

وعادة ما نقوم بتحويل قيم التوزيع الطبيعي إلى توزيع طبيعي معياري باستخدام القانون :

$$X_1 \Rightarrow Z_1 = \frac{x_1 - M}{\sigma}$$

X_1 = متغير عشوائي (حيث كل قيمة X1 في التوزيع الطبيعي تقابلها قيمة Z1

μ= الوسط الحسابي .

التوزيعات الاحتمالية المستمرة الأخرى

يوجد العديد من التوزيعات الاحتمالية المستمرة تعتبر جزءاً من التوزيع الطبيعي. حيث إن هناك ثلاثة من هذه التوزيعات معروفة جداً وتعتبر مفيدة في الدراسات الإحصائية حسب نمط الدراسة وهي توزيع t ستودنت (توزيع t) وتوزيع مربع كاي (χ^2) وتوزيع F هذه التوزيعات سوف تُناقش وتُدرس اختبارات t و X^2 و F في الفصول القادمة إلا أننا سنتعرض للخصائص العامة لكل من هذه التوزيعات .

خصائص توزيع t ستودنت :

يعتبر توزيع t من التوزيعات الاحتمالية المستمرة ومن أهم استخداماته اختبار الفرضيات وقضايا التقدير عندما يتعلق موضوع الدراسة بعينات صغيرة أو عندما تكون قيمة σ^2 (الفرق) في منطقة الدراسة مجهولة فعندئذ نحل محله تباين العينة (فرق العينة) الذي يعطى بالقانون التالي :

$$t = \frac{\bar{x} - M}{s / \sqrt{n}}$$

بإعتبار أن :

n =حجم عينة مأخوذة من مجتمع حيواني ما .

M: الوسط الحسابي لمجتمع حيواني ما غير معروف انحرافه المعياري .

$\bar{x}$ = الوسط الحسابي للعينة .

S = الانحراف المعياري للعينة .

t = المتغير العشوائي لستودنت .

وفيما يأتي نورد أهم خصائص هذا التوزيع :

١- إن توزيع t هو توزيع متناسق ومتناظر حول الوسط الحسابي ولذلك فهو يأخذ شكل جرس الناقوس حيث له قمة وحيدة .

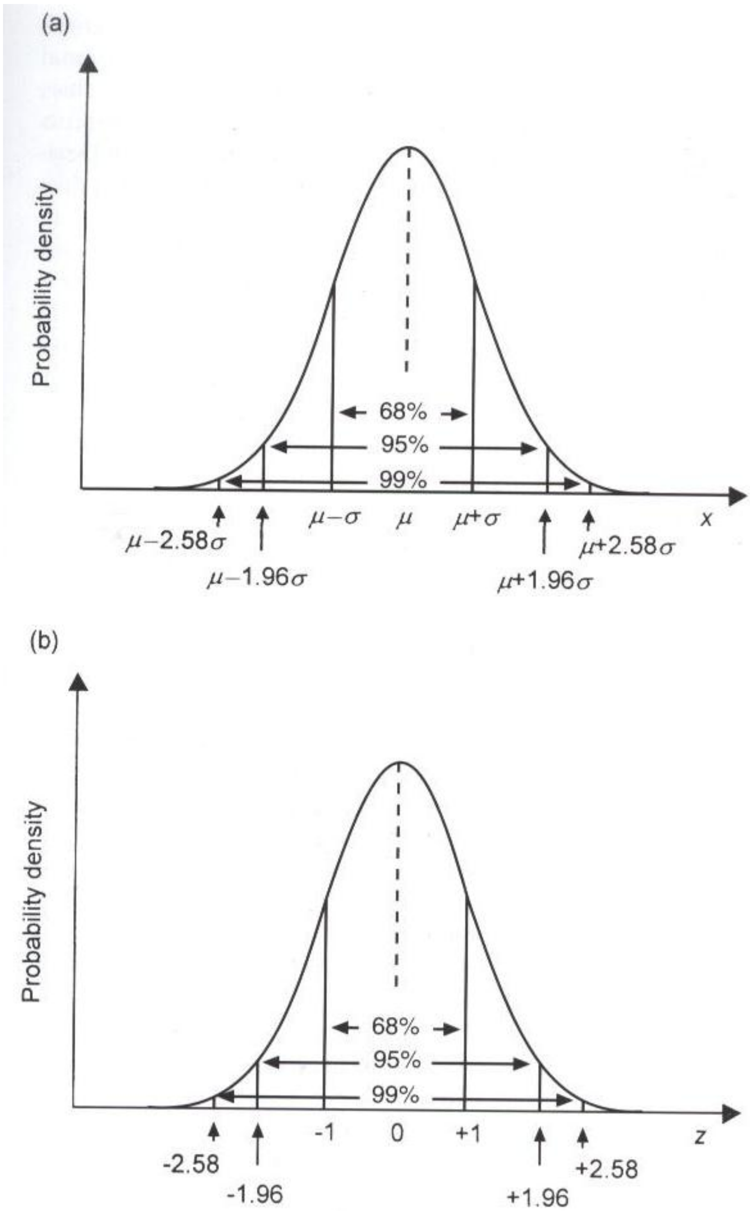
٢- يتميز خاصة بما يدعى بدرجات الحرية degrees of freedom (df) ومعرفتنا بدرجات الحرية هذه تسمح بحساب الاحتمالية لتوزيع t (سيدرس في الفصول القادمة بالتفصيل) .

٣- إن توزيع t هو توزيع لا يميز عن التوزيع الطبيعي المعياري عندما تكون درجات الحرية كبيرة ولكن كلما انخفضت درجات الحرية أصبح توزيع t أكثر فأكثر انتشاراً أو انبساطاً بمنحاه مقارنة بالتوزيع الطبيعي المعياري .

٤- فهو توزيع متناظر حول المستقيم العمودي $t = 0$.

٥- عندما يكون حجم العينة n يكون عدد درجات الحرية n-1 إلا أن هناك حالات تكون فيها عملية الحساب مختلفة بالنسبة إلى درجات الحرية وسنشير إلى ذلك في حينه .

الشكل (١٥) : شكل توزيع t مقارنة بالتوزيع الطبيعي



ولحساب المساحات تحت المنحني t هناك جدول صممه الإحصائيون خصيصاً لحساب المساحات تحت منحني t ستودنت وقد وضع في ملحق هذا الكتاب

. إذ يتم بإستخدامه إيجاد مساحات $\propto$ بعد معرفة قيمة درجات الحرية (df) حيث يتضمن جدول t قيم درجات الحرية وقيم المساحات $\propto$ المرتبة في الصف الأول بينما يشمل الجزء الباقي من الصفوف والأعمدة في الجدول المذكور قيم المتغير العشوائي t .

يرمز لقيم t داخل الجدول بالرمز $t(\propto . df)$.

وبمعرفة قيمة t ودرجات الحرية df تظهر مساحة $\propto$ المطلوبة في أعلى العمود الذي تظهر فيه قيمة t ونكتب $\propto = p[t < (\propto , df)]$

خصائص توزيع مربع كاي $(\chi 2)$

- يمكن أن يأخذ توزيع مربع كاي فقط القيم الإيجابية .
- يتميز أيضاً هذا التوزيع بما يسمى بدرجات الحرية ولذلك تسمح معرفتنا لها بتحديد الاحتمالات المناسبة تحت المنحني .
- كلما زادت درجة الحرية أصبح التوزيع أكثر فأكثر تناظراً من الجهتين وبالحقيقة يتميز بالطبيعية .

خصائص توزيع F

- إن توزيع F هو توزيع تناسبي (مئوي) .
- يتميز أيضاً بدرجات حرية منفصلة ذات جزئين وهذا يرتبط بصورة (بسط) ومخرج (مقام) التناسب المحدد .

مع أن هذا التناسب يمكن أن يأخذ قيمة أكبر أو أصغر من القيمة واحد فإن الاحتمالات المجدولة لتوزيع F ترتبط بالتناسب الذي يكون غالباً أكبر من القيمة (١) . فعلى سبيل المثال يمكن أن يكون البسط (الصورة) في التناسب أكبر من المقام (المخرج) . وهكذا فإن القيم المجدولة تشير فقط إلى الجزء العلوي المرتفع من التوزيع (اختبار F سوف يدرس تفصيلاً في الفصول القادمة)

العلاقات بين التوزيعات

Relationships between Distributions

التقريبات الطبيعية لتوزيعي بواسون والتوزيع ثنائي الحدين الاسمي

Normal approximations of the Binomial and Poisson Distributions

إن التوزيع ثنائي الحدين الاسمي وتوزيعات بواسن تكون منحرفة باتجاه الناحية اليمنى من المنحنى عندما تكون أحجام العينة صغيرة مع أنها تصبح أكثر تناظراً كلما ازدادت أحجام العينات . والحق أن كل توزيع بطريقته يمثل التوزيع الطبيعي بالنسبة لأحجام العينات الكبيرة بما يكفي عندما يكون المنحني الخطي يمثل قيماً احتمالية لقيم الاحتمالية لمتغيرات منفصلة لنأخذ بعين الاعتبار حالة التوزيع الاسمي ثنائي الحدين عندما نشاهد نسبة مئوية (P) عند حدوث حمل في تجارب عددها (n) ، فمن المعقول أن نستخدم التقريب الطبيعي للتوزيع الاسمي ثنائي الحدين إذا كانت كل من قيمة $n \times p$ والقيمة $n \times (1-p)$ هي أكبر من القيمة ٠.٥ وإن الوسط الحسابي والفرق لهذا التوزيع الطبيعي يقدر بالقيمة $(n \times p)$ والقيمة

$n \times p$ $(1-p)$ نسبياً . هذا التقريب يعتبر مفيداً خاصة في الاستنتاج الإحصائي لاختبار فرضية العدم عند حساب حدود الثقة للنسب المئوية (كما سنرى في الفصل التاسع) .

مثال : لنفترض أنه في يوم ما عدد من القطط (n = ١٨) أخذ إلى العيادة البيطرية لتشخيص إصابتها بالبراغيث flees وقد كانت النسبة المئوية للقطط المخموجة هي $6/18 = 0.33$ وعندئذ تكون القيمة : $np = 18 \times 0.33 = 6$

كما أن قيمة : $n(1-p) = 18 \times (1-0.33) = 18 \times 0.67 = 12$

وهكذا فإن التقريب الطبيعي للتوزيع هو مماثل ومناسب لكل جزء من المنحنى وعليه فإن الوسط الحسابي والفرق لهذا التوزيع الطبيعي يقدر بالقيمة :

$np = 18 \times 0.33 = 6$

$np\ (1-p) = 18 \times 0.33\ X\ 0.67 = 3.98$

وهكذا إذا أردنا أن نقيم احتمالية أن عشر قطط أو أكثر يتوقع بأن تصاب بعدوى البراغيث فإننا نحدد القيمة :

$$Z_1 = (10-6)\sqrt{3.98} = 2.01$$

ويشار إلى هذه القيمة في الجدول A1 كما هو مدرج في الملحق .

وبتقسيم الاحتمالية المجدولة بالرقم 2 لأننا مهتمون بدراسة فقط الجزء العلوي من المنحني من هذا التوزيع نجد أن الاحتمالية المطلوبة تقدر بالقيمة 0.02 (هذه القيمة تعتبر مقربة إلى فاصلتين عشريتين 0.0222) .

تمـــــــــارين

س١– بين ما إذا كانت البيانات في أ و ب موزعة طبيعيا و استنتج إذا كان بالإمكان إجراء تحويل لغارتيمي للبيانات لتحويلها إلى توزيع طبيعي تقريبي؟

أ– بيانات مأخوذة من قياس مستوى التتراسيكلين من عينات مأخوذة من العضلات لسمك السلمون (نمط سالمو سللور) حيث الصادات الحيوية أضيفت للمياه ولفترة عشرة أيام لأغراض علاجية: القياسات أخذت التراكيز الموجودة في مناطق العضلات (ملغ/غ) من النسيج العضلي و بعد ثمانية عشر يوماً من إضافة الصاد الحيوي لفحص مستوى الفعالية للجرعة (الجدول رقم ٦)

جدول رقم ٦ : تراكيز الصادات الحيوية الموجودة في مناطق العضلات (ملغ/غ) من النسيج العضلي و بعد ثمانية عشر يوماً من إضافة الصاد الحيوي لفحص مستوى الفعالية للجرعة

١.٨	٠.٥	١.٥	١.٦	١.٣
٢.١	٠.٠	١.٤	٢.٥	١.٩
٠.٨	٠.٧	٠.٣	٠.٤	٢.٢
١.٩	٠.٨	١.٢	٠.١	١.٢
٢.٤	٢.٥	٢.٥	١.٧	٠.٦

ب– البيانات الآتية في الجدول رقم (٧) مأخوذة من مستويات الفوسفات القاعدي في مصل دم لـ ١٢ كلباً بالغاً و بحالة طبيعية (وحدات دولية باللتر)

جدول رقم ٧ : مستويات الفوسفات القاعدي في مصل دم لـ ١٢ كلباً بالغاً وبحالة طبيعية (وحدات دولية باللتر)

٥.٤	٧.٣	٢٠.٣	١٧.٥	٣٥.٩	١٦.٨
٢٨.٦	٥٤.٣	١٠.٠	١٤.٠	١١.٧	٢٤.٣

الفصل السادس
CHAPTER SIX
المعاينة وتوزيع العينات
SAMPLING AND SAMPLES DISTRIBUTION

التميز بين مصطلح العينة والمجتمع الحيواني
The Distinction between the Sample and the Population

عندما نرغب بجمع معلومات من مجتمع ما لا يمكن – إلا ما ندر – أن نفحص كل عضو في المجتمع عند إجراء تقييم إحصائي وبدلاً من ذلك يمكن أن نأخذ مجموعة صغيرة عينةً مختارة من المجتمع ومن ثم نفحص أعضاء هذه المجموعة . وتستخدم النتائج لتقدير بعض مميزات هذا المجتمع .

وهكذا فإن الأهداف الرئيسة لأخذ العينات هي عبارة عن الحصول على البيانات التي تسمح لنا بإجراء الاستنتاجات المتعلقة بدراسة أعداد كبيرة من الحيوانات اعتماداً على فحص العينة ودراسة علاقتها على سبيل المثال بوجود أو عدم وجود المرض الحيواني . وربما تتعلق استنتاجات لإثبات أن المرض غير موجود في هذه المنطقة ، وبالعكس يمكن أن تهدف لإثبات وجود المرض أو تقييم مستوى حدوث المرض . إن الهدف يمكن أن يصف أيضاً إنتاج الحليب في قطعان الأبقار الحلوب والأكثر من ذلك

يمكن أن يقدم تحليلاً وصفياً لنظام الإنتاج الحيواني على سبيل المثال في بلد من البلدان الإفريقية .

الاستنتاج الإحصائي Statistical Inference

يتم الحصول على الاستنتاج الإحصائي من المجتمع الحيواني بأخذ العينة وهذا الاستنتاج يمكّننا من أن نرسم الاستنتاجات حول جوانب معينة من المجتمع الحيواني عندما تكون لهناك مجاميع كبيرة مقسمة إلى مجاميع أخرى أصغر عندئذ تكون فإن العينة ضرورية للتقصي عن حالة معينة فمن المهم جداً أن نعي أن التميز بين العينة والمجتمع الحيواني المأخوذ منه العينات كمكون أساسي للنظرية الإحصائية يتمثل بالاستنتاج الإحصائي هناك نقطتان أساسيتان في الاستنتاج الإحصائي تؤثران جداً في التحليل الإحصائي : وهما طريقة التقدير estimation واختبار النظرية hypothesis testing وسنناقش في هذا الفصل طريقة التقدير بينما يركز اختبار النظرية في قرارنا في اعتماد النتائج التي حصلنا عليها بواسطة العينات المأخوذة أو أن النتائج المحصول عليها لا يمكن اعتمادها نظراً لوجود بعض العوامل أو الأسباب المتداخلة مع المسبب المدروس وهذا ما سنناقشه في الفصل السادس .

تقدير حدود المجتمع الإحصائي باستخدام المعاينة الإحصائية

Estimation of Population Parameters Using the Statistical Sampling

إن الغرض من أخذ العينة هو التعرف على جوانب عدة للمجتمع الإحصائي المدروس وهذه الجوانب تتلخص بعبارة حدود المجتمع الإحصائي population parameters وهي التي تميز توزيع المجتمع الإحصائي المدروس ويمكن وصف توزيع بيانات المجتمع الإحصائي إذا ما عرفنا طبيعة القيم التي حصلنا عليها في هذه البيانات والتي ناقشناها في الفصول السابقة . ومن غير الممكن أن نحدد الوسط الحسابي للمجتمع الإحصائي تماماً عندما نختار فقط العينة من المشاهدات الملاحظة في مجتمع حيواني ما . فعلى سبيل المثال لا يمكننا أن نعرف القيمة الدقيقة لمعدل أعداد الأفراس المشاركة في سباق الخيول عندما تكون في حالة جريان السباق وإنما نعرف فقط النتائج المحصول عليها من العينة المختارة والشيء الأمثل الذي يمكن أن نقوم به هو تقدير هذه القيمة من المعاينة . فمثلاً نحسب العينة الإحصائية والتي لها قيمة قريبة من القيمة الحقيقية ما أمكن لحد المجتمع الحيواني إلا أن تقدير وسط العينة هو أفضل من طريقة تقدير وسط المجتمع الحيواني ففي حين أن وسط العينة هو عبارة عن مجموع أعداد كافة المشاهدات في العينة مقسوماً على أعداد المشاهدات في العينة يكون وسط المجتمع الحيواني هو مجموع كافة المشاهدات في المجتمع الحيواني مقسوماً على أعداد المشاهدات في هذه الحيوانات .

خطأ المعاينة Sampling Error :

ليس من الإمكانية بمكان أن تكون قيمة العينة الإحصائية مساوية تماماً لقيمة حد المجتمع الحيواني التي قدرناها . فيجب علينا أن نوصي أن هناك دائماً احتمالية بوجود خطأ في التقدير لأننا أخذنا العينات من حيوانات القطيع دون النظر إليها جميعها وهذا ما ندعوه بخطأ المعاينة Sampling error ونحتاج إلى تقييم الدقة للعينة الإحصائية كتقدير لحدود حيوانات القطيع ، ومن أجل هذا الغرض نحسب ما يدعى الخطأ المعياري للتقدير Standard error of the estimate سواء كان هذا التقدير المحسوب وسطاً كمعدل إنتاج الحليب في أبقار الفريزيان هولشتاين أو نسبة كنسبة الأبقار المعرضة لداء البريميات سواء كانت تلك الأبقار لديها أو ليس لديها معيار إيجابي للبريميات (وهنا يكون شكل المتغير بيزي أي أن نقول إيجابي أو سلبي) ويمكننا حساب الخطأ المعياري للوسط (والذي يرمز له بـ SEM) .

بالقانون : $\text{SEM} = \dfrac{s}{\sqrt{n}}$

بإعتبار أن القيمة s تحسب بـ : $s = \sqrt{\dfrac{(x-\bar{x})^2}{n-1}}$

بإعتبار أن : s : الانحراف المعياري ، n : عدد المشاهدات

x: قيمة المشاهدة ، $\bar{X}$: متوسط قيمة المشاهدة .

مثال : إن المواسم الإدرارية الأخيرة القياسية (بفترة 305 يوم) لعينة عشوائية من الأبقار عددها 256 بقرة من سلالة الفريزيان هولشتاين لأعداد مختلفة من مواسم

الإدرار قد أعطت تقريراً لمعدل إنتاج حليب قيمته 6414 كغ مع تقدير لانحراف معياري قيمته 2352 كغ ولذلك حسب القانون المدرج أعلاه يحسب الخطأ المعياري للوسط كما يأتي :

$$\frac{2352}{\sqrt{256}} = 147\,\text{Kg}$$

التفريق بين الانحراف المعياري والخطأ المعياري للوسط

The Distinction between the Standard Deviation and the Standard Error of the Mean

لقد ناقشنا في الفصل الثاني الانحراف المعياري SD وسنتناول في هذا الفصل الخطأ المعياري للوسط SEM . وهما قياسان لهما تطبيقات مختلفة عن بعضهما كثيراً ومن المهم هنا أن يكون لدينا فهم واضح للتمييز بين الانحراف المعياري للمشاهدات والخطأ المعياري للوسط وهذا التفريق يجنبنا سوء التفسير للبيانات المدروسة :

إن الانحراف المعياري هو قياس لتشتت المشاهدات . وهو يعطي مؤشراً إلى مدى تقارب المشاهدات لمعدلاتها ويمكن أن يستخدم لإنشاء حد مرجعي Reference Interval والذي يعرف بمدى التشتت للبيانات لقطعان من الحيوانات.

بينما الخطأ المعياري للوسط هو قياس دقة وسط العينة كتقدير لوسط المجتمع الحيواني. وهو يقيم خطأ العينة بإعطاء مؤشر إلى مدى التقارب بين وسط العينة ووسط حيوانات القطيع المأخوذة منه العينات ويستخدم الخطأ المعياري المعدل لإنشاء ما يدعى بحد الثقة Confidence Interval والذي يسمح لنا بأن نحكم على دقة تقديرنا لمعدل حيوانات القطيع المدروس .

حد الثقة للوسط الحسابي

The Confidence Interval for a Mean

مفهوم حدود الثقة Understanding Confidence Intervals

لقد ركزنا دائماً على أن توزيع العينات بالنسبة للوسط هو توزيع افتراضي Hypothetical Distribution . وفي ممارستنا العملية لا نأخذ عينات متكررة من الحيوانات المدروسة وإنما نأخذ عادة عينة واحدة فقط وتستخدم هذه العينة كتقدير لوسط القيم المأخوذة من الحيوانات المدروسة . وهذا يعطينا تصويراً لدرجة الدقة في تقديراتنا لهذه الحيوانات أو القطعان المدروسة . يجب أن ينوه هنا بأن استخدام عينة واحدة من كل حيوان فردي عند التحليل تجنبنا مشكلة أساسية في التحليل الإحصائي والتي تسمى بمشكلة القياسات المتكررة Repeated Measures وهذا يعطي نتائج الدراسة دقة أكثر وهناك العديد من البرامج الإحصائية الحاسوبية التي تجنبنا هذه

المشكلة مباشرة كبرنامج SAS، وغيره من البرامج الإحصائية وإذا كان لدينا قياسات متكررة فيجب أخذ معدل النتائج لدراسة مشكلة ما في هذا القطيع أو ذاك .

وإن أفضل طريقة لتقييم ما إذا كان تقديرنا جيداً هو أن تعرف مدى القيم ضمن معدل القيم الحقيقية للحيوانات المدروسة والتي تحدد ضمن احتمالية معينة ، هذا المدى للقيم يعرف بالحدود العليا والدنيا upper and lower limits (وهو ما يسمى بحدود الثقة) ويدعى حد الثقة بالنسبة للوسط Confidence Interval for the mean

١- إذا كان حد الثقة كبيراً فإن وسط العينة يكون عندها تقديراً غير دقيق بالنسبة لمعدل القطيع المدروس أو الحيوانات المدروسة .

٢- إذا كان حد الثقة صغيراً أو ضيق المدى فيكون عندئذ وسط العينة تقديراً جيداً فهو بشكل آخر ذو تقدير دقيق بالنسبة لمعدل الحيوانات المدروسة .

إن الثقة 95% تمثل أن معدل القيم في الحيوانات المدروسة تشمل ما يدعى بحد الثقة 95% بالنسبة للوسط ونحسب نموذجياً حد الثقة 95% لأي حد من الحدود المراد معرفة دقة الحساب فيها . لكننا يمكن أن نجد أحياناً حدي الثقة 95 %أو 99% مستخدمة في بعض الدراسات . إذ إن حد الثقة 99% هو بالتأكيد أكثر شمولاً ودقة من حد الثقة 95% لأنها أكثر ثقة للحد الذي يشمله ذاك أو هذا الحد . ويعتمد مدى اتساع حد الثقة على :

١- درجة الثقة المطلوبة .

٢- حجم العينـة (فـالحجم الكبيـر للعينـة يتطلـب تقديراً أكثـر دقـة وبالتـالي حد ثقـة بمجال ضيق) .

٣- تـوفر الخاصية المتعلقـة بالتقصـي عن الحالـة (فوجـود متغيرات أكثـر لعـدد من المشاهدات المدروسة يتطلب تقدير أقل دقة وبالتالي حد ثقة بمجال أوسع)

حساب حد الثقة بالنسبة للوسط الحسابي

Calculating the Confidence Interval for the Mean

إن القانون العام لحد الثقة بالنسبة للوسط يحسب كما يلي :

$$\text{CI } 95\% : \ \overline{X} \mp 1.96 \times \text{SEM}$$

بإعتبار أن :

$\overline{X}$: الوسط .

1.96 = ثابت .

SEM: الخطأ المعياري للوسط وبإعتبار أن : $\text{SEM} = \frac{Q}{\sqrt{n}}$

بإعتبار أن σ: الانحراف المعياري .

n: عدد المشاهدات .

فيصبح القانون :

$$\overline{X} \mp 1.96 \times \frac{Q}{\sqrt{n}}$$

$$\left(\overline{X} - 1.96 \times \frac{Q}{\sqrt{n}}, \overline{X} + 1.96 \times \frac{Q}{\sqrt{n}} \right)$$

ويتـراوح مـدى حـد الثقـة الأعلـى والأدنـى مـن خـلال الحصـول علـى القيم المحسوبة في القانون من القوسية المذكورة أعلاه P .

إن الثابت بالنسبة لحد الثقة 95% هو 1.96 (يقرب غالباً إلى القيمة 2)

بينما الثابت بالنسبة لحد الثقة 99% هو 2.58

والثابت بالنسبة لحد الثقة 90% هو 68

وعندما تكون قيمة الانحراف المعياري σ (الانحراف المعياري لقطيع الدراسة) مجهولة نستعيض عنه بالانحراف المعياري الخاص بالعينة والذي يحسب بالقانون الآتي :

$$S = \sqrt{\frac{\sum (X - \overline{X})^2}{n - 1}}$$

إن هناك قوانين أخرى لحساب حد الثقة ولا سيما بالنسبة لتوزيع t ستودنت ولا مجال لتفصيلها وشرحها هنا لأن القانون العام المذكور أعلاه يمكن تطبيقه عندما يكون توزيع القيم للمشاهدات توزيعاً طبيعياً .

توزيع المعاينة للنسبة :

The Sampling Distribution of the Proportion

إن مفهوم توزيع المعاينة للنسبة هو نفس ذلك التوزيع المتعلق بمعدلات العينات المدروسة فهو توزيع افتراضي لخصائص مشاهدات تكون فيها الفوائد مقدرة عندما نريد أن نحصل على استنتاج إحصائي للنسب المئوية لقطيع من القطعان أو مزرعة من المزارع أو مجموعات من حيوانات نريد دراستها .

ويمتاز توزيع المعاينة للنسبة بالخصائص الآتية :

١- يعتبر هذا التوزيع قريباً من التوزيع الطبيعي إذا كان حجم العينة كبيراً. وفي الواقع إن توزيع النسب المئوية هو في الحقيقة توزيع اسمي لكن هذا التوزيع كما شرحنا أقرب إلى التوزيع الطبيعي لعينة كبيرة عدد مشاهداتها n.

٢- إن معدل قيم توزيع المعاينة النسبي هو معدل النسبة لجميع الحيوانات المدروسة وتقدر بالقيمة π وهكذا فإن النسب المئوية للعينة p تحدد بعينة وحيدة مستفردة هو تقدير غير منحرف بالنسبة للنسبة المئوية لجميع الحيوانات المدروسة

٣- إن الخطأ المعياري لتوزيع العينات هو النسبة المئوية التي تحسب بالقانون :

$$\sqrt{\pi(1-n)/n}$$

وهو ما يدعى بالخطأ المعياري للنسبة المئوية وعبارة عن قياس دقة p كتقدير لقيمة π

وهذا الخطأ المعياري يحسب بقيم مشاهدات العينة كما يأتي :

$$SE(P) = \sqrt{\frac{P(1-P)}{n}}$$

بإعتبار أن قيمة P تحسب كما يأتي : $P = \frac{r}{n}$

n: حجم العينة . r = عدد المشاهدات في العينة .

ومع أنه يمكننا تقدير قيمة π بحساب قيمة p فإن توزيع العينة المئوية تبقى توزيعاً قريباً من التوزيع الطبيعي لعينة كبيرة الحجم عدد مشاهداتها n .

ويجب أن نلاحظ أننا إذا استبدلنا تقدير قيمة p بنسبة مئوية p% فإن تقدير الخطأ المعياري يحسب كنسبة مئوية بالقانون :

$$S\,E(P\%) = \sqrt{\frac{P\%(100 - P\%)}{n}}$$

حد الثقة للنسبة المئوية

The Confidence Interval for a Proportion

إن حد الثقة لنسبة حيوانات الدراسة π يحسب بإضافة وطرح النسبة المئوية للعينة (p) مضروباً بالخطأ المعياري (انظر الملحق A2) وهكذا فإن حد الثقة 95% لنسبة حيوانات الدراسة يمكن أن يقدر بالقانون الآتي :

$$P \mp 1.96 \times SE(P) =$$
$$\left\{P - 1.96\sqrt{\frac{P(1-P)}{n}}\,, P + 1.96\sqrt{\frac{P(1-P)}{n}}\right\}$$

ويجب أن نلاحظ أيضاً أننا يمكن أن نعدل القانون المذكور أعلاه إذا أردنا أن تكون القيمة المدروسة على شكل نسب مئوية وذلك باستبدال كل نسبة مئوية بالقيمة المئوية واستبدال القيمة (1) داخل كل جذر تربيعي بالقيمة 100 .

