Statistics

Chapter 1

The Nature of Probability and Statistics

Statistics is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from data.

Main Areas of Statistics

Descriptive statistics consists of the collection, organization, summarization, and presentation of data.

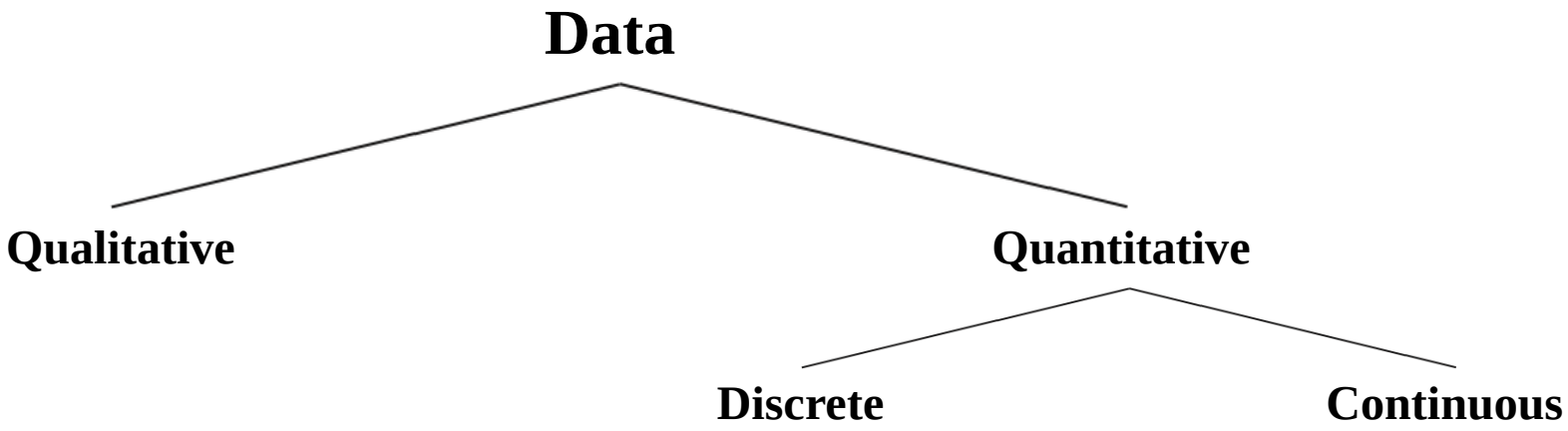
Inferential statistics consists of *generalizing* from samples to populations, performing estimation and hypothesis tests, determining relationships among variables, and making predictions.

Inferential statistics uses **probability** (the likelihood of an outcome occurring) to make conclusions and predictions.

A **population** consists of all subjects that are being studied.

A **sample** is a group of subjects selected from a population.

Data are the values (measurements or observations) that the variables can assume.



A **data set** is a collection of **data values**.

A **variable** is a characteristic or attribute that can assume different values.

A **random variable** is a variable whose values are determined by chance.

Types of Variables

Qualitative variables are variables that can be placed into categories, according to some characteristic or attribute.

Quantitative variables are numeric in nature and can be ordered or ranked. Quantitative variables can be either **discrete** or **continuous**.

A **continuous variable** is a quantitative variable that can assume ANY numerical value between any two specific values.

- Obtained by measuring.
- May include fractions and decimals.

A discrete variable is a quantitative variable that has either a finite number of possible values or a countable number of possible values.

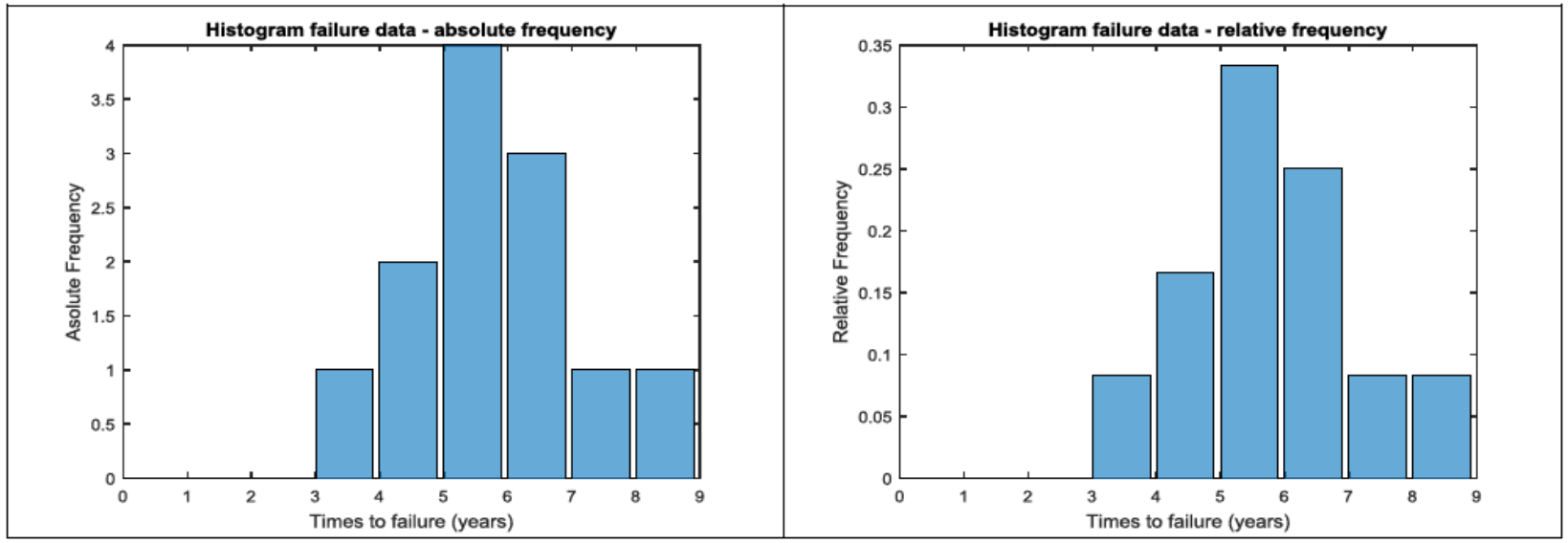
- Countable means that the values result from counting, such as 0, 1, 2, 3…

Histogram

Histogram: set of columns describing the distribution of specific data. Each column represents a number of items or frequency.
This graphic gives an idea of sample shape.

Absolute frequency: It indicates the number of data points.

Relative frequency: I t indicates the percentage of the total number of data points.



The characteristics of any data can be summarized by knowing some statistical properties:

1. The data center is measured using the median or mean.
2. Data spread is measured using variance and standard deviation.
3. Knowing that most of the data is located using the quartiles.
4. Find outlie

Median

When arranging the data in ascending or descending order, the median is the statement in which 50% of the data fall before and 50% after it, and it is calculated:

$$X(1), X(2), ..., X(n-1), X(n)$$

$$\tilde{X} = \begin{cases} X(\frac{n+1}{2}) & \text{if } n \text{ is odd} \\ \frac{X(\frac{n}{2}) + X(\frac{n}{2}+1)}{2} & \text{if } n \text{ is even} \end{cases}$$

Example 1: Find the median for the data listed below 60 72 40 80 63.

