المحاضرة النظرية الرابعة في الإحصاء

الارتباط والانحدار

Correlation & Regression

إعداد . د. حيدر الحسن

1/11/2020

مقدمة:

لا شكَّ أنَّ الطبيعة زاخرة بالظواهر الطبيعية، المعروفة وغير المعروفة، المألوفة والغريبة، وعلى الباحث في أي مجال كان قبل كل شيء؛ أن يتعرَّف على العوامل المؤثّرة على الظاهرة المدروسة وما هي طريقة وشكل وقوّة التأثير. هذا ما يُطلق عليه بشكل عام دراسة العلاقة بين المتغيّرات relationship between variables.

إنَّ معرفة نوع وشكل وقوَّة العلاقة وبالتالي التأثير هي أساس النجاح في التعرّف على الظاهرة المدروسة.

ولحسن الحظ أنَّه تمَّ تطوير وسائل إحصائية كثيرة بناءً على أسس رياضية من أجل دراسة العلاقة بين المتغيّرات relationship between variables. **إنَّ من أهم هذه التحاليل هي التالية:**

- تحليل الانحدار Regression Analysis
- تحليل الارتباط Correlation Analysis
- تحليل التباين Variance Analysis

السؤال الآن الذي يطرح نفسه هو أي من التحاليل السابقة يتوجّب استخدامه من أجل دراسة العلاقة بين المتغيرات، أي العوامل المؤثّرة والظاهرة المدروسة؟

لكن قبل الإجابة على السؤال المطروح علينا أن نقدّم بعض الشروحات الضرورية لتوضيح المسألة:

أنواع المتغيّرات variables: يمكن تقسيم المتغيّرات إلى نوعين وذلك حسب القيم التي يأخذها المتغيّر:

- **متغيّر كمّي Qualitive:** هو المتغيّر الذي يأخذ **قيم كمّية** أي قابلة للقياس كمّيا بوحدات القياس مثل وحدات الوزن، الطول، الحجم ...إلى آخره. وكمثال على ذلك المبيد، السماد، العمر، كمّية الغذاء وهكذا...
- **متغيّر نوعي Quantitive:** هو المتغيّر الذي يأخذ **قيم نوعيّة غير قابلة للقياس كمّيا** بوحدات القياس المذكورة آنفا بل يقاس بوحدات القياس النوعي كالدرجات والتدرجات والمستويات وهكذا ... إلى آخره. وكمثال على ذلك اللون ودرجة النجاح ودرجة المقاومة وغيرها ...

مصادر المتغيّرات variables: يمكن تقسيم المتغيّرات إلى نوعين وذلك حسب نوع التحكّم أو المصدر إلى نوعين:

- **متغيّر احتمالي Random variables:** هو المتغيّر الذي **يأخذ قيم عشوائية محتملة لا يمكن التحكّم بها مسبقا** مثل المتغيّرات الجوية كالحرارة والرطوبة وغيرها.
- **متغيّر غير احتمالي: Controlled variable** المتغيّر الذي **يأخذ قيم محدَّدة يمكن التحكّم بها مسبقا** مثل المبيد، السماد، العمر، كمّية الغذاء وهكذا...

1 - تحليل الانحدار Regression Analysis:

يعرّف تحليل الانحدار بأنَّه أحد التحاليل الإحصائية الهامة التي تُستخدم بشكل عام لدراسة العلاقة بين المتغيّرات relationship between variables

وهو يستخدم في الحالات التالية:

- عندما يكون العامل المؤثّر (إي المتغيّر) من نوع كمّي.
- عندما يكون العامل المؤثّر من النوع المتغيّر غير الاحتمالي Controlled variable.
- عندما يكون التأثير من جهة واحدة أي أنَّ العامل المؤثّر (المتغيّر) يؤثّر على الصفة المختارة للظاهرة المدروسة والعكس غير صحيح.
- عندما تكون الغاية هي معرفة النموذج الرياضي للعلاقة ما بين العامل المؤثّر (أي المتغيّر) والصفة المختارة للظاهرة المدروسة.
- عندما تكون الغاية هي التنبؤ الرياضي (أو التوقع الرياضي) لقيم الصفة المختارة للظاهرة المدروسة.

