الإحصاء والإحتمالات - المحاضرة الثالثة

Statistics and probabilities for engineering- Lecture 3

Dr Fadi KHALIL

Doctor lecturer in statistics and programing

في هذه المحاضرة:

- المقاييس الإحصائية (Statistical measures).
  - العلاقة بين المتوسط الحسابي، الوسيط والمنوال.
  - مفهوم الربيعيات quartiles، العشريات deciles، الميئينات Percentiles.
  - مقاييس التشتت Dispersion measures
- مقدمة عن التوزيعات الإحتمالية Probability distribution

8-المقاييس الإحصائية Statistical measures:

كما تمّ الإشارة له سابقاً بأنّ مقاييس النزعة المركزية تقدم وسيلة لتلخيص الخصائص المتضمنة ضمن مجموعة من البيانات المتاحة. وفي هذا الإطار من المفيد التطرق لمفهوم مهم وهو العلاقة بين أبرز هذه المقاييس وهي المتوسط والوسيط والمنوال. إذ أنّه انطلاقاً من هذه العلاقة يمكن اكتشاف مجموعة من الخصائص كما سبيّنه المثال الآتي:

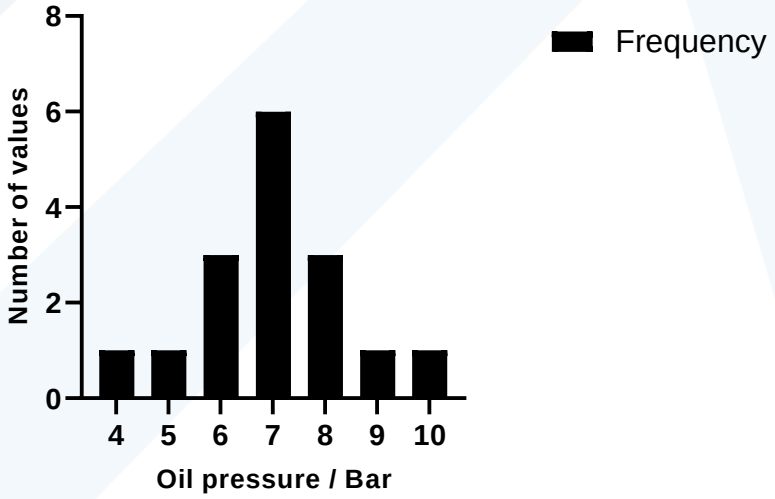
8-2- العلاقة بين الوسط الحسابي والوسيط والمنوال:

لفهم وشرح العلاقة بين هذه المؤشرات فإنّه من الأفضل عرض المثال الآتي:

مثال 12: ليكن لدينا البيانات الآتية عن ضغط الزيت (مقاسة بالبار) في مجموعة من المحركات القديمة:

4, 5, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 9, 10

- من الملاحظ أنّه يمكن تمثيل هذه البيانات بالمدرج التكراري:



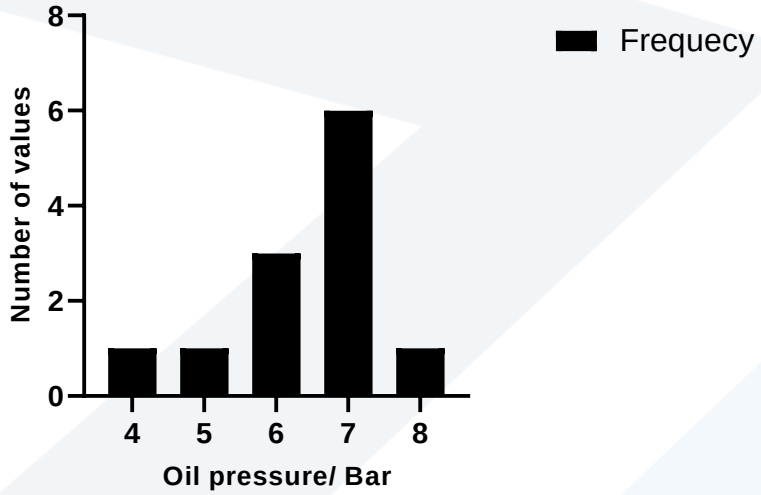
نلاحظ أنّ المدرج التكراري ذو شكل متناظر، أي إذا قمنا برسم خط عمودي من نقطة 7 يمكن تقسيم هذا الشكل إلى قسمين متساويين ومتشابهين (مثل المرآة)، ونلاحظ أيضاً أنّ قيمة كل من الوسط الحسابي والوسيط والمنوال هي قيم متساوية وتبلغ 7.