Sol:

We arrange the data in ascending order

40 60 63 72 80

Since the number of data is odd, the median value (third) represents the median

$\tilde{X} = X\ (5 + 1)/2 = X3 = 63$

Example 2: Find the median for the data listed below 72 60 72 40 80 63.

Sol:

We arrange the data in ascending order

40 60 63 72 72 80

Since the number of data is even, the median value is the average of the two readings in the middle of the data

$\tilde{X}= (63+72)/2 = 67.5$

Statistics

Chapter 1

The Nature of Probability and Statistics

Data Collection and Sampling Techniques

Data is often collected via surveys.

1. Telephone Surveys
2. Mailed Questionnaire
3. Personal Interview
4. Internet Survey

- What are advantages and disadvantages of data collection through surveys?

	Advantages	Disadvantages
Telephone Surveys	1. Less costly than personal interview. 2. People may be more candid.	1.Not everyone has a phone. 2. Cell phones typically not included. 3.Tone of interviewer's voice may affect response.
Mailed Questionnaire	Can cover wider geographic area than phone survey or personal interview 2. Respondents can remain anonymous 3. Less expensive	1.Low number of responses 2.Inappropriate answers to questions 3.Low reading abilities or not understanding questions may create useless responses.

	Advantages	Disadvantages
Personal Interview	In-depth responses to questions	1.Interviewers must be trained in asking questions and recording responses (which is costly). 2. Interviewer may be biased in the selection of participants.
Internet survey	As with telephone and mail surveys, 1. Inexpensive, often free 2. Candor 3. Large geographic coverage 4. Anonymity	1.Will miss demographics without computer access 2. May have inappropriate answers if questions are misunderstood

Sampling Techniques

- Researchers use samples to collect data and information about a particular variable from a population.
- Samples save time, money, and may actually allow a researcher to collect better information.
- Samples need to be representative of the population or they are meaningless in drawing conclusions about the population.
- Sampling must be done in a way that the samples are **unbiased**—that each subject in the population has an equal chance of being in the sample.

Technique	Description
Random sampling	Uses chance methods or random numbers to select the sample. **Everyone or everything from the population has the same chance of being selected for the sample** and it is the best way of obtaining a representative sample.
Systematic sampling	Numbers each subject of the population and then **selects every kth subject.**
Convenience sampling	Selects subjects that are convenient for the researcher. These samples are typically of not statistical value.
Stratified sampling	Divides the population into groups (called strata) according to some characteristic that is important to the study, then randomly samples subject from each group.
Cluster Sampling	Divides the population into groups called clusters by some means such as geographic area or schools in a school district, etc. Then randomly select some of the clusters and use ALL members of the selected clusters.

Observational and Experimental Studies

Observational study - the researcher merely observes what is happening or what has happened in the past and tries to draw conclusions based on these observations.

Experimental study - the researcher manipulates one of the variables and tries to determine how the manipulation influences other variables.

Statistics

The summary of lecture 1 and lecture 2

The importance and concept of statistics

اهمية ومفهوم علم الاحصاء

يعد علم الاحصاء اليوم من اهم العلوم التي تتوقف عليها التنمية السياسية و الاقتصادية و الثقافية .... الخ و للاحصاء حصة اساسية من عمل الدول والمؤسسات و المنظمات السياسية و الاقتصادية والاجتماعية ، و كثيرا ما يرتهن مصير مشاريع او قرارات كبرى بالنتائج االتي يقدمها الاحصاء في مجال معين. و بصورة عامة، فأن افتقاد الجهد الاحصائي، في مجال من المجالات، يمنع من التأكد وتحصيل الضمان في استجابة اي مشروع للواقع، كما يحول دون تحديد مدى نجاحه او اخفاقه، ويجعل في الاقدام عليه شيئا من المخاطر.

ويمكننا فهم الاحصاء بصورة مختصرة كالاتي:

اولا: الاحصاء قادر على توصيف الظواهر العلمية توصيفا رقميا كميا دقيقا و اكثر وضوحا و قربا من الواقع.

ثانيا: يستطيع الاحصاء ان يفسر الظواهر، و ان يحدد مدى تأثير العوامل المفترضة، كما يمكنه التنبؤ بالمستقبل بالمعنى العلمي للكلمة. مثلا يمكن معرفة حالة الطقس في يوم ما بالاعتماد على قاعدة بيانات تظهر حالة الطقس لنفس اليوم بالسنين السابقة.

ثالثا: يعتمد الاحصاء المعادلات الرياضية و حساب الاحتمالات، و يرتكز على اسس علمية رياضية مبرهن عليها.

رابعا: الاحصاء هو منهج و عقل وتفكير و الية تأمل و نمط قراءة. فهو غير محدد بمادة علمية سوى تلك التي تحتوي نظام العينات.

Definition of statistics

تعريف علم الاحصاء

نعني بالاحصاء العلم الذي يكون موضوعه جمع وتنسيق وتصنيف وتعليل الوقائع او المعطيات الرقمية او العددية في كل صنف، بحيث يمكن الحصول على نسبب عددية مستقلة استقلالا ملموسا عن المصادفة و استثناءاتها، و دالة على وجود العلل المنتظمة التي ينعدم حصولها بوجود العلل الفجائية، فهو يتخذ طريقة للتحليل في العلوم الدقيقة و العلوم الاجتماعية و في المشروعات الاقتصادية على اختلافها. وهو يعني بالتنبؤ باحتملات حدوث امر بعينه او حاله بعينها. و تطور علم الاحصاء تطورا كبيرا بعد استحداث الحاسبة الالكترونية التي تتعامل مع كميات من الارقام الضخمة تعاملا سريعا. ان المعنى الحديث لعلم الاحصاء انه العلم الذي يبحث في جمع البيانات، و تنظيمها و عرضها و تحليلها و استقراء النتائج و اتخاذ القرار بناء عليها.

المجتمع الاحصائي (population): هو جميع المفردات و المعلومات التي تتمتع بصيغة ما أو خاصية ما مشتركة ضمن اطار عام واحد ويكون المجتمع الاحصائي مقيدا بزمان و مكان او لا يكون و قد يكون محددا مثل عدد طلاب المرحلة الاولى في كلية الهندسة او يكون غير محدد مثل عدد الاسماك في المحيط الهاديء.

العينة (Sample): و هي جزء من المجتمع الاحصائي محل الدراسة يتم اختياره اما عشوائيا او بصورة علمية خاصة.