يجدر التنويه إلى أنَّه في تحليل الانحدار يمكن أن نأخذ بعين الاعتبار ما يلي:

- يطلق على العامل المؤثّر (أي المتغيّر) اسم آخر وهو **العامل المستقل (X)** بينما يطلق على الصفة المختارة للظاهرة المدروسة **اسم العامل التابع (Y).**
- يمكن أن يوجد في التجربة الحيوية أكثر من عامل مؤثّر (X1,X2,...Xn) بينما لا يمكن أن يوجد أكثر من صفة مختارة واحدة للظاهرة المدروسة.
- في حال وجود أكثر من عامل مؤثّر في التجربة الحيوية فإنَّه لإجراء تحليل الانحدار لابدَّ من تحقيق الشروط السابقة الذكر على كل عامل سيدخل في تحليل الانحدار ولكن في هذه الحالة سيتم **استخدام تحليل الانحدار المتعدد Multi-Regression Analysis**
- أمَّا إذا كان هناك عامل مؤثّر وحيد في التجربة الحيوية فإنَّه في هذه الحالة سيتمّ استخدام تحليل الانحدار البسيط Linear-Regression Analysis

أولا: تحليل الانحدار البسيط Linear-Regression Analysis:

- إذا كان لدينا في تجربة حيوية أو زراعية عامل مؤثّر واحد (X) وندرس صفة مختارة (Y) فإنَّ معادلة الانحدار البسيط هي من الشكل التالي: Y = a + b*X

حيث أنَّ: Y القيمة النظرية للصفة المختارة للظاهرة المدروسة

a ثابت الانحدار - b عامل الانحدار - X العامل المؤثّر (أي المتغيّر)

أمَّا كيفية حساب كل من a وb فهي على الشكل التالي:

حساب a (ثابت الانحدار): باستخدام العلاقة الرياضية التالية: $a = \bar{y} - b * \bar{x}$

حساب b (عامل الانحدار): يمكن حسابه باستخدام العلاقة الرياضية التالية:

$$b = \sum((x - \bar{x}) * (y - \bar{y})) / \sum(x - \bar{x})^2$$

كما ويمكن حساب b عامل الانحدار باستخدام العلاقة الرياضية التالية:

$$b = (\sum(x * y) - (\sum x * \sum(x / n))) / ((\sum(x^2) - (\sum x)^2) / n$$

مثــال: عند دراسة العلاقة ما بين كميَّة السماد المضاف كغ/هـ وصفة الإنتاج لأحد المحاصيل وكان التحليل الإحصائى للبيانات الإحصائية DATA كما يلى:

SUMMARY OUTPUT				السماد	الإنتاج
إحصائية	مؤشّرا ت			X	Y
Regression Statistics				1	5
Multiple R	0.989094			2	13
R Square	0.978307			3	16
Adjusted R Square	0.973968			4	23
Standard Error	2.164651			5	33
Observations	7			6	38
				7	40
ANOVA	تحليل التباين				
	df	SS	MS	F	Significance F
Regression	1	1056.571	1056.571	225.4878	2.37E-05
Residual	5	23.42857	4.685714		
Total	6	1080			
تحليل الانحدار					
	Coefficients	Standard Error	t Stat	P-value	
Intercept a الثابت	0.57143	1.829464	-0.31235	0.767383	
عامل الانحدار b	6.142857	0.409081	15.01625	2.37E-05	

من الجدول السابق الذي يتضمن نتائج تحليل الانحدار باستخدام أحد الطرق السابقة الذكر نلاحظ أنَّ:

b عامل الانحدار = 6.1428

الثابت a = 0.57143

وبالتالي تكون معادلة الانحدار البسيط هي من الشكل التالي: $y' = a + b * x$

وبعد التعويض في المعادلة المذكورة نجد أنَّ:

$$Y = 6.1428 + 0.5714 * X$$

أمَّا تفسير الأرقام الواردة في معادلة الانحدار المذكورة فهو على الشكل التالي:

- أولاً b عامل الانحدار= 6.1428 بما أنَّ b عامل الانحدار موجب فهذا يعني أنَّ الانحدار ايجابي والعلاقة طردية وبالتالي إذا زادت قيمة العامل X ازدادت قيمة العامل التابع وهو الإنتاج وإذا نقصت قيمة العامل X وهو السماد نقصت قيمة العامل التابع وهو الإنتاج.

- تزداد قيمة العامل التابع وهو الإنتاج بمعدَّل قيمة b عامل الانحدار = 6.1428 + (طن/هـ) عندما تزداد قيمة العامل X وهو السماد وحدة واحدة (1 كغ/هـ).

