المحاضرة الثالثة/السنة الثالثة

تطبيق نظرية الاحتمالات والإحصاء الرياضي في الهيدرولوجيا

د.م ميس محمد عليان

علم الإحصاء الرياضي

- يتضمن **علم الإحصاء** تجميع معلومات عن ظاهرة ما ، ثم معالجة هذه المعلومات (القياسات) وجدولتها. فإذاً هناك نوعين من الإحصاء:
  1. الإحصاء الوصفي: يعنى بتجميع البيانات وتنظيمها عن طريق الجداول أو الرسوم البيانية وغيرها.
  2. الإحصاء الاستنتاجي: جزء من الإحصاء يعنى بتحليل البيانات للتوصل إلى التنبؤ أو الاستقراء واتخاذ القرارات.
- أما **الاحتمال** يبحث في إمكانية حدوث ظاهرة أو حادثة ما من عدمها، وذلك اعتماداً على المعطيات الإحصائية التي تم تجميعها ومعالجتها سابقاً.
- تدعى المجموعة الكاملة من العناصر المدروسة والتي تتميز بالانتظام والتجانس في خواصها متحولات أو متغيرات فإذا كان التغير في قيم الظواهر غير خاضع لقانون رياضي معروف دعي المتغير **متغيرا عشوائياً** .
- إذا تم إجراء قياسات عن ظاهرة ما دون التأثير فيها دعيت مجموعة القياسات هذه بالسلسلة العشوائية وسنسميها في الهيدرولوجيا **السلسلة الهيدرولوجية** , مثال ذلك : قياس التصاريف فهو طبيعي حيث تعد هذه القياسات متغيرة عشوائية غير متحيزة .أما إذا تم بناء سد على هذا النهر فإن تصاريف النهر بعد إنشاء السد يمكن التأثير فيها وبذلك تكون متحيزة .
- حيث تعد الظواهر الهيدرولوجية **متحولات عشوائية مستمرة** (تتغير قيمها بشكل مستمر كتغيرات درجات الحرارة مع الزمن)، أو **متحولات عشوائية غير مستمرة** (أي يتغير ضمن مجال محدد مثل عدد أيام الهطول المطري في الشهر أو السنة مثلاً) مع الزمن.
- تأتي أهمية تطبيق الإحصاء والاحتمال في الهيدرولوجيا هو معرفة سلوك ظاهرة هيدرولوجية في المستقبل اعتماداً على نظرية الاحتمالات وذلك بعد تحليل معلومات سابقة لهذه الظاهرة تم تجميعها باستخدام علم الإحصاء. تبدو الحاجة إلى ذلك واضحة عندما يكون المطلوب معرفة قيمة الهطول أو التصريف الأعظمي الذي يحدث مرة كل 100 سنة وعندما تتوافر قياسات لفترة محدودة أقل من مئة سنة، مثلاً ثلاثين سنة .
- سنسمي مجموع القياسات بحجم السلسة n .

التحليل الأولي للقياسات الهيدرولوجية

1- الثوابت الإحصائية الأولية:

- **المتوسط الحسابي X:**

يمثل المتوسط الحسابي القيمة المتوسطة للظاهرة المقيسة خلال فترة القياس ويحسب من العلاقة:

$$X = \frac{\sum_{i=1}^{n} xi}{n}$$

- **الانحراف المعياري σ:**

يمثل مدى تبعثر عناصر السلسلة عن متوسطها الحسابي والذي يحسب من العلاقة:

$$\sigma = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(xi - X)^2}$$

حيث: n : عدد عناصر السلسلة

σ : الانحراف المعياري

- **خواص الانحراف المعياري:**

1 - إن واحدة الانحراف المعياري هي وحدة عناصر السلسلة نفسها .
2 - إذا أضفنا إلى جميع عناصر السلسلة عدداً أو طرحنا منها عدداً فإن قيمة σ لا تتغير .
3 - إذا ضربنا جميع عناصر السلسلة بعدد أو قسمناها على عدد فإن قيمة σ تضرب بالعدد نفسه أوتقسم عليه.

- **معامل التغيير Cv:**

من أجل إجراء مقارنة بين سلسلتين نقوم بحساب معامل التغيير حيث نقوم بقسمة الانحراف المعياري على المتوسط الحسابي وبذلك نتمكن من المقارنة بين عوامل لا بعدية، أي:

$$C_V = \frac{\sigma}{X}$$

$X \neq 0$

- **خواص معامل التغير :**

1. لا تتغير قيمة معامل التغيير عندما نضرب جميع عناصر السلسلة بعدد ثابت أو نقسمها على عدد ثابت .
2 - يتناقص معامل التغيير إذا أضفنا إلى جميع عناصر السلسلة عدداً ثابتاً والعكس صحيح بالنسبة للطرح .
3 - معامل التغيير لا واحدة له .