المتغيير (Variable): خاصية او صفة للافراد او الاشكال تختلف من عنصر الى الاخر. وهو بنوعين:

1- متغيير نوعي (وصفي): ويشمل اي صفة او ظاهرة تتغير نوعيا و تسجل بوصف لفضي مثل ذكر ، انثى، حار ، بارد.

2- متغيير كمي (رقمي): و يشمل صفة او ظاهرة تتغير كميا وتسجل بارقام عددية مثل معدل درجة الحرارة ، منسوب سقوط الامطار. و المتغير الرقمي يقسم الى قسمين:

الف- متغيير كمي متصل (continuous variable): و يكون فيه استمرارية او بمعنى اخر تقبل القيم الكسرية مثل درجة الحرارة والوزن فنستطيع ان نقول 23.5 درجة او 75.3 كغم.

ب- متغيير كمي منفصل (discrete variable): و هي تكون متقطعة او لا تقبل قيم كسرية مثل عدد الطلاب في كلية الهندسة.

The Data

البيانات: هي مجموعة من الحقائق و المعلومات التي تتعلق بظاهرة ما و تشكل مادة خام للاحصاء.

مصادر البيانات الاحصائية:

1- مصادر تاريخية: من السجلات و الوثائق الاصلية و الثانوية والارشيفات الرسمية و الخاصة سواء كانت مبوبة في قواعد بياناتاو عشوائية.

2-مصادر ميدانية: و هي تجمع مباشرة بالمشاهدات الميدانية من وحدات المجتمع الاحصائي.

اساليب جمع البيانات:

1- اسلوب الحصر الشامل: جمع المعلومات من جميع مفردات المجتمع الاحصائي بدون استثناء. (حساب نسبة النجاح للمرحلة الاولى).

2-اسلوب العينة: يتم جمع المعلومات من جزء مختار من المجتمع الاحصائي عشوائيا او غير عشوائي لتمثيل جميع عناصر المجتمع (اي الحكم على كل المجتمع من خلال صفات و خصائص عينة ما مأخوذة منه).

اقسام الاحصاء: يقسم الاحصاء الى قسمين هما:

1- الاحصاء الوصفي: يشمل جمع و تبويب و تصنيف البيانات الاحصائية، و جعل البيانات المتعددة بشكل مفهوم و ذات مدلول يسهل التعامل معها.

2- الاحصاء الاستقرائي (الاستدلالي): يشمل استقراء النتائج و اتخاذ القرارات، و يتعامل مع طرق تحليل و تفسير و استخلاص الاستنتاجات لاتخاذ افضل قرار ممكن عندما تكون المعلومات غير وافية.

Statistics

Lecture 3

Data display methods

طرق عرض البيانات

ان ترتيب وتنظيم البيانات وعرضها بشكل يسهل تمييزها والتعرف على خصائصها يعتبر الحجر الاساس والخطوة الاولى في التحليل الاحصائي وفي بعض الاحيان يكون الغرض من استخدام طرق عرض البيانات هو جذب الانتباه ليس الى الارقام بحد ذاتها وانما الى ميول واتجاه تلك الارقام بالزيادة والنقصان وبالتالي معرفة ماهية او طبيعة الظاهرة المدروسة دون الخوض في تفاصيلها. طرق عرض البيانات هي:

a- Tables Method (طريقة الجداول):

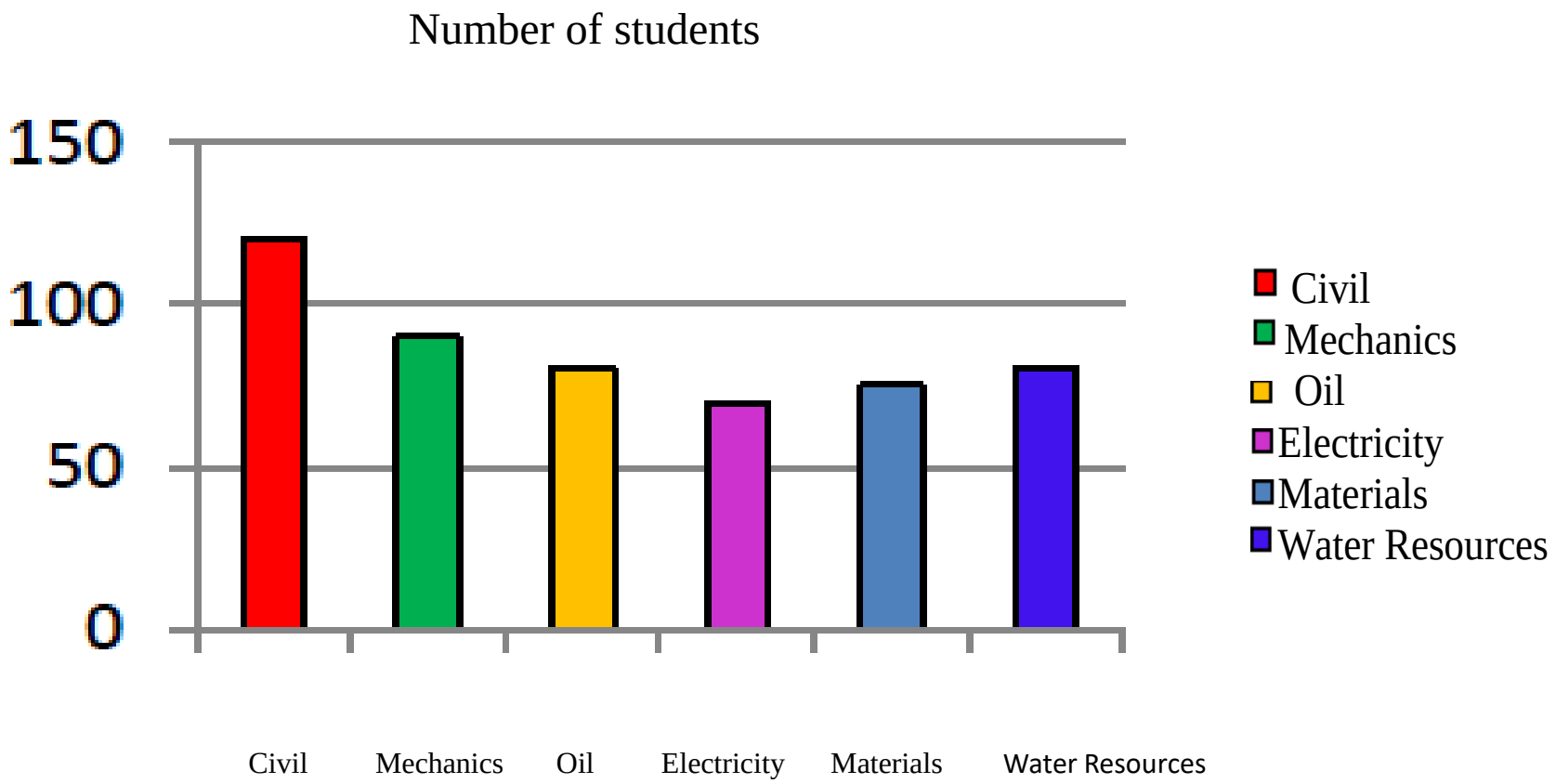
وتشمل وضع البيانات في جداول وتستعمل في عرض تغيير ظاهرة ما مع الزمن او مع مسميات اخرى معينة.