- الثابت a = 0.57143 يعني أنَّ قيمة العامل التابع وهو الإنتاج تساوي الثابت a = 0.57143 (طن /هـ) عندما تكون قيمة العامل X وهو السماد مساوية للصفر أي عندما لا يوجد أي إضافة للعامل X وهو السماد.

- أمَّا على سبيل المثال إذا كانت قيمة (b) **عامل الانحدار قيمة سلبية** فإننا نقول أنَّ الانحدار **سلبي والعلاقة عكسية**، وبالتالي إذا زادت قيمة العامل المستقل X نقصت قيمة العامل التابع وهو الإنتاج وإذا نقصت قيمة العامل X وهو السماد زادت قيمة العامل التابع وهو الإنتاج.

جودة معادلة الانحدار Regression Model:

- إنَّ معادلة الانحدار Regression Model المستخدمة عادةً هي من الشكل الخطي البسيط، لكن في معظم الحالات فإنَّ هذا الشكل الخطي البسيط قد يكون غير مناسباً لتمثيل البيانات الإحصائية DATA وبالتالي كيف يمكن أن نعرف هل هذا النموذج الرياضي أي معادلة الانحدار البسيط مناسب أم لا ؟.
- للإجابة على السؤال السابق يجب التأكّد من أنَّ الفرق بين البيانات التجريبية والبيانات الإحصائية DATA المحسوبة من معادلة الانحدار **هو أقل ما يمكن** أي بتعبير آخر يجب أن يكون المقدار التالي أقل ما يمكن:

$$\sum (y_i - \hat{y}_i)^2$$

- حيث أنَّ:

y_i القيم الحقيقية أو التجريبية.

$\hat{y}_i$ القيم المحسوبة من معادلة الانحدار البسيط.

- كما أنَّ خطأ التناسب والذي يعطى بالعلاقة التالية:

$$ERROR = \sqrt{\sum (y_I - \hat{y}_I)^2 / (n - (k+1))}$$

- حيث أنَّ: k عدد العوامل المستقلة
  n عدد عناصر العينة
  y_i القيم الحقيقية أو التجريبية
  $\hat{y}_I$ القيم المحسوبة من معادلة الانحدار البسيط.

من الطبيعي أنَّه كلما كان المقدار Error، الذي يرمز له عادة S_d، **صغيراً** كلما كانت جودة معادلة الانحدار **أكبر**.

هذا يعني علينا دائما البحث عن معادلة الانحدار الأنسب كما أنَّه يوجد معايير أخرى لجودة معادلة الانحدار ستذكر لاحقا.

2- الانحدار المتعدّد Multi-Regression Analysis:

- إذا كان لدينا في تجربة حيوية أو زراعية **أكثر من عامل مؤثّر واحد** ($X_1, X_2, \ldots X_n$) وندرس صفة مختارة (Y) فإنَّ معادلة الانحدار المتعدّد هي من الشكل التالي:

$$Y = a + b_1 * X_1 + b_2 * X_2 + \ldots . b_n * X_n$$ •

حيث أنَّ: Y القيمة النظرية للصفة المختارة للظاهرة المدروسة

a ثابت الانحدار

b عامل الانحدار

X_1 العامل المؤثر الأول

X_2 العامل المؤثر الثاني

X_n العامل المؤثر الأخير

وعادةً نستخدم **معادلة الانحدار المتعدّد** Multi-Regression Analysis في حال وجود عاملين مستقلين اثنين وتصبح معادلة الانحدار المتعدد من الشكل التالي:

$$Y = a + b_1 * X_1 + b_2 * X_2$$

- أمَّا بالنسبة لطريقة إيجاد عوامل الانحدار المذكورة في المعادلة أعلاه فإنَّ ذلك يتطلَّب الطريقة الرياضية المعروفة في حل جملة معادلات.

- ولكن نظراً لصعوبة هذه الطريقة الرياضية بالنسبة للطلاب غير المتخصصين بالإضافة لتوفر البرامج الإحصائية المتطورة (برامج الحاسوب Software) فلا يوجد أي مبرر لتعلم هذه الطرق الرياضية في هذا الفصل.