مثال : اختيرت عينة عشوائية مكونة من 115 بقرة من قطيع في مزرعة ما في إحدى المناطق . اختبرت عينات الدم المحصول عليها من أبقار هذه العينة للبحث عن وجود أضداد داء الدوران Leptospira وحسب معايير الأضداد المحصول عليها فقد صنفت إلى أضداد إيجابية وأخرى سلبية . ففي هذه العينة وجد أن هناك 36 بقرة تحمل معايير أضداد إيجابية لهذا المرض . إن تقدير النسبة المئوية للأبقار المعرضة لداء الدوران يمكن أن تحسب كما يأتي: 36115 = 0.31

وهكذا يمكن أن يقدر الخطأ المعياري في هذه النسبة كما يأتي :

$$SE_{(P)} = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{0.313(1-0.313)}{115}} = 0.043$$

وبالتالي يمكن أن نحسب حد الثقة لنسبة الحيوانات المعرضة للإصابة بداء الدوران كما يأتي :

$CI95\% = P \mp 1.96\ SE(P)$
$= (0.313 - 1.96 \times 0.043\ ,\ 0.313 + 1.96 \times 0.043)$
$= (0.228\ ,\ 0.398)$

وهكذا فإن ثقتنا 95% للنسبة الحقيقية للأبقار المعرضة للإصابة بداء الدوران تتراوح ما بين 0.23 إلى 0.40 .

الفصل السابع
CHAPTER SEVEN
التصميم التجريبي والتجارب السريرية
EXPERIMENTAL DESIGN AND CLINICAL TRIALS

أنماط الدراسة Types of Study :

إن دراسة الإحصاء في الطب البيطري وعلم الحيوان مرتبط بعلم الوبانيات الذي يقوم بدراسة أنماط المرض ومحدداتها المرضية في حيوانات المزرعة . وفي هذا الفصل سنتعرف بعضَ المفاهيم في علم الوبائيات ولمعرفة المزيد عن هذه المفاهيم يمكننا مراجعة مادة الوبائيات التي سوف تدرس في السنوات اللاحقة .

ثمة ندرة في توافر الحالات المرضية المتعلقة بالدراسات السريرية في حيوانات المزرعة ويمكن أن يعود السبب إلى أن مثل هذه الحالات قد تكون نادرة أو أن تكلفة الحيوانات التي سوف تستخدم في التجربة باهظة الثمن أو أن عامل الوقت يتدخل أحياناً أيضاً في تطبيق الممارسة الحقلية البيطرية الدائمة في دراسة مثل هذه الحالات إذ إن الجميع اليوم منشغل بتطوير صناعة الحيوان (تربية الحيوان) أكثر من انشغاله بالحفاظ على صحة هذا الحيوان .

ولكي نتمكن من تأمين المواد بوفرة كان من الضروري أن نصمم الدراسة بطريقة تخدم إنتاجية الحيوان والمزرعة ككل عموماً . حيث تتوفر هناك طرائق مختلفة وعديدة لدراسة مثل هذه الحالات .

في المرحلة التخطيطية من الدراسة سنواجه بعض الخيارات التي يكون أحدها مناسباً لدراسة هذه المشكلة أو تلك . فمثلاً هل أن الباحث يريد أن يجري اختبارات معينة أو ببساطة سيقوم بزيارات فقط للمزرعة ومشاهدة الحالات المرضية هناك أو هل نريد أن ندرس الحالات بمفردها في وقت محدد من الزمن أو أننا نرغب في متابعة هذه الحالات المرضية خلال فترة زمنية معينة بإستمرار وبلا تحديد . أو هل نريد أن نبدأ بدراسة الحيوانات السليمة من المرض والتأكد بالملاحظة مما إذا كان هناك مرض أو لا أو أننا في الحقيقة نريد أن نبدأ بدراسة الحيوانات المريضة فعلياً ودراسة الأسباب المتوقعة لحدوث مثل هذه الحالات المرضية ؟

التمييز بين الدراسات المعتمدة على المشاهدات السريرية والدراسات التجريبية

The Distinction between Clinical Observational and Experimental Studies

الدراسة المعتمدة على المشاهدات السريرية

في مثل هذه الدراسة نكتفي بمشاهدة الحيوانات المشمولة بالدراسة ومن ثم نسجل قيم القياسات المناسبة حسب الواقع وبذلك لا تتطلب الدراسة إجراء أي اختبار ، على سبيل المثال دراسة المعالجة المثلى لهذه الحيوانات أو دراسة عوامل الخطورة المتعلقة بهذا المرض والتي يمكن أن تؤثر على سير المرض مباشرة أو غير مباشرة ومن الواضح في هذه الدراسة أننا لا نستطيع أن نحدد عشوائياً الحيوانات ونقسمها إلى مجموعات لتقييم المعالجة الدوائية مثلاً وإنما ندرس هذه المشاهدات جماعياً للحصول على قيم خاصة بحدود وعوامل مختلفة متعلقة بدراسة حيوانات الدراسة ويمكن إنجاز هذه الطريقة من الدراسة بإتباع إحدى الطريقتين الآتيتين :

١- إجراء مسح لجميع حيوانات المزرعة أو المنطقة المراد دراستها Apopulation survey ويشمل خصائص حيوانات الدراسة المشمولة في الدراسة كما هو الحال في الدراسات الإحصائية الشاملة أي أن ندرس كل حيوان من هذه الحيوانات إحصائياً .

٢- إجراء مسح يعتمد على أخذ عينات A Sample Survey ويشمل فحص عينة من الحيوانات ممثلة للحيوانات المدروسة ومنها يمكن أن ندون استنتاجاتنا بناءً على نتائج هذه العينة على جميع حيوانات الدراسة .

إن العديد من المسوحات المعتمدة على العينات تمثل دراسة وصفية في محاولة للوصول إلى أن تقدم تقديرات لحدود حيوانات الدراسة هذه الحدود يمكن أن تشمل إنتاج الحليب – العمر الجنس – النوع الحيواني – المنطقة الوراثة ... وغيرها كثير) وربط هذه الحدود بما يحدث لهذه الحيوانات من مشاهدات سريرية. إلا أن بعض هذه الدراسات قد يكون من النموذج التحليلي بسبب أنها تكون معتمدة على دراسة العوامل المترافقة مع حدوث أو عدم حدوث حالة مرضية ما .

إن الدراسة الوبائية هي نوع خاص من مسوحات العينة التحليلية تعتمد على دراسة مسبب المرض بواسطة تحديد ما إذا كانت عوامل الخطورة المختلفة مترافقة مع حدوث وتوزيع حدوث الحالات المرضية أو لا ، فعلى سبيل المثال إن انتشار متلازمة كوشينغ Cushings Syndrome (فرط إفراز في هرمونات الكظر) عند الكلاب هي أكثر شيوعاً عند الكلاب الصغيرة (الحديثة السن) منها عند الكلاب الأخرى في الحقيقة بالدراسات الوبائية يتم تقييم ومقارنة الحالات المرضية بناء على الفروق وغيرها من الاعتبارات الأخرى .

الدراسة التجريبية Experimental Study

ففي الدراسة التجريبية نقوم بالتداخل مع سير الدراسة على سبيل المثال كأن نطبق قسرياً إجراءاً وقائياً لمرض ما كالمعالجة أو خفض تعرض الحيوان لعامل ما كدرجة الحرارة أو غيره من العوامل . وعندئذ نشاهد تأثير هذا الإجراء الوقائي أو العلاجي واستجابة حيوانات الدراسة له مع الأخذ بوجهة نظرنا حول تقييم ما إذا حدث تغير في استجابة حيوانات الدراسة والتي يمكن أن تكون مباشرة على إجرائنا وهناك نوعان من الدراسات التجريبية نوع يدعى بالتجارب المخبرية والآخر يدعى بالتجارب السريرية .

مدخل للتجارب السريرية Introducing to Clinical Trials

إننا نستخدم عبارة تجربة سريرية لوصف أية تجربة مخطط لها والتي تشمل دراسات متعلقة بالإنسان أو الحيوان وهي تصمم لتقييم فعالية واحدة أو أكثر من المعالجات أو الإجراءات الوقائية لأحد الأمراض كإستخدام أحد اللقاحات مثلاً ولقد توسع مفهوم هذه العبارة من مفهوم الطب السريري البشري ليشمل تلك الدراسات المتعلقة بالطب السريري البيطري وعلوم صحة الحيوان ، فعلى سبيل المثال اختبار الفعالية للعوامل الدوائية الجديدة للتحكم بالطفيليات الخارجية في الكلاب والقطط ويجب أن نميز التجربة السريرية clinical trial من التجربة السريرية الحقلية Clinical field trial. فعلى حين أن التجربة السريرية تُجرى ضمن ظروف نظامية

جداً فإن التجربة الحقلية بشكل كامل أو تطبيق جزء من هذه التجارب حقلياً المطبقة تحت الظروف الطبيعية سواء ضمن الظروف.

أهمية تصميم التجربة السريرية

The Importance of Design in the Clinical Trial

إننا نقوم عادة بالتجربة السريرية حتى نقيم الفائدة التي يمكن الحصول عليها من إدخال نوع جديد من المعالجة أو أي تطبيق إجرائي ضمن الظروف المتخذة ضمن هذه التجربة إذ إن أحد أهم اهتماماتنا في التجربة هو النتائج التي نحصل عليها من عينة مأخوذة من مجموعات من الحيوانات تربى ضمن نفس الظروف وتطبيق إجراء علاجي على عدد من حيوانات هذه العينة ومقارنتها مع غيرها من المجموعات التي لم يطبق عليها مثل هذا الإجراء العلاجي . ولكي نتأكد من أن المجموعة التي تلقت مثل هذا العلاج هي أفضل مجموعة تحقيق الفائدة العلاجية ومثل هذه التجربة يجب أن تكون خالية من أي انحراف في النتائج Bios وحتى نقوم بمثل هذه التجربة يجب أن نحدد أمرين اثنين :

١- تحديد الحيوانات التجريبية ضمن المجموعات المعالجة .

٢- تقييم استجابة هذه الحيوانات التجريبية للمعالجة .

ويجب أن يلاحظ أن التجربة المحققة لفرضها يجب أن تشمل ثباتية كل من :

- المقارنـة : Comparative مقارنـة أكثـر مـن مجموعـة علاجيـة . فنكـون عندئـذ قـادرين علـى أن نجعـل تقييمنـا حـول الاسـتجابة لإحـدى المعالجـات أو الإجـراءات المرتبطـة بالاستجابة التـي لـم تحصـل عليها فـي غيـاب المعالجة أو مقارنـة المعالجـة القياسية .

الفصل الثامن
CHAPTER EIGHT
مدخل إلى اختبار الفرضيات
AN INTRODUCTION TO HYPOTHESIS TESTING

مقدمة: Introduction

يمكن أن نقسم عموماً الفرضية أو النظرية الإحصائية إلى جزئيين :

١- هناك ما يسمى بالإحصاء الوصفي descriptive statistics حيث تستخدم فيه أدوات العرض المناسبة كالجداول النموذجية والأشكال والقياسات الرقمية لوصف قاعدة البيانات وتقديم عرض مختصر لتوزيعها .

٢- إضافة إلى ذلك هناك ما يدعى بالإحصاء الاستنتاجي أو الاستدلالي inferential statistics الذي يركز على إعطاء الاستنتاجات لمجتمع ما مستخدماً النتائج التي حصلنا عليها من العينة الممثلة للمجتمع المدروس والمأخوذة من هذا الأخير .

a- وإن إحدى النقاط في الاستدلال الإحصائي هي قضايا التقدير estimation لحدود المجتمع المدروس باستخدام العينة الإحصائية المناسبة (على سبيل المثال ، وسط المجتمع المدروس يمثل بوسط العينة المأخوذة منه) . كما أن

قضايا التقدير تكتمل فقط عندما تكون هناك دقة في عملية التقدير التي تحدد بالخطأ المعياري أو يعطى مؤشر لهذه العملية حد الثقة المدروس .

b- النقطة الثانية في الاستدلال الإحصائي هي اختبار الفرضية Hypothesis testing . وفي هذه الحالة نفحص الفرضية ضمن عبارة واحد أو أكثر من المجتمعات المدروسة . كما أننا نريد أن نعرف ما إذا كانت الفرضية حول هذا المجتمع أو المجتمعات الأخرى يمكن أن تدعم أو تمثل ببيانات العينة .

إن قضايا التقدير للقيم تركز على عملية وصف القيم بينما يركز اختبار الفرضية بلا تحديد على اتخاذ القرار المناسب .

المفاهيم الأساسية لاختبار الفرضية :

The Main Concept of the Hypothesis Test

إن اختبار الفرضية عملية تركز على إجراء الاستنتاجات حول المجتمع المدروس مستخدمين المعلومات المحصول عليها من العينة المأخوذة . ويجب أن نعي أنه من المستحيل أن نكون قادرين على التأكيد المطلق لصحة استنتاجاتنا حول المجتمع . إذ أن عينة واحدة مختارة من المجتمع من غير المحتمل أن تعطي نفس النتائج تماماً كما هي في العينة الثانية .

ولذلك يجب أن نحدد نتائجنا حول المجتمع المدروس باستخدام مفهوم الاحتمالية وهذا يعطي مؤشراً لفرصة الحصول على النتائج المشاهدة إذا كانت الفرضية حقيقية .

وهنا لا بد من استبدال عبارة الحالة المطلقة بعبارة الاحتمالية Probabilistic وهذه الأشكال هي المضمون أو المفهوم لاختبار الفرضية وهكذا يمكن أن يتلخص مفهوم الفرضية الإحصائية بدراسة صحة ادعاء معين عن قيمة معينة ما للمجتمع الإحصائي كالوسط الحسابي أو النسبة أو التباين . كأن نقول مثلاً إن هذا الدواء ناجع بنجاح بنسبة 95% لعلاج التهابات الكبد إلى غيرها من الافتراضات التي نود التأكد من صحتها.

نظرية العدم أو فرضية العدم

The Null Hypothesis H0

ترتكز الدراسات للتقصي عن فرضية معينة أو نظرية علمية حول المجتمع المدروس عادة على عملية مقارنة بالحالة الطبيعية مشمولة ببعض التأثيرات الرقمية لموضوع الدراسة وغالباً ما تكون هذه المقارنة لمعالجات مختلفة وقياس استجابة هذه المعالجة يدعى بالتأثير العلاجي . على سبيل المثال الفروقات بين مختلف العلاجات من حيث الوسط الحسابي .

لنفترض أننا نريد التقصي عن نقص مغنزيوم الدم عند أبقار ترعى على أعشاب الربيع الجديدة وأبقار مرباة تربية مكثفة تجارياً فعندئذ نريد أن نفحص قيم معدل بلازما الدم . وهذا يقترح أن هناك خطورة في كون الأعشاب تحتوي قدراً كافياً أو غير كافٍ من الاحتياجات الطبيعية للمنغنزيوم . وبهذه الحالة يتشكل ما يدعى بنظرية العدم حول معدل القيم الحقيقية لمنغنزيوم البلازما والتي يفترض أن لا تختلف بين المجموعتين .

هذا ما يدعى بنظرية العدم null hypothesis ويرمز لها بالرمز Ho وقد تكون فرضية أو نظرية العدم غير حقيقية وعندئذ تدعى بالفرضية أو النظرية البديلة alternative hypothesis ويرمز لها بالرمز H1 .

نتيجة :

هناك عادة فرضيتان في دراسة أي اختبار ، الأولى منها تدعى بنظرية أو فرضية العدم وهي الفرضية الأولية والتي غالباً ما تؤيد صحة الادعاء ويرمز لها بالرمز Ho والثانية هي الفرضية البديلة والتي هي أساساً لا تؤيد صحة الادعاء ويرمز لها بالرمز H .

عادة تذكر الفرضية البديلة أنه يوجد اختلاف بين قيم الحد لكن الاتجاه لهذا الاختلاف غير معروف . وهذا يقودنا إلى فرضية ذات جانبين أو اختبار الذيلين (two- sided or a two tailed test) . ففي مثالنا السابق على أن تناول

أعشاب حديثة النمو يخفض من تركيز مغنزيوم البلازما ، فإن الفرضية البديلة نقول أنه يوجد معنيين مختلفين لهذا التأثير حيث أنه لا يوجد لسبب لنقول مثلاً أن المعالجة للحيوانات المرباة ضمن حظائر مغلقة هي أفضل طريقة من طريقة استعمال أعشاب الربيع لتأمين الاحتياجات من هذا العنصر المعدني النادر .

- الاختبار الإحصائي وقيمة p

The Test Statistic and the p-Value

من البيانات التي حسبنا منها قيمة الاختبار الإحصائي - (وهو تعبير جبري معين للفرضية التي نختبرها) ، وعادة يستخدم الحاسوب لعملية الحساب هذه وأحياناً تحسب يدوياً - ويرفق بكل قيمة الاختبار الإحصائي ما يدعى بالاحتمالية Probability وهو ما يدعى بالقيمة p– (p-value) . وهي وهذه القيمة تصنف فرضية الحصول على تأثير القيم المشاهدة (وبعبارة أخرى أكثر توضيحاً) إذا كانت فرضية العدم حقيقية .

اتخاذ القرار باستخدام قيمة p

Making a Decision Using the p-Value

١- إذا كانت النتائج المشاهدة غير مطابقة لما نتوقع وعندما تكون نظرية العدم حقيقية ، فإننا نستنتج أنه يوجد دلالة كافية لرفض reject نظرية أو فرضية

العـدم . ونقـول أن نتيجـة الاختبـار إحصـائياً هـي معنويـة Statistically significant .

٢- وإذا كانت النتائج المشاهدة ثابتة مع ما نتوقع إذا كانت نظرية العدم غير حقيقية ونستنتج أنه لا يوجد دلالة كافية لتكون حقيقية وعندها لا نرفض فرضية (نظرية) العدم . ونقول أن نتيجة الاختبار غير معنوية non – significant . وتسمح لنا قيمة (p) بأن نحدد فيما إذا كنا نملك دليلاً كافياً لرفض فرضية العدم بالمقارنة مع الفرضية البديلة .

- إذا كانت قيمة (p) صغيرة جداً فعندئذٍ من غير المحتمل أن نستطيع الحصول على النتائج المشاهدة إذا كانت فرضية العدم حقيقية ولذلك ترفض فرضية العدم Ho .

- أما إذا كانت قيمة p كبيرة جداً فعندئذ ثمة فرصة كبيرة للحصول على النتائج المشاهدة إذا كانت فرضية العدم حقيقية وعندئذ ترفض فرضية العدم لا ترفض . وبوضوح إن التمييز بين قيم (P) الصغيرة والكبيرة هو موضوع متروك لاتخاذ القرار المناسب . فيجب أن نحدد قبل البدء بعملية جمع البيانات المكونة لقيم (p) سواء بقيمة صغيرة أم كبيرة وهنا لا بد من استخدام مصطلح مستوى المعنوية Significance level للاختبار . كما أن اختبار مستوى المعنوية يعتمد على طبيعة البيانات والظروف المتعلقة بالتقصي عن الحالة إذ هناك إمكانية تجعلنا

مثلاً نختار قيمة منخفضة جداً لمستوى المعنوية كأن نقول المستوى 0.01 ، إذا كنا نريد أن ننظر إلى الجانب السببي لرفض نظرية العدم وهذا يعني أنه في مثالنا نحصل على قيمة (p) والتي هي أقل من 0.01 فإننا نرفض نظرية العدم ونقول إن النتيجة معنوية عند المستوى 1% .

على سبيل المثال نريد التقصي عما إذا كانت المعالجة بالصادات الحيوية مكلفة فإننا نريد أن نكون واثقين جداً من فوائدها خلال فترة المعالجة وكذا تكلفتها المادية . وهنا يمكن أن نختار القيمة الأكثر ارتفاعاً فنقول 010 في بداية الاختبار لفعالية اللقاح الكامن الجديد كبديل للشفاء أو الوقاية من المرض المعدي والذي يسبب خسائر فادحة وهذا يؤكد أن الفوائد المحصول عليها غير مؤكدة وتحتاج إلى قياس معنويتها . إن اختيار الفاصل القاطع 0.05 Cut off point هو غالباً اختيار لمستوى المعنوية مثلاً إذا كانت قيمة p أصغر من 0.05 فعندئذ ترفض نظرية العدم ، أما إذا كانت قيمة p أكبر أو تساوي 0.05 فعندها نقبل فرضية العدم . وهناك تمييزات إضافية تستخدم أحياناً بواسطة الإشارة النجمية :

(***) تمثل p (p< 0.0001) وهي معنوية واضحة جداً .

(**) تمثل قيمة (P<0.001) وهي معنوية واضحة .

(*) تمثل قيمة (0.01 >p<0.05) وهي معنوية وعندما لا توجد معنوية يرمز لها بالرمز (NS) وعندها تكون قيمة p أكبر من 0.05 .

وبشكل آخر المعنوية الواضحة جداً تكون ضمن المجال [0.0000-0.0001]

المعنوية الواضحة تكون ضمن المجال [0.001-0.0001]

معنوية فقط [0.01-0.001]

ونذكر كلما صغرت قيمة (p) كان هناك دليل أوضح ومؤكد ضد نظرية العدم . ومن المهم الإشارة هنا إلى أن قرار الأخذ بصحة الفرضية أو رفضها لا يعني القطيعة في الحكم لأننا نقبل بالفرضية أو نرفضها باحتمال ثقة معين يفترض أن يكون كبيراً ومقنعاً مما يعني أننا قد نرفض فرضية وهي صحيحة أو نقبل بفرضية وهي خاطئة إلا أن احتمال وقوعنا بالخطأ سيصغر مع ارتفاع احتمال الثقة .

اشتقاق قيمة(p): Deriving the p-Value

نموذجياً نحصل على قيمة p من التوزيعات المعروفة للاختبار الإحصائي وهي التوزيعات الطبيعية ، وتوزيع t ، F ، α (والتي سوف تدرس في فصول لاحقة) . وإذا أنجز التحليل بالحاسوب فإن قيمة p التي نطلبها نحصل عليها مباشرة بواسطة نتائج الحاسوب وإذا كان من الضروري يمكن أن نحصل على قيمة p بالإشارة إلى الاختبار الإحصائي من الجدول المدرج بالتوزيعات المذكورة أعلاه .

درجات الحرية للاختبار الإحصائي

The Degrees of Freedom of the Test Statistic

سوف نجد ن عبارة درجات الحرية degrees of freedom (df) مستخدمة بإستمرار في التحليل الإحصائي . فإذا ما قمنا باستخدام جداول لربط قيمة الاختبار الإحصائي إلى قيمة p فإننا عادة يجب أن نعرف درجات الحرية للتوزيع المناسب لاختبارنا الإحصائي المستخدم . إن درجات الحرية للاختبار الإحصائي هي عبارة عن عدد المشاهدات المستقلة المساهمة بتشكيل الاختبار الإحصائي فعلى سبيل المثال أعداد المشاهدات لتقييم الاختبار الإحصائي ناقص عدد المشاهدات المحددة بقيمها بحدود معينة . إذ إن أسهل طريقة لحساب درجات الحرية لأي اختبار إحصائي هو أن نأخذها من حساب الفرق بين عدد المشاهدات في العينة وعدد الحدود التي يجب أن نقدرها لكي نقيم هذه المشاهدات إحصائياً .

ولذلك لنفرض مثلاً أن تقديرنا لفرق المجتمع بكامله σ^2 لمتغير x في حجم عينة n بواسطة عينة إحصائية S^2 يعطى بالقانون : $S^2 = \frac{\sum (X - \overline{X})^2}{(n-1)}$ (الفرق) يجب أن نقدر المعدل لكي نقيم قيمة الصورة ولذلك تكون درجات الحرية لهذا الفرقS^2 هي (n-1) بإعتبار أن n هي عدد قيم المشاهدات .

ملخص لإجراء اختبار الفرضية / النظرية / :

إن إجراء اختبار الفرضية :يتلخص في ست مراحل :

١- حدد فرضية العدم Ho أو الفرضية البديلة المطلوبة في الاختبار .

٢- اجمع البيانات وانظر نظرة شاملة إذا كان هناك إمكانية للتقصي عن توزيعات هذه البيانات فهناك العديد من الاختبارات تحدد الافتراضات التوزيعية حول البيانات : افحص التوزيعات المحددة لهذا الاختبار .

٣- على الحاسب ، اختر الاختبار المناسب ، أو باستخدام الحساب اليدوي المناسب لاختيار الاختبار الإحصائي المناسب مستخدمين بيانات العينة .

٤- اربط القيمة المحسوبة بالاختبار الإحصائي بقيمة p .

٥- اعتبر قيمة p لمحاكمات أو تحكيم فيما إذا كانت البيانات قابلة للتحكيم باستخدام فرضية العدم ومن ثم قرر ما إذا كانت البيانات ترفض أو تقبل فرضية العدم .

٦- إلا أنه من الأجدى والأنسب أن نحسب حد الثقة لتأثير موضوع الدراسة .

حيث أن نحدد المعنوية لنتيجة الدراسة في ضوء الإجراء السابق ولذلك يدعى هذا أحياناً باختبار المعنوية Significance test .

وإن اختبار الاختبار – كما سنرى لاحقاً – ليس سهلاً دائماً بل إنه يعتمد على طبيعة البيانات .

الخطأ نمط I والخطأ نمط II Type I and Type II Errors

عندما نجد نتيجة الاختبار معنوية نرفض فرضية العدم في مستوى المعنوية المدرجة . فإذا كان الاستنتاج غير صحيح و حقيقة المعدلين المدروسين متساوية فعندئذ نرفض فرضية العدم وعندما لا ترفض (عندما تكون حقيقية ..) نكون قد حددنا نمط الخطأ I .

وبشكل آخر أو بديل ، عندما نجد نتيجة الاختبار غير معنوية لا ترفض فرضية العدم عند مستوى المعنوية المدرج ، ولذلك لا يمكن أن نستدل إحصائياً على معدلات مجتمعين مختلفين .

وهنا تكون الحالة غير صحيحة والحقيقة لمعدلين مختلفين وعندئذ لا ترفض فرضية العدم وعندما ترفض الفرضية (عندما تكون خاطئة) نحدد نمط الخطأ II كما هو موضح في الجدول ٨ .

جدول رقم ٨ : الأخطاء في فرضية العدم

لا ترفض فرضية العدم Ho	رفض فرضية العدم Ho	الوصف
القرار صحيح	نمط الخطأ I	فرضية العدم حقيقية Ho
نمط الخطأ II	القرار صحيح	فرضية العدم خاطئة Ho

CHAPTER NINE
اختبارات الفرضية -١- اختبار t
مقارنة متوسط وحيد أو متوسطين
HYPOTHESIS TESTS 1 – THE T TEST COMPARING ONE OR TWO MEANS

متطلبات اختبارات الفرضية لمقارنة المتوسطات

Requirements for Hypothesis for Comparing Means

١ – طبيعة البيانات The . Nature of the Data

الوسط الحسابي هو أحد مقاييس الميول المركزية لمتغيرات كمية كما ذكرنا سابقاً . لنفترض هنا أننا نريد دراسة قياس تأثير الإجهاد على الأبقار أثناء السفر وحيث إن الكورتيزون يحرر من غدة الادرينالين استجابةً لحالات الإجهاد الشديدة وفي هذه الحالة نرغب بتحديد ما إذا كان قياس معدل الكورتيزون في بلاسما الدم في أثناء فترة النقل عند الأبقار الحلوب يختلف عن تلك الأبقار التي هي في حالة راحة ، ولقياس هذه الفروقات باستعمال الطرق التي سنتطرق إليها في هذا الفصل يجب أن نأخذ بعين الاعتبار طبيعة البيانات

أ– هناك متغير وحيد مراد دراسته (على سبيل المثال في حالتنا هذه هو عبارة عن كورتيزول البلاسما) .

ب- يجب أن تكون البيانات المقيسة ذات طبيعة قياس كمي (رقمي) على سبيل المثال قياسات الكورتيزون تقاس بوحدة ng / ml .

ج- الجانب الهام افتراض في مثل هذه الدراسة هو أن يكون المتغير المراد التقصي عنه ذا توزيع طبيعي إلا أن هذا يكون غير منطقي في المقاييس الحيوية كما هو الحال في كورتيزون البلاسما . لكن عندما تلاحظ أن حجم العينة المدروسة كبير فيقرب معدل العينة تقديرياً أنه موزع توزيعاً طبيعياً.

٢- تطبيقات حجم العينة The Implication of Sample Size

إذا كان حجم العينة يشير إلى إجمالي أعداد المشاهدات فإن التقصي عندئذ يرتكز على:

١- توزيع المتغير يكون صعباً لتقييم عينات صغيرة جداً لمقارنات تشمل على سبيل المثال 5 أو 6 مشاهدات . وأيضاً فإن العينة الصغيرة الحجم لا تمثل المجتمع الحيواني المدروس .

٢- وإذا كان حجم العينة صغيراً (لنقل عدد المشاهدات أقل من ٣٠) فإن الاختبار الإحصائي يتبع توزيع Students t Distribution (t ويعتمد فيه على توزيع طبيعي تقريبي (الفروقات تكون متساوية لمقارنة عينتين) .

٣- إذا كان حجم العينة كبيراً فإن توزيع الاختبار الإحصائي مقرب طبيعي لعينة ذات مشاهدات أكثر من (٣٠) إذا كانت البيانات موزعة طبيعياً . وبالمقابل يمكن

للعينة الكبيرة يمكن أن تزيد عن ٤٠٠ مشاهدة إذا كانت البيانات غير موزعة طبيعياً . إلا أن وجود الانحراف عن الطبيعي سوف يؤثر على تعريف مصطلح /عينة كبيرة/ .