$$Mod = Me = \bar{x}$$

- ولكن في حال كانت البيانات كما يأتي:

4, 5, 6, 6, 6, 7, 7, 7, 7, 7, 7, 8

حينها سيكون المدرج التكراري :



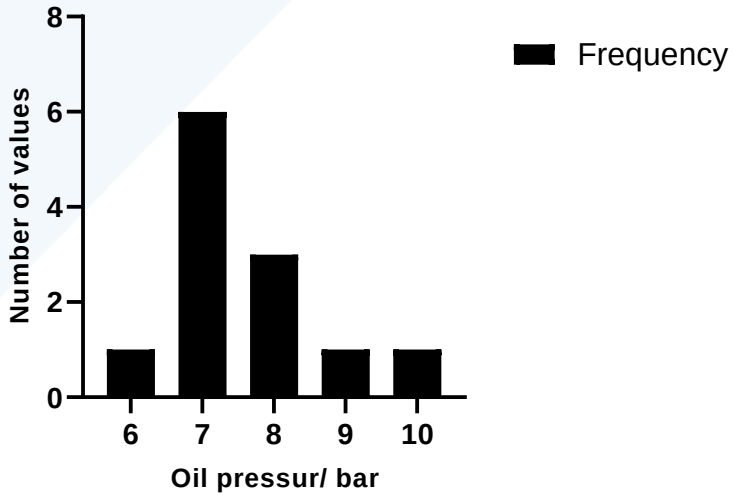
هنا نلاحظ أنّ المدرج التكراري لم يعد متناظر، ونقول انّه ملتو نحو اليسار. ونلاحظ أيضاً أن قيمة المتوسط هي أقل الوسيط ومن الموال. حيث تبلغ قيمة الوسيط $Me = 6.5,\ Mod = 7, \bar{x} = 6.3$

$$\bar{x} \leq Me \leq Mod$$

- بالمقابل إذا كان لدينا البيانات الآتية:

$$6, 7, 7, 7, 7, 7, 7, 8, 8, 8, 9, 10$$

في هذه الحالة يكون المدرج التكراري ملتو نحو اليمين، كما هو واضح في الشكل الآتي:



هنا تكون قيمة المتوسط هي الأكبر يليها الوسيط ومن ثم المنوال.

$$Me = 7.5,\ Mod = 7, \bar{x} = 7.7$$

$$Mod \leq Me \leq \bar{x}$$

3-8- الربيع الأول والثاني *First and second quartiles*:

مفهوم الربيعيات *Quantiles* هو تعميم لفكرة الوسيط *Median*. فالوسيط هو القيمة التي تقسم السلسلة إلى قسمين متساويين. كذلك، الربيع يقسم البيانات إلى قسمين ولكن بنسب أخرى. على سبيل المثال الربيع الأول *quantile 25%* هو القيمة التي تقسم البيانات إلى قسمين، بحيث يكون على الاقل 25% من القيم تكون أقل أو تساوي قيمة الربيع الأول، وعلى الأقل 75% من القيم هي أكبر أو تساوي من قيمة الربيع الأول.