- مثـــال: لدينا البيانات التالية التي تبيّن العلاقة ما بين السماد (X_1) بـ الكغ/هـ والري (X_2) وصفة الإنتاج (Y) طن/هـ من أحد المحاصيل الزراعية حيث أجري **تحليل الانحدار المتعدّد Multi-Regression Analysis كما يلي:**

SUMMARY OUTPUT			X1	X2	Y
إحصائية	مؤشرات		1	2	5
Regression Statistics			2	5	13
Multiple R	0.98962932		3	3	16
R Square	0.97936618		4	8	23
Adjusted R Square	0.96904927		5	6	33
Standard Error	2.36032433		6	5	38
Observations	7		7	9	40
تحليل التباين					
ANOVA					
	df	SS	MS	F	Significance F
Regression	2	1057.715476	528.8577	94.92825	0.000426
Residual	4	22.28452381	5.571131		
Total	6	1080			
الانحدار	تحليل				
العوامل	Coefficients	Standard Error	t Stat	P-value	
Intercept الثابتa	-0.0547619	2.29767487	-0.02383	0.982127	
X1	6.36428571	0.661612967	9.619349	0.000653	
X2	-0.25833333	0.570072486	-0.45316	0.673928	

من الجدول السابق الذي يتضمَّن نتائج تحليل الانحدار **باستخدام أحد برامج جاهزة** نلاحظ أنَّ:

b_1 عامل الانحدار = 6.36428571

b_2 عامل الانحدار =-0.25833333

الثابت a = -0.0547619

وبالتالي تكون معادلة الانحدار المتعدّد هي من الشكل التالي:

$$y' = a + b_1 * x_1 + b_2 * x_2$$

وبعد التعويض في المعادلة المذكورة نجد أنَّ:

$$Y = -0.0547619 + 6.36428571\ X_1 - 0.25833333\ X_2$$

إختبار (ت) T- test

هو أحد الأساليب الاحصائية التي تستخدم للمقارنة بين المجموعات.

في هذا الاختبار تتم المقارنة بين مجموعتين فقط. هناك أسلوب إحصائي آخر يستخدم للمقارنة بين أكثر من مجموعتين، هذا الاسلوب الإحصائي يُطلق عليه (تحليل التباين الأُحادي One Way ANOVA).

- **من الحالات التي يُستخدم فيها اختبار (ت):**

1. المقارنة بين مجموعتين مستقلتين ومن الأمثلة عليها المقارنة بين الذكور والإناث، المقارنة بين طلاب الكليات العلمية وطلاب الكليات الإنسانية، المقارنة بين طلاب السنة الأولى والسنة الثانية، المقارنة بين دخل أعضاء الهيئة التدريسية والإداريين في الجامعة.

تذكّر أن الأفراد في المجموعتين مختلفتين أي بمعنى أنهم ليسوا نفس الأفراد.

2- المقارنة بين المجموعات المرتبطة ومن الأمثلة عليها المقارنة بين وزن الأطفال عند عمر شهر وعند عمر ثلاثة شهور، المقارنة بين معدَّل الطلاب في الثانوية العامة في الفصل الأول والفصل الثاني، المقارنة بين مسافة الوثب الطويل من الثبات قبل وبعد إخضاع المفحوصين لبرنامج تدريبي مُقترح لتنمية القوة العضلية للطرف السفلي، المقارنة بين علامات الطلاب قبل وبعد أخذ مادة الإحصاء في كلية الزراعة، المقارنة بين نمو محصول البطاطا قبل التسميد وبعده. **تذكَّر أن الأفراد في المجموعتين هم نفس الأفراد.**

- **اختبار (ت) مبني على أربعة افتراضات:**

1. أن يكون مجتمع الدراسة (المجتمع الذي سُحبت منه عينات المجتمع) ذو توزيع طبيعي.
2. يجب أن تكون العينات مسحوبة بشكل عشوائي من مجتمع الدراسة، إذا لم تكن العينات مسحوبة بشكل عشوائي من مجتمع الدراسة **لا نستطيع تعميم النتائج التي حصلنا عليها من العينة على مجتمع الدراسة.**
3. عند المقارنة بين مجموعتين يجب أن يكون **تشتّت المجموعتين متقارب**، بمعنى آخر أن يكون تباين المجموعتين متقارب.
4. يجب أن تكون البيانات نسبية.