وهكذا من الصعوبة بمكان أن نوضح الفروقات بين العينات الكبيرة والصغيرة ضمن إطار مفهوم اختبار t . وربما نجد العديد من النصوص الإحصائية التي تميز بين مصطلحات العينة الكبيرة والصغيرة بربطها إلى الاختبار الإحصائي كما هو الحال في توزيع t (بالنسبة للعينات الصغيرة ، الفروقات متساوية عند مقارنة عينتين) أو إلى توزيع طبيعي (بالنسبة للعينات الكبيرة ، فروقات متساوية أو غير متساوية) .

٣- تصميم الدراسة Study Design

إن النقطة الأكثر أهمية في تصميم الدراسة هي عملية جمع البيانات يستخدم الاختبار لمقارنة متوسطات لمجموعات فردية أو مجموعتين من المشاهدات هذه المجاميع المقارنة إما أن تكون مستقلة عن بعضها أو تكون على شكل مشاهدات شفعية إلا أننا نرغب أحياناً بأن تقارن المشاهدات لأكثر من مجموعتين ، فإن الخيار عندئذ لتحليل مثل هذه البيانات هو تحليل التباين Analysis of Variance .

اختبارات t لعينة واحدة One Sample t - Tests :

أحياناً قد نهتم بدراسة معدل المشاهدات لمجموعة مفردة تأخذ قيمة معينة . فعلى سبيل المثال مجموعة من الخراف في إحدى الحظائر وفي إحدى المزارع أظهرت انخفاضاً في الكسب الوزني اليومي مقارنة بمعدل النمو الطبيعي لهذه المزرعة . وما نريد إنجازه هنا أن نختبر ما إذا كان معدل الكسب الوزني الحي لهذه الخراف في هذه الحظيرة يدعم الفرضية أن تؤيد أن نمو هذه الخراف في المعدل المتوقع للنمو ضمن هذه المزرعة .

– الافتراضات Assumptions

يفترض اختبار t ستودنت لعينة واحدة أن بيانات العينة هي مجموعة مختارة عشوائياً من قيم لحيوانات ذات توزيع طبيعي .

وأحياناً قد تكون البيانات موزعة توزيعاً غير طبيعي أو منحرفة قليلاً عن التوزيعات الطبيعية بالنسبة للناظر . فعندئذ نقربها إلى الشكل الطبيعي بإجراء ما يسمى بالتحويل المناسب appropriate transformation .

نموذجياً يكون التحويل اللغاريتمي هو الإجراء الأمثل لهذه الحالة إحصائياً .

أو إننا يمكننا أن نلجأ إلى إحدى الاختبارات اللا معلمية المناسبة .

appropriate non – parametric test . كاختبار الإشارة Sign test أو اختبار كالموغاروف . سمير نوف Kalmogorav – Smir nov أو اختبار الأدوار Runs test .

The Approach الطريقة

سنعرض هذه الطريقة عرضاً عاماً بدون الدخول في تفاصيل معقدة وسنوضح ذلك بمثال الخراف في المقطع السابع .

١- نحدد نظرية أو فرضية العدم null hypothesis فإن معدل الخراف المدروسة للمتغير المراد دراسته (معدل النمو) مساوٍ لقيمة محددة Mo . بينما الفرضية أو الطريقة البديلة alternative hypothesis . فإن المعدل هو غير مساوٍ لقيمة محددة وهذا يقودنا إلى اختبار الفرضية البديلة ذات الذيلين مثل حالة الوسط الحسابي $H_1 : M \neq M_0$

كأن نقصد هنا بالفرضية البديلة أن معدل النمو عند الخراف زائد أو ناقص عن الطبيعي من القيم الطبيعية المتوقعة .

٢- نجمع البيانات ونعرضها على شكل خط نقطي line plot ، أوعلى شكل مستطيلات بسيطة Simple Histogram أو أي نمط من أنماط التوزيعات الشكلية البسطية مثل بوكس وويسكر Box and Whisker bar وغيرها ...

وبالاستعانة بالشكل الملاحظ نفحص افتراض أن البيانـات موزعـة توزيعاً تقريبياً طبيعياً .

٣- احسب الاختبار الإحصائي statistic test كالفرق بين متوسط العينـة $(\overline{X})$ – والقيمـة المحددة لمتوسط القطيـع $(\overline{X})$ أو حيوانـات المزرعـة فـي مثالنـا الحـالي وضمن شروط فرضية العدم H_0 , $M = M_0$ ، مقسوماً على متوسط الخطأ المعياري .

وكما هو معروف لدينا سابقاً متوسط الخطأ المعياري Standard error of the mean ,

يحسب كما يلي : $SEM = \dfrac{S}{\sqrt{n}}$.

بإعتبار أن :

S : هي الانحراف المعياري المحسوب .

n : حجم العينة / عدد المشاهدات / .

وبهذه الحالة فإن اختبار العينة واحدة .

$$Test_1 = \frac{\overline{X} - Mo}{S / \sqrt{n}}$$.

مع الأخذ بعين الاعتبار أن درجة الحرية هي $n - 1$.

٤- نحصل على قيمة (P) بحساب قيمة الاختبار الإحصائي والنظر إلى توزيع قيم t حسب المحصول عليها في الاختبار كما هو مدرج في الملحق (A3).

٥- بحسب قيمة P يمكن أن نحكم ما إذا كانت البيانات متوافقة مع نظرية العدم . وعندئذ نقرر قبول أو رفض فرضية العدم وعادة نرفض نظرية العدم إذا كانت قيمة $P < 0.05$.

٦- لا بد أن هناك حد حساب حد الثقة Corfidence Interval الخاص بالمتوسط يسمح لنا بالحكم على أهمية النتائج المدروسة لاسيما درجة الثقة ٩٥ % والتي تحسب كما يأتي :

$\overline{X} - t_{0.05}\frac{S}{\sqrt{n}}$	إلى	حيث :	t	القيمة المحسوبة
$\overline{X} - t_{0.05}\frac{S}{\sqrt{n}}$	المجال		٠.٠٥	

بإعتبار :

قيمة t المحسوبة $t_{0.05}$ بالجدول .

بإعتبار أن $t_{0.05}$ هي قيمة حرجة Critical value يحصل عليها من جدول توزيع t ستودنت (الملحق A3) مع درجة حرية $n - 1$ وهي تعطي إجمالي المنطقة الذيلية للتوزيع باحتمالية ٠.٠٥ .

مثال :

يظهر الجدول رقم ٩ معدل الكسب الوزني اليومي لعينة عشوائية لـ ٣٦ خروفاً خاصاً بالتربية في إحدى وحدات التربية . يتوقع معدل نمو في وحدة التربية هذه ٦٠٧ غ باليوم لـه١هـ المرحلة من النمو اعتماداً على المؤشرات العملية المنجزة

حالياً عند هذه الخراف . هل هذه متوافقة مع معدل الكسب الوزني المتراوحة في الوزن ٦٠٧ غ باليوم .

الجدول رقم ٩ : معدل الكسب الوزني اليومي لخراف (الأوزان بالغرام)

٥٧٧	٥٩٦	٦١٢	٦٠٠	٥٨٤	٦١٨
٦٢٧	٥٨٨	٦٠٦	٥٥٩	٥١٦	٦٠٧
٦٠٨	٥٩١	٥٦٥	٥٨٦	٦٢١	٦٢٣
٥٩٨	٦٠٢	٦٣١	٥٧٠	٥٩٥	٦٠٣
٦٠٥	٦١٦	٥٧٤	٥٧٨	٦٠٠	٥٩٦
٦١٩	٦٣٦	٥٨٩	٥٩٤	٦٠١	٥٨١

الوسط الحسابي $\overline{X}$ = ٥٩٩.٠٩٤

الانحراف المعياري S = ١٨.٦٥٦

الخطأ المعياري $\frac{S}{\sqrt{n}} = 18.656/\sqrt{36}$

١- بالنسبة لفرضية العدم ، هل هناك فرق جوهري مع المتوسط المتوقع M_O حيث المتوسط الحقيقي للكسب الوزني هو ٦٠٧ غ يومياً . هل الفرضية البديلة يمكن أن تستخدم هنا ؟

٢- برسم مخطط باستخدام المستطيلات البسطية أو بوكس وويسكر نجد أن مثل هذه البيانات موزعة طبيعياً تقريباً .

٣- إن الوسط الحسابي للعينة هو ٥٩٩.٠٩٤ غ يومياً مع تقدير للانحراف المعياري بالقيمة ١٨.٦٥٦ غ يومياً . إن اختبار t الإحصائي هو كما يأتي :

$$Test_1 = \frac{X^{-}M}{S/\sqrt{n}} = \frac{599.94 - 607}{18.656/\sqrt{36}} = \frac{599.94 - 607}{3.109} = 205$$

مع درجة تقدر بـ n = 36 .

درجة حرية n – 1 = degree of freedom .

Df = 36 – 1 = 35 .

٤- كما نلاحظ في الجدول A3 أن قيمة 0.02 > P > 0.01 وإذا حسبنا هذه القيمة بالحاسوب فسوف نحصل على قيمة p = 0.017 ولذلك لدينا فرصة أقل من ٢ % للحصول على متوسط كسب وزني حي أقل من ٥٩٩.٢ غ يومياً إذا كانت فرضية العدم حقيقية .

٥- ولذلك من غير المحتمل أن تكون فرضية العدم حقيقية . فإننا نرفض فرضية العدم Ho ونستنتج أن قيم البيانات هي غير متوافقة inconsistent مع معدل يومي للكسب الوزني ٦٠٧ غ يومياً .

٦- إن حد الثقة (CI : 95 %) لمعدل حقيقي للكسب الوزني الحي هو :

$$\overline{X} \mp t_{0.05} \frac{S}{\sqrt{n}}$$

أو X ±SEM

قيمة $\overline{X} \mp t$

عندها P = 0.05 .

وباعتبار $\overline{X}$ (الوسط الحسابي) . ٥٩٩.١٩٤ .

t0.05 (حسبت بالجدول الخاص بتوزيع t) = ٢.٠٤٥ # ٢.٠٣ .

CI : 95 % = 599.194 ∓ 2.03(3.109) = (592.88 , 605.51) .

طريقة التباينات المتساوية

The Approach – Equal Variances

حدد فرضية العدم التي تعتمد على أن معدلات المجموعتين المدروستين متساوية وعموماً تفترض الفرضية البديلة أن هذه المعدلات غير متساوية وهو ما يعتمد عليه اختبار الفرضية ذات الذيلين Two Tailed Test .

اجمع البيانات واعرضها في مخطط تخطيطي وإذا كان حجم العينة في كل مجموعة صغيراً فأنجز مخططاً نقطياً ، وإذا كان حجم العينة كبيراً في كل مجموعة فعندئذ من السهولة بمكان رسم وتوضيح أمكنة توضع الوسيط الحسابي وقيم المدى للعينة المدروسة . أو يمكننا فحص التوزيع الطبيعي التقريبي بواسطة مخطط مستطيلات بسيط . كما نفحص التباينات المتساوية باستخدام اختبار F (سيوضح في الفصل القادم) علماً أن قيمة F دائماً سوف تظهر عند إنجاز الاختبار باستخدام الحاسوب .

الاختبار الإحصائي ينجز بواسطة تقسيم الفروقات في معدلات العينة بين المجموعتين على الأخطاء المعيارية لكل مجموعة . البرنامج الحاسوبي سينجز الحساب . لكن من المهم أن نتعلم هذا الحساب يدوي وحسابي ، وعندئذ فإن الاختبار الإحصائي لتوزيع (t) سيجري كما يأتي :

$$Test_2 = \frac{\overline{X}1 - \overline{X}2}{5x^2(\frac{1}{n1} + \frac{1}{n2})}$$

وتحسب درجة الحرية هنا كما يأتي :

. n1+ n2 – 2

وبإعتبار أن :

ni = عدد المشاهدات .

$\overline{Xi}$ = متوسط العينة .

Si : الانحراف المعياري المقدر والمحسوب كما يأتي :

$$S^2 = \frac{(n_1 - 1)S_1{}^2 + (n_2 - 1)S_2{}^2}{n1+n2-2}$$

عادة نحصل على قيمة P من النتائج في الحاسوب ولكن يمكن حساب قيمة الاختبار الإحصائي يدوياً ونحصل على هذه القيمة ؛ بالإشارة إلى جدول توزيع t (الجدول الملحق A3) .

وعندئذ يمكن أن تستخدم قيمة P للحكم على معنوية وثقة البيانات بالنظر إلى فرضية العدم وما إذا كان من الممكن أن نقبل أو نرفض فرضية العدم . وعادة نرفض الفرضية إذا كانت قيمة P أقل من ٠.٠٥ ($P < 0.05$) .

– احسب درجة الثقة المناسبة (على سبيل المثال الاختلافات والفروقات بين متوسطات المجموعتين) وإذا لم نحصل على درجة الثقة ٩٥ % باستخدام الحاسوب يمكننا حسابها كالآتي :

$$\boldsymbol{CI}\,45\% = (\overline{\boldsymbol{X}}1 - \overline{\boldsymbol{X}}2) \pm {}^{\boldsymbol{t}}{}_{0.05}\ \mathrm{X}\boldsymbol{SE}(\overline{\boldsymbol{X}}1 - \overline{\boldsymbol{X}}2)$$

وبإعتبار أن :

$$SE = \frac{S}{\sqrt{n}}$$

$$CI95\% = (\overline{X}1 - \overline{X}2) \pm {}^{t}{}_{0.05}\chi\sqrt{S^2(1/n_1 + 1/n_2)} \; .$$

وبإعتبار أن قيمة $t_{0.05}$: نحصل عليها من جدول توزيع t .

درجة الحرية : $n_1 + n_2 - 2$

S^2 : الفرق المقدر (على افتراض أن الفروقات بين العينتين متساويتان تقريباً في هذا الاختبار)

وإن احتمالية ذات الذيلين لتوزيع (t) تكون قريبة أو مساوية لقيمة ١.٩٦ (وهي تقرب عادة إلى القيمة ٢) عندما تكون درجات الحرية كبيرة جداً ، لنقل مثلاً أكبر من ٢٠٠ .

لنقل إننا نريد مقارنة متوسط وزن الجسم في أوقات السِّفاد لمجموعة من النعاج المستوردة (مع وضع تغذية جيدة لها خلال عدة أسابيع قبل السفاد) والمجموعة الأخرى توضع شاهداً (نعاج محلية) .

- **فرضية العدم :**

١- متوسط وزن الجسم في النعاج المدروسة في كلتا المجموعتين متساوية .

٢- الفرضية البديلة : هذا المتوسط مختلف في كلتا المجوعتين .

البيانات التي حصلنا عليها في كلتا المجموعتين كما هو في الجدول رقم ١٠.

إن حجم العينتين المستوردة والشاهد كانت ٥٤ عينة جمعت عشوائياً كما ظهر من الجدول السابق لأوزان الجسم في كلتا العينتين (٢٤ عينة مستوردة و ٣٠ عينة شاهداً) .

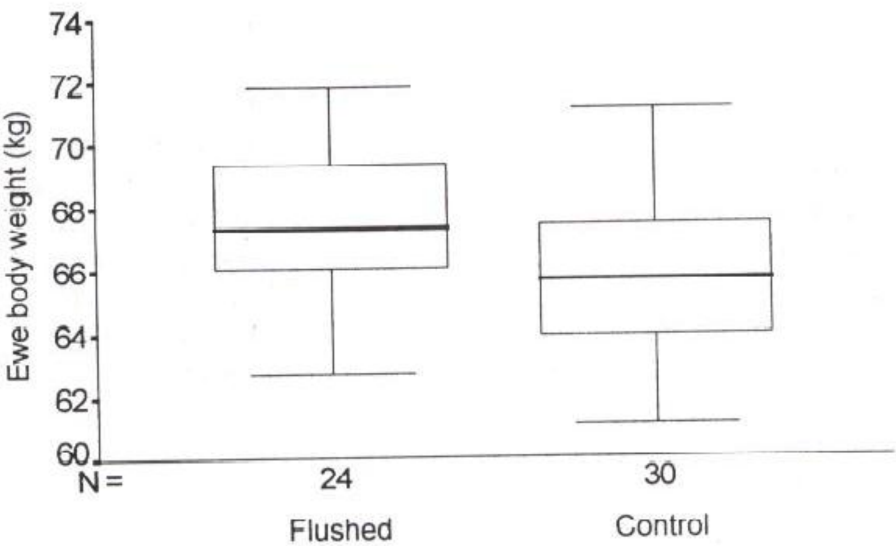
من خلال توزيع بوتس وويسكر في الشكل التالي نجد أن توزيع البيانات ذات توزيع طبيعي تقريباً (الشكل ١٦) .

الجدول رقم ١٠: متوسط وزن الجسم في النعاج المدروسة

أوزان مجموعة النعاج الشاهد		أوزان مجموعة النعاج المستوردة
٦٢.٥	٧١.١	٧٠.٧
٦٦.٨	٦٧.٠	٧١.٨
٦٩.٥	٦٧.٥	٦٤.٩
٦٤.١	٦٨.٥	٦٨.٢
٦٥.٣	٦٣.٦	٦٩.٤
٦٥.٦	٦٤.٨	٦٤.٤
٦٦.٤	٦٧.٤	٦٦.٩
٦٦.١	٦٦.٠	٦٩.٤
٦٨.٦	٦٩.٢	٦٧.٨
٦٢.٥	٦٢.٦	٦٦.٨
٦٣.٩	٦١.١	٦٧.٠
٦٥.٧	٦١.٨	٦٧.١
٦٧.٢	٦٩.٦	٦٧.٦
٦٥.٢		٦٦.١
٦٣.٥		٦٢.٧
٦٥.١		٦٤.٦

٦٩.٥		
٦٨.١		
٦٦.٠		
٦٩.٤		
٦٩.٨		
٦٧.٩		
٦٦.٢		
٦٤.٢		

الشكل رقم (١٦) : توزيع بوكس وويسكر (عينات الشاهد Control) و العينات المستوردة (Flushed)



يمكن إجراء هذا الاختبار بواسطة أي برنامج إحصائي في الحاسوب أو

يمكن حساب قيمة توزيع (t) الإحصائي لعينتين كما يأتي :

$$Test2 = \frac{Variance\ of\ means}{SE\ (Variance\ of\ means)}$$

حيث : الصورة تمثل التباين في المعدلات والمخرج يمثل حاصل ضرب الخطأ المعياري بتباين المعدلات :

$$= \frac{1.5433}{0.655} = 2.43$$

درجة الحرية : $df = 24 + 30 - 2 = 52$

ونحسب قيمة المخرج والتي هي: SE (تباين المتوسطات) كما يلي:

$$\sqrt{\frac{23(2.52)^2 + 29(2.447)^2}{24+30-2}(\frac{1}{24}+\frac{1}{30})4}$$

إن قيمة P التي يمكن أن تظهر لدينا (للمعنوية ذات الذيلين) والتي يمكن أن تحسب أوتوماتيكياً بواسطة الحاسوب هي $P = 0.018$ وهي مؤشر على أن هناك فرصة للحصول على فرق بين المعدلات ويقدر بـ ١.٥٩ .

وإذا حسبنا هذه القيمة يدوياً فإن قيمة توزيع (t) يشار إليها بالقيمة (2.43) في الجدول (A3) الملحق . مع درجة حرية 52 ونجد أن قيمة (P) تقع بين المجالين $0.01 < P < 0.02$.

وبالنظر إلى فرضية العدم ، نجد أنه لا توجد فروق في متوسط أوزان الجسم في كلتا المجموعتين ، فمن غير المحتمل أن تكون هذه الفروق حقيقية ولذلك ترفض فرضية العدم وبالمقابل بالنظر إلى الفرضية البديلة (H1) نجد أنه هناك فروقات في متوسط أوزان الجسم . وهكذا فإن متوسط أوزان الجسم معنوية الفروقات بالمقارنة مع

متوسط أوزان الجسم في النعاج المستوردة وبمعدل 1.59 كغ أكثر قيمة من تلك الأوزان عند نعاج الشاهد المحلية .

– درجة الثقة 95 % تحسب كما يأتي :

$$CI : 95\ \% = 1.59 \pm 2.007$$
$$= (\ 0.28 - 2.41\ )$$

ونحصل على القيمة 2.007 من جدول توزيع t (الملحق A3)

اختبار (t) المعدل – التباينات غير المتساوية:

Modified t – Test – Unequal Variances

إذا كانت الفروقات في كلتا المجموعتين غير متساوية فإن الفرق المقدر S^2 عندئذ ، يستخدم في مخرج /مقام/ القانون الخاص بالاختبار الإحصائي في المجموعتين Test 2 . وعندما يمكن حساب اختبار t المعدل عندئذ تكون التباينات غير متساوية كما يأتي :

$$Test_3 = \frac{\overline{X}1 - \overline{X}2}{\sqrt{\frac{S1^2}{n_1} + \frac{S2}{n_2}}}$$

وعلى أية حال فإن هذا الاختبار لا يتبع توزيع t ولذلك لا يعتبر تقييم قيمة P طريقة مباشرة . إلا أنه في العينات الكبيرة الحجم (أكثر مدة من 50 عينة) لا يتبع اختبار Test3 توزيع الطبيعي التقريبي Approximately Normal Distribution . ويمكن عندئذ أن نحصل على قيمة P المناسبة من جدول التوزيع

الطبيعي المعياري (في الملحق A.1) . وإذا كان حجم العينة صغيراً فإنه يجب إما أن نقوم بعملية تحويل للبيانات لإنجاز طريقة الفروق المتساوية أو تطبيق الإخبارات اللامعلمية المناسبة Appropriate non Parametric مثل اختبار مجموع ترتيب ويلكوسون Wilcoxon Rank Sum Test والذي سيشرح في الفصول القادمة والخاص باختبار t لعينتين Tow Sample t Test .

اختبار t الزوجي The Paired t – Test

يستخدم اختبار t الشفعي عندما يكون لدينا عينتان ممثلتان ومختارتان من الحيوانات المدروسة لمقارنة مشاهدات غير مستقلة أو زوجية (شفعية) ن على سبيل المثال عندما تقارن مستويات الغلوكوز في المصل عند الكلاب بالربط مع كمية الأنسولين كمتغير غير مستقل عند الإصابة بداء السكري من خلال تغذية هذه الكلاب بوجبات فيها ألياف قليلة أو كبيرة بأخذ عينة عشوائية في أثناء مراحل التجربة .

الافتراضات : Assumptions

إن صلاحية اختبار t الشفعي يعتمد على الافتراضات الآتية :

١- يوجد اختبار عشوائي **Random Selection** للعينة .

٢- مكان التجربة أيضاً مختار عشوائياً .

٣- الاختلافات بين أزواج (أشفاع) المشاهدات موزعة توزيعاً طبيعياً تقريبياً **Approximately Normally Distribution** مع أن المشاهدات الأولية في كل مجموعة يمكن أن لا تكون موزعة طبيعياً .

الطريقة : The Approach

بنفس الطريقة التي استخدمت في اختبار العينة الواحدة تطبق هنا ضمن الافتراضات المنوه بها سابقاً .

باستخدام الحاسوب يمكن الحصول على النتائج لتوزيع t الشفعي مباشرة أو يمكن تطبيق الاختبار الإحصائي بطريقة مشابهة لذاك المستخدم في اختبار العينة الواحدة . وهو يتبع توزيع (t) وهو يعطى كما يأتي :

$$Test_4 = \frac{\bar{d}}{SE(\bar{d}) = \dfrac{\bar{d}}{SJ / \sqrt{n}}}$$

إذ تحسب درجة الحرية كما يأتي :

$$\text{df} = \text{n} - 1$$

بإعتبار :

n : هو عدد الأشفاع في العينة (الأزواج في العينة إذ يختار كل حيوان نبين على شكل زوجين مع بعض في كل عملية (اختبار لكل مجموعة)) .

J : متوسط الفروقات في العينة .

$SE\bar{d}$: الخطأ المعياري المقدر للتباين /الفروقات/ .

Sd : الانحراف المعياري المقدر للتباين بين أزواج أو أشفاع العينة .

نحصل على قيمة (P) إما من الحاسوب مباشرة باستخدام اختبار t الشفعي أو نحصل عليها باستخدام قيمة توزيع (t) في الملحق (A.3) لنقرر إما رفض أو قبول فرضية العدم وعادة نرفض فرضية العدم إذا كانت قيمة P < 0.05 .

وتحسب درجة الثقة ٩٥ % لمعدل الاختلاف الحقيقي كما يأتي :

$$\bar{d} \pm t_{0.05} SE\ (\bar{d}) = \bar{d} \pm t0.05\ (Sd\sqrt{n})$$

حيث نحصل قيمة $^{t}0.05$ بواسطة جدول توزيع (t) في الملحق مع حساب درجة الحرية = (n – 1) بالمطابقة إلى الاحتمالية ذات الذيلين مع مستوى بيتا β (٠.٠٥) للاحتمالية .

تمـــــــــارين

اختبارات t :

تم قياس حمض البولـة والكرياتنين في بلازما الدم روتينياً لتقييم وظائف الكليـة في قطط مريضة و سليمة ظاهرياً فحصلنا على النتائج الآتية من المخبر المختص :

حمض البولة 7.5 (mmol) ، وفي عينة عشوائية أخرى شملت 140قطـة سليمة ظاهريا، فإن متوسط قيمة حمض البولـة المقيسة كانت 9.7 (mmol) ، مع خطأ معياري 0.22 (mmol) .

هل يوجد دليل أو مؤشر على التغير في اختلافات النتائج بين المخبر والعينة العشوائية

الفصل العاشر
CHAPTER TEN
اختبار الفرضية رقم 2- اختبار F لمقارنة متوسطين أو أكثر
HYPO THESIS TEST 2 – THE F TEST COMPARING TWO MEANS OR MORE THAN TWO MEANS

مدخل Introduction :

أخذنا بالحسبان اختبارات الفرضية في الفصل التاسع اعتماداً على توزيع (t) والتي استخدمت لمقارنة المتوسطات . فعلى سبيل المثال أنجزنا اختبار t ستودنت لعينتين لمقارنة متوسط أوزان الجسم للنعاج المستوردة والشاهد المحلية . هذا الاختبار يكون صالحاً فقط لهذا إلا أننا نريد أن نتأكد أن كلتا المجموعتين لهما نفس التباين أو متشابهة بتبايناتها .

في هذا الفصل سنقوم بمراجعة بعض الاختبارات التي تتبع باختباراتها إحصائياً توزيع (F) والتي تقارن الفروقات بين مجاميع الدراسة . ويمكن أن تستخدم اختبارات F استخداماً واسعاً ضمن هذا المفهوم كنوع من الفروق لمقارنة متوسطين أو أكثر ليست هي غير شائعة كحالة من حالات دراسات علم الحيوان والطب البيطري .

اختبار F لتباينين متساويين

The F – Test for the Equality of two Variances

يدعى اختبار -F (F-test) غالباً باختبار تناسب التباين Variance Ratio ويمكن أن يستخدم للتقصي عن نمطين من البيانات المتجانسة ذات القيم المستمرة بينما يستخدم اختبار ليفانس Levenes test لتحليل فرقين أو أكثر في مجاميع الدراسة .

الافتراضات في توزيع -F : Assumptions of F- Distribution :

اختبار الفرضية هذا يعتمد على افتراضات أن العينات مستقلة بتركيبها وهي مختارة عشوائياً من حيوانات ذات بيانات موزعة طبيعياً .

الطريقة : The a Approach

١- حدد فرضية العدم – بأن نقول إن فرقين من حيوانات الدراسة ذات انحرافات معيارية متساوية $\sigma^2{}_1 , \sigma_2{}^2$ثم حدد الفرضية البديلة التي ترتكز عادة على أن التباينات بين حيوانات الدراسة متساوية .

٢- جمع وفحص البيانات ومن ثم تلخيصها برسوم خطية نقطية أو على شكل مستطيلات بسيطة أو باستخدام بوكس وويسكر وباستخدام هذه الرسومات نفحص التوزيع الطبيعي للبيانات حسب افتراضات توزيع -F .

٣- نختار الاختبار المناسب بالحاسوب أو نحسب يدوياً الاختبار الإحصائي .

وهنا نشتق الفروقات المقدرة بين حيوانات الدراسة بواسطة تناسب الانحراف المعياري في العينة ذات القيمة الأكبر (S_1^2) على الانحراف المعياري في العينة ذات القيمة الأصغر هذا التناسب يتبع توزيع -F ويعطى بالتناسب الآتي :

$$Test_5 = \frac{S_1^{\ 2}}{S_2^{\ 2}}$$

مع قيمة درجة حرية $n_1 - 1$, $n_2 - 1$.

بإعتبار أن درجة الحرية $n_1 - 1$ تمثل درجة الحرية للانحراف المعياري ذي القيمة الأكبر في /مقام/ التناسب بينما درجة الحرية $n_2 - 1$ تمثل درجة الحرية للانحراف المعياري ذي القيمة الأصغر في /مخرج بسط/ التناسب .

وبإعتبار أن n_1 و n_2 تمثلان حجمي عينتين .

٤- حدد قيمة P الاحتمالية : يمكن أن نحصل على هذه القيمة مباشرة من الحاسوب أو يمكن الحصول عليها باللجوء إلى جدول -F (الملحق A.5 ، جدول) . ففي هذا الجدول توجد عدة صفوف أو أعمدة تتطابق مع درجات الحرية للصورة /المقام/ /والمخرج/البسط/ . وهو ما يمثل تناسب -F (F ratio) أو قيمة F بالمعنى الإحصائي وبالنتيجة نحصل على نفس الجدول على قيمة P .