أمّا الربيع الثالث *quantile 75%* هو القيمة التي تقسم البيانات إلى قسمين، بحيث يكون على الاقل 75% من القيم تكون أقل أو تساوي قيمة الربيع الثالث، وعلى الأقل 25% من القيم هي أكبر أو تساوي من قيمة الربيع الثالث. ويلاحظ أنّ الوسيط هو الربيع الثاني أي *quantile 50%*

ويمكن تعميم فكرة الربيعيات لصياغة أي قيمة يمكن أن تقسم سلسلة البيانات بعد ترتيبها لأقسام عدة، مثلاً، ليكن المقدار $(\alpha * 100) * \%$:

- إذا كانت $\alpha = 0.1, 0.2, \ldots\ldots 0.9$ فإنّه يطلق على هذه القيم بالعشريات *deciles*.
- إذا كانت $\alpha = 0.01, 0.02, \ldots\ldots 0.99$ فإنّه يطلق على هذه القيم بالميئيات *percentile*.
- كذلك إذا كانت $\alpha = 25\ , 50, 75$ يطلق عليه كما تم عرضه سابقاً بالربيعيات *quartiles*.

ولإيجاد أي قيمة من هذه القيم، يتم تطبيق ما يأتي:

ليكن عدد من البيانات n مرتبة ترتيباً تصاعدياً، بالتالي لحساب قيمة $\alpha * 100\ \%$ أو X_α يتم التمييز بين حالتين:

إذا لم يكن $n * \alpha$ عدد غير صحيح فإنّ $\tilde{X}_\alpha$ هو أول قيمة ترتيبها أكبر من $n * \alpha$.

أمّا إذا كان $n * \alpha$ عدد صحيح فإنّ X_α هو الوسط الحسابي للقيمة $X_{n\alpha}$ والقيمة $X_{n\alpha+1}$ أي :

$$\tilde{X}_\alpha = \frac{X_{n\alpha} + X_{n\alpha+1}}{2}$$

ولتوضيح ما ورد أعلاه، لنأخذ المثال الآتي:

مثال:

لتكن البيانات الآتية عن درجات الحرارة في اللاذقية في شهر تشرين الأول وهي مرتبة كالآتي:

C°	21	22	22	23	24	24	24	25	25	25	25	25	26	26	26	26
i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
C°	27	27	28	28	28	29	29	29	29	29	30	30	30	30	31	
i	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	

المطلوب: تحديد الربيع الأول 25%، الوسيط (الربيع الثاني) 50%، الربيع الثالث 75%.

الحل:

في الربيع الأول 25%، يكون $\alpha = 0.25$، و $n = 31$ بالتالي $n\alpha = 31 * 0.25 = 7.75$ ، وبما إنّ 7.75 هو عدد غير صحيح يتم النظر إلى الجدول السابق واختيار أول قيمة ترتيبها أكبر من 7.75 وهو 8، أي أنّ الربيع الأول هو القيمة التي ترتيبها 8، أي $\tilde{X}_{0.25} = X_8 = 24$. ولتفسير هذه القيمة يمكن القول أنّه خلال الشهر الماضي 25% من الأيام كانت درجة حرارتها 24 فأقل، وأنّ 75% من الأيام كانت درجة حرارتها أكبر من 24.

في الربيع الثاني 50% أو الوسيط، يكون $\alpha = 0.5$ ، و $n = 31$ بالتالي:

$n\alpha = 31 * 0.5 = 15.5$ ، وبما إنّ 15.5 هو عدد غير صحيح، يتم النظر إلى الجدول السابق واختيار أول قيمة ترتيبها أكبر من 15.5 وهو الرقم 16، أي أنّ الربيع الثاني (أو الوسيط) هو القيمة التي ترتيبها 16، أي $\tilde{X}_{0.5} = X_{16} = 26$. ولتفسير هذه القيمة يمكن القول أنّه خلال الشهر الماضي 50% من الأيام كانت درجة حرارتها 26 فأقل، وأنّ 50% من الأيام كانت درجة حرارتها أكبر من 24.

ملاحظة: لإيجاد الوسيط يمكن أيضاً اتباع الطريقة التي تم عرضها سابقاً، وهي بما إنّ عدد القيم $n = 31$ هو عدد فردي بالتالي الوسيط هو القيمة التي ترتيبها $\frac{n+1}{2} = \frac{31+1}{2} = 16$ أي أنّ الوسيط هو القيمة التي ترتيبها 16، أي $X_{16} = 26$.