Ex: Represent the distribution of the students of the engineering college department that admitted for the year 2013-2014 among the scientific departments:

Scientific department	Civil	Mechanics	Oil	Electricity	Materials	Water Resources
Number of students	120	90	80	70	75	80

b- Rectangles or Columns Method (Bar charts) (طريقة المستطيلات او الاعمدة):

وتشمل وضع المسميات على المحور الافقي او المحور العمودي ورسم مستطيل على كل مسمى بحيث يكون ارتفاع كل مستطيل ممثلا للقيمة المقابلة لذلك المسمى وذلك باستعمال مقياس رسم المناسب. ان هذه الطريقة هي من اكثر الطرق شيوعا ويمكن عرض عدة ظواهر مع الزمن.



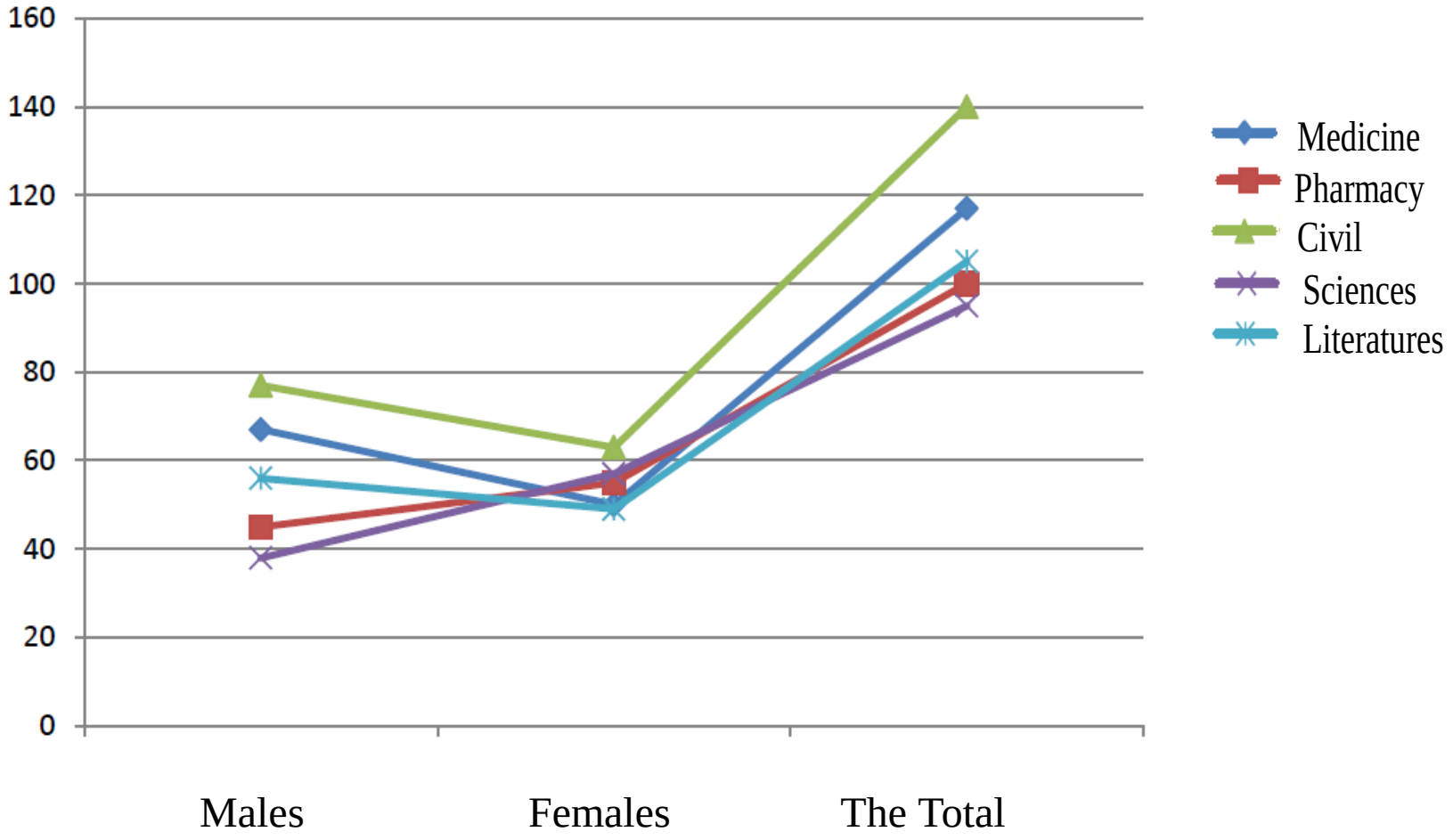
c- Included line method (Broken) (طريقة الخط المدرج او المتكسر):

وتشمل عرض البيانات الناتجة عن تغيير ظاهرة او عدة ظواهر مع الزمن مثل تغيير عدد طلبة المدارس الثانوية مع السنوات، تغيير درجة الحرارة للمريض خلال ساعات، مجموع ارصدة البنوك الخاصة مع السنوات.... الخ. وعادة ما تستخدم من قبل خبراء البورصة في متابعة صعود وهبوط الاسهم بشكل يومي او كل ساعة.

Ex: The following table represents the distribution of graduates from Baghdad University. Represent the following data in the table in the included line method.

Specialization	medicine	Pharmacy	civil	Sciences	Literatures
Males	67	45	77	38	56
Females	50	55	63	57	49
The Total	117	100	140	95	105

Sol:


160
140
120
100
80
60
40
20
0
Males
Females
The Total
Medicine
Pharmacy
Civil
Sciences
Literatures

d- Curved line method (طريقة الخط المنحني):

تماثل طريقة الخط المتدرج و نحصل عليها بتمديد الخط المتكسر ليصبح منحني، و تستعمل هذه الطريقة عندما تتغير الظاهرة على فترات قصيرة و كثيرة.

e- Pictorial method: (الطريقة التصويرية)

وتستعمل كه بديل لطريقة المستطيلات وتستعمل في الاعلانات والتقارير الحكومية وكتب الاطفال.

f- Divided circuit method (bie-charts) (طريقة الدائرة المقسمة) :

وتكون بتمثيل البيانات بدائرة كاملة مقسمة الى أجزاء بحسب الظواهر، واكثر ما تستخدم في المجالات الاقتصادية و تستخدم عندما يكون الهدف ابراز اجزاء الظاهرة المدروسة.

Ex: The following table represents the qualifications of the working team of a contracting company. Represent the following using the bie-charts method.