٥- ثم نستخدم قيمة P للحكم ما إذا كانت البيانات المدروسة متوافقة مع فرضية العدم. ومن ثم نقرر ما إذا كنا نرفض أو نقبل فرضية العدم وعادة ترفض فرضية العدم إذا كانت قيمة أقل من (٠.٠٥) ($P < 0.05$) .

وعادة ننجز هذا الاختبار لتقييم التساوي في الفروقات لتصل إلى نتيجة حول تطابق الافتراض في الاختبارات الأخرى مثل اختبارات (t) لعينتين . ولذلك من غير المحبب والممكن أننا سنحتاج إلى تطبيق درجة الثقة لتناسب فرقين . إذ أن درجة الثقة لتناسب الفرق تعتمد على توزيع -F .

مثال :

قارنا في مثالنا السابق باختبار (t) لعينتين متوسط معدل أوزان عينتين مختارتين عشوائياً لكل من الأغنام المستوردة والأغنام المحلية قبل فترة التزاوج .

١- من الرسم البياني باستخدام بوكس وويسكر وجد أن كلتا العينتان موزعتان طبيعياً .

٢- نريد أن نفحص فرضية التباينات المتساوية بحساب تناسب كلا الفرقين المقدرين وبإعتبار أن الانحراف المعياري للكسب الوزني لكلا النوعين من الأغنام المستوردة والشاهدة المحلية هي كما يأتي :

٢.٢٥٢ كغ و ٢.٤٩٧ كغ وكذا كان الفرقان المقدران لكلا النوعين من الأغنام هما :

$(٥.٠٧٢)^٢$ كغ و $(٦.٢٣٥)^٢$ كغ ولذلك فإن تناسب التباين الأكبر على التباين الأصغر كما يأتي :

$$\frac{6.235}{5.072} = 1.23$$

وبإعتبار أن عدد الأغنام المحلية الشاهد = ٣٠ فإن درجة الحرية

$$29 = 30 - 1 = n - 1$$

وكذا أمر الأغنام المستوردة = ٢٤ رأساً ولذا تكون درجة الحرية

$$23 = 24 - 1 = n - 1$$

وهكذا بمعرفة قيمة F ١.٢٣٢ في الجدول الملحق ((A.5(b) مع درجة حرية ٢٩ و ٢٣ نلاحظ أنه لا يوجد عمود لدرجة الحرية ٢٩ في البسط (الصورة) أو صف لدرجة الحرية ٢٣ في المقام (المخرج) فإننا نجد أن قيمة F = ١.٢٣ أقل من القيمة المجدولة (٢.٠٩) لدرجة حرية غير محدودة (∞) في البسط و ٢٠ درجة حرية في المخرج لقيمة $P / 2 = 0.025$ وهكذا فإن قيمة P أقل من ٠.٠٥

اختبار ليفيز لتساوي فرقين أو أكثر

Levene's Test for the Equality of Two or more Variances

عندما نستخدم الحاسوب يمكن أن نجد اختباراً بديلاً للفروقات المتجانسة يدعى اختبار ليفين Levenes test . ويستخدم لمقارنة فرقين . إنه مفيد بهذه الحالة في مفاهيم تحليل الفرق من اختبار الإحصائي المحسوب لاختبار ليفيز يمكن الحصول على توزيع -F .

ونلاحظ من البيانات المذكورة في مثال الكسب الوزني للأغنام وباستخدام البرنامج الإحصائي SPSS أنه يوجد في نتائج اختبار (t) لعينتين تضمين اختبار

ليفين للفرق المتساوي . وهذا يعطي قيمة $P = 0.617$ وهي متوافقة مع النتيجة المحصول عليها حساب تناسب التباينات .

تحليل التباين للمتوسطات المتساوية

The Analysis of Variance (ANOVA) for the Equality of Means

بعد تحليل التباين The analysis of variance (ANOVA) ليستخدم لوصف تقنيات إحصائية لمقارنة معدلات مجموعتين أو أكثر بالبحث والتقصي عن التباينات المناسبة فيما بينها . إن تحليل التباين يعتمد على اختبار تناسب التباين كاختبار -F الذي يقارن بين فرقين بفحص تناسبهما وربطهما بتوزيع - F .

ومع أن المفاهيم الأساسية لتحليل التباين يمكن أن نعبر عنها بعبارات بسيطة نسبياً فإن التفاصيل الرياضية لهذا الحساب يمكن أن نجدها في العديد من الكتب الإحصائية للمستويات المتقدمة في علم الإحصاء التي توضح جميعها نظرية تحليل التباين تحليلاً رياضياً مسهباً . إلا أننا من المهم في أيامنا هذه أن نكون على دراية بفهم النتائج التي نحصل عليها من الحاسوب والتي تعطينا نتائج تحاليل الفروق مباشرة .

فمن التحليل أي من البرامج الإحصائية ستُعرض لدينا المعلومات الآتية :

- جدول تحليل التباين Analysis of Variance Table الذي يوضح مصادر الاختلافات في المتغيرات أو العوامل المدروسة لمجاميع أو مجموعة محددة .
- كما هو مبين في الجدول رقم (١١) الذي يوضح مقارنة الاختلاف في تركيب نمطين من الخلطات العلفية بالمقارنة بالشاهد .

الجدول رقم ١١ : مقارنة الاختلاف في تركيب نمطين من الخلطات العلفية بالمقارنة بالشاهد

المجموعة المدروسة		
خلطة علفية رقم (٢)	خلطة علفية رقم (١)	شاهد
×	×	×
×	×	×
×	×	×
×	×	×
×	×	×
×		×
		×
		×

ومن تحليل هذه البيانات مثلاً سوف نحصل على القيم الآتية :

- مجموع المربعات Sum of Squares
- درجات الحرية Degress of Freedon

ومن ثم فإن مجموع المربعات هذه تقسم على درجات الحرية لنحدد قيمة متوسط المربعات meam square الذي يقدم إلينا قيمة متوسط التباينات المناسبة ، ومن ثم فإن متوسط المربعات المحصول عليها تقارن باستخدام اختبار -F كما سنشرحه مفصلاً في هذا الفصل لاحقاً . ومن أجل كل عامل مدروس تحدد فرضيةُ العدم معدلاتِ الحيوانات المدروسة للمجاميع المعرفة على افتراض أنها متساوية بمستويات محددة لذاك العامل المدروس .

أنماط تحليل التباين Particular Forms of AOV

هناك عدة أنماط لتحليل الفرق :

١- تحليل التباين وحيد الإتجاه One – Way Analysis of Variance :

وهو أبسط شكل من أنماط تحليل التباين لعامل واحد ، كما هو الحال في اختبار (t) لعينتين عندما يطلب منا تحليل مقارنة معدلات أكثر من مجموعتين.

٢- تحليل التباين وحيد الإتجاه لقياسات متكررة :

One – Way Repeated Measures

وهو ما يشابه اختبارات (t) الزوجية أو الشفعية ، مثلاً لمقارنة ثلاثة أنماط أو أكثر من أنماط من المعالجات الدوائية . فعندئذ مثلاً نقول إن هناك قياسات متكررة لكل مستوى من الجرعات المتكررة ولنفس العامل المدروس كما هو موضح في الجدول الآتي الذي يبين تحليل التباين الوحيد الإتجاه لقياسات متكررة لمجاميع

مـــن الحيوانـــات المعالجـــة بجرعـــات مختلفـــة بالمقارنـــة بالشـــاهد (الجدول رقم ١٢) .

جدول رقم (١٢) : مقارنة بين المجموعات المعالجة

مجموعات المعالجة			
الجرعة رقم (١)	الجرعة رقم (٢)	جرعة الشاهد	أرقام الحيوانات المعالجة
×	×	×	١
×	×	×	٢
×	×	×	٣
			٤
			٥
			K

٣- تحليل التباين باتجاهين Two – Way Analysis of Variance

وهي تقوم بفحص تأثير عاملين تحت استجابة معينة عندما يكون لدى هذه العوامل مستويان أو أكثر من التأثير . فعندئذ نقوم بإنشاء جدول الطريقة الثنائية حيث يمثل في كل صف وعمود مستويات لكل من العاملين المدروسين .

وعندما يكون هناك تداخل بين أحد المستويات مع أحد العاملين نقوم بعملية ضرب تداخلي لكلا المستويين وهو ما يدعى بالفعل المتبادل (التدافع) interaction بين العوامل المدروسة وهذا يكون فقط عندما تكون الاختلافات في مستويات أحد العاملين متشابهة مع المستويات المختلفة للعامل الآخر .

لنفترض على سبيل المثال أن هناك نمطين من الخلطات العلفية المخصصة لتحقيق أفضل نمو للحيوانات المدروسة وهي ما يراد مقارنته .

بإعتبار أن عدد الحيوانات التي أجريت عليها هذه الدراسة (٤٢) حيواناً وتم تقديم كل خلطة من هاتين الخلطات وبكميات مختلفة في اليوم الواحد بثلاثة مقادير مختلفة عن بعضها وبالتالي فإن لدينا ست مجاميع للدراسة كل مجموعة مكونة من سبعة حيوانات ومن ثم نقوم بمراقبة الكسب الوزني لهذه المجاميع خلال فترة ثلاث أشهر مثلاً باستخدام طرق تحليل التباين فإن من الممكن أن نقارن – ما إذا كان هناك اختلاف في المقادير الثلاث هذه من خلال الكسب الوزني للحيوانات – بين كل من نمطي الخلطات العلفية المدروسة .

هذا المثال يمكن أن يوضح بالجدول رقم (١٣) الممثل بالبيانات المتعلقة بأنماط التغذية في مجاميع الدراسة وبمقادير مختلفة .

تحليل التباين وحيدة الإتجاه One – Way AOV:

تقوم هذه الطريقة فكرة التركيز علـى تحليل مستويات عديدة لعامل وحيد .
وفـي كل مستوى تعرف مجموعـة من المشـاهدات فعلـى سبيل المثال يمكن أن تمثل المستويات معالجاتٍ مختلفة كما ذكرنا في مثالنا السابق .

إن الافتراض فـي تحليل التبـاين وحيد الإتجـاه أنـها تقـول إن العينـات الممثلـة لمجـاميع الدراسـة هـي مختـارة عشـوائياً ومسـتقلة عـن بعضـها Randomly and Independently وذات بيانـات موزعة طبيعياً Normally Distributed إلا أنـه يمكـن تحليل بيانـات ذات توزيـع طبيعـي تقريبـي مـع أنـه يتوقـع أن يكـون هنـاك تـأثير بسيط في النتائج المحللة .

جدول رقم (١٣) : **البيانات المتعلقة بأنماط التغذية في مجاميع الدراسة وبمقادير مختلفة**

الخلطات العلفية اليومية المقدمة			
كمية العلف (٣) كغ	**كمية العلف (٢) كغ**	**كمية العلف (١) كغ**	
××× ××××	××× ××××	××× ××××	**الخلطة العلفية رقم (١)**
××× ××××	××× ××××	××× ××××	**الخلطة العلفية رقم (٢)**

وإذا ما قمنا بالالتزام بتطبيق فرضية الاختبار فيجب أن نقوم بإجراء عملية تحويل للبيانات إلى المستويات المطلوبة Appropriate Transformation أو استخدام الطريقة البديلة للطريقة اللامعلميةلتحليل التباين وحيد الإتجاه (Kruskal – Wallis) أو الطريقة الثنائية لفريدمان (Friedman) . (Friedman two way)

طريقة تحليل التباين وحيد الإتجاه The Approach to one – Way AOV :

– حدد فرضية الاختبار (فرضية العدم أو الفرضية البديلة) المختارة .

اجمع البيانات واعرضها بنفس الطريقة المستخدمة باختبار (t) لعينتين باستثناء أنه لدينا أكثر من مجموعتين . ومن ثم اختبر فرضية الاختبار (التوزيع الطبيعي والتجانس في الفروق) .

استخدم الحاسوب لحساب الاختبار الإحصائي الذي نجده في جدول AOV وهو عبارة عن التناسب بين المجموعتين إلى مجاميع متوسط المربعات (أو ما يدعى بتناسب F) (F-ratio) . بإعتبار أن تناسب F يتبع توزيع F مع درجة حرية (K-1) و (n-K) . بإعتبار أن K : هو عدد المجموعات و N : هو إجمالي عدد المشاهدات في العينة .

انظر بعدئذ بقيمة (P) : وعادة تعطى قيمة P في جدول AOV باستخدام التحليل في الحاسوب ، أو يمكن أن تستخدم جدول توزيع F في الملحق ((A.5(a)) إذ نحدد في البداية درجة الحرية في جدول F ثم احسب قيمة (K-1) في الصورة

(البسط) (بين المجموعات) ثم استخدم القيمة K – n لحساب المخرج (المقام) (درجات حرية ضمن المجموعات) وهي ما تسمى درجات الحرية للخطأ الممكن حدوثه (Residual) .

استخدم قيمة P لمحاكاة البيانات برفض أو قبول فرضية العدم ومن ثم اتخذ القرار بناء على قيمة (P) إذ نرفض فرضية العدم Ho إذا كانت قيمة P أقل من ٠.٠٥ فإذا رفضنا فرضية العدم فيمكن أن نحتاج إلى أن ننشئ القواعد على التباين بين متوسطات مجاميع الدراسة .

– احسب درجة الشفعة للفروقات بين متوسطات المجاميع بنفس الطريقة كما هو الحال في اختبار (t) لعينتين .

مثال رقمي لطريقة تحليل التباين الوحيد الإتجاه Example of One – Way AOV

غذيت مجموعة من الكلاب بوجبات مجففة ومغطاة بعوامل يعتقد أنها تؤثر على تشكل حصيات وترسبات في الأسنان . وتقاس عادة هذه الترسبات في هذه الدراسة بمعايير تقدر النسبة المئوية للأسنان المغطاة بالترسبات وسماكة هذه الترسبات . فقد وجد أن (٢٦) كلباً تم اختيارهم عشوائياً واستخدمت في ثلاث معالجات وهي كالآتي :

المجموعــة الشــاهد ومجموعــة عولجــت باســتخدام بيروفوســفات (P_2O_7)

والمجموعة الثالثة عولجت باستخدام هيكساميتافوسفات (HMP) .

ومن ثم قياس سماكة هذه الترسبات لكل كلب تحت ظروف التجربة وبعد

أربعة أسابيع بعد تلقيه المعالجة كما هو موضح في الجدول رقم ١٤.

درجـة الثقـة ٩٥ % قـد حسـبت لكـل معـدل (الخـاص بالترسـبات النسـبية

مستعملين معدل المربعات المتبقي residual mean square = ٣٥٣ من جدول

AOV والمحسوب بتقدير التباين) كما هو موضح في الشكل رقم (١٧) .

الحل :

١- بالنسبة لنظرية العدم ، هل أن المعدل الحقيقي للترسبات في المعالجات الثلاث

متساوية وفي الفرضية البديلة هذه الترسبات ليست متساوية عند جميعها .

٢- من الشكل السابق نجد أن البيانات موزعة طبيعياً تقريباً .

٣- تناسب F في جدول (AOV) والذي سيعرض فيما يأتي نجد أن قيمة $P = 0.005$

.

٤- ولذلك فمن غير المحتمل أن يكون معدل كامل هذه المجموعات الثلاث متساوياً ولذلك نرفض فرضية العدم .

٥- المعدلات ودرجات الثقة المترافقة ستستعرض في جدول لاحق .

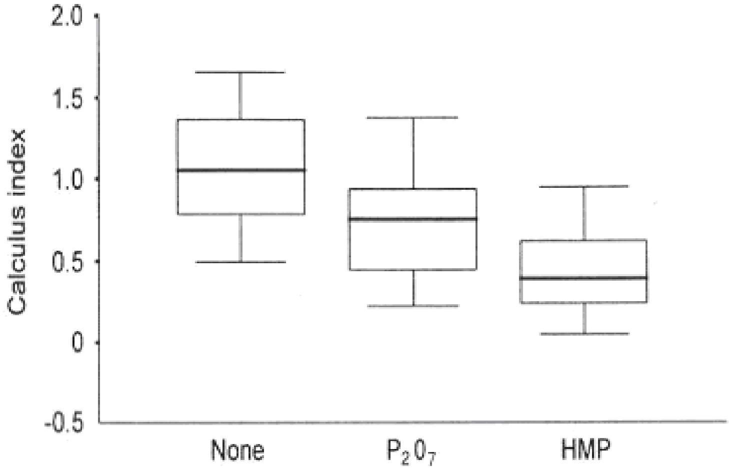
حيث نجد أن الترسبات كانت بشكل عند مجموعة كلاب الشاهد ويمعنوية كبيرة P = 0.004 (مع معدل ثقة للترسبات ١.٠٩ مع درجة ثقة ٩٥ % تراوحت بين (١.٤١ – ٠.٧٦) منه من مجموعة الكلاب والمعالجة بـ HMP بمعدل ثقة للترسبات ٠.٤٤ ، درجة ثقة ٩٥ % تراوحت بين (٠.١٩ – ٠.٦٨) . قدرنا التباين بين هذه المتوسطات فكـان ٠.٦٥ ومـع درجـة ثقـة ٩٥ % للفـروق الحقيقيـة للمعـدلات فكـان (٠.١٩ – ١.١١) المقارنتان الأخيرتان للمتوسطات لم تكن معنوية (قيمة P< 0.05) (أي ما بين المعالجـة بـ P_2O_7 وكلتـا المجمـوعتين الأخيـرتين أو بـين HMP وكلتـا المجمـوعتين الأخيرتين كما هو موضح في الجداول ١٤ ذات أرقام و ١٥ و ١٦ و ١٧ و ١٨ و الشكل رقم ١٧.

الجدول رقم (١٤): تمثيل التجربة لتغذية الكلاب بالبيانات التالية

السماكة في مجموعة HMP	رقم الكلب في المجموعة الثالثة (P_2O_7)	السماكة في مجموعة P_2O_7	رقم الكلب في المجموعة الثانية (P_2O_7)	السماكة عند كلاب التجربة	رقم كلب التجربة (الشاهد)
٠.٣٤	١٩	٠.٣٤	١٠	٠.٤٩	١
٠.٠٥	٢٠	٠.٧٦	١١	١.٠٥	٢
٠.٥٣	٢١	٠.٤٥	١٢	٠.٧٩	٣
٠.١٩	٢٢	٠.٦٩	١٣	١.٣٥	٤
٠.٢٨	٢٣	٠.٨٧	١٤	٠.٥٥	٥
٠.٤٥	٢٤	٠.٩٤	١٥	١.٣٦	٦
٠.٧١	٢٥	٠.٢٢	١٦	١.٥٥	٧
٠.٩٥	٢٦	١.٠٧	١٧	١.٦٦	٨
		١.٣٨	١٨	١	٩

مجموعة HMP	مجموعة P_2O_7	المجموعة الشاهد	
٩	٩	٩	حجم العينة
٠.٤٤	٠.٧٥	١.٠٩	(المتوسط الحسابي)
٠.٢٩	٠.٣٧	٠.٤٢	SD (الانحراف المعياري)
٠.٠١	٠.١٢	٠.١٤	متوسط الخطأ المعياري SEM
(٠.٧٤ – ٠.١٣)	(١.٠٣ – ٠.٤٦)	(١.٣٧ – ٠.٨١)	درجة الثقة : CI ٩٥ % *

الشكل رقم (١٧): توزيع بوكس وويسكر لمجموعات التجربة في الكلاب



جدول رقم (١٥) : النتائج المحصول عليها من الحاسوب لتحليل التباين

لبيانات الكلاب المعالجة ضد الترسبات السنية لثلاث مجموعات

درجة الثقة ٩٥ % الحد الأدنى	درجة الثقة ٩٥ % الحد الأعلى	الخطأ المعياري (Std . Error)	الانحراف المعياري SD	معدل الترسبات السنية	N العدد		
٠.٧٦٤١	١.٤١٣٧	٠.١٤٠٨	٠.٤٢٢٥	١.٠٨٨٩	٩	شاهد	المعالجة
٠.٤٦٢٦	١.٠٣٠٧	٠.١٢٣٢	٠.٣٦٤٥	٠.٧٤٦٧	٩	P_2O_7	
٠.١٩٤٥	٠.٦٨٠٥	٠.١٠٢٨	٠.٢٩٠٧	٠.٤٣٧٥	٨	HMP	
٥٩٠٩	٠.٩٤٩١	٨.٦٩٧E-20	٠.٤٤٣٥	٠.٧٧٠٠	٢٦	الإجمالي	

لاحظ أن E – 20 تعني أن القيمة مضروبة $10X^{-2}$

جدول رقم (١٦) : اختبار التجانس للفروقات

	اختبار ليفين Levene statistic	df 1	df 2	المعنوية Sig
الترسبات السنية	0.855	2	23	0.438

جدول رقم (١٧) : تحليل الفرق (AOV) .

	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F	المعنوية Sig
الترسبات السنية بين المجموعات	١.٨٠٥	٢	٠.٩٠٢	٦.٦٦٨	٠.٠٠٥
الترسبات السنية ضمن المجموعات	٣.١٢٢	٢٣	٠.١٣٥		
الإجمالي	٤.٩١٧	٢٥			

جدول رقم (١٨) : المقارنات المتعددة

المجموعة I،المجموعة J	معدل الفرق (J-I)	الخطأ المعياري	المعنوية α*	درجة الثقة ٩٥%	
				الحد الأدنى	الحد الأعلى
شاهد P_2O_7	٠.٣٤٢٢	٠.١٧٩	٠.١٨٢	٠.١٠٥٥	٠.٧٨٩٩
HMP	٠.٦٥١٤ *	٠.١٧٩	٠.٠٠٤	٠.١٨٩٩	١.١١٢٩
شاهد P_2O_7	٠.٣٤٢٢	٠.١٧٣	٠.١٨٢	٠.٧٨٩٩	٠.١٠٥٥
HMP	٠.٣٠٩٢	٠.١٧٩	٠.٢٩١	٠.١٥٢٣	٠.٧٧٠٧
شاهد P_2O_7	٠.٦٥١٢ *	٠.١٧٩	٠.٠٠٤	١.١١٢٩	٠.١٨٩٩
HMP	٠.٣٠٩٢	٠.١٧٩	٠.٢٩١	٠.٧٧٠٧	٠.١٥٢٣
* = تعني معدل الاختلاف في المعنوية عند مستوى المعنوية ٠.٠٥ = α					

تمـــــــــارين

تظهر البيانات التالية أوزان الكبد (كغ) مأخوذة لعينة عشوائية من الأبقار في مزرعتين جنوب غرب بريطانية خلال فترة جائحات المتورقات الكبدية عند الأبقار.

١- أوجد أوجه التشابه في الأوزان في كلا المزرعتين باستخدام اختبار الفرضية المناسب

٢- هل التغير في كلا المشاهدات عند كلا المزرعتين متشابهين

المجموعة الأولى:

١٨ – ١٨.٥ – ١٨.٩ – ١٨.٢ – ١٧.٩ – ١٥.٩ – ١٦.٨ – ١٨.٢ – ١٧.٣

– ١٧.٥ – ١٧.٧ – ١٧.٨ – ١٧.١ – ١٧ – ١٦.٣

المجموعة الثانية:

١٤.٣ – ١٣.٢ – ١٧.٣ – ١٤.٩ – ١٦.٤ – ١٦ – ١٨.٦ – ١٧.٣ – ١٥.٥

– ١٦.٨ – ١٥.٧ – ١٨ – ١٥.٢

الفصل الحادي عشر

ELEVEN CHAPTER

اختبارات الفرضية -3- اختبار مربع كاي – مقارنة النسب المئوية

HYPOTHESIS TESTS 3- THE CHI – SQUARE TEST – COMPANING PROPORTIONS

مدخل Introduction :

ناقشنا في الفصلين السابقين اختبارات الفرضية للوسط الحسابي مع مقاييس الميول المركزية والتشتت للمتغير الكمي . أما في هذا الفصل فإننا سنقوم بوصف بعض اختبارات الفرضية المتعلقة بالنسب المئوية وذاك الحد الذي يلخص المشاهدات للمتغير الثنائي الحد (بينري) Binary variable بالإضافة إلى ذلك إلى أننا سنقوم بشرح كيف نقوم بتحليل بيانات عندما يكون المتغير الذي له أكثر من فئتين .

وسنعيد تسمية أن المتغير الثنائي الحد (بينري) هو متغير نوعي (فئوي) عندما تكون هناك فئتين فقط لاستجابة متغير ما ، وهاتين الفئتين إما أن تنسب إلى حالة نجاح (الحالة) Success والثانية هي حالة (فشل) Failure .

على سبيل المثال نريد التقصي عن داء البريميات المسبب بجراثيم البريميات Leptospira لحدوث الإجهاض عند الأبقار . إن الاختبار المستخدم

للتقصي عن وجود أضداد جراثيم البريمية يلخص على شكل استجابة متغير ثنائي الحد ، فإما أن تكون هناك حالات إيجابية أي حالة (نجاح) أو أن تكون نتيجة الاختبار سلبية أي حالة (فشل) .

وكذلك مقارنة النسب المئوية للأبقار الإيجابية بداء البريميات و الموجودة ضمن مجموعتين، المجموعة الأولى أجهضت والمجموعة الثانية ولدت عجولاً طبيعية .

اختبار الفرضية لنسب مئوية مفردة

Testing A Hypothesis about A Single Proportion

١- الطريقة The Approach

من خصائص توزيع العينة التي تكون على شكل نسبة مئوية نعرف أن مثل هذا التوزيع طبيعي تقريبي . فإذا كان حجم العينة (n) كبيراً فإن النسبة المئوية (P) هي عبارة عن تقدير غير متحيز /منحرف/ unbiased لنسبة الحيوانات المدروسة (π) والخطأ المعياري للنسبة المئوية في العينة يقدر كما يأتي :

$$\sqrt{P(1-P)/n}$$

من هذه المعلومات نختبر فرضية النسبةِ المئوية للنجاحات في مجموعة أو مزرعة أو قطيع مفرد من الحيوانات مستعملين الطريقة الآتية :

١- حدود فرضية العدم للنسبة المئوية لعدد النجاحات في الحيوانات المدروسة أنها مساوية لقيمة محدودة (1 π) . حدد الفرضية البديلة غير مساوية للقيمة (1 π) .

٢- اجمع البيانات باختبار عينة عشوائية من حيوانات الدراسة وصنف كل قيمة مفردة من حالات النجاح أو الفشل .

٣- احسب الاختبار الإحصائي test statistic مع افتراض وجود توزيع طبيعي تقريبي

$$Test_6 = \frac{|p-\pi_1| - \frac{1}{2n}}{\sqrt{\frac{\pi_1(1-\pi_1)}{n}}}$$

بإعتبار أن :

P : هي النسبة المئوية المشاهدة لعدد حالات النجاح في التجربة .

n : هي عدد أفراد الدراسة في العينة المأخوذة .

الخطوط العمودية (| |) لكلا الجانبين للاختلاف في النسب المئوية تشير إلى تجاهل علاقة الاختلاف .

طرحنا القيمة 1/(2n) – وتدعى عامل التصحيح الثابت (Continuity Correction) – من هذا الفرق لنضع الافتراض بين أيدينا أننا نتعامل مع بيانات ذات توزيع طبيعي

ولنقربْ البيانات الاسمية التوزيع ، وهذا التأثير لعامل التصحيح يُتجاهل عندما يكون حجم العينة كبيراً .

٤- حدد قيمة (P) باستعمال جدول التوزيع الطبيعي المعياري Stondard Normal Distribution (الجدول A1) .

٥- اجعل القرار اعتماداً على رفض أو قبول فرضية العدم حسب قيمة (P) .

٦- احسب درجة أو حد الثقة للنسب المئوية لحالات النجاح .

حد الثقة ٩٥ % لقيمة π تحسب بالعلاقة

$$P \mp 1.96\sqrt{P(1-P)/n}$$

مثال Example :

لنفترض أننا نريد التقصي عن تناسب الجنس في الأرانب البرية في الفصل الحالي مثلاً ، أشرنا إلى أن تناسب الجنس للولادات الحية محسوب بناء على تعداد الإناث . سجلاتنا تظهر بواسطة عينة عشوائية لـ ٢٩٧ ولادة منها π ١٦٧ أنثى ، السؤال هنا هل هناك دليل لبعض العوامل (كالعوامل الوراثية وهي عوامل خطورة لارتفاع نسبة نفوق الأجنة الذكور) تؤثر على تناسب الجنس ؟

١- فرضية العدم : النسب المئوية للولادات الحية للإناث والذكور في هذه الأرانب البرية معرفة بأنها مساوية للقيمة ٠.٥ بالنسبة للفرضية البديلة تتقصى عن الحالة

العكسية أي أن الولادات الحية من الذكور والإناث في هذه الأرانب البرية معرفة بأنها غير مساوية للقيمة ٠.٥ .

٢- النسبة المئوية للإناث المأخوذة من العينة العشوائية الإجمالية (٢٩٧) ولادة حية هي :

$$\frac{167}{297} = 0.562$$

الاختبار الإحصائي في هذه الحالة يحسب كما يأتي :

$$Test_6 = \frac{|0.562 - 0.5| - 1(594)}{\sqrt{0.5(1-05)/(297)}} = 2.08$$

من القيمة المدرجة أعلاه في العلاقة المنوه بها ٢.٠٨ = ٢ وبالإشارة إلى الجدول (A.1) نجد أن قيمة P = 0.0375 (ويمكن تقريبها إلى فاصلتين عشريتين فتكون قيمة P المقربة (P = 0.04) .

٥- حد الثقة ٩٥ % للنسبة المئوية الحقيقية لولادات الإناث الحية هي :

$$\begin{aligned}
&= 0.562 \mp 1.9 \times SE \\
&= 0.562 \mp 1.96 \times \sqrt{0.562(1-0.562)/(297)} \\
&= 0.51 \\
&\text{أو} \\
&= 0.62
\end{aligned}$$

أي أن حد الثقة ٩٥ % يتراوح من (٠.٥١ إلى ٠.٦٢) .