في الربيع الثالث 75% ، يكون $\alpha = 0.75$، و $n = 31$ بالتالي:

$n\alpha = 31 * 0.75 = 23.25$ ، وبما إنّ 23.25 هو عدد غير صحيح، يتم النظر إلى الجدول السابق واختيار أول قيمة ترتيبها أكبر من 23.25 وهو الرقم 24، أي أنّ الربيع الثالث هو القيمة التي ترتيبها 24، أي $\tilde{X}_{0.75} = X_{24} = 29$. ولتفسير هذه القيمة يمكن القول أنّه خلال الشهر الماضي 75% من الأيام كانت درجة حرارتها 29 فأقل، وأنّ 25% من الأيام كانت درجة حرارتها أكبر من 29.

4-8- مقاييس التشتت (Spread measues):

إنّ مقاييس النزعة المركزية التي تمّ تناولها سابقاً تعطي معلومة مهمة عن الظاهرة المدروسة ولكنّها غير كافية للحصول على وصف دقيق لهذه الظاهرة. فمثلاً بفرض لدينا علامات ثلاث طلاب في صفين دراسيين:

A : 60 70 80

B : 50 70 90

نلاحظ أنّ لكلا الصفين نفس المتوسط الحسابي وهو 70 (كمقياس من مقاييس النزعة المركزية) ولكن هذه القيمة غير كافية للقول أن مستوى الطلاب في كلا الصفين هو مستوى متقارب، حيث نرى أنّ مستوى الطلاب في الصف A متقارب أكثر من مستوى الطلاب في الصف B، نتيجة لأحد الأسباب. وبالتالي نستنتج أنّه لا بد من دراسة مقدار تشتت أفراد العينة عند دراسة ظاهر ما، وهذا ما يتم قياسه عن طرق مقاييس التشتت.

من مقاييس التشتت نذكر ما يأتي:

8-4-1- المدى (Range):

يعبر عن المدى عن الفرق بين أكبر وأصغر قيمة في العينة:

$$R = X_{max} - X_{min}$$

لنأخذ المثال السابق (13) الذي يتعلق بنسبة الكثافة السكانية، لمجموعة من المدن:

المدينة	عدد السكان في 10 متر
دمشق	6
حلب	4
حمص	3
اللاذقية	5
درعا	2
السويداء	1

نلاحظ أنّ المدى للقيم في هذا المثال يبلغ:

$$\bar{x} = X_{max} - X_{min} = 6 - 1 = 5$$

8-3-2- التباين (Variance):

يعرف التباين بأنّه الوسط الحسابي لمربعات انحرافات القيم عن متوسطها الحسابي، أي يعطى بالعلاقة الآتية:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

أيضاً استمراراً للمثال السابق، نذكر أن قيمة الوسط الحسابي هي 3.5، وبالتالي لحساب التباين:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - 3.5)^2}{6}$$

$$\sigma^2 = \frac{(6-3.5)^2 + (4-3.5)^2 + (3-3.5)^2 + (5-3.5)^2 + (2-3.5)^2 + (1-3.5)^2}{6} = 2.83$$

هنا نلاحظ أنّ قيمة التباين مقارنة بقيمة المتوسط هي قيمة متوسطة نسبياً بمقارنة بقيمة المتوسط الحسابي.

انطلاقاً من خاصية الوسط الحسابي بأنّ مجموع انحرافات القيم عن وسطها الحسابي يكون مساوياً للصفر، يتضح لنا فائدة أخذ مربعات هذه القيم بعين الاعتبار. وهناك طريقة أخرى، وهي أخذ القيم المطلقة للانحرافات وبالتالي نحصل على ما يسمى بالانحراف المتوسط (Absolute deviation man):

$$MD = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$$

$$\sigma^2 = \frac{|6-3.5| + |4-3.5| + |3-3.5| + |5-3.5| + |2-3.5| + |1-3.5|}{6} = 1.5$$

8-4-3- الانحراف المعياري (Standard deviation):

ويعرف بأنّه الجذر الموجب للتباين، أي

$$SD = \sqrt{\sigma^2}$$

من المثال السابق، يتضح ان قيمة الانحراف المعياري هي :

$$SD = \sqrt{\sigma^2} = \sqrt{2.83} = 1.6$$

8-4-4- معامل الاختلاف (Coefficent variation):