- **معامل عدم التناظرCs :**

- يعبر معامل عدم التناظر عن مدى تناظر الانحرافات الموجبة والسالبة لعناصر السلسلة بالنسبة لمتوسطها الحسابي، حيث تتصف السلاسل الهيدرولوجية غالباً بأنها لا تتوزع بشكل متناظر حول المتوسط .

- قد يكون لسلسلتين قيمة المتوسط الحسابي نفسها وقيمة معامل التغيير نفسها ومع ذلك تكونان مختلفتين ويظهر الاختلاف من خلال حساب عدم التناظر:

$$C_s = \frac{\sum_1^n (Ki - 1)^3}{(n-1) * {C_V}^3} \qquad K_i = \frac{xi}{X}$$

طرائق عرض البيانات:

1. الجداول:
2. طريقة المستطيلات والأعمدة
3. طريقة الخط المنكسر:
4. طريقة الخط المنحني:
5. طريقة التوزيع التكراري:

تعد طريقة التوزيعات التكرارية أهم الطرق في تنظيم المعطيات والبيانات الكثيرة ، فعند توافر عدد كبير من البيانات لا تستطيع المقارنة بين مفردات البيانات عند استخدام إحدى طرق العرض السابقة ويصعب عليك فهمها لذلك لابد من تلخيصها وعرضها بطريقة مبسطة تسهل عليك فهمها من هذه الطرق التوزيعات التكرارية.

حيث تعتمد طريقة إنشاء التوزيع التكراري على تقسيم مدى البيانات إلى فئات وحصر عدد البيانات ضمن كل فئة.

طريقة إنشاء التوزيع التكراي:

مثال 1:

بيانات قيم الهطل المطري الشهري في إحدى محطات الرصد الجوي

44	48	32	27	48	25
24	42	29	33	46	23
25	46	23	23	49	27
25	37	48	41	36	28
36	39	47	43	41	31
43	41	33	48	43	35
48	36	36	46	38	28
47	28	22	34	32	27
	33	39	24	24	41
43	23	44	46	45	44

- بفرض عدد الفئات k=6
- تحديد المدى = القيمة العظمى –القيمة الصغرى M=49-22=27
- نحسب طول الفئة L=$\frac{M}{K}$=$\frac{27}{6}$ =4.5 (البيانات معطاة إلى أقرب رقم صحيح (لايوجد كسور) نقرب طول الفئة إلى الأعلى لأقرب عدد صحيح فيصبح 5 .
- نرتب البيانات ترتيبا تصاعديا أو تنازلياً (ليكن في مثالنا تنازلياً)
- نعين الحد الأعلى للفئة الأولى (أكبر قيمة 49) الحد الأعلى الفعلي للفئة الأولى 49+0.5=49.5:
- الحد الأدنى الفعلي للفئة الأولى:49.5-5=44.5 الحد الأدنى للفئة الأولى :44.5+0.5=45
- هكذا بالنسبة لباقي الفئات بطرح طول الفئة من كل حد ثم نحدد الحدود الدنيا والعليا الفعلية بالطريقة نفسها

تكرار التجاوز

حدود الفئة	الحدود الفعلية للفئات	مركز الفئة	التكرار	التكرار النسبي	التكرار النسبي المئوي	التكرار المتجمع
49-45	49.5-44.5	47	13	0.22	22	13
44-40	44.5-39.5	42	12	0.2	20	25
39-35	39.5-34.5	37	10	0.17	17	35
34-30	34.5-29.5	32	7	0.12	12	42
29-25	29.5-24.5	27	10	0.17	17	52
24-20	24.5-19.5	22	8	0.13	13	60
المجموع			60	1	100	
ملاحظة	مجموع التكرارت يجب أن يساوي عدد البيانات					

مثال 2: نفرض أن لدينا مجموعة من الأعداد تمثل علامات 50 طالباً في أحد المقررات وهي مرتبة بشكل تصاعدي كما في الجدول الآتي:

العلامة	م	العلامة	م	العلامة	م	العلامة	م
60	40	43	27	30	14	0	1
63	41	44	28	31	15	11	2
64	42	46	29	33	16	12	3
6768	43	48	30	34	17	15	4
72	44	51	31	35	18	17	5
75	45	51	32	35	19	18	6
77	46	53	33	37	20	21	7
80	47	54	34	38	21	22	8
82	48	55	35	40	22	22	9
98	49	55	36	41	23	25	10
	50	56	37	42	24	26	11
		57	38	42	25	28	12
		59	39	43	26	29	13

وتتراوح العلامات بين الصفرو 98، فإننا نقسم هذه العلامات إلى فئات عددها بين 5 و 15 (أويساوي الجذر التربيعي لعدد البيانات) هذا وتوجد عدة وجهات نظر من أجل التقسيم إلى مجالات بعضها ما يطلب عدداً أصغرياً من المجالات K، ينصح ألا تتجاوز قيمة عرض المجال عن(0.6σ)

وسوف تعتبر هذه الفئات في هذه المسألة تساوي عشر فئات:

- تحديد المدى = القيمة العظمى –القيمة الصغرى M=98-0=98
- نحسب طول الفئة L=$\frac{M}{K}$=10
- نعين الحد الأدنى للفئة الأولى (أصغر قيمة في البيانات=صفر)
- الحد الأدنى الفعلي للفئة الأولى :0.5-=0.5- 0
- الحد الأعلى الفعلي للفئة الأولى:9.5= 10+0.5-
- الحد الأعلى للفئة الأولى :9=0.5- 9.5

التكرار النسبي المئوي	التكرار النسبي	عدد الطلاب في كل فئة (التكرار)	مراكز الفئات	الحدود الفعلية للفئات	الحدود النظرية للفئات
2	1/50	1	4.5	0-9.5	0-9
10	5/50	5	14.5	9.5-19.5	10-19
14	7/50	7	24.5	19.5-29.5	20-29
16	8/50	8	34.5	29.5-39.5	30-39
18	9/50	9	44.5	39.5-49.5	40-49
18	9/50	9	54.5	49.5-59.5	50-59
10	5/50	5	64.5	59.5-69.5	60-69
6	3/50	3	74.5	69.5-79.5	70-79
4	2/50	2	84.5	79.5-89.5	80-89
2	1/50	1	94.5	89.5-99.5	90-99
100		50			

قد يكون من المفيد معرفة الطلاب الذين علاماتهم أقل أو تساوي علامة معينة لذا نشكل الجدول الآتي (نرتب ترتيب تصاعدي) :

العلامة	0	9.5	19.5	29.5	39.5	49.5	59.5	69.5	79.5	89.5	99.5
التكرار	0	1	5	7	8	9	9	5	3	2	1
عدد الطلاب الذين علاماتهم أقل أو تساوي (تكرارعدم التجاوز) (التكرار المتجمع)	0	1	6	13	21	30	39	44	47	49	50
النسبة المئوية	0	2	12	26	42	60	78	88	94	98	100

وإذا أردنا معرفة الطلاب الذين علاماتهم أكثر من علامة معينة لذا نشكل الجدول الآتي (نرتب ترتيب تنازلي) :

العلامة	99.5	89.5	79.5	69.5	59.5	49.5	39.5	29.5	19.5	9.5	0
التكرار	0	1	2	3	5	9	9	8	7	5	1
عدد الطلاب الذين علاماتهم أكثر أو تساوي (تكرار التجاوز) (التكرار المتجمع)	0	1	3	6	11	20	29	37	44	49	50
النسبة المئوية	0	2	6	12	22	40	58	74	88	98	100

التكرار المطلق:

يتردد في الاقتصاد المائي مفهوم التكرار ، حيث يعرف التكرار المطلق (Xi) h بأنه عدد المراقبات والتي من أجلها Xi<X<Xi +ΔXi حيث ΔXi: هو عرض المجال. يمكن الحصول عملياً على التكرار المطلق بتقسيم كامل مجال تحول القيم إلى عدد من المجالات K عرض كل منها ΔXi ومن ثم نقوم بعد القيم Xi ضمن كل مجال .

التكرار النسبي:

إذا قمنا بتقسيم قيم التكرار المطلق على عدد عناصر السلسلة نحصل على التكرار النسبي، ويعطى بالعلاقة التالية:

$$N_{(X_i)} = \frac{h_{(X_i)}}{n} \quad \%$$

المدرج التكراري:

عندما نوقع قيم التكرار التي نحصل عليها على المحور الشاقولي (محور التراتيب) مع الحدود الفعلية للفئات على المحور الأفقي (محور الفواصل) وذلك بشكل مستطيلات قاعدتها هي الحدود الفعلية وارتفاعها هو التكرار المقابل لكل فئة فنحصل على المدرج التكراري ويسمى توزيع التكرار.