Qualification	Engineer	Supervisor	Expert worker	Worker
The number	3	6	5	20

Sol:

Total summation for team = 3+6+5+20 = 34

نحسب الزوايا المركزية لكل قيمة من لبيانات:

Central angle for the engineers qualification = (3/34)*360° = 31.764°

Central angle for the supervisor qualification = (6/34)*360° = 63.530°

Central angle for the expert worker qualification = (5/34)*360° = 52.941°

Central angle for the worker qualification = (20/34)*360° = 211.765°

الان يتم رسم دائرة تقسيم البيانات.

Frequency distributions

التوزيعات التكرارية

وتشمل تنظيم البيانات الكثيرة بحيث لا تفقد هذه البيانات من أهميتها الا الشيء القليل او لا تفقد ابدا. ويقسم التوزيع التكراري الى نوعين:

a- Frequency distributions (توزيع تكراري بسيط):

Ex: The following data represent the marks of 22 students in one of the exams. Represent the following data in a simple frequency table.

(19, 20, 19, 17, 15, 20, 18, 14, 12, 15, 18, 15, 17, 19, 11, 20, 14, 15, 13, 20, 15, 18)

Sol:

في التوزيع التكراري البسيط نحتاج الى:

1- قيم الفئات او القياسات (x_i). (Class values).

2- التكرارات المقابلة لهذه الفئات (f_i). (Frequencies).

الدرجة (X_i)	11	12	13	14	15	16	17	18	19	20	
التكرار (f_i)	1	1	1	2	5	0	2	3	3	4	$\sum f_i = 22$

ملاحظة: ان مجموع التكرارات يجب ان تساوي عدد البيانات الاصلي.

b- Distribution of classified frequency (توزيع تكراري ذو فئات) :

يستخدم هذا النوع من التوزيعات التكرارية في تنظيم البيانات ذات العدد الكبير جدا و ذات المدى (الفرق بين أعلى وأقل قيمة بالبيانات) كبير.

Ex: The following data represent the results of the compression test for fifty concrete cubes in units (MPa). Arrange the following data in the frequency table with the classes.

(29 30 32 39 38 38 44 28 33 31 35 37 42 49 25 34 30 31 35 37 40 26 32 33 33 39 44 45 26 31 34 31 36 36 41 27
30 31 32 37 40 39 25 34 31 30 38 35 43 48)

Sol:

1- نفرض عدد فئات (Number of classes) مناسب (عدد الفئات المفروض في اي جدول تكراري ذو فئات يجب ان لا تقل عن 5 فئات و لا يزيد عن 15 فئة).

في هذا المثال سنفرض عدد الفئات = 5 .

2- يستخرج مدى البيانات (rang of the data) = أكبر قيمة بالبيانات – أقل قيمة بالبيانات = 49 – 25=24

3- نحسب طول الفئة (Class length)الواحدة = $\frac{\text{المدى}(rang)}{\text{عدد الفئات}(\text{Number of classes})}$

$= \frac{24}{5} = 4.8 = 5$ (يقرب الى اعلى عدد صحيح قريب)

4- نفرض الحد الادنى للفئة الاولى (The minimum limit for the first class) = اصغر قيمة بالبيانات = 25

5- نحسب الحد الادنى الفعلي للفئة الاولى(The actual minimum for the first class) = الحد الادنى للفئة الاولى – 0.5 = 24.5

6- نحسب الحد الاعلى الفعلي للفئة الاولى(The actual upper limit for the first class) = الحد الادنى الفعلي للفئة الاولى + طول الفئة = 24.5 + 5 = 29.5

7- نحسب الحد الاعلى للفئة الاولى (The upper limit for the first class) = الحد الاعلى الفعلي للفئة الاولى – 0.5 = 29.5- 0.5 = 29

8- يتم تحديد الحدود الدنيا و العليا لباقي الفئات بأضافة طول الفئة لكل حد.

9- تعيين مراكز الفئات (center of classes) (X_i) = $\frac{\text{الحد الادنى للفئة+الحد الاعلى للفئة}}{2} = \frac{29+25}{2} = 27$

10- يتم حساب مراكز الفئات الباقية = مركز الفئة السابقة + طول الفئة

11- جمع التكرارات (Collect of Frequencies) و الناتج يجب ان يساوي عدد البيانات الاصلي.

Class limits حدود الفئات	Actual limits for classes الحدود الفعلية للفئات	Centers of classes مراكز الفئات	Frequencies التكرارات
25 - 29	24.5 – 29.5	27	7
30 - 34	29.5 – 34.5	32	19
35 - 39	34.5 – 39.5	37	14
40 - 44	39.5 – 44.5	42	7
45 - 49	44.5 – 49.5	47	3
			$\sum f_i = 50$

Other types of frequency distributions

انواع اخرى من التوزيعات التكرارية

a- Congregate frequency distribution (Cumulative) ((التوزيع التكراري المتجمع(التراكمي)):

التوزيع التكراري المتجمع (التراكمي): وهو بنوعين التوزيع التكراري المتجمع الصاعد و التوزيع التكراري المتجمع النازل.

يتم حساب التوزيع التكراري المتجمع الصاعد (distribution of up word combined frequency)= تكرار الفئة + تكرار الفئات السابقة.

Same as the previous example, calculate the up word combined frequency.

Actual limits for classes	Frequencies (f_i)	The up word combined frequency
Less than 24.5	0	فئة غير موجودة 0
24.5 – 29.5	7	7
29.5 – 34.5	19	26
34.5 – 39.5	14	40
39.5 – 44.5	7	47
44.5 – 49.5	3	The total number of data (n) = 50

اما لحساب التوزيع التكراري المتجمع النازل (distribution of down word combined frequency)= تكرار الفئة – تكرار الفئات السابقة.

For the seam example:

Actual limits for classes	Frequencies (f_i)	The down word combined frequency
24.5 – 29.5	7	50
29.5 – 34.5	19	43
34.5 – 39.5	14	24
39.5 – 44.5	7	10
44.5 – 49.5	3	3
More than 49.5	0	لانها فئة غير موجودة 0

b- Distribution of Relative Frequency (p) (التوزيع التكراري النسبي) :

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

f: تكرار الفئة

n: مجموع التكرارات

ملاحظة مجموع التكرارات النسبية يجب ان يساوي 1.

c- Distribution of percentile Frequency (التوزيع التكراري المئوي):

Distribution of percentile Frequency = P*100% التوزيع التكرار المئوي

For the seam example:

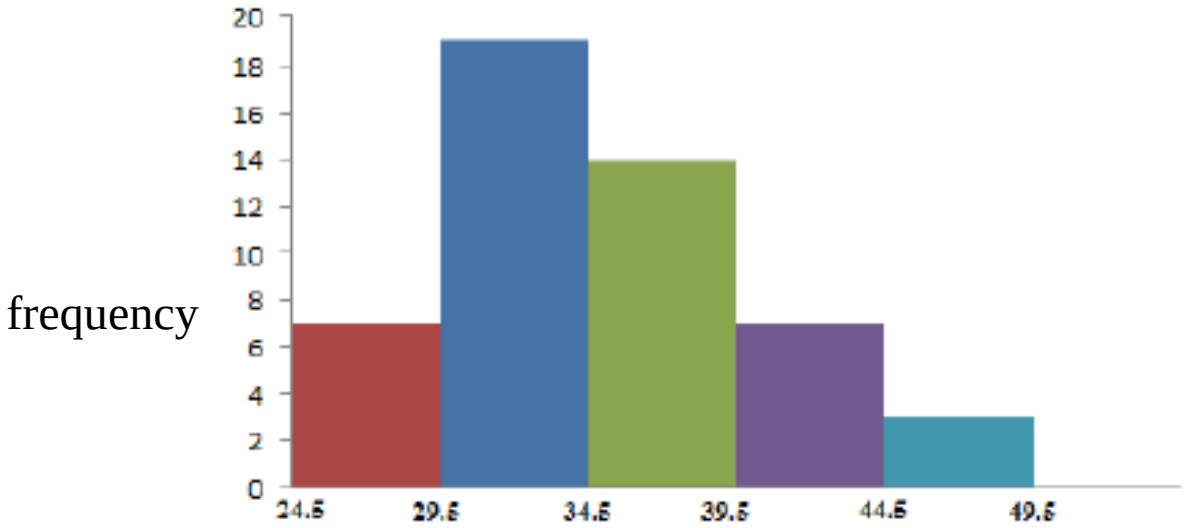
Actual limits for classes	Frequencies (f_i)	Relative Frequency	percentile Frequency
24.5 – 29.5	7	7/50 = 0.14	14
29.5 – 34.5	19	19/50 = 0.38	38
34.5 – 39.5	14	14/50 = 0.28	28
39.5 – 44.5	7	7/50 = 0.14	14
44.5 – 49.5	3	3/50 = 0.06	6
		Σ =1	Σ =100

Graphically represent the frequency distributions

تمثيل التوزيعات التكرارية بيانيا

1- طريقة المدرج التكراري (Frequency Histogram): حيث يتم تمثيل تكرار كل فئة بمستطيل حدود قاعدته يمثل الحدود الفعلية لتلك الفئة و ارتفاعه يتناسب مع تكرارها.

Returning to the previous example, the histogram distribution is represented by the frequency histogram method.



2- المضلع التكراري (Frequency polygon): و هو مضلع مغلق نحصل عليه بتصنيف الاضلاع العلوية للمستطيلات في المدرج التكراري و توصيل نقاط التصنيف مع بعضها و من ثم اغلاق المضلع مع المحور الافقي.

3- المنحني التكراري (Frequency Curve): و يرسم بتمديد خطوط المضلع التكراري لجعله منحنيا، و يستخدم في حالة البيانات الكبيرة من النوع المتصل مثل بيانات الزمن والوزن.

4- المضلع التكراري المتجمع (Accumulated Frequency Polygon): و يرسم بمد خطوط بين قيم التكرار المتجمع للفئات في مناطق نهايات الفئات الحقيقية.

Statistics

Lecture 5

Ex1: The following data shows the educational level of twenty female workers in a factory:

the educational level	uneducated	primary	Secondary 1	Secondary 2
The number	5	8	1	6

Represent the following using the bie-charts method.

Sol:

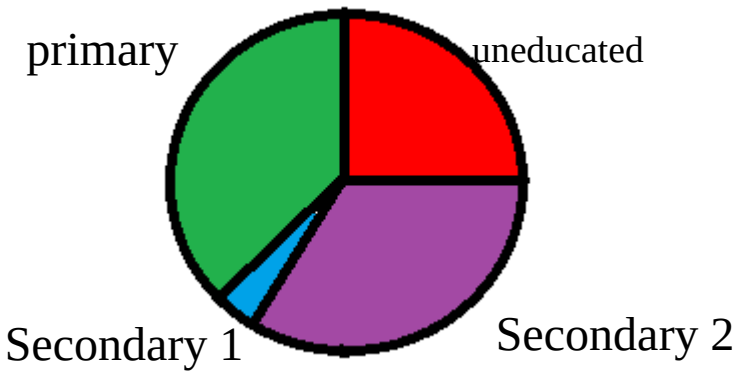
Total summation for team = 5+8+1+6=20

Central angle for the uneducated = (5/20)*360° = 90°

Central angle for the primary = (8/20)*360° = 144°

Central angle for the Secondary 1 = (1/20)*360° = 18°

Central angle for the Secondary 2 = (6/20)*360° = 108°

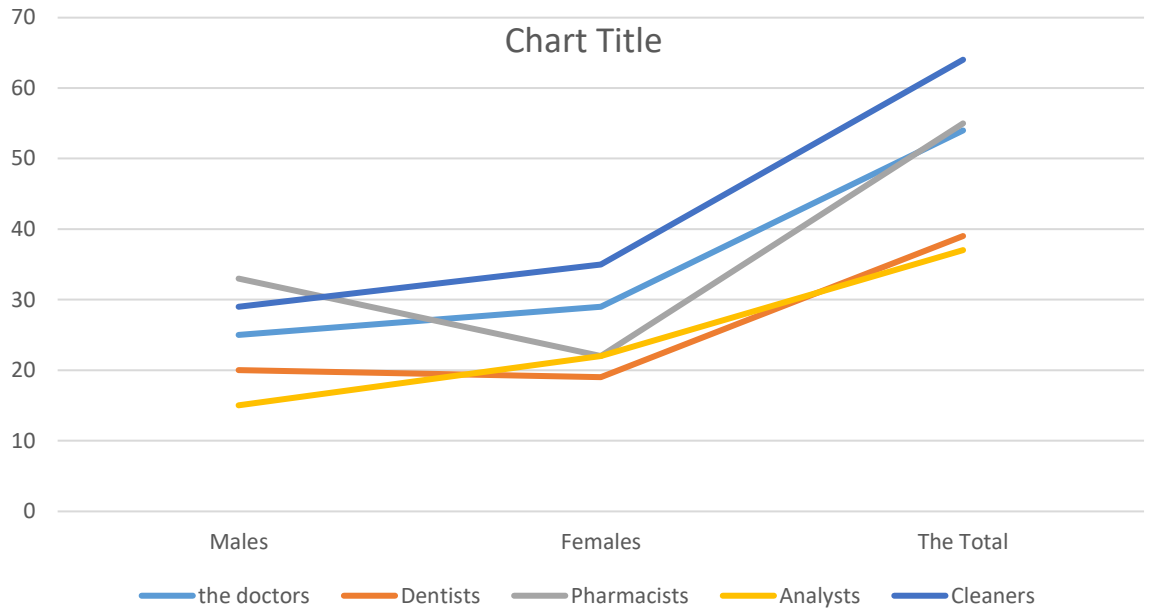


Ex2: The following data represent the number of male and female employees at Al-Hayat Hospital for all categories from doctors to cleaning workers:

Categories	the doctors	Dentists	Pharmacists	Analysts	Cleaners
Males	25	20	33	15	29
Females	29	19	22	22	35
The Total	54	39	55	37	64

Represent the following data in the table in the included line method.