كما رأينا سابقاً أنّه للحكم على مقدار تشتت البيانات في عينة ما لا بدّ من حساب مقاييس التشتت المختلفة، ولكن لتفسير قيم هذه المقاييس هنا مقياس معامل الاختلاف (CV) الذي يعبر عن النسبة المئوية للانحراف المعياري مقارنة

بالوسط الحسابي، ويفيد أيضاً هذا المقياس بمقارنة مقدار التشتت بين عينتين مختلفتين من ناحية الحجم (أو عدد القيم) ومن ناحية قيمة المتوسط الحسابي.

يحسب معامل الاختلاف (CV) كما يأتي :

$$CV = \frac{SD}{\bar{x}} * 100 = \frac{1.6}{3.5} * 100 = 48\%$$

يمكن تفسير هذا المعامل بأنّ نسبة تشتت العينة بالنسبة للمتوسط الحسابي هو 48% وهو أقل من 50% إذا يمكن القول أنّ التشتت في المستوى المتوسط.

9- التوزيعات الاحتمالية Probability distribution:

9-1- مفهوم الاحتمالات:

يرتبط مفهوم الاحتمالات بمصطلح عدم التأكّد، فالعالم مليء بظاهرة عدم التأكد، و الحوادث الطبيعية المستقبلية التي لا يمكن التأكد من وقوعها، أو أنّنا لا نستطيع أن نعرف بشكل مسبق ما هي القيمة التي سيأخذها متغير ما. مثلاً عند تجربة رمي القطعة النقدية فإنّنا لا نعرف فيما إذا كانت النتيجة هي الحصول على صورة أو شعار، كذلك في حال جنس المولد أيضاً.

في هذا السياق اهتمّ الكثير من الإحصائيين بالفكرة القائلة بأنّ عدم التأكد لوقوع حدث ما لا يمنع من الاستفادة من التجارب السابقة للتنبؤ باحتمال وقوع هذا الحدث، مثلاً نعود لتجربة رمي القطعة النقدية، إذا أعدنا التجربة عدد كبير جداً من المرات فإننا نلاحظ ان احتمال الحصول على صورة أو شعار يقترب من (2/1)، وهذا ما يّسمّى **بقانون الاعداد الكبيرة**، الذي يمكن استخدامه لتقديم صيغة محددة للتنبؤ باحتمال وقوع الأحداث الغير مؤكّدة.

عموماً، ومن خلال التجربة يمكن تعريف احتمال وقوع حدث ما (P) من ضمن مجموعة من الحالات الكلية (n):

$$P = \frac{k}{n}$$

حيث، يعبر k عن عدد الحالات الممكنة.

في تجربة حجر النرد يمكن حساب احتمال أن يظهر الرقم 2 على الوجه العلوي هو $(\frac{1}{6})$، لأنّ عدد الوجوه التي تحمل الرقم 2 في حجر النرد هو وجه واحد، في حين أن حجر النرد يحوي على 6 وجوه، إذاً عدد الحالات الممكنة هو (1) وعدد الحالات الكلية هو 6.

مثال 14:

إذا كان لدينا صندوق يحوق على 5 كرات، 2 صفراء، و 3 حمراء، وبالتالي إذا سحبنا كرة بشكل عشوائي من هذا الصندوق، يكون احتمال أن تكون هذه الكرة صفراء هو $(\frac{2}{5})$، حيث يبلغ عدد الحالات الممكنة (2) والحالات الكلية هو (5).

خصائص قيمة الاحتمال:

[1] من الملاحظ مما سبق أن قيمة الاحتمال تتراوح ضمن المجال [0, 1]، حيث يسمى الحدث الذي يكون احتمال وقوعه (1) بالحدث الأكيد، والحدث الذي يساوي احتمال وقوعه (0) بالحدث المستحيل. أمّا الحدث الذي لا يكون أكيد ولا مستحيل فتتراوح قيمة احتمال حدوثه بين الصفر والواحد.