α

- تحسب قيمة t من خلال القانون التالي:

$$t = \frac{\overline{X} - \mu}{S_{\overline{x}}}$$

- حيث: $\overline{X}$= هي متوسط العينة

μ = متوسط المجتمع

$S_{\overline{x}}$ = الخطأ القياسي للعينة

وبناء على خطوات اختبار المعنوية بـ t فخطوات هذا الاختبار كالآتي:

1-أ: النظرية الفرضية (HT) تفترض أنَّ هذه العينة تنتمي إلى المجتمع ذي المتوسط u أي تكون النظرية الفرضية: متوسط المجتمع المعلوم = μ : HT

1- ب: النظرية البديلة :(HA) العينة من مجتمع مختلف معنوياً عن المجتمع المذكور وبالتالي فتكون النظرية البديلة: متوسط المجتمع المعلوم ≠ μ : HA

2- تحديد مستوى المعنوية المستعمل

3- إيجاد قيمة t الجدولية عند درجات الحرية df حيث (df = n - 1) ومستوى المعنوية

4- حساب t وفق القانون السابق.

$t\,(\alpha_{\infty}, df\,(n-1)$ =

5- المقارنة بين قيمة t المحسوبة وقيمة t الجدولية

6- القرار الإحصائي:

- إذا كانت قيمة t المحسوبة واقعة داخل منطقة قبول النظرية الفرضية تقبل النظرية الفرضية وترفض النظرية البديلة

- بينما اذا كانت t المحسوبة واقعة خارج منطقة قبول النظرية الفرضية أي داخل منطقة رفض النظرية الفرضية ترفض النظرية الفرضية وتقبل النظرية البديلة وذلك عند مستوى المعنوية المستعمل في الاختبار.

7- القرار التطبيقي: حيث يطبق القرار الإحصائي على السؤال المطروح في الاختبار.

مثال: تنتج إحدى شركات الزيوت الغذائية زيت من زيوت دوار الشمس (زهرة الشمس) وتقول في بياناتها أنَّ متوسط تركيز الأحماض الدهنية المشبعة بهذا الزيت = 14غ/100مل، تمَّ تحليل عدة عينات سحبت بطريقة عشوائية من هذا الزيت وحصلنا على النتائج التالية لتركيز الأحماض الدهنية المشبعة في الزيت (غ/100مل)

19 – 18 – 17 – 18 – 22 – 16 – 18 – 19 – 16 – 21 – 22 -20

- هل هذا الزيت ينطبق عليه ما كتب على عبواته؟ وهل الشركة صادقة في ادعائها بأنَّ محتوى الزيت من الأحماض الدهنية المشبعة = 14غ/100مل وذلك باحتمال 0.99
- الحل:

النظرية الفرضية $HT : \mu = 14$

النظرية البديلة $HA : \mu \neq 14$

قيمة t الجدولية عند درجة حرية (11) و = 0.01

$t(0.01 , 11) = 3.11$

$$t = \frac{\bar{X} - \mu}{S_{\bar{x}}}$$

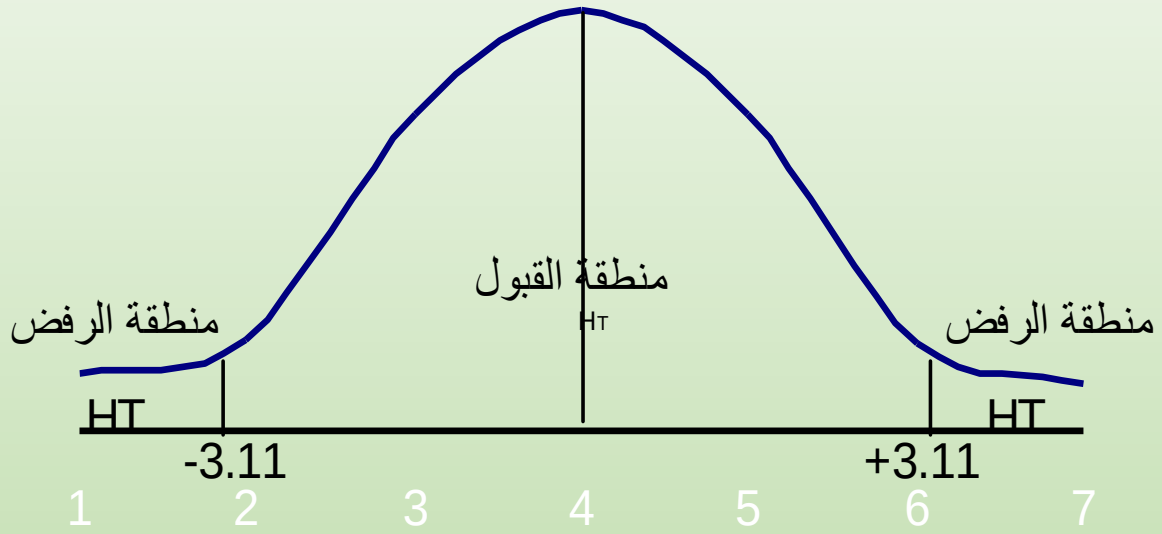
$$\bar{X} = \frac{226}{12} = 18.83$$

$$S^2 = 4.36$$

$$S_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{4.36}{12}} = \sqrt{0.363} = 0.602$$

$$t = \frac{18.83 - 14}{0.602} = 8.02$$

المقارنة : t المحسوبة تقع في منطقة رفض النظرية الفرضية



القرار الإحصائي: ترفض النظرية الفرضية وتقبل النظرية البديلة باحتمال 0.99.

القرار التطبيقي: الزيت الذي سحبت منه العينة لا ينتمي إلى الزيت الذي متوسط محتواه من الأحماض الدهنية المشبعة 14غ/100مل، أي أنَّ الشركة غير صادقة في ادعائها.

- **كيفية حساب درجات الحرية:**
- في حالة المجموعات المستقلة: درجات الحرية = ($ن_1$-1) + ($ن_2$-1)
- أمَّا في حالة المجموعات المرتبطة فإنَّ درجات الحرية = ن - 1
- فإذا كانت قيمة (ت) المحسوبة أكبر من قيمة (ت) الجدولية عند درجات الحرية المناسبة، نستنتج وبمستوى ثقة معين (0.10، 0.05، 0.01) أنَّه توجد فروق ذات دلالة إحصائية بين المجموعتين (في حالة المجموعات المستقلة) أو بين القياسين (قبل وبعد) في حالة المجموعات المرتبطة، ولصالح المجموعة أو القياس الافضل.

- **على سبيل المثال:**

أذا أردنا المقارنة بين أوزان 10 طلاب و10 طالبات من كلية الزراعة.

فإذا كانت قيمة (ت) المحسوبة تساوي (5) وكانت قيمة (ت) الجدولية عند درجات الحرية المناسبة (18) وعند مستوى ثقة 0.01 تساوي (2.878) فإننا نستنتج أنه توجد فروق ذات دلالة إحصائية بين أوزان الطلاب (80كغم) وأوزان الطالبات (70 كغم) ولصالح الطلاب (لأن متوسط الطلاب أعلى من متوسط الطالبات).

وبناءً عليه لدينا ثقة أكبر من 99% أنَّ الفروق بين الجنسين هي فروق حقيقية ولا تُعزى للصدفة (هرمون التستستيرون لدى الرجال يُساعد على بناء كتلة عضلية أكبر لدى الرجال مقارنة مع السيدات) وإن ثقتنا بأنَّ هذه الفروق تُعزى للصدفة أقل من 1%.

3- الانحدار اللاخطي Nonlinear-Regression Analysis:

- لدى دراسة علاقة ما لا بدَّ أن نستقرئ أولاً فيما إذا كان يوجد علاقة أصلا بين المتغيّرات (X و Y) وبمعنى آخر هل تقدّم البيانات الإحصائية DATA المتوفرة دليلا كافيا على وجود علاقة خطية بين المتغيّرات.
- إنَّ القول بأنَّ العامل X والعامل Y لا يرتبطان ببعضهما خطياً يكافئ القول بأنَّ b عامل الانحدار الخطي يساوي الصفر والعكس صحيح
- أي إذا أثبتنا أنَّ b عامل الانحدار الخطي يساوي الصفر أو لا يتميز عن الصفر بشكل معنوي فهذا يعني أن العلاقة بين المتغيّرات ليست خطية أو العامل X والعامل Y لا يرتبطان ببعضهما خطياً.
- إذن إذا كنا نرغب في دراسة العلاقة بين المتغيّرات لمعرفة هل العلاقة خطيَّة أم لا علينا أن نختبر الفرضية الابتدائية (H0: b=0) مقابل الفرضية البديلة (H1: b ≠ 0) .
- من أجل هذا الغرض فإنَّ الاختبار المناسب **هو اختبار ت T-TEST** عند مستوى المعنوية 5% ودرجة الحرية DF=N-2.