مع أن قيمة P = 0.04 تعطينا نتيجة برفض فرضية العدم ، فإن الحد الأدنى لدرجة الثقة لا يزيد عن ٠.٥٠ وهو مساوٍ لنسبة الذكور والإناث ، وهنا يجب أن نكون حذرين في تركيزنا كثيراً على استنتاج أن هناك بعض العوامل تؤثر على تناسب الجنس . هذا المثال يعطينا أيضاً تصور كم يجب أن نتوخى الحذر والدقة بالنسبة لتعقيدات وتفسيرات القيمة الاحتمالية لقيمة (P) .

مقارنة نسبتين مئويتين – مجموعات مستقلة

Comparing Two Proportions – Independent Groups

نريد على سبيل المثال أن نجري تقصياً عن الإجهاضات عند الأغنام وعن تأثيرها على انتشار البروسيلة الغنمية عند الأغنام .

بهذه الحالة اختيرت عينة عشوائية من الأغنام المجهضة سواء تلك المجهضة حديثاً (غير المعالجة) أو تلك المعالجة بالصادات الحيوية .

وهنا يراد إجراء مقارنة بين هاتين المجموعتين المعالجة وغير معالجة وبعد فترة زمنية معينة بالنسبة لتلك الحيوانات المشخص لديها وجود حالات إيجابية بالاختبارات المصلية للأغنام المدروسة (حالات النجاح أو الحالات الإيجابية أو الحالات المخموجة) في هاتين المجموعتين المستقلتين من حيث المشاهدات .

لتحليل هذا النوع من الدراسات نعتبر أن حيوانات الدراسة من مجاميع حيوانية مستقلة عن بعضها ونستخدم النسب المئوية للعينة P1 و P2 ونقدر النسب المئوية لحيوانات

الدراسة $\pi 1$ و $\pi 2$. وعندئذ نختبر فرضية أن النسب المئوية لحيوانات الدراسة متشابهة بطريقة واحدة أو بطريقتين .

١- يمكننا استخدام اختبار مربع كاي (χ^2) الذي سنورد شرحه بعد قليل والذي يعتمد على توزيع مربع كاي .

٢- الطريقة الثانية أن نقوم باستخدام التقريب الطبيعي إلى التوزيع ثنائي الحدين Binomial Distribution ونشتق الاختبار الإحصائي مع تقريب للتوزيع الطبيعي .

- وباستخدام جدول 2 × 2 الاحتمالي يمكن أن نلخص النتائج في مثالنا السابق بواسطة عرض التكرارات بما يدعى بطريقة التكرار ذي الإتجاهين Two – Way frequency أو ما يدعى بجدول 2 × 2 .

مقارنة نسبتين مئويتين في جدول 2 × 2 وباستخدام اختبار مربع كاي

Comparing Two Proportions in a 2 × 2 Table Using the Chi – Square Test

إذا لم يكن هناك توافق بين الاستنتاج والمجموعة المدروسة ، نتوقع نسباً من حالات النجاح بأن تكون متشابهة في المجموعة المدروسة الثانية .

وهكذا يمكننا أن نقارن بين النسبتين بالتقصي للتوافق بين العاملين واللذان يعرفان جدول الاحتمالية (2 × 2) .

كلمة عامل (a Factor) في هذا المفهوم هو ذاك المتغير مع واحد أو أكثر من الفئات ضمن الأفراد المصنفة . وفرضية العدم في هذه الحالة تحدد فيما إذا كان لا يوجد توافق بين العاملين (النتيجة والمجموعة) وإثبات أن التشابه لنسبتي المجموعتين هي متساوية .

ولكي نحسب الاختبار الإحصائي باستخدام هذه الطريقة يجب أن نقارن تكرار ما نشاهده في كل خلية لجدول الاحتمالية (2×2) للتكرار . وتلك الخلية وإثبات من خلال لفرضية العدم أن نسب النجاحات في كلتا المجموعتين متساوية فإذا كانت فرضية العدم حقيقية فإننا نتوقع نسبة إجمالية من عدد النجاحات وهي $(a + b) / n$ وتطبق على كلتا المجموعتين (1 , 2) إن النجاحات المتوقعة تحت شروط فرضية العدم هي :

المجموعة (١) : $(a + c) \times (a + b) / n$

المجموعة (٢) : $(b + d) \times (a + b) / n$

الأعداد المتبقية في كل مجموعة يتوقع أن تحسب من عدد حالات الفشل ومما سبق يمكننا أن ننشئ جدولاً التكراراتُ المتوقعة في كل خلية فيه مطابقة لكل تكرار مشاهد كما هو موضح في الجدول ١٩ و ٢٠.

فإذا وجدنا الفروق بين التكرارات المشاهدة والمتوقعة كبيرة نرفض فرضية العدم .

نقرر فيما إذا كانت الفروق كبيرة بحساب الاختبار الإحصائي المناسب الذي يقرب بواسطة توزيع مربع كاي .

وأشرنا بالاختبار الإحصائي هنا إلى الجدول (A.4) لتحديد قيمة (P) .

الجدول رقم (19) : التكرارات المشاهدة.

الصف			**المجموعات**	
المجموع الإجمالي	٢	١		
a + b	b	a	نجاح	نتائج
c + d	d	c	فشل	
الأعمدة			المجموع الكلي	
المجموع الإجمالي	b + d	a + c		
		n = a + b + c + d		
$P1 = \frac{a}{a+c}$		$P2 = \frac{b}{b+d}$ نسب النجاحات المشاهدة		

الجدول رقم (٢٠) : التكرارات المتوقعة في الخلايا الأربعة من الجدول

المجموعة رقم (٢)	**المجموعة رقم (١)**	**النتائج**
$\frac{(b+d)(a+b)}{n}$	$\frac{(a+c)(a+b)}{n}$	**النجاحات**
$\frac{(b+d)(c+d)}{n}$	$\frac{(a+c)(c+d)}{n}$	**الفشل**

الافتراضات Assumptions :

نفترض في اختبار مربع كاي للتوافق في جدول الاحتمالية مع صفين وعمودين أن الأفراد مختارة عشوائياً من المجتمع الحيواني المدروس .

– من اختبار مربع كاي (χ^2) للتوافق في جدول (2×2) هو غير صالح إذا كانت التكرارات المتوقعة في أي واحدة من الخلايا الأربعة أقل من ٥ تكرارات وفي هذه الحالة نطبق اختبار فيشر Fishers Exact Test وهذا يتضمن حساب الاحتمالية الدقيقة لجدولنا المنوه به .

وفي هذه الحالة يفضل استخدام الحاسوب للحصول على نتيجة سريعة .

الطريقة The Approach :

– حدد فرضية العدم أن كلتا النسبتين لمجموعتي الدراسة متساوية أو معادلة بأن تقول لا يوجد هناك توافق بين كلا العاملين المدروسين .

وتحدد الفرضية البديلة عموماً بأن تقول إن كلتا النسبتين غير متساوية أو يوجد توافق بين عاملي الدراسة .

– اجمع البيانات واعرض التكرارات المشاهدة في جدول الاحتمالية (2×2).

– استخدم الطريقة المناسبة لاختبار مرب مربع كاي (χ^2) Chi – Squared test في الحاسوب .

إذا أردنا الحساب يدوياً فعندئذ نحسب التكرارات المتوقعة في كل خلية في الجدول إذا كانت فرضية العدم حقيقية وهنا تحسب قيمة الاختبار الإحصائي لمربع كاي كما يأتي :

$$Test_7 = \frac{|O - E|}{E} - 0.5$$

مع درجة حرية واحدة .

بإعتبار أن :

O : التكرار المشاهد .

E : التكرار المتوقع .

القيمة المطلقة الموجودة بين التكرار المشاهدة والمتوقعة لنتجاهل الإشارة الموضوع للفرق بينهما .

تحسب العلاقة المدرجة أعلاه لكل خلية في الجدول .

أما القيمة (0.5) فهي تعرف بما يسمى عامل تصحيح ياتس Yates correction وهو عامل يضاف لتجنب الانحراف قيمة الحساب . هذا الانحراف يمكن أن يظهر لأننا نقرب توزيع $\chi^٢$ إلى توزيع طبيعي مع وجود بعض التوزيعات المشكلة على فئات أو ترتيبات .

- إن العلاقة المدرجة أعلاه لحساب الاختبار الإحصائي والذي يعرف بالاختبار 7 (Test 7) عندما يكون جدول الاحتمالية (2 × 2) يتضمن صفين وعمودين لكن لتقييم الشكل أسرع يحسب هذا .

$$Test_7 = \frac{n\,(/\,ad - bc\,/ - 0.5n)^2}{(a+b)\,(c+d)\,(a+c)(b+d)}$$

- احصل على قيمة (P) ، على العموم تحسب بواسطة الحاسوب لكن يمكن أن تشتق من الإشارة إلى القيمة المحسوبة بالاختبار الإحصائي في جدول توزيع مربع كاي (χ^2) ، (في الجدول A.4) .

والاختبار الإحصائي هذا يشمل درجة حرية واحدة وهي تعادل عدد الصفوف مطروحاً منها الرقم ١ (مثلاً ٢ – ١ = ١) وكذا عدد الأعمدة مطروح منها الرقم (١) (مثلاً ٢ – ١ = ١) .

- اتخذ قراراً برفض أو قبول فرضية العدم .

عادة كما تعرفنا سابقاً ترفض فرضية العدم عندما لا يوجد هناك توافق عندما تكون قيمة P أقل من ٠.٠٥ .

- احسب درجة الثقة للاختلاف في النسب المئوية . وإذا لم تتمكن من الحصول على درجة الثقة المشار إليها بالحاسوب يمكننا حساب درجة الثقة ٩٥ % كما يأتي:

$$(P_1 - P_2) \mp 1.96 Se\,(P_1 - P_2)$$

$$(P_1 - P_2) \mp 1.96\ x\ \sqrt{\frac{P1\,(1-P_1)P_2\ (1-P_2)}{n-1}}$$

اختبار الترافق في جدول الاحتمالية R X C

Testing Associations in an RXC Contingency Table

يمكننا أن نتوسع في اختبار مربع كاي (χ^2) للتوافق في جدول 2 × 2 في جدول الاحتمالية والذي يتصف بوجود صفوف (r) وأعمدة عددها c ونفترض أنه يوجد سوا . في الصفوف (r) أو (c) أو كليهما قيم أكبر من القيمة (2) . وفي هذه الحالة يتركز الاهتمام في تحديد ما إذا كانت القيم في الصفوف أو الأعمدة مرتبطة ببعضها بطريقة أو بأخرى . ونختبر في هذه الحالة فرضية العدم لعدم التوافق بحساب مربع كاي كاختبار إحصائي لهذه الحالة.

الافتراضات : Assumption

نفترض أن كلا المتغيرين هو على شكل فئات ونعتبر أن البيانات الموجودة في جدول الاحتمالية هذا عبارة عن قيم لتكرارات معينة . ونفترض أن البيانات مستقلة بمعنى أن أي قيمة لأي حيوان فردي غير مكررة لأكثر من مرة واحدة .

الطريقة العامة : The General Approach

الطريقة لاختبار فرضية العدم أنه لا يوجد توافق في حيوانات الدراسة بين المتغيرين الفئويين المدروسين المعرفين في جدول الاحتمالية RXC المطابق لجدول 2 × 2.

١- حدد فرضية العدم والفرضية البديلة .

٢- اجمع البيانات ومثلها في جدول الاحتمالية .

٣- اختر أيقون اختبار مربع كاي Chi – Squared test في الحاسوب أو حساباً يدوياً بحساب الاختبار الإحصائي الذي يقدر توزيع مربع كاي والذي يمثل بالعلاقة الآتية .

$$Test_8 = \sum \frac{(O-E)^2}{E}$$

تحسب درجات حرية كما يأتي : (r-1) (C-1) .

بإعتبار أن :

O : تمثل التكرارات المشاهدة .

E : تمثل التكرارات المتوقعة .

٤- حدد قيمة (P) إما بالحاسوب أو بجدول مربع كاي $\chi^2 = 169$.

٥- قرر قبول أو رفض فرضية العدم مع اعتبار أن قيمة (P) ترفض فرضية العدم إذا كانت قيمة (P) أقل من ٠.٠٥ ولاحظ هنا أننا طالما لم نقدر تأثير المتغيرات بعضها على بعض لا نستطيع أن نحسب درجة الثقة الخاصة بها .

مثال :

التدريب الآتي في تقنيات التلقيح عند الأبقار . بإعتبار أن مركز التلقيح الاصطناعي قارن ثلاث طرق للتدريب حققت معدلات نجاح أدرجت في الجدول رقم (٢١) .

جدول رقم (٢١) : مقارنة ثلاث طرق للتدريب في تقنيات التلقيح الإصطناعي

المجموع	الطريقة الثالثة	الطريقة الثانية	الطريقة الأولى	
٧٢٨	٢٦١ (٢٨١.٥)	١٩٢ (١٨٧.٧)	٢٧٥ (٢٥.٨٨)	حامل
٢٦٥	١٢٣ (١٠٢.٥)	٦٤ (٦٨.٣)	٧٨ (٩٤.٢)	غير حامل
٩٩٣	٣٨٤	٢٥٦	٣٥٣	المجموع
٠.٧٣	٠.٦٨	٠.٧٥	٠.٧٨	النسب المئوية للحامل
	٠.٦٣ إلى٠.٧٣	٠.٦٩ إلى ٠.٨٠	٠.٧٣ إلى ٠.٨٢	**CI: 95%** النسب المئوية الحقيقية

السؤال هنا هو هل يوجد أي دليل للاعتقاد أن طرق التدريب الثلاثة تظهر معدلات مختلفة من النجاح ؟

١- فرضية العدم تكون حقيقية بالنسبة لمعدلات الحمل وتكون متشابهة في طرق التدريب الثلاث . وأي طريقة أخرى للتعبير عن فرضية العدم هذه تبين أنه

لايوجد توافق بين طرق التدريب الثلاث وكذا الأمر بالنسبة لحالة الحمل عند الأبقار . الفرضية البديلة هي في هذه الحالة أن معدلات الحمل غير متساوية في الطرق الثلاث .

٢- اجمع البيانات (وهي معروضة في الجدول السابق) .

٣- التكرارات المتوقعة والمطلوبة : إذا أنجز الاختبار يدوياً (موضحة بالأرقام الموضوعة بين قوسين في الجدول السابق) . فالقيم المتوقعة للأبقار الحوامل للطريقة الأولى هي :

$$(353\times 728) / (993) = 25880$$

وهكذا بالنسبة لطرق التدريب الأخرى .

ينجز هذا الاختبار الإحصائي كما يأتي :

$$Test_8 = \frac{(C275-258.80)^2}{258.80} + \frac{(78-94.20)^2}{94.20} + \frac{(192-187.68)^2}{187.86}$$

$$+\frac{(64-68.32)^2}{68.32} = \frac{(261-281.52)^2}{281.52} + \frac{(123-102.48)^2}{102.48}$$

$$= 1.015 + 2.787+0.099 + 0.273 + 1.496 + 4.110 = 7.8$$

٤- القيمة المشار إليها هذه هي قيمة توزيع مربع كاي (χ^2) والتي نجدها في الجدول (A.4). وهذه تتوافق مع درجات حرية ٢ = (١ - ٢) (١ - ٣) فإننا نحصل على قيمة P محصورة بين المجالين $0.001 > P < 0.01$ وبالتحليل الإحصائي باستخدام الحاسوب نحصل على قيمة ($0.008 = P$) .

٥- وهكذا فإننا نملك دليلاً لرفض فرضية العدم لأن معدلات النجاح هي متشابهة في طرق التدريب الثلاث .

مقارنة نسبتين مئويتين لزوجين من المشاهدات

Comparing Two Propositions
Paired Observations

يتركز اهتمامنا أحياناً في مقارنة نسبتين مئويتين عندما يكون لدينا زوجان من النتائج على شكل متغيرات من النوع البينيري (ثنائية الحد) لنفترض أننا نريد دراسة عينة عشوائية عددها (m) من الحيوانات وتُقُصي في كل حيوان بطريقتين . نلاحظ استجابة كل عضو من النتائج الشفعية في حالة نجاحها أو فشلها إذ تستخدم كلتا الحالتين على شكل متغيرات من نوع بينيري .

فعلى سبيل المثال إذا أردنا مقارنة فعالية التشخيص لكلتا الطريقتين للكشف عن تريكوموناس الجنينية في الثيران فإن العينة المأخوذة من كل ثور تصنف بكل من الطريقتين نتيجةً إيجابية (حالة نجاح) أو نتيجةً سلبية (حالة فشل) .

تُحديد أعداد الأزواج في كلتا الطريقتين . وفيما يأتي نبين التكرارات المشاهدة في نماط النتائج المحصول عليها من تحليل كلتا العينتين .

التكرارات المشاهدة (A)

نمط الحالة	النتائج باستخدام كلتا الطريقتين		للتكرار
	١	٢	
١	نجاح	نجاح	E
٢	نجاح	فشل	F
٣	فشل	نجاح	G
٤	فشل	فشل	H
m			المجموع

جدول الاحتمالية لكلتا الطريقتين للتكرارات المشاهدة(B)

		الطريقة (٢)		مجموع أعداد الأزواج
		نجاح	فشل	
الطريقة (١)	نجاح	e	f	e + f
	فشل	g	h	g + h
				m = e + f + g + h

استخدمنا اختبار ماك نيمارس McNemars لاختبار فرضية العدم التي تعتمد على أن النسب المئوية الحقيقية لعدد النجاحات باستخدام كلتا الطريقتين متساوية . وهذا من العلاقة

(في الطريقة الأولى) $P1 = (e + f) / m$

(في الطريقة الثانية) $P2 = (e + g) / m$

ويكون فرق التقدير في كلتا الطريقتين على الشكل الآتي :

$$P1 - P2 = (f - g) /m$$

وهنا يُركز فقط على الأزواج (الأشفاع) المختلفة (أي ليست من نوع نتيجة واحدة أي إما إيجابي سلبي أو سلبي إيجابي) بينما تُتجنب النتائج من نوع واحد أي إيجابي مع إيجابي (نجاح – نجاح) أو سلبي (فشل مع فشل) كما هو المبدأ في الاختبار الإحصائي مك نيمارس .

يعتمد اختبار ماك نيمارس Mc Nemars على التكرارات المشاهدة f و g وتطابقها مع التكرارات المتوقعة تحسب تحت شروط فرضية العدم .

الافتراضات : Assumptions

يفترض أن يكون كلا المتغيرين (حالة النجاح والفشل) تابعين لمتغيري الدراسة ويشاهد كل متغير عند كل حيوان مختار عشوائياً على شكل شفعي .

الطريقة : The Approach

١- حدد فرضية العدم التي تعتمد على أن النسب المئوية لحالات النجاح في كل حيوانات الدراسة متساوية . حدد الفرضية البديلة وهي تعتمد عموماً على أن هناك نسبتين مئويتين غير متساويتين .

٢- اجمع البيانات واعرضها في جدول تكراري .

٣- اختر أيقون اختبار مك نيمارس Mc Nemars عند استخدام الحاسوب أو تحسب يدوياً باستخدام الاختبار الإحصائي الذي يقرب توزيع مربع كاي وهو كما يأتي :

$$Test_{(9)} = \frac{(|f-g|-1)^2}{f+g}$$ (اختبار رقم ٩)

مع درجة حرية واحدة .

يمكن أن تجد أحياناً أن بعضهم يستخدم مك نيمارس Mc Nemars بالاختبار الإحصائي $\sqrt{Test9}$ الذي يقرب إلى التوزيع الطبيعي المعياري

كما أن القيمة (١) في الصورة (البسط) في اختبار رقم (٩) هو عامل ثابت يطرح من القيمة المطلقة للفرق المقدر (دون النظر إلى الإشارة) بين القيمة f و g ويقرب إلى التوزيع ثنائي الحد باستخدام توزيع مربع كاي المستمر .

٤- أحصل على قيمة (P) إما باستخدام الحاسوب أو بالإشارة إلى الاختبار الإحصائي في جدول توزيع مربع كاي (χ^2) ، (الجدول A.4) مع درجة حرية واحدة .

٥- قرر قبول أو رفض فرضية العدم مع الأخذ بالاعتبار قيمة (P) عادة نرفض فرضية العدم إذا كانت قيمة أقل من ٠.٠٥ .

٦- احسب درجة الثقة للاختلاف (للتباينات) في النسب المئوية للنجاحات المقدرة بالعلاقة .

P1 = (e + F) / m (العينة الأولى) .

وكذلك العلاقة

P2 = (e + g) / m (في العينة الثانية) .

حد الثقة التقريبي ٩٥ % للاختلافات الحقيقية في النسب المئوية يعطى بالعلاقة .

$$\frac{F-g}{m} \mp 1.96 \frac{1}{m} \sqrt{F+g-\frac{(F-g)^2}{m}}$$

مثال :

يعتبر ارتفاع عدد الخلايا الجسمية في الحليب مؤشراً دليلاً لوجود التهاب في غدة الضرع حددت إحدى الدراسات أن قيمة عدد الخلايا الجسمية (Somatic cell Counts) عندما تزيد على ٣٠٠ خلية / مل تعتبر مؤشراً دليلاً إيجابياً للالتهاب في غدة الضرع وما ينقص عن هذا العدد يعتبر مؤشراً دليلاً سلبياً لوجود التهاب الضرع ثم مقارنة النتائج المدرجة أعلاه باستخدام اختبار كاليفورنيا الخاص بالتهاب الضرع (California Mastitis test) واختبار الفوسميتك الإلكتروني (Fosmatic test) ثم الحصول على نتائج إيجابية في (٦١) عينة من أصل (٨٣) عينة باستخدام اختبار كاليفورنيا على (٧٣) حالة إيجابية من أصل (٨٣) عينة باستخدام اختبار الفوسميتك .

١- اختبار الفرضية العدم : النسب المئوية الحقيقية لكشف الحالات الإيجابية متشابهة في كل من اختبار كاليفورنيا واختبار الفوسميتك .

اختبـار الفرضـية البديلـة : النسـب المئويـة للحـالات الإيجابيـة مختلفـة باسـتخدام كلتـا الطريقتين.

٢- بيانات هذه الدراسة يمكن تمثيلها في الجدول رقم (٢٢) :

استنتاجاتنا تشير إلـى أن طريقـة الفوسـمتيك لهـا قـدرة علـى كشـف الحـالات الإيجابية كشفاً أفضل من طريقة كاليفورنيا كنسبة مئوية لحالات التهابات الضرع .

٦- تقـدر النسـب المئويـة للحـالات الإيجابيـة للالتهـاب بالضـرع باسـتخدام اختبـار كاليفورنيا واختبار الفوسمتيك كما يأتي :

$$\frac{61}{83} = 0.735$$ (كاليفورنيا)

$$\frac{73}{83} = 0.880$$ (اختبار الفوسميتك)

الجدول رقم (٢٢): نتائج قيم الخلايا الجسمية في الحليب

	SCC(+) (>٣٠٠×١٠٠٠)	SCC(-) (> <٣٠٠×١٠٠٠)	المجموع
اختبار كاليفورنيا	٩	٤١	٣٣
اختبار الفوسميتك	٢	٨	١٠
المجموع	٦١	٢٢	٨٣

حد الثقة المقدر للاختلاف الحقيقي في النسب المئوية لكشف حالات التهابات الضرع باستخدام كلتا الطريقتين آخذين بالاعتبار الشفعي (زوجي المقارنة) كما يأتي :

$$\frac{14-2}{83} \mp 1.96 \frac{1}{83}\sqrt{14+2-\frac{(14-2)^2}{83}} = 0.1446 \mp 0.0892 = 0.055 - 0.234$$

بمعنى آخر ، مع أن اختبار كاليفورنيا قد حدد نسبة حالات إيجابية ٧٤% من العينات المدروسة حدد اختبار الفوسمتيك نسبة حالات إيجابية ٨٨ % ، فإن النسبة المئوية للفرق بين هاتين النسبتين كانت (١٤ % = ٨٨ – ٧٤) . هذا الفرق قد يكون في حده الأدنى (٦%) وفي حده الأعلى (٢٣%) . وهنا يجب أن نحكم على هاتين الطريقتين وضمن شروط معينة ما إذا كانت النسبة ٦ % تؤخذ بعين الاعتبار كاختلاف مهم ، في هذه الدراسة

اختبار التوافق بواسطة مربع كاي (χ^2)

The Chi – Squared Goodness of Fit – Test (χ^2)

إن التكررات المشاهدة في كل فئة لكل استجابة (على سبيل المثال الصنف الجيني أو حد فاصل لمتغير مستمر) يمكن أن تقارن بالعدد المتوقع في تلك الفئة إذا اتبعت في توزيع البيانات بالتوزيع النظري .

وهذا يعطي الإمكانية لاختبارها إحصائياً بترتيب توزيع مربع كاي (χ^2).

لاحظ أنه في هذا النص لا نفرض أسئلة على التوزيعات الاسمية الثنائية الحد أو توزيع بواسون من الأعداد المتوقعة والتي يمكن استثناؤها .

الافتراضات : Assumptions

في مفهوم اختبار التوافق نفترض أن العينة مختارة عشوائياً من حيوانات الدراسة والاستجابات المدروسة مستقلة عن بعضها وتصنف الفئات إلى صفوف مميزة أو حدود فواصل محدودة . تقريب توزيع مربع كاي (χ^2) هو شكل ضعيف إذا كان التكرار المتوقع أقل من (٥) في أكثر من ٢٠ % من الفئات المدروسة .

الطريقة : The Approach

- فرضية العدم : توزيع المتغير في حيوانات الدراسة يتبع توزيعاً نظرياً معيناً .
- الفرضية البديلة : توزيع المتغير في حيوانات الدراسة لا يتبع توزيعاً نظرياً معيناً .

٢- اجمع البيانات واعرضها في جدول التكرار .

٣- احسب التكرار المتوقع في كل فئة وجدد الاختبار الإحصائي إما باستخدام أيقون (α) في الحاسوب أو يدوياً باستخدام قانون (α) كما يأتي :

$$Test_8 = \sum \frac{(O-E)^2}{E}$$

بإعتبار أن :

O : تمثل التكرارت المشاهدة .

E : تمثل التكرارات المتوقعة .

$\sum$: مجموع كامل الفئات .

هذا الاختبار الإحصائي يسلك تقريب مربع كاي (χ^2) **مع درجات حرية تقدر كما يأتي .**

(بمجموع الفئات) – (عدد الحدود التي يجب أن تقدر لكي نحسب القيم المتوقعة) –١

على سبيل المثال ، الوسط الحسابي هو حد يجب أن يقدر فقط باستخدام توزيع بواسون .

٤- احصل على قيمة (P) إما من الحاسوب أو بالنظر إلى جدول (A.4).

٥- اتخذ القرار بقبول أو رفض فرضية العدم . وعادة لا توجد هناك ضرورة لرفض فرضية العدم إذا كانت قيمة (P) أقل من (٠.٠٠٥) . وبما أن فرضية العدم لا ترتبط بتأثير العنصر المدروس ، والذي نقوم بتقديره من بيانات العينة فلا داعي لأن تحسب درجة أو حد الثقة .

مثال نموذجي باستعمال اختبار مربع كاي (χ^2) :

في تطبيق مكون من (٥٣٥) رأس بقري – وجد أن هناك (١٥) حالة إجهاض ناجمة عن داء البريميات نمط Hardjo – وذلك بإجراء اختبار الاليزا النوعي وقد أدرجت هذه النتائج في الجدول رقم (٢٣) .

جدول رقم ٢٣: مثال عن نتائج اختبار الاليزا

	نتائج الاختبار المصلي الإيجابي	نتائج الاختبار المصلي السلبي	المجموع
الإجهاض	٦	٩	١٥

حمل طبيعي	٧٨	٤٤٢	٥٢٠
المجموع	٨٤	٤٥١	٥٣٥

السؤال : ما احتماليةُ التوافق المشاهدة لحالة الإجهاض مع حمل طبيعي ؟

المنظورة المتوقعة لحالات الإجهاض الإيجابية

١٥/٥٣٥ = ٠.٠٢٨٦٣

٠.٠٢٨٠٣ × ٨٤ = ٠.٣٥٥

٢.٣ = مصل إيجابي مع حالة إجهاض

٨١.٦٤ = ٢.٣٦- ٨٤ (مصل إيجابي مع عدم وجود إجهاض)

١٢.٦٤ = ٤٨٨.٣٦- ٤٥١ احتمالية إجهاض مع وجود مصل سلبي

يمكن تلخيص القيم المتوقعة في الجدول رقم (٢٤) كما يأتي :

جدول رقم ٢٤: القيم المتوقعة لاختبار الاليزا

	نتائج الاختبار المصلي الإيجابي	نتائج الاختبار المصلي السلبي	المجموع
حالة الإجهاض	٢.٣٦	١٢.٦٤	١٥ (٢.٨ %)

حمل طبيعي	٤٣٨.٣٦	٨١.٦٤	٥٢٠ (٩٧.٢ %)
المجموع	٤٥١	٨٤	٥٣٣ (١٠٠ %)

وبتطبيق علاقة مربع كاي (χ^2)

$$X^2 = \sum \frac{(O-E)^2}{E}$$

ثابت ياتس لجداول الاحتمالية 2 × 2

$$X^2 = \sum \frac{(1O - E1 - 0.5)^2}{E}$$

بإعتبار أن :

O : هي القيمة المشاهدة في كل خلية في الجدول .