[2] إذا قمنا بجمع قيم احتمالات جميع الحوادث الممكنة فإننا نحصل على القيمة (1)، إذا عدنا للمثال (14) نلاحظ أن احتمال الحصول على كرة صفراء هو $(\frac{2}{5})$، أمّا احتمال الحصول على كرة حمراء هو $(\frac{3}{5})$، وبالتالي فإنّ احتمال الحصول على كرة صفراء أو حمراء يساوي $(\frac{2}{5}+\frac{3}{5}=1)$.

2-9- أشهر التوزيعات الاحتمالية:

يعتبر التوزيع الاحتمالي من الأدوات الإحصائية التي تستخدم لوصف الظاهرة المدروسة عن طريق حساب الاحتمالات المصاحبة للقيم التي من الممكن أن تأخذها هذه الظاهرة.

يشبه التوزيع الاحتمالي مبدأ التوزيع التكراري ففي الوقت الذي يتضمن التوزيع التكراري قيمة تكرار محدد لكل قيمة يأخذها المتغير، فإنّ التوزيع الاحتمالي أيضاً يعطي (أو يوزّع) لكل قيمة يأخذها المتغير قيمة احتمالية محددة. ويرتبط بمفهوم التوزيع الاحتمالي قانون التوزيع الاحتمالي الذي يقوم بحساب احتمال أن يأخذ المتغير قيمة ما (في حال كان المتغير منقطع) أو قيمة محصورة ضمن مجال ما (إذا كان المتغير مستمر).

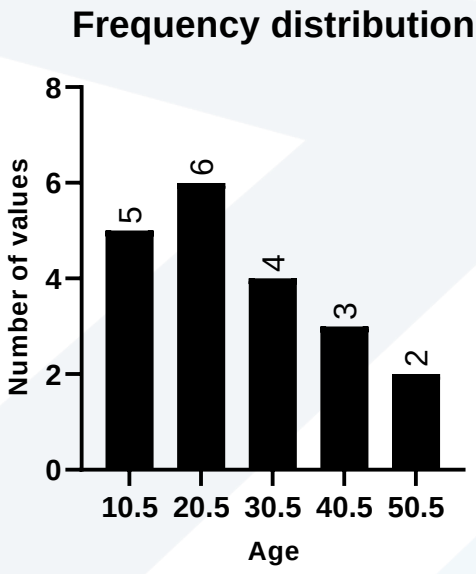
أحد أشكال التوزيع الاحتمالي هو ما يسمى بتابع التوزيع التراكمي (Cumulative probability distribution)، ويعطي احتمال أن يأخذ المتغير قيمة أصغر من قيمة محددة. ويرتبط تابع التوزيع بفكرة النسب التراكمية التصاعدية التي تعرفنا عليها في الجدول التكراري، ولكي نفهم هذا الأمر لنتناول المثال الآتي:

مثال 15: ليكن لدينا الجدول الذي يحتوي على تكرار الفئات العمرية لـ 20 شخص في أحد الأحياء:

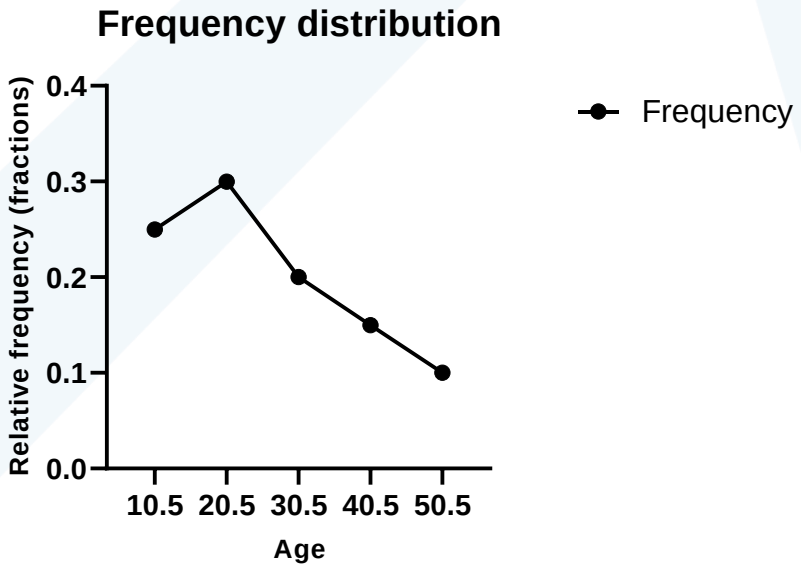
الفئات	التكرار	النسبة التكرارية	النسبة التجميعية التصاعدية
[6-15]	5	%25	%25
[16-25]	6	%30	%55
[26-35]	4	%20	%75
[36-45]	3	%15	%90
[46-55]	2	%10	%100
المجموع	20	%100	