Sol:



Ex3: In the process of building a building with bricks, which is transported from the factory to the work site by vehicles especially. The time required to transport each cargo was measured in minutes, and the results were as follows:

30	18	17	13	24	20	20	16	24	25
19	26	28	23	23	28	17	18	10	18

Arrange the following data in the frequency table with the classes.

Sol:

Assume that the number of classes= 7.

rang of the data = 30 -10 = 20

Class length = $\frac{rang}{\text{Number of classes}} = \frac{20}{7} = 2.85 = 3$

The minimum limit for the first class = 10

The actual minimum for the first class = 10- 0.5 = 9.5

The actual upper limit for the first class = 9.5 + 3 = 12.5

The upper limit for the first class = 12.5 – 0.5 = 12

The lower and upper limits of the rest of the classes are determined by adding the length of the category to each boundary.

Center of classes (X_i) = $\frac{10+12}{2}$ = 11

Center of classes of the rest = Center of the previous category+ Class length

Collect of frequencies (The result must be equal to the original data count).

Class limits	Actual limits for classes	Centers of classes	Frequencies
10 - 12	9.5 – 12.5	11	1
13 – 15	12.5 – 15.5	14	1
16 – 18	15.5 – 18.5	17	6
19 – 21	18.5 – 21.5	20	3
22 – 24	21.5 – 24.5	23	4
25 – 27	24.5 – 27.5	26	2
28 - 30	27.5 – 30.5	19	3
			$\sum f_i = 20$

Ex4: The following data represents the number of customers of a commercial store in a city daily during two consecutive months:

(40 55 47 49 50 48 41 49 45 46 54 51 45 46 48 54 58 62 47 48 47 58 51 49 48 46 44 39 36 43 41 44 43 38 39 61 48 47 43 46 40 44 52 50 60 35 36 56 53 44 39 41 38 45 50 52 55 60 57 49)

Show this data in a frequency distribution table of 6 classes of equal length.

Sol:

Assume that the number of classes= 6.

rang of the data = 62 – 35 = 27

Class length = $\frac{rang}{\text{Number of classes}} = \frac{27}{6} = 4.5 = 5$

The minimum limit for the first class = 35

The actual minimum for the first class = 35- 0.5 = 34.5

The actual upper limit for the first class = 34.5 + 5 = 39.5

The upper limit for the first class = 39.5 – 0.5 = 39

The lower and upper limits of the rest of the classes are determined by adding the length of the category to each boundary.

Center of classes (X_i) = $\frac{35+39}{2}$ = 37

Center of classes of the rest = Center of the previous category+ Class length

Collect of frequencies (The result must be equal to the original data count).

Class limits	Actual limits for classes	Centers of classes	Frequencies
35 - 39	34.5 – 39.5	37	8
40 – 44	39.5 – 44.5	42	12
45 – 49	44.5 – 49.5	47	20
50 – 54	49.5 – 54.5	52	10
55 – 59	54.5 – 59.5	57	6
60 - 64	59.5 – 64.5	62	4
			$\sum f_i = 60$

Ex5: For the example 4:

1- Calculate the up word combined frequency.

2- Calculate the down word combined frequency.

3- Calculate the distribution of percentile Frequency.

Sol:

1-

Actual limits for classes	Frequencies (f_i)	The up word combined frequency
Less than 34.5	0	0
34.5 – 39.5	8	8
39.5 – 44.5	12	20
44.5 – 49.5	20	40
49.5 – 54.5	10	50
54.5 – 59.5	6	56
59.5 – 64.5	4	The total number of data (n) = 60

2-

Actual limits for classes	Frequencies (f_i)	The down word combined frequency
34.5 – 39.5	8	60
39.5 – 44.5	12	52
44.5 – 49.5	20	40
49.5 – 54.5	10	20
54.5 – 59.5	6	10
59.5 – 64.5	4	4
More than 64.5	0	0

3-

Actual limits for classes	Frequencies (f_i)	Relative Frequency	percentile Frequency
34.5 – 39.5	8	(8/60)=0.14	14
39.5 – 44.5	12	(12/60)=0.2	20
44.5 – 49.5	20	(20/60)=0.34	34
49.5 – 54.5	10	(10/60)=0.16	16
54.5 – 59.5	6	(6/60)=0.1	10
59.5 – 64.5	4	(4/60)=0.06	6
		Σ =1	Σ =100

Statistics

Lecture 10

Measures of dispersion

In statistics, the measures of dispersion help to interpret the variability of data i.e. to know how much homogenous or heterogeneous the data is. In simple terms, it shows how squeezed or scattered the variable is.

في الإحصاء ، تساعد مقاييس التشتت في تفسير تباين البيانات ، أي معرفة مقدار البيانات المتجانسة أو غير المتجانسة. بعبارات بسيطة ، يوضح مدى ضغط أو تشتيت المتغير.

Types of measures of dispersion:

1. Range. (المدى)

2. Quartile deviation. (الانحراف الربيعي)

3. Average of the absolute deviations. (متوسط الانحرافات المطلقة)

4. Variance . (التباين)

5. Standard deviation. (الانحراف المعياري)

In addition to **the coefficient of variation** that is used to compare two or more sets of data in terms of Dispersion.

بالإضافة إلى **معامل الاختلاف** الذي يستخدم لمقارنة مجموعتين أو أكثر من البيانات من حيث التشتت.

1. Range. (المدى)

The range is the difference between the largest and smallest value in the data, it is the period that

In whichType equation here. the variable under study changes and is denoted by the symbol (R).

A. Non- classified data:

ويحسب المدى في حالة البيانات غير المبوبة كما يلي :

R = largest value - smallest value

R = اصغر قيمة – اكبر قيمة

B. Classified data:

أما في حالة البيانات المبوبة نعتبر أكبر قيمة هي الحد الأعلى للفئة الأخيرة وأصغر قيمة هي الحد الأدنى للفئة الأولى مع مراعاة أن تكون فئات الجدول مرتبة ترتيباً تصاعدياً ، أي يحسب المدى في حالة البيانات المبوبة كما يلي :

R = The upper limit for the last class - The minimum limit for the first class

R = الحد الادنى للفئة الاولى – الحد الاعلى للفئة الاخيرة

إذا كان المدى صغيراً فيعنى ذلك أن البيانات منتشرة في فترة قصيرة أي قريبة من بعضها وتشتتها صغير ، أما إذا كان المدى كبيراً فيعني ذلك أن البيانات منتشرة في فترة طويلة أي متباعدة عن بعضها وتشتتها كبير.

Ex1: Using the term compare the dispersion of the three groups.

G1: 6, 6, 6, 6, 6.

G2: 4, 5, 6, 7, 8.

G3: 0, 1, 6, 10, 13.