E : القيم المتوقعة في كل خلية .

$$X^2 = \frac{(6-2.36)^2}{2.36} + \frac{(9-12.64)^2}{12.64} + \frac{(78-81.64)^2}{81.64} + \frac{(442-438.36)^2}{438.36}$$

$$X^2 = \frac{3.64^2}{2.36} + \frac{-3.64^2}{12.64} + \frac{-3.64^2}{81.64} + \frac{3.64^2}{438.63}$$

$$X^2 = 5.61 + 1.05 + 0.16 + 0.03$$

$$X^2 = 6.85$$

والآن ماذا بعد :

ننظر الآن إلى قيمة مربع كاي (χ^2) في جدول توزيع X^2 .

الحد الأدنى لقيم (α) العمودية في المستوى المشار إليه كما هو ممثل في الجدول رقم (٢٥).

جدول رقم ٢٥: مستوى المعنوية و درجة الحرية لاختبار الاليزا

درجة الحرية	مستوى المعنوية (قيمة P)		
	٠.٠٠٥	٠.٠١	٠.٠٠١
١	٣.٨٤	٦٦.٣	١٠.٨٣
٢	٥.٩٩	٩.٢١	١٣.٨٢
٣	٧.٨١	١١.٤٣	١٦.٢٧
٤	٩.٤٩	١٣.٢٨	١٨.٤٧
٦	١٢.٥٩	١٦.٨١	٢٢.٤٦
١٠	١٨.٣١	٢٣.٢١	٢٩.٥٩

ولنحسب درجة الحرية نتبع العلاقة المذكورة سابقاً في هذا الفصل وهي

$$Df = (R-1) \times (C-1)$$

R : عدد الصفوف (باستثناء الصفوف المحتوية على مجموع العمود) .

C : عدد الأعمدة (باستثناء الأعمدة المحتوية على مجموع الصف) .

وفي مثال فإن درجة الحرية بتطبيق العلاقة السابقة .

$$df = (2-1) \times (2-1) = 1$$

ولذلك تنحصر قيمة (P) بين القيمة ٠.٠١ والقيمة ٠.٠٠١

$$0.01 > P > 0.001$$

ولذلك ترفض فرضية العدم والنتائج المحصول عليها حقيقية .

الفصل الثاني عشر
CHAPTER TWELVE
الارتباط والانحدار الخطي
LINEAR CORRELATION AND REGRESSION

مدخل لفهم الارتباط والانحدار الخطي

Introducing Linear Correlation and Regression

شرحنا في الفصل السابق العلاقة بين المتغيرات الفئوية مع الأخذ بالحسبان أن اختبار مربع كاي (χ^2) يعتمد على مبدأ في فرضية العدم أنه لا يوجد توافق بين متغيرين .

في هذا الفصل سنصف التقنيات الإحصائية التي تستخدم في دراسة التوافق بين متغيرين مستمرين .

على سبيل المثال نريد دراسة العلاقة بين الوزن الحي ونمط التغذية عند الأغنام في دراسة مثل هذه العلاقة مثلاً سيكون لكل من الارتباط والانحدار الخطي تعريفه الخاص.

ومثال آخر : دراسة العلاقة بين الوزن الحي عند الأبقار أو الدواجن مع إضافة كمية معينة من عنصر غذائي معطى في علائق الأعلاف إلى آخر ما هنالك من تساؤلات لا نرغب في الإجابة عنها .

ومع ميولنا في كثير من الأحيان للاعتقاد بوجود علاقة ارتباطية بين ظاهرتين كدراسة العلاقة بين مكونات قيم الدم (خلايا بيضاء وخلايا حمراء وغيرها) البرهان عملياً على صحة هذا الأمر بدراسة الارتباط يعطي البرهان عملياً يجعلنا نؤمن بوجود تلك العلاقة بدلاً من الميل إلى الاعتقاد بوجوده .

الارتباط الخطي Linear Correlation :

يذكر الارتباط الخطي لتحديد ما إذا كانت هناك علاقة خطية بين متغيرين من النوع المستمر وقياس درجة تلك العلاقة . وهنا نريد أن نتعرف كيف يمكن أن نصف التوافق الخطي بين نمطين من المتغيرات . ومن أجل ذلك نشتق مقياساً يدعى بمعامل الارتباط Correlation Coefficient والذي يعكس وجود النقاط المنحرفة عن الخط المستقيم لرسم العلاقة الخطية .

وهكذا فإن المهمة الأساسية لفرضية الارتباط هي التركيز على وجود علاقة أو عدم وجود هذه العلاقة بين متغيرين (X) و (Y) لظاهرتين معينتين .

فإذا وجدت هذه العلاقة فإن فرضية الارتباط تقيس عندئذ شدة التوافق لهذا الارتباط . إذ إنه في كثير من الأحيان لا يكفي أن نقول إن هناك علاقة بين ظاهرتين ما دون إعطاء أية تقييمات أخرى لشدة هذه العلاقة .

معامل الارتباط The Correlation Coefficient :

إذا كنا نعتقد وجود علاقة خطية بين متغيرين كميين مع تغير في إحدى قيم المتغيرين والمترافقة مع تغير قيم المتغير الآخر ، فإننا نهتم بتحديد قوة العلاقة المدروسة ، إذ إننا لا نقوم فعلياً برسم خط مستقيم في تحليل فرضية الارتباط (وهذا جزء من تحليل الانحدار) ، لكننا يمكن أن نتخيل خطاً على نحو تقديري تقريبي للبيانات المدروسة .

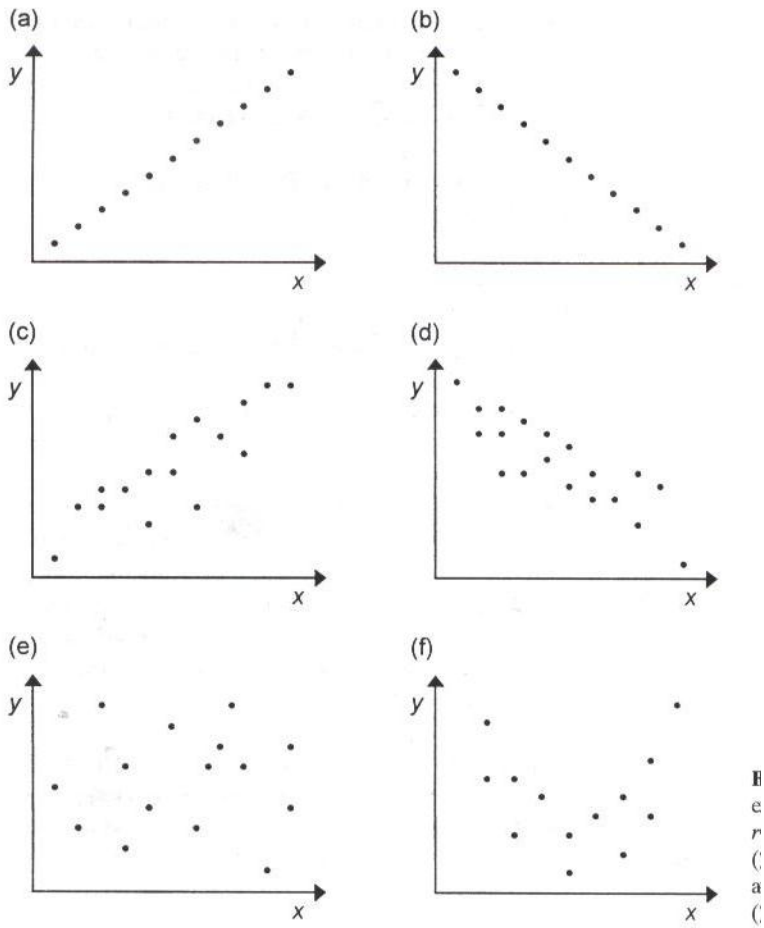
بمعنى آخر هل تلك النقاط الممثلة لكل قيمة لكل من المتغيرين في مخطط يمثل قيمة , (X)) (قريبة من هذا الخط أو أنها بعيدة عنه .

وهكذا فالعلاقة الخطية بين كلا المتغيرين فالنقاط القريبة من الخط تمثل أقوى توافق خطي بين المتغيرين .

نقيس درجة أو شدة هذا التوافق بحساب ما يدعى بمعامل ارتباط بيرسون Correlation Coefficient وهو يأخذ القيم ضمن المجال بين [−١ والقيمة +١]

نقول إن هناك ارتباطاً تاماً Perfect Correlation إذا كانت كافة النقاط مندرجة على الخط . في هذه الحالة تكون قيمة معامل الارتباط هي العظمى (١) إما موجبة أو سالبة (+١) أو (−١) .

الشكل رقم (١٨) الارتباط الخطي



نقول إن هناك ارتباطاً إيجابياً إذا كانت هناك إشارة موجبة لمعامل الارتباط وعندئذ توجد علاقة مباشرة بين كلا المتغيرين ولذلك تزداد قيمة أحد المتغيرين بينما تنخفض قيمة الآخر .

نقول عن الارتباط إنه السلبي عندما تكون إشارة معامل الارتباط سلبية وعندئذ توجد هناك علاقة عكسية بين كلا المتغيرين ولذلك تزداد قيمة أحد المتغيرين وتنخفض قيمة المتغير الآخر .

نقول إنه لا يوجد توافق خطي (على سبيل المثال كأن يكون هناك عدم وجود ارتباط) إذا كان معامل الارتباط بقيمة الصفر وعندئذ يشير المخطط النقطي إلى عدم وجود علاقة خطية بين المتغيرات (الشكل ١٨ (e)). ويجب أن تلاحظ أن العلاقة غير خطية بين المتغيرات يمكن أن تعطي أيضاً معامل ارتباط بقيمة الصفر

وهكذا فإن القيمة الأقرب للخط المستقيم لأن معامل الارتباط يكون بقيمته العظمى وتشير إلى علاقة أقوى بين المتغيرات وتفسر إلى قيم أقرب إلى الخط المستقيم للمثل للعلاقة المدروسة .

ويوجد هنا علاقة لحساب معامل الارتباط أو أننا نحصل على هذه القيمة باستخدام الحاسوب أو الآلة الحاسبة .

فإذا أخذنا عينة عشوائية مستقلة الإشفاع ، أعدادها (n) لقيم المشاهدات (X_1 , Y_1) و (X_2 , Y_2) و (X_3 , Y_3) (X_n , Y_n) . لنمطين من المتغيرات المستمرة (X) و (Y) وعندئذ فإن تقدر قيمة معامل الارتباط P (وهو الحرف اليوناني وهو rho) في قطيع من الحيوانات لعينة ما معامل الارتباط الذي يعطى بالعلاقة الآتية :

$$\Gamma = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2\Sigma(Y-\overline{Y})^2}}$$

وهنا نلاحظ ما يأتي :

١- معامل الارتباط مستقل بالنسبة لوحدات القياس لمتغيرين .

٢- يمكننا إجراء تبادل داخلي بين (X) و (Y) بدون التأثير على قيمة معامل الارتباط .

٣- معامل الارتباط تبقى صلاحيته فقط ضمن حدود معينة للبيانات المدروسة .

اختبار الفرضية عندما تكون قيمة معامل الارتباط الصفر

Testing a Hypothesis That The Correlation Coefficient is Zero

يعرض معامل الارتباط قياساً لقوة التوافق بين كلا المتغيرين المدروسين وإذا لم تكن هناك علاقة توافق خطي بين المتغيرين فإن قيمة معامل الارتباط تساوي الصفر ولذلك فإن اختبار الفرضية (P = O) وهذا يفسر أنه لا يوجد دليل لإثبات علاقة سببية بين المتغيرين ومن المستحيل أن يشير ذلك إلى وجود اختلاف بين كل منهما .

الافتراضات Assumption :

- قيست التغيرات اعتماداً على عينة عشوائية للحيوانات الفردية .

- إن كلا المتغيرين (X) و (Y) من النوع الكمي .

- يكون اختيار الفرضية حقيقياً بالنسبة لمعامل الارتباط عندما يساوي القيمة الصفر وهذا يتطلب وجود متغير واحد على الأقل من كلا المتغيرين قيمة موزعة طبيعياً .

الطريقة The Approach :

– حدد فرضية العدم التي تعتمد على معامل الارتباط لحيوانات الدراسة (P) يساوي القيمة (١) . وتتبنى الفرضية البديلة عموماً أن معامل الارتباط لا يساوي القيمة الصفر .

– اجمع البيانات واعرضها في مخطط نقطي يمثل ما إذا كانت هناك علاقة خطية بين المتغيرين وتأكد أن الافتراض محقق في هذه العلاقة أنه يوجد على الأقل متغير واحد من المتغيرين المدروسين موزع طبيعياً .

– احسب معامل ارتباط العينة ، ومن المفضل أن تحسب باستخدام الحاسوب . ويمكن أن تجد أن الحاسوب يحسب الاختبار الإحصائي الذي يعتمد على (توزيع t) مع درجات حرية (2- n) ويحسب الاختبار الإحصائي بالعلاقة المدرجة أعلاه :

$$Test_{10} = \Gamma\sqrt{\frac{(n-2)}{(1-\Gamma^{2})}}$$

بإعتبار أن :

r : تمثل معامل ارتباط العينة .

n : عدد أشفاع المشاهدات في العينة .

ليس صعباً أن نحسب الاختبار الإحصائي رقم (١٠) Test يدوياً لكن هذا يتطلب وجود جدول (جدول A.6) يربط قيم (r) مباشرة بقيم (P) .

– احصل على قيمة (P) ، عموماً تحسب بسهولة باستخدام الحاسوب . ويمكن الحصول على قيمة (r) مباشرة باستخدام الجدول الإحصائي (A.6) على نحو بديل

يمكن استخدام الاختبار الإحصائي (Test 10) من الجدول الإحصائي اعتماداً على توزيع (t) ، الجدول (A.3) مع درجات الحرية (n – 2) .

– قرر ما إذا كنت ترفض أو تقبل فرضية العدم اعتماداً على قيمة (P) وعموماً فإننا نرفض فرضية العدم إذا كانت قيمة (P) أقل من ٠.٠٥ نحسب درجة الثقة عندما يكون معامل الارتباط حقيقياً . مع أن توزيع العينة لقيمة (r) ليس طبيعياً فإن التوزيع المتغير يحول إلى القيمة .

$$Z = 0.5 Loge\{(1+r)(1-r)\}$$

يتبع التوزيع الطبيعي ونستخدم هذه المعلومات لنتمكن من حساب درجة الثقة لمعامل الارتباط (P) .

إن حدود الثقة ٩٥ % التقريبية للقيمة Z تحسب كما يأتي :

$$Z1 = Z - 1.96\sqrt{n-3}$$
$$Z2 = Z - 1.96\sqrt{n-3}$$

وعندئذ نحول البيانات بحساب اللغاريتم الطبيعي لنحصل على درجة الثقة لمعامل الارتباط (P) وهكذا فإن Z1 درجة الثقة ٥٩ % لمعامل الارتباط (P) وتقريبياً تحسب كما يأتي :

من $\dfrac{e^{2Z1}-1}{e^{2Z2}+1}$

إلى $\dfrac{e^{2Z2}-1}{e^{2Z2}+1}$

الانحدار الخطي البسيط Simple Linear Regression

سنصف عند دراستنا لتحليل الانحدار الخطي البسيط العلاقةَ بين متغيرين كميين X و Y بواسطة تحديد ورسم الخط المستقيم الذي يقدر البيانات النقطية في المخطط النقطي بما يقرب أحدهما من الآخر . فإننا نربط المتغير X كواحد من القيم التي يمكن أن تقاس بدون خطأ أو أن المجرب يحددها مسبقاً ولذلك على سبيل المثال ، يمكن أن تمثل الجرع الدوائية أو الأعمار أو الأوزان أو المركزات بقيم مسبقة .

ومن جهة أخرى فإن المتغير (Y) هو متغير عشوائي موضوع للاختلاف التجريبي كضغط الدم النظامي وتركيز الهيموغلوبين أو الكثافة اللونية ، فإننا نفترض أن المتغير Y هو متغير غير مستقل ومقام على المتغير X ولذلك إذا غيرنا من قيمة (X) فهذا سوف يقودنا إلى التغيير في قيمة (Y) ولكن هذا التغير لن يكون متناظراً .

ندعو المتغير (Y) بالمتغير غير المستقل أو المستجيب أو الناتج ونمثله بمحور عمودي في المخطط النقطي .

– ندعو المتغير (X) بالمتغير المستقل أو المتغير المفسر أو المتغير الانحداري أو متغير الشرح ونمثله بمحور أفقي .

ويمكن أن ترسم عيانياً ماذا يمكن أن نعتقد عن أفضل تطابق للخط المستقيم لكن هذا سيكون طريقة موضوعية ولن يكون مرضياً على نحو كامل . وبديلاً عن ذلك نستخدم

قانوناً (معادلة رياضية) ليصف العلاقة للخط المستقيم بين المتغيرين (X) و(Y) فإن المعادلة الرياضية تعطى بِ :

$$\mathrm{Y}_{POP} = \alpha + \beta \times$$

بإعتبار أن :

$\mathrm{Y}_{\mathrm{POP}}$: هي القيمة المفسرة أو معدل (Y) لقيم معطاة عند المحور (X) .

α : (القيمة ألفا) : وهي تمثل قيمة الثابت والتي هي عبارة عن تقاطع المحورين (inter Cept) أي هي قيمة (Y) عندما تكون قيمة (X) تساوي الصفر .

β : (بيتا) : وهي قيمة الانحدار (Slope) الخطي والتي تمثل معدل تغير في قيمة (Y) لوحدة التغير في (X) فعلى سبيل المثال إن قيمة β تصف كم تتغير قيم (Y) بمتوسطاتها عندما تزداد قيمة (X) بوحدة واحدة .

إن كلاً من ألفا وبيتا عبارة عن حدود تعرف الخط المستقيم وهي ما يدعى بمعامل الانحدار (Regression Coefficient) .

ويجب أن نقدر كلا الحدين α و β بالترتيب بواسطة عينتين عشوائيتي الأزواج من المشاهدات عددها (n) .

$$\{(X_1, Y_1), (X_2, Y_2), (X_3, Y_{3)}, ...(x_n, Y_n)\}$$

بطريقة تطابق فيها الخط الوهمي المرسوم بالنسبة للنقاط المدرجة وعلى نحو قريب من هذا الخط ما أمكن .

وهكذا من خلال ملاحظة الثنائيات المدرجة أعلاه تحقق معادلة المستقيم المعطى وبالتالي فهي تقع عليه أو بتعبير آخر يمر المستقيم منها ولن نخرج أباً من هذه الثنائيات عن المستقيم .

في المواقع لدراسة العلاقة الارتباطية $Y= f (X)$ التي تدل على الاتجاه العام يلزم التمييز بين المتغير المستقل الذي جرت العادة على تسميته (X) والمتغير التابع والذي يرمز له عادة بالرمز (Y) حيث تكون الظاهرة أو الخاصية (X) والتي يخصص لها المحور السيني الأفقي هي الظاهرة أو الخاصية المؤثرة وتكون الظاهرة (Y) والتي يخصص لها المحور الصادي العمودي هي الظاهرة المتأثرة .

وهكذا فإن أفضل تطابق للخط المستقيم لأنه سيحدد قيم الانحرافات الفعلية من تلك القيم المتوقعة بواسطة معادلة المستقيم .

ولكن نتجنب الانحرافات الإيجابية والسلبية عن الخط المستقيم نتجاهل كلاً منها . ونقوم بتربيع هذه القيم قبل عملية جمعها ونختار أفضل تطابق خطي والذي يحد من مجموع مربعات الانحرافات للقيم المشاهدة للمحور الصادي (Y) عن القيم المتوقعة . وهذا ما يعرف بالمبدأ الأساسي لطريقة المربعات الصغرى Principle of Least Squares .

وهذا ما يعبر عنه رياضياً بالمعادلة الآتية :

$$SSE = \sum_{i=1}^{n} (Y_{observed} - Y_{pridicted})^2$$

بإعتبار أن :

SSE : تمثل مجموع مربعات الانحرافات أو ما يدعى بمجموع مربعات للخطأ (Sum of Squares of error) .

إن القيم الرقمية لألفا (α) وبيتا (β) التي تحد من قيم SSE يمكن الحصول عليها المعادلة الآتية :

$$\beta = \frac{SP_{XY}}{SS_X}$$

$$\alpha = \overline{Y} - \beta \overline{X}$$

بإعتبار أن :

$$SP_{XY} = \sum_{i=1}^{n} (Xi - \overline{X)(Yi - \overline{Y})}$$

$$SS_X = \sum_{i=1}^{n} (X_1 - \overline{X})^2$$

علماً أن القيم :

SPXY = تمثل مجموع قيم الانحرافات للثنائيات X و Y عن معدلاتها المحددة (Sum of the Products) .

SSX = تمثل مجموع مربعات الانحرافات لقيم X عن معدل قيم X (Sum Squared)

وعندما تحسب قيمة ألفا ∝ وبيتا β يمكن إيجاد معادلة المستقيم الخطي لنحصل على المعادلة المتوقعة للمربعات الصغرى (Least Squares Prediction Equation) أو ما يدعى بالانحدار الخطي (Regression Line) .

لنمثل ذلك بمثال رقمي في الجدول رقم (٢٦) . لنفترض أننا نزود متممات علفية لمجموعة مكونة من خمسة حيوانات خلال فترة عدة أشهر وقد استهلاك كل حيوان الكميات الآتية :

الجدول رقم ٢٦ : استهلاك المتممات العلفية لمجموعة مكونة من خمسة حيوانات خلال فترة عدة أشهر

كمية المتممات العلفية المقدمة (كغ)	رقم الحيوان
٧	١
٩	٢
١١	٣
١١	٤
١٢	٥

وخلال نفس الفترة يكون الكسب الوزني المحصول عليه لحيوانات الدراسة الخمسة مسجلاً كما في الجدول رقم (٢٧) كما يأتي :

الجدول رقم ٢٧ : الكسب الوزني لكل حيوان من مجموعة التجربة للمتممات العلفية

الكسب الوزني لكل حيوان (كغ)	**رقم الحيوان**
١٧	١
٧٩	٢

٩٦	٣
١٠٣	٤
١١٦	٥

يحتوي الجدول الآتي (رقم ٢٨) البيانات والكميات التي نحتاجها في عملية الحساب الكمية لنحصل على معادلة الانحدار الخطي الذي يسمح لنا بتقدير الكسب الوزني المتوقع بالارتباط بكمية المتممات المستهلكة .

جدول رقم (٢٨): البيانات والكميات المطلوبة للحصول على معادلة الانحدار الخطي

رقم الحيوان	**العلف المستهلك (x)**	**الكسب الوزني (y)**	**XI-X**	**$XI-X^{٢}$**	**YI-Ỹ**	**XI-X YI-Ỹ**
١	٧	٧١	-٣	٩	-٢٠	
٢	٩	٧٩	-١	٨	-١٢	
٣	١١	٩٦	١	١	٥	
٤	١١	١٠٣	١	١	١٢	
٥	١٢	١٠٦	٢	٢	١٥	
المجموع	٥٠	٤٥٥	-	-	-	
الوسط الحسابي	١٠	٩١	-	-	-	

وهكذا يمكننا أن نحسب قيمة بيتا β_1 التي تمثل الانحدار و هي بعبارة أخرى تمثل معامل الانحدار كما يأتي :

$$\beta 1 = SPx\, y / SSx$$

$$\beta1 = 119/16$$
$$\beta1 = 7.4375$$

بينما قيمة تقاطع المحورين (β_0) و هي أيضا تمثل ثابت الانحدار (regression constant) ويحسب كما يأتي :

$$B_0 = y - \beta1\ x$$
$$B_0 = 91 – 7.4374$$
$$B_0 = 16.625$$

وهكذا فإن معادلة الانحدار الخطي الصغرى والتي سوف نستخدمها لتقدير الكسب الوزني مع الارتباط بالكمية المقدمة التي يستهلكها الحيوان :

$$Y = B_0 + B_1\ x$$
$$Y = 16.625 + 7.4375x$$

ومن ذلك يمكن الآن أن نتوقع الوزن المكتسب للحيوان والذي استهلك كمية ٨ كغ من العلف المقدم كمتممات علفية . وبضرب القيمة (٨) للمتغير (x) في المعادلة يمكننا الحصول على :

$$Y = 16.625 + 7.4375\ (8)$$
$$Y = 16.625 + 59.5$$
$$Y = 76.125$$

وهكذا فإننا نتوقع أن يحقق الحيوان الذي استهلك بعلفه ٨ كغ من المتمم العلفي كسباً وزنياً ٧٦.١٢٥ خلال فترة تجربتنا . وعلى أية حال يجب أن نعي أن الحيوان ما من هذه المجوعة من غير المحتمل أن يحقق كسباً وزنياً ٧٦.١٢٥ كغ

بسبب الاختلافات المتنوعة بين الحيوانات. ويجب أن يُعتبر وزن ٧٦.١٢٥ معدلَ كسب وزني للحيوانات المستهلكة ٨ كغ من المتمم العلفي بينما يتراوح الكسب الوزني للحيوانات الفردية حول المعدل . وفي بياناتنا المدرجة في مثالنا . فإن الحيوانين رقم ٣ و ٤ يستهلكان قرابة ١١ كغ من المتمم العلفي وعلى أية حال فإن الكسب الوزني لهما لم يكن متساويا بينما الكسب الوزني للحيوان رقم 3 هو قرابة ٩٦ كغ بينما الحيوان رقم ٤ يكسب ١٠٣ كغ وباستخدام معادلة الانحدار فإن الكسب الوزني المتوقع لاستهلاك الحيوان ١١ كغ من المتمم العلفي خلال تجربتنا وهو سوف يحقق ٩٨.٤٣٧ كغ . وهكذا فإن كلا القيم المشاهدة فعليا تختلف فعليا عن القيمة المتوقعة ، وعلى أية حال فإن معدل كلا القيمتين يمكن أن يحسب كما يأتي :

$$99.5 = \frac{103 + 96}{2}$$

وهذه القيمة المذكورة أعلاه قريبة من القيمة المتوقعة .

رأينا مما ذكر سابقاً أن مجموع مربعات الانحرافات للقيم المشاهدة من القيم المتوقعة تدعى بمجموع مربعات الخطأ (SSE) . وفي حالتنا هذه فإن مجموع مربعات الخطأ (SSE) قد أخذ القيمة ٥٢.٩٣٨ كما هو مشاهد في الجدول رقم (٢٩) التالي .

الجدول رقم ٢٩ : حساب القيم المتوقعة للكسب الوزني

رقم الحيوان	العلف المستهلك (x)	الكسب الوزني (y)	قيمة Yالمتوقعة	قيمة Y المشاهدة – قيمة Y المتوقعة	(قيمة Y المشاهدة – قيمة Y المتوقعة)

١	٧	٧١	٦٨.٦٨	٢.٣١٢٥	٥.٣٤٧٧
٢	٩	٧٩	٨٣.٥٦	٤.٥٦٢٥-	٢٠.٨١٦٤
٣	١١	٩٦	٩٨.٤٣٧	٢.٤٣٧٥-	٥.٩٤١٤
٤	١١	١٠٣	٩٨.٤٣٧	٤.٥٦٢٥	٢٠.٨١٦٤
٥	١٢	١٠٦	١٠٥.٨٧٥	٠.١٣	٠.٠١٦٩
المجموع	٥٠	٤٥٥			٥٢.٩٣٨٨
الوسط الحسابي	١٠	٩١			

وإن القيمة الصغرى لقيمة SSE مرتبطة بإجمالي مجموع المربعات (TSS) والتي هي عبارة عن مجموع المربعات للانحرافات للقيم المشاهدة من المعدل العام . إذ إن الاختلاف الأكثر يفسر بواسطة نموذج الانحدار . وإن قيمة TSS يمكن الحصول عليها بالقانون :

$$TSS = SUM\ (Yi - \dot{Y})^2$$

ومن مثالنا نحسب حساباً مسبقاً قيمة ((Yi – average Y˙)) لكي نحصل على معامل الانحدار وهكذا فإن قيمة TSS تحسب كما يأتي :

$$TSS = (-20)^2 + (-12)^2 + 5^2 + 12^2$$

إن الاختلاف بين قيمة TSS وقيمة SSE تمثل قيم مجاميع المربعات للانحدار (SSR) وهكذا فإن قيمة TSS تساوي عندئذ :

$$TSS = SSR + SSE$$

بإعتبار أن :

$$SSR = SUM\ (Y\ Predicated - Average\ Y\ )^2$$

بإعتبار أن :

Y : هي القيمة المتوقعة للمتغير Y

average Y: هي معدل المتغير Y

ومن الجدول الثاني يمكن أن نحسب التباينات بين الكسب الوزني المتوقع لكل حيوان ومعدل الكسب الوزني الكلي لِ ٩١ كغ . فإذا قمنا بعملية تربيع هذه القيم مع مجموع القيم المربعة فإننا نحصل على :

$$SSR = (68.6875 - 91\ )^2 + (83.5625 - 91)^2 + (98.4375 - 91)^2 + (98.4375 - 91\ )^2 + (105.875 - 91)^2$$

$$SSR = 497.84766 + 55.3164 + 221.26563$$

$$SSR = 885.625$$

$$938\ (TSS) = 885.0625\ (SSR) + 52.9388\ (SSE)$$

تقسيم إجمالي مجموع مربعات المتغير Y

Partitioning the Total Sum of Squares of Y

Total SS = SSR + SSE

وتحسب درجة الحرية لكل حد من الحدود السابقة كما يأتي :

Df. for SS = n-1

Df. for SSR = عدد المتغيرات المستقلة

Df. for SSE = n-1 – regression Df.