المضلع التكراري:

ولرسم المضلع التكراري: نقوم بوصل منتصفات الأضلاع العلوية للمدرج التكراري (أو بين مراكز الفئات)، ولكي نغلق المضلع نمدده إلى منتصف فئتين وهميتين تكرار كل منهما يساوي الصفر على يمين الفئات ويسارها.

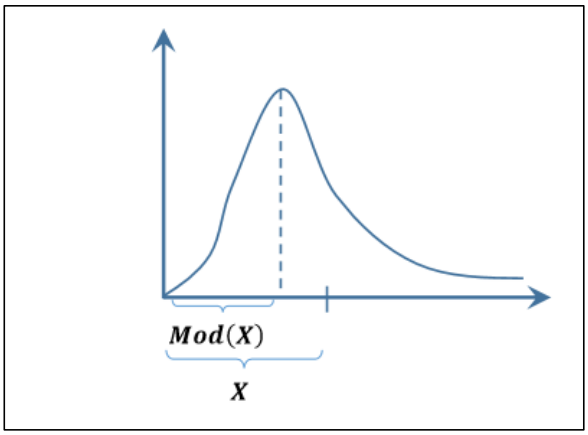
إذا كان عرض المجال صغيرا نحصل على منحني مستمر يسمى منحني التكرار وذلك عندما يكون عدد العناصر كبير جداً.

منحني كثافة الاحتمال (منحني التكرار النسبي):

إذا أوقعنا قيم التكرار النسبي مع الحدود الفعلية للفئات نحصل على منحني كثافة الاحتمال .

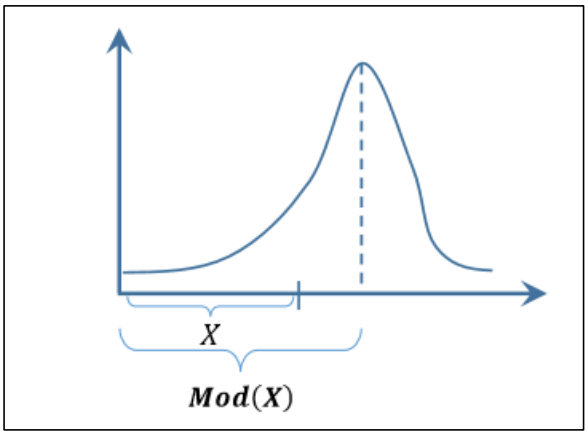
استناد ا إلى منحني التكرار أو التكرار النسبي يمكن الحصول على قيم هامة مثل معيار التوزيع($M0$أو $Mod(X)$ وهي القيمة الأكثر تكرارا .

✓ كما يمكن أيضاً تعيين انحراف توزيع التكرار أو منحني التكرار النسبي حيث يمكن الحديث عن سلسلة ذات انحراف يساري عندما :



$$Mod_{(X)} < \underset{\text{متوسط حسابي}}{X}$$

أو سلسلة ذات انحراف يميني عندما:



$$Mod_{(X)} > X$$

-غالباً ما نصادف في الهيدرولوجيا سلاسل ذات انحراف يساري، تحسب قيمة الانحراف S من العلاقة التالية :

$$S = \frac{X - Mod(x)}{\sigma}$$

كما يمكن استناداً إلى مخطط التكرار تعريف وسطي التوزيع (Me) أو(Median) أي هي قيمة (X) التي تقسم مخطط التكرار إلى قسمين متساويين .

المنحني التكراري التجميعي (منحني الاستمرار):

نحصل عليه بأن نرسم محورين متعامدين ونمثل الحدود العليا للفئات على المحور الأفقي والتكرار المتجمع على المحور الشاقولي

- نشير هنا إلى أنه عندما نقوم بتجميع قيم التكرار بدءا من القيم الدنيا نحصل على قيم التكرار لعدم التجاوز (يسمى المنحني منحني الاستمرار لعدم التجاوز) ، وعندما نبدأ بتجميع قيم التكرار بدءاً من القيم العظمى نحصل على قيم التكرار للتجاوز (يسمى المنحني منحني الاستمرار للتجاوز).
- عندما نقسم قيم منحني الاستمرار للتجاوز أو لعدم التجاوز على عدد عناصر السلسلة n نحصل على ما يسمى تابع توزيع الاحتمالات للتجاوز أو لعدم التجاوز .
- لقيمة ما Xi يكون مجموع التكرار لعدم التجاوز ومجموع التكرار للتجاوز يساوي مجموع عناصر السلسلة أي:

$$D_{Xi} + \overline{D_{xi}} = n$$

- D_{Xi} : مجموع التكرار لعدم التجاوز.