- مثـــال: لدى دراسة العلاقة بين المتغيّرين السماد X (كغ/هـ) والإنتاج Y(طن/هـ) حصلنا على البيانات الإحصائية DATAالواردة فى الجدول التالى بالإضافة إلى تحليل الانحدار ومن ثم تطبيق اختبار ت T-TEST :

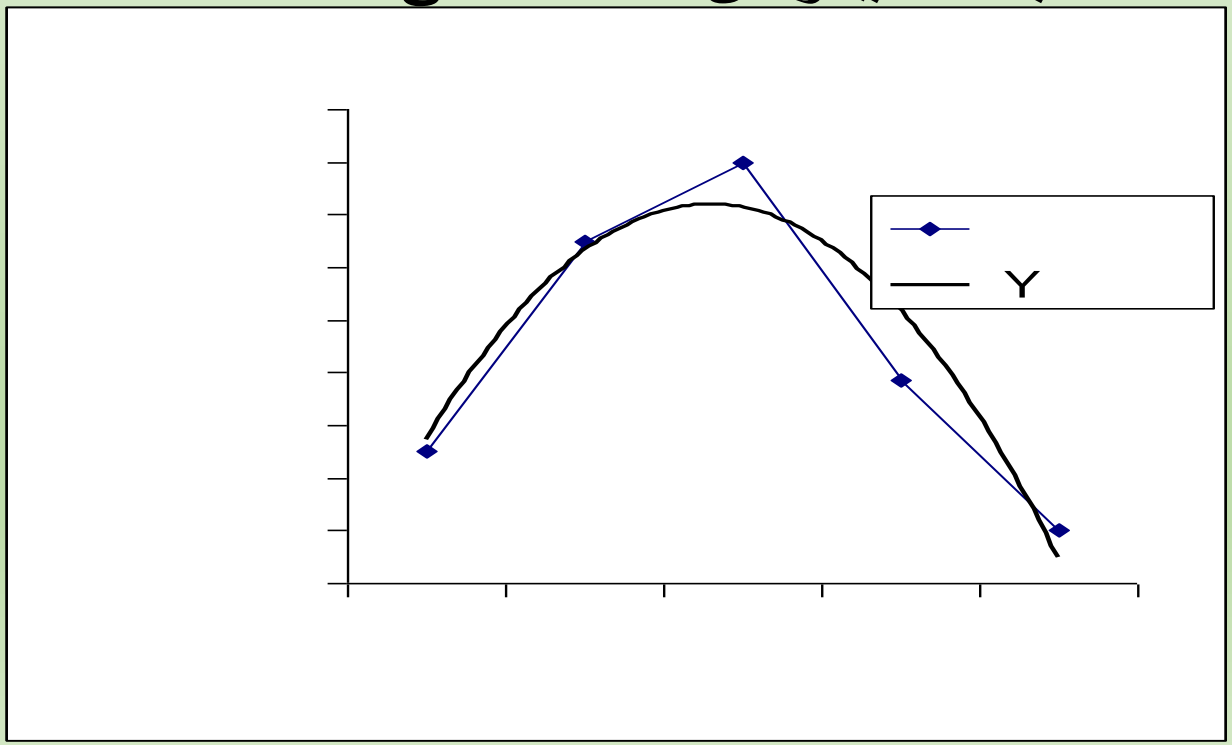
SUMMARY OUTPUT		السماد	الإنتاج	
الانحدار	تحليل	X	Y	
Regression Statistics		10	5	
Multiple R	0.387298335	20	13	
R Square	0.15	30	16	
Adjusted R Square	-0.133333333	40	4	
Standard Error	6.519202405	50	2	
Observations	5			
ANOVA				
	df	SS	MS	F
Regression	1	22.5	22.5	0.529412
Residual	3	127.5	42.5	
Total	4	150		
	Coefficients	Standard Error	t Stat	P-value
Intercept	a = 12.5	6.837397166	1.828181	0.164969
X	b = -0.15	0.206155281	-0.72761	0.519498

من الجدول السابق أعلاه نلاحظ أنّ: $t\ Stat = -0.727$

$P\text{-}value = 0.519$

نستنتج من قيم اختبار ت T-TEST المطبقة على عامل الانحدار b لمعرفة هل العلاقة خطية أم لا أي اختبار مقابل الفرضية البديلة (H1: b ≠ 0) نجد ان الفرضية الابتدائية (H0: b=0) مرفوضة وبالتالي نقبل الفرضية البديلة (H1: b ≠ 0)