بإعتبار أن SSR = مجموع مربعات الانحدار

والتي هي عبارة عن الكمية الكلية للاختلاف المفسر بالمتغير المستقل (x) كما أن قيمة SSE تساوي مجموع مربعات الخطأ والتي هي عبارة عن الكمية الكلية للاختلاف غير المفسر بالمتغير (x)

إن تقسيم قيمة SS إلى SSR و SSE يدعى بتحليل الفرق . كما أن النسبة المئوية لإجمالي التباينات للقيمة TSS والمحسوبة بواسطة نموذج الانحدار (SSR) تشير إلى ما يدعى بمعامل التحديد ويرمز له بالرمز (R^2).

وتحسب قيمته كما يأتي :

$$R^2 = SSR/ TSS$$

وفي مثالنا السابق فإن :

$$R^2 = 885.0625/ 938 = 0.9436$$

إن النسبة المئوية ٩٤ للفرق في الكسب الوزني في خمسة حيوانات يمكن أن تحسب فقط بواسطة كمية المتممات العلفية المستهلكة .

وفي الحقيقة هذه النسبة كبيرة جدا فإنه لم يكن لدينا أدنى شك في أنه هناك علاقة حقيقية بين كمية المتمم العلفي المستهلك والكسب الوزني . وللتأكد من دقة هذه النتائج لابد لنا من حساب درجة المعنوية الإحصائية . وكما هو الحال في جميع الاختبارات الإحصائية تقاس المعنوية الإحصائية باستخدام نظرية العدم التي تثبت وجود علاقة حقيقية بين المتغير (y) والمتغير (x) وإن كمية الاختلاف المفسر بواسطة SSR هي عبارة عن مجرد نتيجة احتمالية . وهذا يمكن أن يظهر فيما إذا كانت نظرية العدم حقيقية فإن القيم عندئذ كالآتي :

$$SSR/df\ Regression = SSE/df\ error$$

حيث أن:

df regression: درجة الحرية الخاصة بالانحدار

df error: درجة الحرية للخطأ

هذه التقديرات الكمية تدعى بالوسط الحسابي للمربعات و تحسب كما يأتي :

MSR = SSR / df. Regression

MSE = SSE / df. Regression Error

بإعتبار أن :

MSR : ويدعى بالوسط الحسابي لمربع الانحدار

MSE : ويدعى بالوسط لمربع الخطأ

df. Regression : وهي تساوي عدد المتغيرات المستقلة الداخلة بالقانون

df. Error : وهي تساوي (n-1) – df. Regression بإعتبار أن n هي عبارة عن عدد المشاهدات الملحوظة .

وكلما كبرت قيمة MSR بالارتباط بقيمة MSE كان هناك احتمالية أقل لتكون نظرية العدم حقيقية . وإن الاختبار الإحصائي المستخدم لتقييم هذه الاحتمالية هو اختبار F الإحصائي والذي يحسب بالتناسب الآتي :

F = MSR/MSE = 885.06/17.69= 50.16

هذه العملية تسمى بتوزيع F وإن درجة الحرية للصورة في هذا التناسب هي عبارة عن درجة الحرية الخاصة بالانحدار df. Regression بينما درجة الحرية للمخرج في هذا التناسب هي عبارة عن قيمة درجة الحرية للخطأ df. Error

وفي مثالنا المذكور أعلاه لدينا متغير مستقل وحيد وهو عبارة عن كمية المتمم العلفي المستهلك :

(df. Regression = ١) و لذلك فإن :

MSR = SSR /1 =885.0625

و إن درجة الحرية للخطأ في مثالنا هي:

DF = (5-1) –1 =3

و هكذا فإن قيمة :

$$MSE = 52.9388/3 = 17.6463$$

$$\longrightarrow \quad F= 885.0625/17.6463 = 50.16$$

و إذا نظرنا إلى الاحتمالية المحصول عليها من قيمة F ، فكلما كانت هذه القيمة كبيرة كانت نظرية العدم حقيقية . وإذا نظرنا إلى جدول توزيع F مع قيمة درجات الحرية ١ و ٣ فإننا نجد أن هناك احتمالية أقل من ٠.٠١ وهذا يعني أنه هناك أقل من ١% فرصة للحصول على قيمة F بقيمتها المتوقع أن تكون كبيرة والتي قدرت بـ ٥٠.١٦ إذا لم يكن هناك علاقة بين كمية المتمم العلفي المستهلك والكسب الوزني . ولذلك يمكننا الحصول على درجة ثقة ٩٩% أن نظرية العدم خاطئة وأنه يوجد علاقة بين كلا المتغيرين .

تحليل الانحدار اللوغاريتمي

LOGISTIC REGRESSION ANALYSIS

في كثير من الدراسات فإن المتغير غير المستقل لموضوع الدراسة يمثل وجود أو عدم وجود لحالات معينة وهذا المتغير يمثل نوعاً من المتغيرات الفئوية Outcome Variable is Categorical ولذلك لا نستطيع استخدام الانحدار

المتعدد العادي لهذا النوع من البيانات لكن هناك إمكانية لإعادة فرز مثل هذه البيانات بتقنيات أكثر تناسباً من حيث نمط تحليلها .

وتعتبر تقنية مانتل هنسنزل Mantel – Hanszel (MH) . مناسبة لتحليل العلاقة بين المتغيرات الفئوية في بيانات يكون فيها تعداد صفوف العامل المدروس (على شكل فئة) محدداً . تحسب تقنية مانتل هنسنزل تناسبات الأفضلية كفئة معينة Stratum – Specific Odds ratios لكل نمط من مستويات المتغير المتداخل وعندئذ نحسب تناسب الأفضلية العام وزنياً . وإن استخدام تقنية (MH) ليصبح محدود الاستعمال أيضاً عندما تكون أعداد جداول 2 × 2 المطلوب تحليلها كبيرة أو علاوة على ذلك فإن من غير الممكن تنسيق والتعامل مع متغيرات مستمرة .

إن الطريقة المناسبة والبديلة هي تحليل الانحدار باستخدام العملية اللوغاريتمية Regression Analysis Using The Logistic Function .

إن المبدأ الأساسي لتقنية الانحدار اللوغاريتمي مشابه جداً لمربعات الانحدار الصغرى Least Squares Regression ومن المساوئ لهذه الطريقة أننا نحصل بالحساب على تناسب الأفضلية بينما في بعض الدراسات يكون حساب الخطورة النسبية Relative Risk والذي يمكن أن يكون أكثر تناسباً من استخدام تناسب الأفضلية إلا أنه لا يستخدم في بعض الدراسات الوبائية العملية وخاصة في دراسة حالة التحكم مع حالات تجريبية أخرى . إلا أن تناسب الأفضلية يستخدم في كافة

الدراسات الوبائية سواء كانت مسوحات وبائية أو دراسات وبائية تطبيقية تجريبية كما ستدروسون ذلك في مقرر الوبائيات البيطرية .

نظرية الانحدار اللوغاريتمي The Theory of Logistic :

إن الاختلاف الأساسي بين مربعات الانحدار الصغرى والانحدار اللوغاريتمي هو في طبيعة المتغير غير المستقل . ففي انحدار مربعات الصغرى يكون المتغير غير المستقل من نمط مستمر (على سبيل المثال وزن الجسم) بينما في الانحدار اللوغاريتمي يكون المتغير غير المستقل هو من نمط بينري (ثنائي Binary) (على سبيل المثال متغير فئوي مع وجود فئتين فقط) . فعلى سبيل المثال المصل الإيجابي مع المصل السلبي ، أو نقول بقرة مصابة . وبقرة غير مصابة) فإذا أعطي للفئتين قيم رقمية صفر (0) وواحد (1) فإنه عادة ما نمثل القيمة لا (بالنتيجة السلبية) والقيمة نعم (بالنتيجة الإيجابية) نسبياً . وعندئذ يعطي معدل هذه القيم في عينة من الحيوانات نفسَ نسبة نموذج الانحدار المناسب سوف يتنبأ نسبة الأفراد الحيوانية المميزة بموضوع الدراسة أو يعادل احتمالية الأفراد الحيوانية المميزة لأي من هذه المجاميع في النموذج . ومن الناحية التدريبية الإحصائية من المفضل الطريقة باستخدام تحويل هذه النسبة المئوية ولسبب وحيد فقط هو أن النموذج الإحصائي

يمكـــن أن يتنبـــأ قيمـــاً احتماليـــة خـــارج المـــدى صـــفر (0) والقيمـــة (1) .

والأيقون المستخدم عموماً في عملية التحويل يدعى بالتحويل اللوغاريتمي Logit Trans formation ويكتب Logit (P) والذي يعطي قيمة اللوغاريتم الطبيعي للتناسب لفرد من الحيوانات في حالة معينة . بينما القيمة (P) هي نسبة الأفراد المميزة بخصائص معينة متعلقة بموضوع الدراسة .

فعلى سبيل المثال إذا كانت قيمة (P) تمثل الاحتمالية لحيوانات إيجابية مصلياً للبروسيلة المجهضة Brucella Aborts فعندئذ تمثل القيمة (Y – P) احتمالية أن يصبح الحيوان إيجابياً للإصابة بالبروسيلة المجهضة . ويجب أن نتذكر أن التناسب (Y – P) / P يدعى بتناسب الأفضلية أو التراجمية Odds وعندئذ تعطى القيمة اللوغاريتمية لقيمة (P) كما يأتي :

$$Logit(P) = Log(\frac{P}{Y-P})$$

= وهي تمثل عموماً اللغاريتم التراجمي Log odds وبالمقارنة بقيمة (P) والتي يجب أن تأخذ القيمة بين (0) و (1) فإن القيمة اللغاريتمية لـ (P) : Logit (P) يمكن أن تأخذ أي قيمة بين المدى (-∞ , + ∞) .

إن الاختلاف بين الانحدار الخطي والانحدار اللغاريتمي هو أن طرق التشابه العظمى Maximum Likelihood (ML) تستعمل بدلاً من المربعات الصغرى

التقليدية لتقدير الحد . ومن ضمن محاسن طرق التشابه الأعظمي ليس من الضروري أن يكون المتغير غير المستقل من نمط التوزيع الطبيعي مما يعني أن هناك إمكانية لتقدير المعادلات الخطية المعقدة وكذا النماذج غير الخطية أيضاً .

وعلى نحو مشابه للانحدار العادي يقدم الانحدار اللغاريتمي يقدم وصفاً نموذجياً المجاميع الخطية للمتغيرات المستقلة مع المتغيرات غير المستقلة .

$$Odds = \frac{P}{1-P} : P = \frac{Odds}{1+Odds}$$

وإذا كانت قيمة :

$$\text{Log odds} = \beta_0 + \beta_1 \times$$

ولذلك الاحتمالية لقيمة (P) تساوي :

$$:P\ (y\ 1\ /x\ ) = {}_e\ \beta_0 + \beta_1\ x\ /\ 1 + {}_e\ \beta_0 + \beta_1\ x$$

وهكذا فإن طريقة الانحدار تمثل أولاً حساب قيمة L1 (و هي تساوي القيمة التالية)((p_1) Logit) للمجموعة الأولى) و L2 (وهي تساوي (p_2) Logit) للمجموعة الأخرى) وعندئذ يكون حساب الاختلاف في الاحتمالية للحيوانات المصابة مثلاً بالبروسيلة المجهضة كما يأتي :

$$l_1 - l_2 = Log\left(\frac{P_1}{1-P_1}\right) - Log\left(\frac{p_2}{L-P_2}\right)$$
$$= Log\left(\frac{P_1(1-p_2)}{p_2(1-P_1)}\right)$$

والتي تساوي تمثيل لغاريتم تناسب الأفضلية Odds Ratio وبمعنى آخر :

القيمة $\beta 1$: تمثل اللغاريتم المقدر لتناسب الأفضلية (التراجحية) مع الأخذ بالاعتبار

التعرض للعامل 1$(OR = {}_e \beta_1)$.

القيمة $\beta 0$: وهي تقدر لغاريتم الأفضلية للمرض في حال عدم التعرض لعوامل

الخطورة

ومن البرامج الإحصائية المستخدمة لطرق الانحدار اللغاريتمي :

. SPPSS , STATISTIX , SAS , GLIM , EGRET , BMDP, S -Plus

الفصل الثالث عشر
CHAPTER THIRTEEN
الاختبارات الإحصائية اللامعلمية
NON-PARAMETRIC STATISTICAL TESTS

مقدمة Introduction

درسنا الاختبارات المعلمية في الفصول السابقة، و التي تتضمن افتراضات معينة لِتوزيع المشاهدات ، فعلى سبيل المثال عند استخدام اختبارات ستودنت ذي

العينيتين افترضنا أن البيانات موزعة طبيعياً والتباينات في العينات المشاهدة في المجموعتين متشابهتان . إلا أنه في بحوث الحيوان تكون نتائج البيانات أقل درجة من حيث الشمولية والدقة في أعداد المشاهدات أو توزيع البيانات . ولذلك من الممكن حسب الافتراضات المطبقة في هذه الاختبارات أن تكون نتائج هذه الاختبارات اللامعلمية والتي ليس فيها أي افتراضات لتوزيع البيانات . ولهذا السبب ندعوها بالاختبارات الخالية من التوزيع (Distribution –Free Tests) .

إن عمليات التحليل في الاختبارات اللامعلمية عادة تعتمد على ترتيب البيانات فمثلا عدد النجاحات / الأخماج/ في البيانات المشاهدة ترتب تصاعديا أو تنازلياً منه من تحليها كبيانات خام غير مجرى عليها أية طريقة . وببساطة يمكن أن نضع بعض المقارنات بين الاختبارات المعلمية ونظيراتها في الاختبارات اللامعلمية.

لن نتمكن من دراسة كامل الاختبارات اللامعلمية نظراً لضيق الفترة الزمنية لمنهج الاحصاء المقرر للسنة الاولى ونعطي فكرة عن أهم اختبار من الاختبارات المعلمية وهو اختبار الإشارة . بقية الاختبارات يتم دراستها في مناهج الإحصاء للدراسات العليا حسب المناهج الإحصائية في العلوم الطبية لانها تسنخدم غالبا عندما تكون لدينا عينات صغيرة الحجم.

الجدول رقم ٣٠: مقارنة الاختبارات الحدية مع الاختبارات اللامعلمية

الاختبارات المعلمية	الاختبارات اللامعلمية
اختبارات لعينة واحدة	اختبار الإشارة
One sample t test	Sign test
اختبار –ت– الشفعي	اختبار الإشارة، اختبار رتب ويلكسون
Paired t test	Sign test , Wilcoxon Signed rank Test
اختبارات ذي العينيتين	اختبار رتب ويلكسون
Two sample t test	Wilcoxon rank test
تحليل التباين وحيد الإتجاه	كرسكال– واليز الوحيدة الاتجاه
One way -ANOVA	Kruskal –Wallis One way ANOVA
تحليل التباين باتجاهين	طريقة فريدمان الثنائية
Two way ANOVA	Fridman Two –way ANOVA
معامل ارتباط بيرسون	معامل ارتباط رتب سبيرمان
Pearson correlation coefficient	Spearman rank correlation coefficient

اختبار الإشارة The Sign Test

ويتمثل الاختبار الإحصائي :

$$Test12 = \left(\left| p - 0.5 \right| - 1/(2n) \right) \Big/ \sqrt{0.5^2 / n}$$

بإعتبار أن :

n : عدد التباينات غير مشتملة على القيمة صفر – كافة الأزواج المرتبطة ببعضها تستثنى من التحليل .

P = k/n : وهي نسبة المشاهدات الإيجابية (أو السلبية) للأزواج أو الأشفاع لقيم أزواج n الموحدة .

1/(2n) : وهو عبارة عن معامل تصحيح مستمر يسمح لنا باستخدام توزيع طبيعي مستمر لتقريب التوزيع الثنائي الاسمي المتقطع أو غير المستمر .

تعتبر بقية الاختبارات لمرحلة دراسية تخصصية في الإحصاء والتي أُشير إليها في الجدول السابق .

المـراجع العـربيـة

١- دركزنلـي، محمـد سـمير (٢٠٠٠)، الإحصـاء الحيـوي، منشـورات جامعـة حلـب. سوريا

٢- كنجو، أنيس اسماعيل (١٩٩٣). الإحصاء و الاحتمال. منشورات جامعة الملك سعود، المملكة العربية السعودية.

المراجـع الأجنبية
REFERENCES

1- Altman, D. G. (1992). How large a sample. In: Stastics in Practice. (ed. SM. Gore & D. G. Altman). British Medical Association.

2- Armitage, P. & Berry, G. (1994). Statistical Methods in Medical Research, 3rd edn. Blackwell Scientific Publication, Oxford.

3- Bland, J. M. & Kerry, S. M. (1997). Sample Size in Cluster Randomisation . British Medical Journal, Vol. 315, No. 600.

4- Burton, S. A., Lemke, K. A., Ihle, S. L., & MackKenzie, A.L.(1997). Effects of Medetomiddine on Serum Insulin and Plasma Glucose Concentrations in Clinically Normal Dogs. American

5- Cochran, W. G. (1977). Sampling techniques, 3rd edn. Wiley, New York.

6- Cox, D., R. & Oakes, D. (1984). Analysis of Survival Data. Chapman and Hall, London.

7- Donner, A. (1987). Statistical Methodology for Paired Cluster Design. American Journal of Epidemiology, 126, 972-9.

8- Everitt, B. S. (1995). The analysis of Repeated Measures: A practical Review with Examples. The Statistician, 44, 113.

9- Gardner, M. J. & Altman, D. G. (eds)(1989). Statistics with Confidence: Confidence Intervals and Statistical Guidelines. British Medical Journal, London.

10- Haber, M., Logini, I. M., & Halloran, M. E. (1991).International Journal of Epidemiology, 20 300-10.

11- Hseih, F. Y. (1988). Sample Size Formulae for Intervention Studiers with the Cluster as Unit of Randomisation. Statistics in Medicne, 8, 1195-201.

12- Jekel, J. F., Elmore, J. G. & Katz, D. L. (1996). Epidemiology, Biostatistics and Preventive Medicine. Saunders, Philadelphia.

13 - Kendall, M. G. (1980). Multivariate Analysis, 2d edn. Griffin, London.
14- Littele, T. W. A ., Richards, M. S., Hussaini, S. N. & Jones, T. D. (1980). The Significance of Leptospiral
15- Antibodies in Calving and Aborting Cattle in South West England. Veterinary Record, 106, 221-4
16- Manly, B. F. J. (1986). Multivariate Statistical Methods, A primer, Chapman and Hall , London.
17- McCoy, M. A., Goodall, E. A. & Kennedy, D. G. (1996). Incidence of Bovine Hypomagnesaemia in Northern Ireland and Methods of Magnesium Supplementation. Veterinary Record, 138, 41-3.
18- Nada, A. S., Dodson, H. & Ward, W. R. (1990). Relationship between an increase in Plasma Cortisol during Transport-Induced Stress and Failure of Oestradiol to induce a Luteinising Hormone Surge in Dairy Cows. Research Vterinary Science, 49, 25-8.
19- Nicholas, F. W. (1987). Veterinary Genetics. Oxford University Press, Oxford.
20- Pearson, R. A. & Ouassat, M. (1996). Estimation of the live weight and body condition of Working Donkeys in Morocco. Veterinary Record, 138. 229-33.
21- Schroeder, L. D., Sjoquist, D. L. & Stephan, P. E. (1986). Understanding Regression Analysis: An Introductory Guide. Sage University Paper Series on Quantitative Applications in the Social Science, Series No. 07-057. Sage University Press, Beverly Hills, California.
22- Siegel, S. & Castellan, N. J. (1988). Non-parametric Statistics for the Behavioral Science, 2nd edn. McGraw-Hill, New York.
23- Yates, F. (1981). Sampling Methods for Cenuses and Surveys, 4th edn. Griffin, London.

البرامج الإحصائية المشار إليها خلال سرد التقنيات الإحصائية

Computer Packages Mentioned

Borenstein, M., Rothstein, H. & Cohen, J. (1997). Sample Power 10.0. SPSS., Chicago, USA.

N-Query Advisor 2.0. Statistical Solutions Ltd. SPSS, version 10.0. SPSS Inc. Chicago, USA.

STATISTIX, (1998). Analytical Software, Manual Guide, Version 2.0, New York, USA.

S-Plus, (1998). User' Manual. MathSoft, Version 4.0. StatSci. Division, MathSoft, Inc. Seattle, Washington, USA.

تم بعون الله

الملحقات الإحصائية

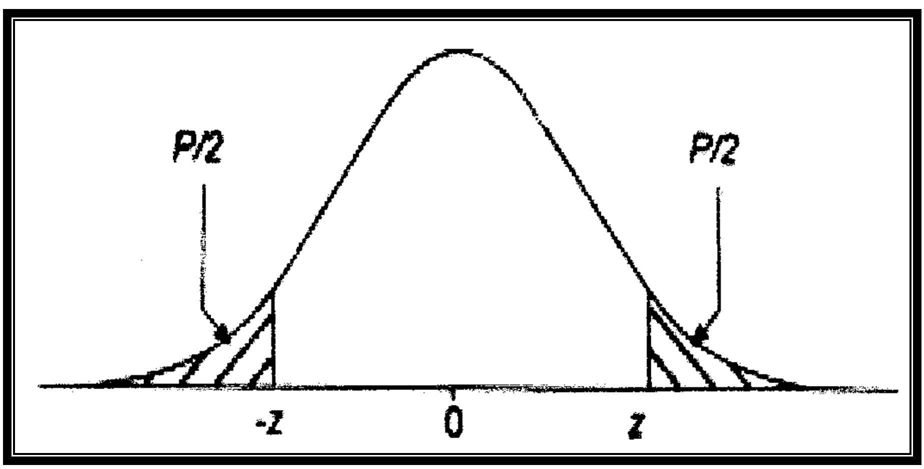
الجداول الإحصائية

إن القيمة المجدولة هي عبارة عن قيمة (P) ذات القيمة المعبرة عن ذات الذيلين للتوزيعات الطبيعية المطابقة لقيمة محددة للإنحراف الطبيعي المعياري .

$Z=X-M/Q$

بإعتبار أن قيمة (X) هي عبارة عن المتغير ذي التوزيع الطبيعي ، مع متوسط = (M) وانحراف معياري (Q) .

على سبيل المثال إذا كانت قيمة (Z) = ١.٩٦ فعندئذ تكون قيمة (P) ذات الذيلين = ٠.٠٥ (P= 0.025) في كل جانب أو ما يسمى بالمصطلح الإحصائي في كل ذيل كما هو مبين أدناه :



ملحق A1 : التوزيع الطبيعي المعياري

The Standard Normal Distribution
(Two-tailed P-Values from Values of z, the SND)

Z	P	Z	P	Z	P	Z	P	Z	P	Z	P
0.00	1.0000	0.53	0.5961	1.04	0.2983	1.56	0.1188	2.08	0.0375	2.59	0.0096
0.01	0.9920	0.54	0.5892	1.05	0.2937	1.57	0.1164	2.09	0.0366	2.60	0.0093
0.02	0.9840	0.55	0.5823	1.06	0.2891	1.58	0.1141	2.10	0.0357	2.61	0.0091
0.03	0.9761	0.56	0.5755	1.07	0.2846	1.59	0.1118	2.11	0.0349	2.62	0.0088
0.04	0.9681	0.57	0.5687	1.08	0.2801	1.60	0.1096	2.12	0.0340	2.63	0.0085
0.05	0.9601	0.58	0.5619	1.09	0.2757	1.61	0.1074	2.13	0.0332	2.64	0.0083
0.06	0.9522	0.59	0.5552	1.10	0.2713	1.62	0.1052	2.14	0.0324	2.65	0.0080
0.07	0.9442	0.60	0.5485	1.11	0.2670	1.63	0.1031	2.15	0.0316	2.66	0.0078
0.08	0.9362	0.61	0.5419	1.12	0.2627	1.64	0.1010	2.16	0.0308	2.67	0.0076
0.09	0.9283	0.62	0.5353	1.13	0.2585	1.65	0.0989	2.17	0.0300	2.68	0.0074
0.10	0.9203	0.63	0.5287	1.14	0.2543	1.66	0.0969	2.18	0.0293	2.69	0.0071
0.11	0.9124	0.64	0.5222	1.15	0.2501	1.67	0.0949	2.19	0.0285	2.70	0.0069
0.12	0.9045	0.65	0.5157	1.16	0.2460	1.68	0.0930	2.20	0.0278	2.71	0.0067
0.13	0.8966	0.66	0.5093	1.17	0.2420	1.69	0.0910	2.21	0.0271	2.72	0.0065
0.14	0.8887	0.67	0.5029	1.18	0.2380	1.70	0.0891	2.22	0.0264	2.73	0.0063
0.15	0.8808	0.68	0.4965	1.19	0.2340	1.71	0.0873	2.23	0.0257	2.74	0.0061
0.16	0.8729	0.69	0.4902	1.20	0.2301	1.72	0.0854	2.24	0.0251	2.75	0.0060
0.17	0.8650	0.70	0.4839	1.21	0.2263	1.73	0.0836	2.25	0.0244	2.76	0.0058
0.18	0.8572	0.71	0.4777	1.22	0.2225	1.74	0.0819	2.26	0.0238	2.77	0.0056
0.19	0.8493	0.72	0.4715	1.23	0.2187	1.75	0.0801	2.27	0.0232	2.78	0.0054
0.20	0.8415	0.73	0.4654	1.24	0.2150	1.76	0.0784	2.28	0.0226	2.79	0.0053
0.21	0.8337	0.74	0.4593	1.25	0.2113	1.77	0.0767	2.29	0.0220	2.80	0.0051
0.22	0.8259	0.75	0.4533	1.26	0.2077	1.78	0.0751	2.30	0.0214	2.81	0.0050
0.23	0.8181	0.76	0.4473	1.27	0.2041	1.79	0.0735	2.31	0.0209	2.82	0.0048
0.24	0.8103	0.77	0.4413	1.28	0.2005	1.80	0.0719	2.32	0.0203	2.83	0.0047
0.25	0.8026	0.78	0.4354	1.29	0.1971	1.81	0.0703	2.33	0.0198	2.84	0.0045
0.26	0.7949	0.79	0.4295	1.30	0.1936	1.82	0.0688	2.34	0.0193	2.85	0.0044
0.27	0.7872	0.80	0.4237	1.31	0.1902	1.83	0.0672	2.35	0.0188	2.86	0.0042
0.28	0.7795	0.81	0.4179	1.32	0.1868	1.84	0.0658	2.36	0.0183	2.87	0.0041
0.29	0.7718	0.82	0.4122	1.33	0.1835	1.85	0.0643	2.37	0.0178	2.88	0.0040
0.30	0.7642	0.83	0.4065	1.34	0.1802	1.86	0.0629	2.38	0.0173	2.89	0.0039
0.31	0.7566	0.84	0.4009	1.35	0.1770	1.87	0.0615	2.39	0.0168	2.90	0.0037
0.32	0.7490	0.85	0.3953	1.36	0.1738	1.88	0.0601	2.40	0.0164	2.91	0.0036
0.33	0.7414	0.86	0.3898	1.37	0.1707	1.89	0.0588	2.41	0.0160	2.92	0.0035
0.34	0.7339	0.87	0.3843	1.38	0.1676	1.90	0.0574	2.42	0.0155	2.93	0.0034
0.35	0.7263	0.88	0.3789	0.39	0.1645	1.91	0.0561	2.43	1.0151	2.94	0.0033
0.36	0.7188	0.89	0.3735	0.40	0.1615	1.92	0.0549	2.44	0.0147	2.95	0.0032
0.37	0.7114	0.90	0.3681	1.41	0.1585	1.93	0.0536	2.45	0.0143	2.96	0.0031
0.38	0.7039	0.91	0.3628	1.42	0.1556	1.94	0.0524	2.46	0.0139	2.97	0.0030
0.39	0.6965	0.92	0.3576	1.43	0.1527	1.95	0.0512	2.47	0.0135	2.98	0.0029
0.40	0.6892	0.93	0.3524	1.44	0.1499	1.96	0.0500	2.48	0.0131	2.99	0.0028
0.41	0.6818	0.94	0.3472	1.45	0.1471	1.97	0.0488	2.49	0.0128	3.00	0.0027
0.42	0.6745	0.95	0.3421	1.46	0.1443	1.98	0.0477	2.50	0.0124	3.10	0.00194
0.43	0.6672	0.96	0.3371	1.47	0.1416	1.99	0.0466	2.51	0.0121	3.20	0.00137
0.44	0.6599	0.97	0.3320	1.48	0.1389	2.00	0.0455	2.52	0.0117	3.30	0.00097
0.45	0.6527	0.98	0.3271	1.49	0.1362	2.01	0.0444	2.53	0.0114	3.40	0.00067
0.46	0.6455	0.99	0.3222	1.50	0.1336	2.02	0.0434	2.54	0.0111	3.50	0.00047
0.47	0.6384	1.00	0.3173	1.51	0.1310	2.03	0.0424	2.55	0.0108	3.60	0.00032
0.48	0.6312	1.01	0.3125	1.52	0.1285	2.04	0.0414	2.56	0.0105	3.70	0.00022
0.49	0.6241	1.02	0.3077	1.53	0.1260	2.05	0.0404	2.57	0.0102	3.80	0.00014
0.50	0.6171	1.03	0.3030	1.54	0.1236	2.06	0.0394	2.58	0.0099	3.90	0.00010
0.51	0.6101			1.55	0.1211	2.07	0.0385			4.00	0.00006
0.52	0.6031										

الملحق A2 : التوزيع الطبيعي المعياري (قيم Z , SND من قيم P)

The Standard Normal Distribution
(Values of z, the SND, from P- Value)