- $\overline{D_{xi}}$: مجموع التكرار للتجاوز.

التحليل الاحتمالي للمعطيات الهيدرولوجية:

عند إجراء التحليل الاحتمالي للمعطيات الهيدرولوجية نطبق نظرية الاحتمالات على قيم القياسات لذلك لابد من إيضاح المفاهيم التالية:

1- الاحتمال:

يعرف الاحتمال بشكل بدائي كما يلي:

$$\boldsymbol{p} = \frac{\text{عدد االحالات االمناسبة}}{\text{عدد الحالات الكلية}}$$

وبتطبيق ذلك على حجر الزهر نجد مثلاً أن احتمال ظهور العدد (4) هو:

$$p(x = 4) = \frac{1}{6}$$

أما إذا طرحنا سؤالاً آخر : ما هو احتمال ظهور عدد أقل أو يساوي (4):

$$P_u(x \leq 4) = \frac{4}{6}$$

نسمي هنا P_u احتمال عدم التجاوز وهو يعطي احتمال الوصول إلى قيمة محددة ل xi أو أقل منها.

- احتمال ظهور قيمة أكبر من (4) فهو:

$$\mathrm{P_{\overline{u}}(x > 4)} = \frac{2}{6}$$

نسمي هنا $P_{\overline{u}}$احتمال التجاوز

- حديثاً تم تعريف الاحتمال بالشكل الآتي:

$$p(xi) = \lim_{n\to\infty}(\frac{h(xi)}{n})$$

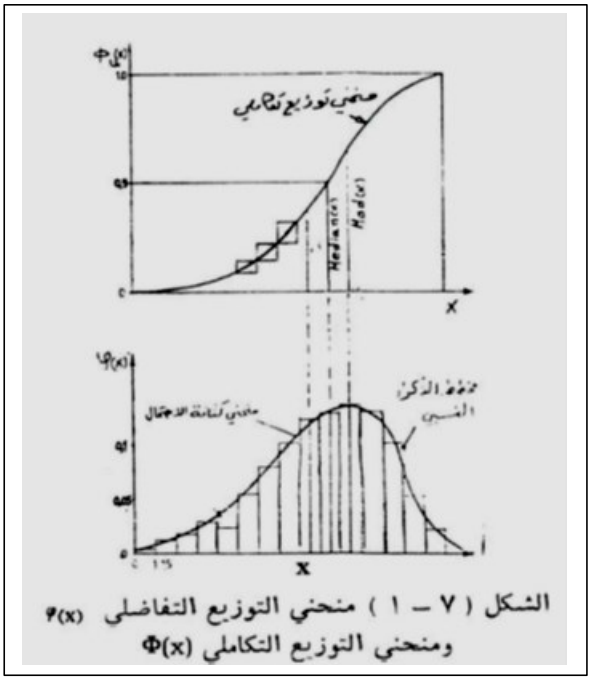
أي P(xi) هي نسبة التكرار المطلق إلى المجموع الكلي لعدد العناصر n

- عندما يكون عدد عناصر السلسلة n لانهائي يتحول مخطط التكرار النسبي إلى منحني كثافة الاحتمال أو مايسمى منحني توزيع الاحتمال التفاضلي الذي يكتب بالشكل التالي:

$$\varphi_{(\mathrm{xi})} = \frac{\mathrm{h(xi)}}{\mathrm{n}}$$

ومنه نجد:

$$\int_{-\infty}^{+\infty} \varphi_{(x)} * dx = \int_{-\infty}^{+\infty} \frac{h_{(xi)}}{n} dx * = \frac{n}{n} = 1$$


الشكل (٧ ــ ١) منحني التوزيع التفاضلي φ(x)
ومنحني التوزيع التكاملي Φ(x)