من المثال السابق يتضح أنّ احتمال أن نختار شخص عمره بين [5-15] هو 25% أو 0.25، كذلك احتمال أن نختار شخص عمره 35 فأقل هو 75% أو 0.75.

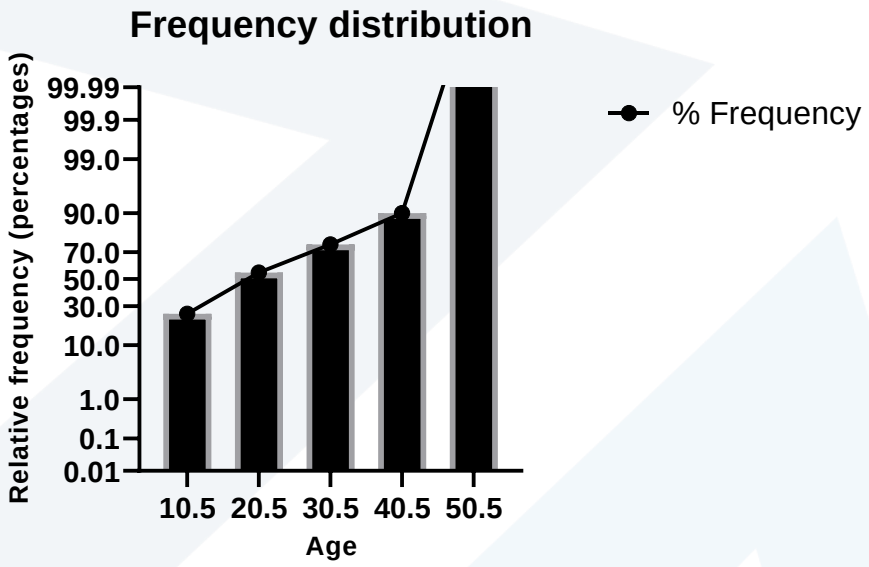
هنا يمكن رسم تابع التوزيع الاحتمالي:



يمكن تمثيل الشكل السابق بشكل المضلع (Polygon) كالآتي:

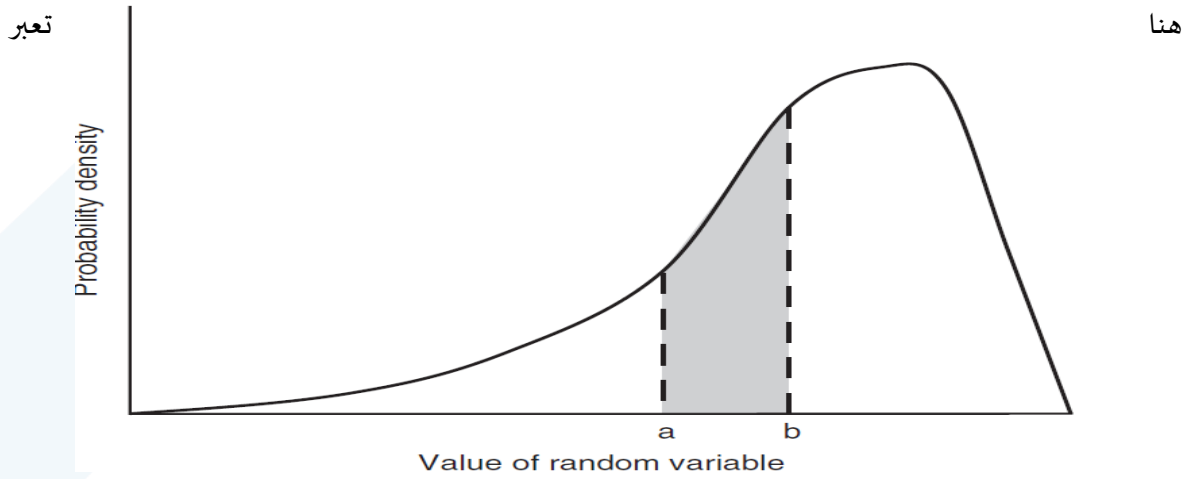


أما تابع التوزيع الاحتمالي التراكمي فيأخذ الشكل الآتي:



نلاحظ أن الفئات العمرية السابقة تعبر عن متغير منقطع، ولكن في حال كان التوزيع مستمر، حيث يكون هناك عدد كبير من القيم المستمرة، فإنّ التوزيع الاحتمالي يأخذ شكل منحني مستمر يعبر عن احتمال أن يأخذ المتغير قيمة تقع بين قيمتين محددتين (a, b).

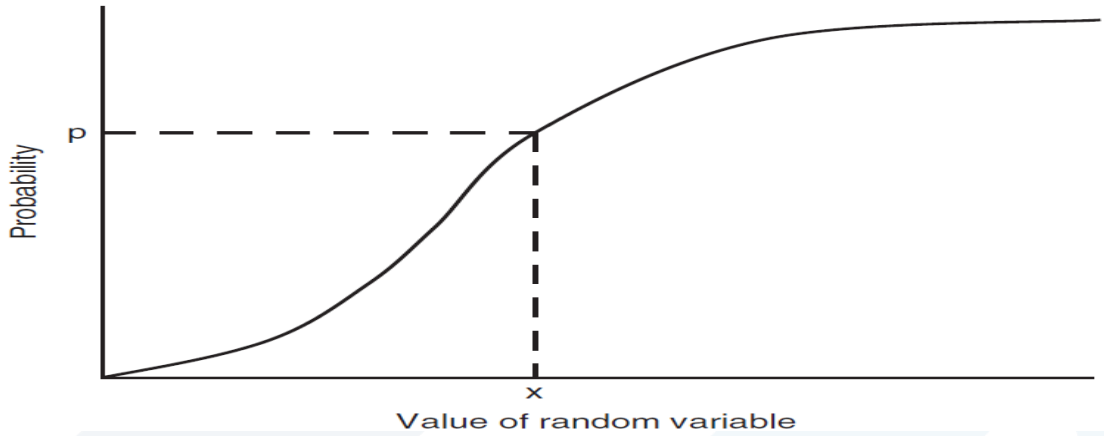
مثلاً، قد يكون من الصعب احتمال أن يصل إنتاج القمح في أحد الحقول إلى 1.23 طن بالضبط، ولكن قد يكون من الممكن حساب أن يقع إنتاج القمح في هذا الحقل بين 1 طن و 2 طن. وهنا يطلق على اسم تابع التوزيع اسم تابع الكثافة الاحتمالية (density of the probability)، فكما لاحظنا سابقاً أنّ المدرج التكراري يعبر شكل التوزيع الاحتمالي للتوزيع المنقطع، فإنّ شكل تابع التوزيع للمتغير المستمر يمكن أن يكون على الشكل الآتي:



المساحة المظللة على احتمال أن يقع قيمة المتغير المستمر بين قيمتين (a, b)، كما في المثال السابق على احتمال أن يكون إنتاج القمح في أحد المحاصيل يتراوح بين 1 طن و 2 طن.

تجدر الإشارة هنا على أن إجمالي المساحة الواقعة تحت المنحني تكون مساوية للواحد، لأنّها تعبر عن احتمال الحدث الأكيد وهو احتمال أن يأخذ المتغير المستمر قيمة ما.

أمّا تابع التوزيع التراكمي يأخذ الشكل الآتي:



يمكن تصنيف التوزيعات الاحتمالية حسب المتغيرات إلى توزيعات احتمالية للمتغيرات المنقطعة توزيعات احتمالية للمتغيرات المستمرة.