هذا يؤكد **أن عامل الانحدار الخطي غير معنوي** ولا يتميز عن الصفر بشكل معنوي، هذا يقود للقول أنّه توجد علاقة لا خطية كما يبدو من الشكل التالي:



نلاحظ من الشكل الأخير أنَّه لا يمكن تمثيل العلاقة السابقة بخط مستقيم. إذن لدى دراسة العلاقة بين المتغيّرين السماد X(كغ/هـ) والإنتاج Y (طن/هـ) وحصلنا على البيانات الإحصائية DATA الواردة في الجدول السابق فإنَّ معادلة الانحدار اللاخطي التي توافق الشكل البياني السابق هي من الشكل التالي :

$$Y = a + b_1*X+b_2*X^2$$

حيث أنَّ : Y القيمة النظرية للصفة المختارة للظاهرة المدروسة

a ثابت الانحدار

b_1 عامل الانحدار الخطي

b_2 عامل الانحدار التربيعي

وعادةً نستخدم معادلة الانحدار اللاخطي في حال **وجود عاملين مستقلين اثنين** وتصبح معادلة الانحدار اللاخطي من الشكل التالي:

$$Y = a + b_1*X_1+b_2*X_2 +b_3*X^2+b_4*X^2+b_5*X1*X2$$

- أمَّا بالنسبة لطريقة إيجاد عوامل الانحدار المذكورة في المعادلة أعلاه فإنَّ ذلك يتطلب الطريقة الرياضية المعروفة في حل جملة معادلات،
- ولكن نظرا لصعوبة هذه الطريقة الرياضية بالنسبة للطلاب غير المتخصصين بالإضافة لتوفر البرامج الإحصائية المتطورة (برامج الحاسوب Software) فلا يوجد أي مبرر لتعلم هذه الطرق الرياضية في هذا الفصل.
- وهنا تبرز مسألة هامة وهي اختيار معادلة الانحدار اللاخطي المناسبة من بين مجموعة النماذج السابقة الذكر وهذه قضية تشكل تحدٍ كبير خصوصا بالنسبة لغير المتخصصين وهو ما يطلق عليه PEST FIT, لذا فإننا سنقدم وسيلة مساعدة سهلة لاختيار النموذج المناسب وهي التالية :
- لدينا مقياسين أو مؤشّرين إحصائيين اثنين هما:
- عامل التحديد R^2
- الخطأ القياسي للانحدار Standard Error

ويمكن للباحث أن يعتمد على كلا المؤشرين الإحصائيين المذكورين لاختيار النموذج المناسب، إذا أمكن ذلك، ولكن يمكن الاعتماد على أحدهما إذا كان الآخر متوفر أو غير كاف أيضا.

فبالنسبة إلى المؤشّر الأول **عامل التحديد** R^2 فهو عبارة عن النسبة المئوية للعامل المؤثر الذي أمكن كشفه من خلال النموذج المقترح ويمكن القول أن قيمته تتراوح بين 0% -100% وكلما كانت أكبر كلما كان النموذج المقترح أكثر تناسبا مع البيانات الإحصائية DATA .

أمَّا بالنسبة إلى المؤشّر الثاني **الخطأ القياسي للانحدار** فيمكن القول أنّ قيمته ليس لها مجالا محدودا **وكلما كانت أصغر كلما كان النموذج المقترح أكثر تناسبا مع البيانات الإحصائية DATA .**

لذلك عند المقارنة بين نموذجين أو أكثر يمكن استخدام كلا المؤشرين الإحصائيين أو أحدهما كأساس لاختيار النموذج المناسب أو الأكثر تناسبا.

- **اختبار معنوية الانحدار:** إنَّ المقصود من اختبار معنوية الانحدار هو التأكّد إحصائياً من أنَّ

$$b_i \neq 0$$

- ولتحقيق ذلك فانه يتوجب استخدام اختبار ت (t-test) وهو أحد اختبارات المعنوية المعروفة وذلك حسب العلاقة التالية:

$$t = b / s_b$$

- حيث أن: b عامل الانحدار
- s_b الخطأ القياسي للانحدار

إلى اللقاء في المحاضرة القادمة