- الشكل التالي يبين منحني التوزيع التفاضلي $\varphi_{(x)}$
وكذلك منحني التوزيع التكاملي.
- بمكاملة تابع الكثافة الاحتمالي $\varphi_{(x)}$ نحصل على
تابع توزيع الاحتمال أو منحني توزيع الاحتمال التكاملي:

$$\emptyset_{(x)} = \int_{-\infty}^{xi} \varphi_{(x)} * dx = \int_{-\infty}^{+\infty} \frac{h_{(xi)}}{n} dx *= \frac{n}{n} = 1$$

كما نلاحظ أن

$$\emptyset_{(x_n)} = 1$$

بالنظر لأن حدود التكامل تبدأ من القيم الدنيا حتى القيمة التي نحسب لها الاحتمال نكون هنا قد حسبنا احتمال عدم التجاوز والذي يؤمز له ب:

$$\emptyset(x \le xi) = \emptyset_{u(xi)} = \int_{-\infty}^{xi} \varphi_{(x)} * dx$$

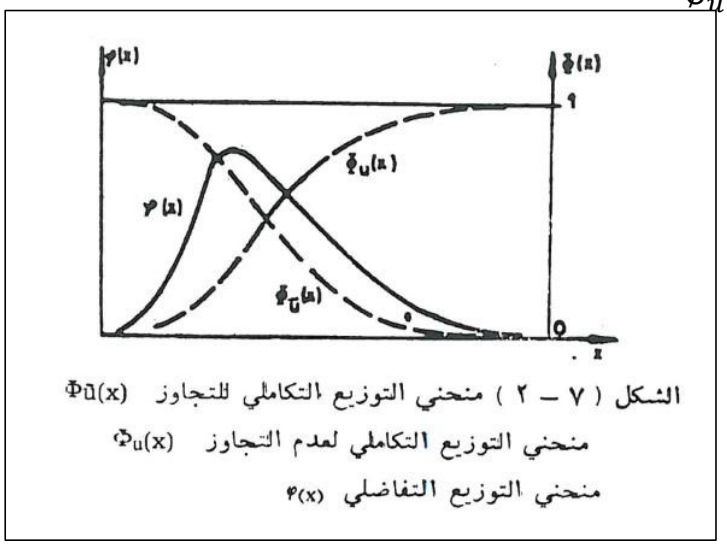
بالطريقة نفسها نحسب احتمال التجاوز من المعادلة التالية:

$$\emptyset(x > xi) = \emptyset_{\overline{u}(xi)} = \int_{xi}^{+\infty} \varphi_{(x)} * dx$$

نلاحظ هنا أنه تمت مكاملة تابع كثافة الاحتمال من القيمة Xi حتى القيمة العظمى .

- الشكل التالي يظهر أحد منحنيات التوزيع التفاضلية بالإضافة إلى منحني التوزيع التكاملي للتجاوز ومنحني التوزيع التكاملي لعدم التجاوز:

$$\emptyset_{u(xi)} + \emptyset_{\overline{u}(xi)} = 1$$


الشكل (٧ ــ ٢) منحني التوزيع التكاملي للتجاوز Φū(x)
منحني التوزيع التكاملي لعدم التجاوز Φu(x)
منحني التوزيع التفاضلي φ(x)

غالبا مايسمى احتمال التجاوز باحتمال الضمان أو اختصاراً الضمان

فترة التكرار:

لنفرض أنه تم قياس التصاريف السنوية الأعظمية لأحد الأنهار (أعلى قيمة للتصريف كل سنة) ووجد أن نسبة قيم التصاريف التي هي أعلى من $120m^3/s$ تبلغ 5% أي:

$$\phi_{\overline{u}(120)} = 5\%$$

- من هنا يمكن أن نستنتج أنه وسطياً تصادف القيمة للتصريف 120m3/s أو أعلى منها مرة كل عشرين سنة، بمعنى آخر إن قيمة التصريف $Q \geq 120\ m3/s$ تتكرر مرة كل عشرين سنة T= year
- بهذا نسمي T فترة التكرار ويمكن أن تكتب كالآتي:

$$Q_{(T)} = Q_{(20)} = 120m^3$$

- نعرف فترة التكرار بأنها الزمن الوسطي لتكرار حادث ما ويرمز له ب T كما تسمى أحياناً فترة الرجوع وتقدر بالسنة.
- مما سبق يمكن أن نستنتج أنه من العينة العشوائية والتي تمثل عناصرها قيمة التصريف الأعظمي من كل عام فإن فترة التكرار تحسب من العلاقة:

$$T = \frac{1}{\phi_{\overline{U}}} = \frac{1}{1 - \phi_U}$$

انتهت المحاضرة