Two-tailed probanility	SND-z	One-tailed probability
1.00	0.00	0.50
0.90	0.13	0.45
0.50	0.67	0.25
0.25	1.15	0.125
0.20	1.28	0.10
0.15	1.44	0.075
0.10	1.64	0.50
0.05	1.96	0.025
0.02	2.33	0.01
0.01	2.58	0.005
0.005	2.81	0.0025
0.001	3.29	0.0005

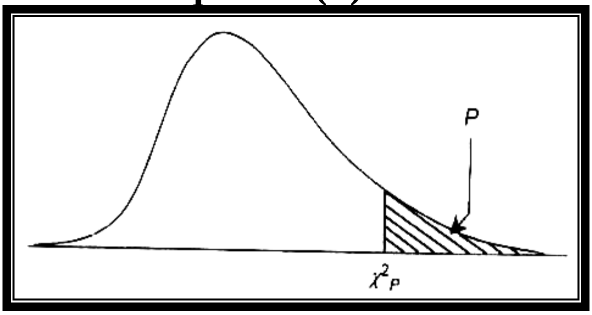
Percentage of the t- Distribution الملحق A3 : النقاط المئوية لتوزيع ت ستودنت

df	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.001
1	0.158	0.325	0.510	0.727	1.000	1.376	1963	3.078	6.314	12.206	31.821	63.667	636619
2	0.142	0.289	0.445	0.617	0.816	1.061	1386	1.886	2.920	4.303	6.965	9.925	31593
3	0.137	0.277	0.424	0.584	0.765	0.978	1290	1.638	2.353	3.182	4.541	5.841	12924
4	0.134	0.271	0.414	0.569	0.741	0.941	1490	1.533	2.132	2.776	3.747	4.604	8610
5	0.132	0.267	0.408	0.559	0.727	0.920	1456	1.476	2.015	2.571	3.366	4.032	6869
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.449	1.943	2.447	3.143	3.707	5.959
7	0.130	0.265	0.402	0.549	0.711	0.896	1.119	1.415	1.825	2.365	2.995	3.499	5.408
8	0.130	0.262	0.393	0.546	0.706	0.889	1.108	0.397	1.86	2.366	2.896	3.355	5.011
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.381
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.396	1.782	2.179	2.381	3.055	4.318
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.128	0.258	0.392	0.534	0.689	0.863	1.069	1.333	1.741	2.110	2.567	2.898	3.965
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.056	2.528	2.845	3.850
21	0.127	0.257	0.391	0.532	0.696	0.859	1.063	1.323	1.721	2.090	2.518	2.831	3.819

22	0.127	0.256	0.390	0.532	0.696	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.214	2.069	2.900	2.817	3.767
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.311	2.064	2.492	2.397	3.745
25	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.127	0.256	0.390	0.531	0.681	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.127	0.256	0.389	0.531	0.681	0.855	1.067	1.314	1.303	2.062	2.473	2.771	3.690
28	0.127	0.256	0.389	0.530	0.681	0.855	1.056	1.313	1.301	2.048	2.457	2.763	3.674
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.699
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.319	1.697	2.042	2.457	2.790	3.646
40	0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
90	0.126	0.255	0.388	0.528	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.497
100	0.126	0.254	0.386	0.526	0.677	0.845	1.042	1.291	1.661	1.984	2.364	2.626	3.391
200	0.126	0.254	0.386	0.525	0.676	0.843	1.039	1.286	1.653	1.972	2.345	2.601	3.340
∞	0.126	0.253	0.385	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

الملحق A4 : توزيع مربع كاي

The Chi Squared (α) Distribution



Two-taited P-value							
df	0.500	0.250	0.100	0.050	0.025	0.010	0.001
1	0.45	1.32	2.71	3.84	5.02	6.63	10.83
2	1.39	2.77	4.61	5.99	7.38	9.21	13.82
3	2.37	4.11	6.25	7.81	9.35	11.34	16.27
4	3.36	5.39	7.78	9.49	11.14	13.28	18.47
5	4.35	6.63	9.24	11.07	12.83	15.09	20.52
6	5.35	7.84	10.64	12.59	14.45	16.81	22.46
7	6.35	9.04	12.02	14.07	16.01	18.48	24.32
8	7.37	10.22	13.36	15.51	17.53	20.09	26.12
9	8.34	11.39	14.68	16.92	19.02	21.67	27.88
10	9.34	12.55	15.99	18.31	20.48	23.21	29.59
11	10.34	13.70	17.28	19.68	21.92	24.72	31.26
12	11.34	14.85	18.55	21.03	23.34	26.22	32.91
13	12.34	15.98	19.81	22.36	24.74	27.69	34.53
14	13.34	17.12	21.06	23.68	26.12	29.14	36.12
15	14.34	18.25	22.31	25.00	27.49	3058	37.70
16	15.34	19.37	23.54	26.30	28.85	32.00	39.25
17	16.34	20.49	24.77	27.59	30.19	33.41	40.79
18	17.34	21.60	25.99	28.87	31.53	34.81	42.31
19	18.34	22.72	27.20	30.14	32.85	36.19	43.82
20	19.34	23.83	28.41	31.41	34.17	37.57	45.32
21	20.34	24.93	29.62	32.67	35.48	38.93	46.80
22	21.34	26.04	30.81	33.92	36.78	40.29	48.27
23	22.34	27.14	32.01	35.17	38.08	41.64	49.73
24	23.34	28.24	33.20	36.42	39.36	42.98	51.18
25	24.34	29.34	34.38	37.65	40.65	44.31	32.62
26	25.34	30.43	35.56	38.89	41.92	45.64	54.05
27	26.34	31.53	36.74	40.11	43.19	46.96	55.48
28	27.34	32.62	37.92	41.34	44.46	48.28	56.89
29	28.34	33.71	39.09	42.56	45.72	49.59	58.30
30	29.34	34.80	40.26	43.77	46.98	50.89	59.70
40	39.34	45.62	51.80	55.76	59.34	63.69	73.40
50	49.33	56.33	63.17	67.50	71.42	76.15	86.66
60	59.33	66.96	74.40	79.08	83.30	88.38	99.61
70	69.33	77.58	85.53	90.53	95.02	100.42	112.32
80	79.33	88.13	96.58	101.88	106.63	112.33	124.84
90	89.33	98.64	107.56	113.14	118.14	124.12	137.21
100	99.33	109.14	118.50	124.34	129.56	135.81	149.45

الملحق A5a : النقاط المئوية لتوزيع F (P=0.05 , P=0.01)

Percentage Points of the F – Distribution (P=0.05 , P=0.01)

		df numerator. V1									
df denominator v2	P	1	2	3	4	5	6	7	8	12	24
1	0.05	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	243.9	249.1
	0.01	4052	5000	5403	5625	5764	5859	5928	5981	6106	6235
2	0.05	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.41	19.45
	0.01	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.42	99.46
3	0.05	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.74	8.64
	0.01	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.05	26.60
4	0.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.91	5.77
	0.01	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.37	13.93
5	0.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.68	4.53
	0.01	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	9.89	9.47
6	0.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.00	3.84
	0.01	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.72	7.31
7	0.05	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.57	3.41
	0.01	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.47	6.07
8	0.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.28	3.12
	0.01	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.67	5.28

9	0.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.07	2.90
	0.01	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.11	4.73
10	0.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.91	2.74
	0.01	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.71	4.33
12	0.05	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.69	2.51
	0.01	9.33	6.93	5.95	5.411	5.06	4.82	4.64	4.50	4.16	3.78
14	0.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.53	2.35
	0.01	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	3.80	3.43
16	0.05	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.42	2.24
	0.01	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.55	3.18
18	0.05	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.34	2.15
	0.01	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.37	3.00
20	0.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.28	2.08
	0.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.23	2.86
30	0.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.09	1.89
	0.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	2.84	2.47
40	0.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.00	1.79
	0.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.66	2.29
60	0.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	1.92	1.70
	0.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.50	2.12
120	0.05	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.2	1.83	1.61
	0.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.34	1.95

الملحق A5b : النقاط المئوية لتوزيع F (P=0.05 , P=0.01)

Percentage Points of the F – Distribution (P=0.05 , P=0.01)

Df denominator V2	P	1	2	3	4	5	6	7	8	12	24	∞
1	0.025	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	976.7	997.2	1018
	0.005	16211	20000	21615	22500	23056	23437	23715	23925	24426	24940	25465
2	0.025	38.51	.39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.41	39.46	39.50
	0.005	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.5	199.5
3	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.34	14.12	13.90
	0.005	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.39	42.62	41.83
4	0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.75	8.51	8.26
	0.005	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	20.70	20.03	19.32
5	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.52	6.28	6.02
	0.005	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.38	12.78	12.14
6	0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.37	5.22	4.85
	0.005	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.03	9.47	8.88
7	0.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.67	4.42	4.14
	0.005	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.18	7.65	7.08
8	0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.20	3.95	3.67
	0.005	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.01	6.50	5.95
9	0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	3.87	3.61	3.33
	0.005	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.23	5.73	5.19

10	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.62	3.37	3.08
	0.005	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.66	5.17	4.64
12	0.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.28	3.02	2.72
	0.005	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	4.91	4.43	3.90
14	0.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.05	2.79	2.49
	0.005	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.43	3.96	3.44
16	0.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	2.89	2.63	2.32
	0.005	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.10	3.64	3.11
18	0.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.77	2.50	2.19
	0.005	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	3.86	3.40	2.87
20	0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.68	2.41	2.09
	0.005	9.94	6.99	5.82	5.7	4.76	4.47	4.26	4.09	3.68	3.22	2.69
30	0.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.41	2.14	1.79
	0.005	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.18	2.73	2.18
40	0.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.29	2.01	1.64
	0.005	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	2.95	2.50	1.93
60	0.025	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.17	1.88	1.48
	0.005	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	2.74	2.29	1.69
120	0.025	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.05	1.76	1.31
	0.005	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.54	2.09	1.43
∞	0.025	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	1.94	1.64	1.00
	0.005	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.36	1.90	1.00

الملحق A6 : القيم المحددة لعامل ارتباط بيرسون

Critical Values of Prearson's Correlation Coefficient (r)

	Two-tailed P-value				Two-tailed P-value		
Sample size	0.05	0.01	0.001	Sample size	0.05	0.01	0.001
3	0.9969	0.9999	1.0000	23	0.4132	0.5256	0.6402
4	0.9500	0.9900	0.9990	24	0.4044	0.5151	0.6287
5	0.8783	0.9587	0.9911	25	0.3961	0.5152	0.6177
6	0.8114	0.9172	0.9741	26	0.3882	0.4958	0.6073
7	0.7545	0.8745	0.9509	27	0.3809	0.4869	0.5974
8	0.7067	0.8343	0.9249	28	0.3739	0.4785	0.5880
9	0.6664	0.7977	0.8983	29	0.3673	0.4705	0.5790
10	0.6319	0.7646	0.8721	30	0.3610	0.4629	0.5703
11	0.6021	0.7348	0.8471	35	0.3338	0.4296	0.5322
12	0.5760	0.7079	0.8233	40	0.3120	0.4026	0.5007
13	0.5529	0.6835	0.8010	45	0.2940	0.3801	0.4742
14	0.5324	0.6614	0.7800	50	0.2787	0.3610	0.4514
15	0.5139	0.6411	0.7604	55	0.2656	0.3445	0.4317
16	0.4973	0.6226	0.7419	60	0.2542	0.3301	0.4143
17	0.4821	0.6055	0.7247	70	0.2352	0.3060	0.3850
18	0.4683	0.5897	0.7084	80	0.2199	0.2864	0.3611
19	0.4555	0.5751	0.6932	90	0.2072	0.2702	0.3412
20	0.4438	0.5614	0.6788	100	0.2172	0.2830	0.3569
21	0.4329	0.5487	0.6652	150	0.1603	0.2097	0.2660
22	0.4227	0.5368	0.6524				

الملحق A7 : معامل ارتباط رتب سبيرمان

Spearman's Rank Correlation Coefficient (rs)

Sample size	0.1	0.05	0.02	0.01	0.002
4	0.8000	-	-	-	-
5	0.8000	0.9000	0.9000	-	-
6	0.7714	0.8286	0.8857	0.9429	-
7	0.6789	0.7450	0.8571	0.8929	0.9643
8	0.6190	0.7143	0.8095	0.8571	0.9286
9	0.5833	0.6833	0.7667	0.8167	0.9000
10	0.5515	0.6364	0.7333	0.7818	0.8667
11	0.5273	0.6091	0.7000	0.7455	0.8364
12	0.4965	0.5804	0.6713	0.7273	0.8182
13	0.4780	0.5549	0.6429	0.6978	0.7912
14	0.4593	0.5341	0.6220	0.6747	0.7670
15	0.4429	0.5179	0.6000	0.6536	0.7464

الملحق A8 : قيم P ذات الذيلين لاختبار الإشارة

Two tailed P – values

	K									
N	0	1	2	3	4	5	6	7	8	9
4	0.125	0.624	1.00	-	-	-	-	-	-	-
5	0.062	0.376	1.00	-	-	-	-	-	-	-
6	0.032	0.218	0.688	1.00	-	-	-	-	-	-
7	0.016	0.124	0.454	1.00	-	-	-	-	-	-
8	0.008	0.070	0.290	0.726	1.00	-	-	-	-	-
9	0.004	0.040	0.180	0.508	1.00	-	-	-	-	-
10	0.002	0.022	0.110	0.344	0.754	1.00	-	-	-	-
11	0.001	0.012	0.066	0.226	0.548	1.00	-	-	-	-
12	-	0.006	0.038	0.146	0.388	0.774	1.00	-	-	-
13	-	0.004	0.022	0.092	0.266	0.582	1.00	-	-	-
14	-	0.002	0.012	0.058	0.180	0.424	0.790	1.00	-	-
15	-	-	0.008	0.036	0.118	0.302	0.608	1.00	-	-
16	-	-	0.004	0.022	0.076	0.210	0.554	0.804	1.00	-
17	-	-	0.002	0.012	0.050	0.144	0.332	0.630	1.00	-
18	-	-	0.002	0.004	0.030	0.096	0.238	0.480	0.814	1.00
19	-	-	-	0.004	0.020	0.064	0.168	0.360	0.648	1.00
20	-	-	-	0.002	0.012	0.042	0.116	0.264	0.504	0.824

الملحق A9 : اختبار رتب وليكوكسون سايند
The Wilcoxon Signed Rank Test

Two-tailed P-value					
N	0.1	0.05	0.02	0.01	0.001
4	-	-	-	-	-
5	0-15	-	-	-	-
6	2-19	0-21	-	-	-
7	3-25	2-26	0-28	-	-
8	5-31	3-33	1-35	0-36	-
9	8-37	5-40	3-42	1-44	-
10	10-45	8-47	5-50	3-52	-
11	13-53	10-56	7-59	5-61	0-66
12	17-61	13-65	9-69	7-71	1-77
13	21-70	17-74	12-79	9-82	2-89
14	25-80	21-84	15-90	12-93	4-101
15	30-90	25-95	19-101	15-105	6-114
16	35-101	29-107	23-113	19-117	9-127
17	41-112	34-119	28-125	23-130	11-142
18	47-124	40-131	32-139	27-144	14-157
19	53-137	46-144	37-153	32-158	18-172
20	60-150	52-158	43-167	37-173	21-189
21	67-164	58-173	49-182	42-189	26-205
22	75-178	66-187	55-198	48-205	30-223
23	83-196	73-203	62-214	54-222	35-241
24	91-209	81-219	69-231	61-239	40-260
25	100-225	89-236	76-249	68-257	45-280

الملحق A10b : القيم المحددة لاختبار مجموع رتب وليكوكسون .

قيمة الإحتمالية ذات الذيلين = ٠.٠٥

Critical Values for the Wilcoxon Rank Sum Test. Twotailed P=0.05

n_1	4	5	6	7	8	9	10	11	12	13	14	15
n_2												
4	10-26	16-34	23-43	31-53	40-64	49-77	60-90	72-104	85-119	99-135	114-	130-
5	11-29	17-38	24-48	33-58	42-70	52-83	63-97	75-112	89-127	103-	152	170
										144	118-	134-
6	12-32	18-42	26-52	34-64	44-76	55-89	66-104	79-119	92-136		162	181
7	13-35	20-45	27-57	36-69	46-82	57-96	69-111	82-127	96-144	107-		
8	14-38	21-49	29-61	38-74	49-87	60-102	72-118	85-135	100-	153	122-	139-
9	14-42	22-53	31-65	40-79	51-93	62-109	75-125	89-142	152	111-	172	191
10	15-45	23-57	32-70	42-84	53-99	65-115	78-132	92-150	104-	162	127-	144-
									160	115-	181	201
11	16-48	24-61	34-74	44-89	55-105	68-121	81-139	96-157	107-	171	131-	149-
12	17-51	26-64	35-79	46-94	35-110	71-127	84-146	99-165	169	119-	191	211
13	18-54	27-68	37-83	48-99	60-116	73-134	88-152	103-		180	136-	154-
14	19-57	28-72	38-88	50-104	62-122	76-140	91-159	172	111-	124-	200	221
15	20-60	29-76	40-92	52-109	65-127	79-146	94-166	106-	177	188	141-	159-
								180	115-		209	231
16	21-63	30-80	42-96	54-114	67-133	82-152	97-173	110-	185	128-		
17	21-67	32-83	43-101	56-119	70-138	84-159	100-	187	119-	197	145-	164-
18	22-70	33-87	45-105	58-124	72-144	87-165	180		193	132-	219	241
19	23-73	34-91	46-110	60-129	74-150	90-171	103-	113-	123-	206	150-	169-
20	24-76	35-95	48-114	62-134	77-155	93-177	187	195	201	136-	228	251
							107-	117-	127-	215	155-	174-
							193	202	209	141-	237	261
							110-	121-		223	160-	179-

							200	209 124- 217 128- 224	131- 217 135- 225 139- 233 143- 241 147- 249	145- 232 150- 240 154- 249 159- 257 163- 266 167- 275	246 164- 256 169- 265 174- 274 179- 283 184- 292 188- 302	271 184- 281 190- 290 195- 300 200- 310 205- 320 211- 329

الملحق A10a : القيم المحددة لاختبار مجموع رتب وليكوكسون .

قيمة الإحتمالية ذات الذيلين = ٠.٠٥

Critical Values for the Wilcoxon Rank Sum Test. Twotailed P=0.05

n_1	4	5	6	7	8	9	10	11	12	13	14	15
n_2												
4	_	_	21-45	28-56	36-67	46-80	57-93	68-108	81-123	94-140	109-	125-
5	_	15-40	22-50	29-62	38-74	48-87	59-	71-116	84-132	98-149	157	175
							101				112-	128-
6	10-	16-44	23-55	31-67	40-80	50-94		73-125	87-141	101-	168	187
7	34	16-49	24-60	32-73	42-86	52-	61-	76-133	90-150	159		
8	10-	17-53	25-65	34-78	43-93	101	109	79-141	93-159	104-	116-	132-
9	38	18-57	26-70	35-84	45-99	54-	64-	82-149	96-168	169	178	198
10	11-	19-61	27-75	37-89	47-	108	116	84-158	99-177	108-	120-	136-
	41				105	56-	66-			178	188	209
11	11-	20-65	28-80	38-95		115	124	87-166	102-	111-	123-	140-
12	45	21-69	30-84	40-	49-	58-	68-	90-174	186	188	199	220
13	12-	22-73	31-89	100	111	122	132	93-182	105-	115-	127-	144-
14	48	22-78	32-94	41-	51-		71-	96-190	195	197	209	231
15		23-82	33-99	106	117	61-	139	99-198	109-		131-	149-
	12-			43-	53-	128			203	118-	219	241
16	52	24-86	34-	111	123	63-	73-	102-	112-	207		
17	13-	25-90	104	44-	54-	135	147	206	212	122-	135-	153-
18	55	26-94	36-	117	130	65-	76-	105-	115-	216	229	252
19	13-	27-98	108		56-	142	154	214	221	125-	139-	157-
20	59	28-	37-	46-	136	67-	79-	108-		226	239	263
	14-	102	113	122		149	161	222	119-	129-	143-	162-

	62 15-65 15-69 16-72 16-76 17-79 18-82		38-118 39-123	47-128 49-133 50-139 52-144	58-142 60-148 62-154 64-160 66-166	69-156 72-162 74-169 76-176 78-183 81-189	81-169 84-176 86-184 89-191 92-198 94-206 97-213	111-230 114-238	229 122-238 125-247 129-255 132-264	235 133-244 136-254 140-263 144-272 148-281 152-290	249 147-259 151-269 155-279 160-288 164-298 168-308 172-318	273 166-284 171-294 175-305 180-315 184-326 189-336 193-347

الملحق A11 : جدول الأرقام العشوائية
Table of Random Numbers

77267	67258	38499	94709	46989	44360	46788	62666	67551	79212
72309	70484	25843	72251	82013	70561	14058	38073	53571	91594
54395	89438	92622	45780	29108	53340	85537	50232	28477	93512
98270	62867	44084	98370	59635	25367	30528	58516	78666	83753
66032	31218	29309	26890	34700	43168	09914	47240	51526	51115
57577	70054	60345	84988	24257	19358	39083	63075	67491	55733
85981	87059	50122	80180	98114	64749	75696	59666	43806	52538
75254	50278	61364	84524	18067	94064	42011	21085	79258	44419
71820	17948	38074	48411	63605	34244	96320	36384	80985	79176
55759	77728	41765	61731	27045	81464	44584	11390	85593	69342
92725	91260	25468	94632	44972	96413	93134	29630	70497	71787
18169	44658	95643	71214	61018	90640	59106	76377	90625	15455
55710	88227	84684	33948	29576	57306	96961	90832	52720	38631
49556	24412	93967	63006	69252	52089	29551	62555	54033	39961
87891	87778	61646	24558	16210	81147	16734	24214	45062	64957
18699	87766	21808	40788	77612	97617	34199	86693	51631	17475
96353	24916	71714	97492	42680	14894	87091	51667	10183	28272
58932	54779	81765	17127	76773	52970	52430	56064	62116	48515
60376	97508	91787	84684	39870	12608	11299	30277	87317	61131
71521	34632	82603	56428	71537	10548	10765	51679	45875	12404
17760	37556	52225	68445	18626	67414	62242	51329	80427	11747
36901	23375	21348	32148	47612	60511	24558	91901	50626	65405
17488	34113	69144	24953	13842	90301	38518	59852	96747	96478
16435	63514	78929	62326	89294	48853	35503	43729	89186	36601
79145	37322	17054	61899	74394	58695	77454	81735	98688	91397
96388	61117	31714	58107	85666	47675	33123	08943	75625	06598
38147	23339	32981	80989	96940	44860	39707	84883	26243	59861
95893	06491	95520	91538	35285	17192	80784	32664	49226	25919
53544	31391	23798	36857	16786	19639	93659	66776	34108	74268
03513	34015	78337	46158	92198	99481	70804	73939	39152	44116
92737	89927	81721	33548	78029	62464	53482	54191	95898	66099
10688	61502	73817	63841	87058	23377	24045	99470	17509	26636
51658	59565	61280	48120	38438	57832	25639	84632	38523	89459
53916	57066	46906	18657	79932	93039	62470	22405	78427	92145
40912	63211	63856	61644	18635	02946	30842	24031	36992	37917
75159	14888	59932	74222	39075	33201	33747	53800	79883	26609
25224	72513	58746	52366	73436	74699	80799	36699	16557	58671
82703	53196	34797	28093	97105	56797	39992	67944	00310	49311
51627	41127	86363	48078	27726	37269	21629	21785	25822	95264
02574	68647	82762	80442	70966	95743	56140	58213	78202	90038

جدول المصطلحات العلمية

A	
Absolute value	القيمة المطلقة
Analysis of variance	تحليل التباين
Approximation	تقريب
Arithmetic mean	الوسط الحسابي
Ascending order	ترتيب تصاعدي
B	
Binomial distribution	توزيع ثنائي
C	
Chi – square	مربع كاي
Coding	ترميز
Coefficient	معامل
Coefficient variability	معامل الاختلاف (C.V)
Coefficient of determination	معامل التحديد (R2)
Confidence interval	حد الثقة
Continuous random variable	متغير عشوائي مستمر
Correlation	ارتباط
Correlation Coefficient	معامل الارتباط
Cumulative frequency	تكرار توزيعي
D	
Data	البيانات
Design	التصميم
Degree of freedom	درجة الحرية
Descriptive of data	وصف البيانات

Descriptive statistics	الاحصاء الوصفي
Deviation	إنحراف
Deviation mean	متوسط الانحراف
Distribution	توزيع
Discrete	منفصل
Discrete probability distribution	توزيع احتمالي منفصل
Discrete random distribution	توزيع عشوائي منفصل
E	
Effect	تأثير
Equations	معادلات
Error	خطأ
Error (Experimental, Random, Standard, Systematic)	خطأ (تجريبي ، عشوائي ، قياسي ، نظامي)
Estimate	تقدير
Event	حادثة
Expect	المتوقع أو النظري
Experiment	تجربة
F	
Factor	عامل
Frequency (distribution, curve, table)	تكرار (توزيع ، منحنى التكرار، جدول التكرار)
F- test	اختبار ف
F distribution	توزيع ف
G	
Graphic	بياني

Graphic presentation	عرض بياني
H	
Histogram	مخطط مستطيلات
Hypothesis	الفرضية
Hypothesis testing	اختبار الفرضية
I	
Independent	الاستقلال
Independent events	حوادث مستقلة
Independent trails	تجارب مستقلة
Inspection	كشف
Inter-quartile range	مدى الربيع
Interval estimation	تقدير بفترة (تقدير مجالي)
Introduction	مقدمة
K	
Kurtosis	تفرطح
L	
Level of significance	مستوى المعنوية
Least significant difference	اختبار أقل معنوية
Least significant rang	اختبار أقل مدى معنوي
Linear	خطي
Linear correlation	الارتباط الخطي
Linear regression	الانحدار الخطي
Line graph	خط بياني
Lower confidence bound	حد الثقة الأدنى
Lower quartile	الربيع الأدنى
M	

Mathematical expectation	التوقع الرياضي
Mean	متوسط (وسط)
(Arithmetic, Geometric)	(وسط حسابي ، وسط هندسي)
Mean absolute	متوسط الانحراف المطلق
Mean of squares (MS)	متوسط المربعات (التباين)
Measure	قياس (مقياس)
Measures of central tendency	مقاييس الميول المركزية
Measures of dispersion	مقاييس التشتت
Measures of reliability	اختبـارات درجـة الإعتماديـة علـى النتائج
Median	الوسيط
Methods of data classification	طرق ترتيب البيانات
Methods of data presentation	طرق عرض البيانات
Missing values	القيم المفقودة
Mode	المنوال
Model (Linear, Mathematical)	نموذج (خطي ، رياضي)
N	
Nature order	ترتيب طبيعي
Nominal data	بيانات وصفية (اسمية)
Normal approximation to binomial	تقريب طبيعي إلى توزيع ثنائي
Normal curve	منحنى طبيعي
Normal distribution	توزيع طبيعي
Null hypothesis	فرضية العدم

O	
Observations	مشاهدات
Observed frequencies	تكرارات مشاهدة
One sided test	اختبارات من جهة واحدة
Ordinal data	بيانات ترتيبية
Outcome	نتيجة
p	
Parameters	حد (معلمة)
Partial	جزئي
Partition	تجزئة
Percentage	نسبة مئوية
Point estimation	التقدير النقطي
Population	المجتمع
Population mean	الوسط الحسابي للمجتمع
Population parameters	حدود (معالم)المجتمع
Population size	حجم المجتمع
Population standard deviation	الانحراف المعياري للمجتمع
Population variance	تباين المجتمع
Poisson distribution	توزيع بواسون
Probability	احتمال
Probability distribution	التوزيع الاحتمالي
Properties	خواص
Proportion	نسبة

Q	
Quartiles	الربيعات
R	
Random	عشوائي
Random numbers	أرقام عشوائية
Random samples	عينات عشوائية
Random variables	متغيرات عشوائية
Range	مدى
Rank correlation	ارتباط الرتب
Ratio	تناسب
Raw data	البيانات الخام
Real	حقيقي
Real numbers	أرقام حقيقية
Regression	انحدار
Regression coefficient	معامل الانحدار
Regression equation	معادلة الانحدار
Regression line	مستقيم الانحدار
Rejection region	منطقة الرفض
Relative frequency	التكرار النسبي
Rule	قاعدة
Rules of probability	قواعد الاحتمال
S	
Sample	عينة
Sample mean	الوسط الحسابي للعينة
Sample size	حجم العينة

Sample theory	نظرية العينة
Sampling	معاينة
Sampling distribution	توزيع المعاينة
Scatter diagram	مخطط الانتشار
Significance level	مستوى المعنوية
Simple regression	الانحدار البسيط
Spearman 's coefficient of correlation	معامل ارتباط سبيرمان
Standard deviation	الانحراف المعياري
Standard error	الخطأ المعياري
Statistical hypothesis	الفرض (النظرية) الإحصائية
Statistical inference	الاستدلال الإحصائي
Stratified sampling	معاينة طبقية
Student's distribution	توزيع ستودنت
Subset	مجموعة جزئية
Sum of squares	مجموع المربعات
T	
Test of hypothesis	اختبار الفرضيات
Test statistic	إحصائية الاختبار
Test one sided	اختبار الجهة الواحدة
Test two sided	اختبار من جهتين
U	
Union	اتحاد
Universal set	مجموعة شاملة
Upper limit	الحد الأعلى

Uniform distribution	توزيع متناظر
V	
Variable	متغير
Variance	تباين
Venn -diagram	مخطط فن
Vital statistics	الإحصاء الحيوي
W	
Weighted mean	المتوسط المرجح
Weights	أوزان