Sol:

R(G1) = 6 – 6 = 0

R(G2) = 8 – 4 = 4

R(G3) = 13 – 0 = 13

بما أن مدى المجموعة الأولى يساوي صفراً ، فيعني ذلك أنه لا يوجد اختلاف أو تشتت بين قيم هذه المجموعة ، أي أن كل القيم داخل هذه المجموعة متساوية.

ونلاحظ أن مدى المجموعة الثانية أصغر من مدى المجموعة الثالثة، وهذا يعني أن تشتت القيم داخل المجموعة الثانية أقل من تشتت القيم داخل المجموعة الثالثة.

H.W1: Compare the dispersion of the statistics subject's scores for three groups of students:

G1: 50, 62, 30, 70, 45, 80.

G2: 0, 20, 70, 55, 65, 82, 35, 77.

G3: 49, 54, 85, 50, 60, 63, 75, 65.

Ex2: Calculate the range for the following data showing weights in kilograms of 100 students:

Weight (kg)	Number of students
60 – 63	5
63 – 66	15
66 – 69	40
69 – 72	28
72 – 75	12

Sol:

R = The upper limit for the last class - The minimum limit for the first class

= 75 – 60 = 15.

Range properties:

1. Range is a measure that is easy to calculate and simple in concept and significance.
2. It is concerned with only two values in the data and neglects the rest.
3. It depends in its calculation on the smallest and largest value only, and therefore it is considered a misleading measure, because it is when the largest or minor values, or both, are outliers, then the range is large while the group values are not divergent, for example if we have the following data:

   55, 14, 11, 9, 8, 12, 10.

   R = 55 – 8 = 47

   فالمدى كبير ، ويشير إلى وجود تشتت كبير في المجموعة في حين أن القيم متقاربة ، ولذلك فهو مقياس مضلل ، وسبب ذلك هو اعتماده على القيم المتطرفة ، فلو حذفنا القيمة المتطرفة وهي القيمة 55 فنجد أن قيمة المدى تساوي :

   R = 14 – 8 = 6

   وهي قيمة صغيرة وواقعية .

2. Variance(التباين)

The variance is defined as the arithmetic mean of the squares of the deviations of values from their arithmetic mean (S^2).

A. Non- classified data:

- نحسب الوسط الحسابي للبيانات.
- نحسب انحراف كل قيمة عن الوسط الحسابي ،

حيث : انحراف القيمة عن الوسط الحسابي = القيمة - الوسط الحسابي

- نوجد مربع انحراف كل قيمة عن الوسط الحسابي .
- نوجد مجموع مربعات انحرفات القيم عن الوسط الحسابي .
- نحسب قيمة التباين باستخدام القانون التالي :

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

Ex3: Calculate the variance for data representing 8 students' grades:

7, 6, 10, 9, 8, 5, 5, 6.

Sol:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \frac{56}{8} = 7$$

Value (X_i)	($\overline{X}$ - X_i)	($\overline{X}$ - X_i)^2
7	0	0
6	-1	1
10	3	9
9	2	4
8	1	1
5	-2	4
5	-2	4
6	-1	1
	Σ	24

$$S^2 = \frac{24}{7} = 3.43$$

B. Classified data:

في حالة البيانات المعروضة في جداول تكرارية نتبع الخطوات التالية :

- نحسب مراكز الفئات.
- نحسب الوسط الحسابي للتوزيع .
- نحسب انحراف كل مركز عن الوسط الحسابي .
- نوجد مربع انحراف كل مركز عن الوسط الحسابي .
- يضرب مربع انحراف كل مركز عن الوسط الحسابي لفئة في تكرار هذه الفئة .
- نجمع حواصل الضرب التي تحصلنا عليها في الخطوة السابقة .
- نحسب قيمة التباين باستخدام القانون التالي :

$$S^2 = \frac{\sum_{i=1}^{k} f_i \left(x_i - \bar{x} \right)^2}{\sum_{i=1}^{k} f_i - 1}$$

حيث : x_i : مركز الفئة ، f_i : تكرار الفئة ، $\bar{x}$: الوسط الحسابي

Ex4: Calculate the variance for the following data, which represent the distribution of 200 workers by overtime spend on the job at the factory every month:

Overtime (h)	Number of Workers
0 – 10	20
10 – 20	80
20 – 30	50
30 – 40	40
40 – 50	10
Σ	200

Sol:

$$S^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{k} f_i - 1}$$

class	f_i	X_i	$X_i f_i$	$(X_i - \bar{x})$	$(X_i - \bar{x})^2$	$f_i(X_i - \bar{x})^2$
0 – 10	20	5	100	-17	289	5780
10 – 20	80	15	1200	-7	49	3920
20 – 30	50	25	1250	3	9	450
30 – 40	40	35	1400	13	169	6760
40 – 50	10	45	450	23	529	5290
Σ	200		4400			22200

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} = \frac{4400}{200} = 22$$

$$S^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{k} f_i - 1}$$

$$= \frac{22200}{200-1} = \frac{22200}{199} = 111.56\ h^2$$

ونلاحظ أن التباين وحداته هي مربع وحدات القياس الأصلية، وكثيراً ما تكون غير ذات معنى فمثلاً في هذا المثال التباين = $111.56\ h^2$ فهنا ساعة تربيع ليس لها أي معنى، وهذا هو عيب التباين ؛ ولذلك نُرجع الوحدات إلى أصلها بأخذ الجذر التربيعي للتباين ، ويسمى المقياس الجديد بالانحراف المعياري .

3. Standard deviation (الانحراف المعياري)

Standard deviation is the measure of dispersion of a set of data from its mean. It measures the absolute variability of a distribution; the higher the dispersion or variability, the greater is the standard deviation and greater will be the magnitude of the deviation of the value from their mean.

It is the positive square root of the mean of the squares of deviations of values from their mean arithmetic, that is, it is the positive square root of the variance, and it is denoted by a symbol (S).

هو الجذر التربيعي الموجب للوسط الحسابي لمربعات انحرافات القيم عن وسطها الحسابي ، أي: هو الجذر التربيعي الموجب للتباين.

A. Non- classified data:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n - 1}}$$

B. Classified data:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{k} f_i (x_i - \overline{x})^2}{\sum_{i=1}^{k} f_i - 1}}$$

For Ex :

$$S = \sqrt{3.43} = 1.85$$

For Ex :

$$S = \sqrt{111.56} = 10.56$$

إذا لم يكن الوسط الحسابي عدداً صحيحاً فإن حساب التباين ومن ثم الانحراف المعياري باستخدام الصيغ السابقة الذكر ، يصبح أمراً غير سهلٍ ، ولذلك اشتقت من الصيغة الأساسية للتباين والتي تعتمد على الانحرافات عن الوسط الحسابي ، صيغة أخرى تعتمد على القيم مباشرة ، وذلك لتسهيل العمليات الحسابية ، وبالطبع الصيغتان تعطيان نفس النتيجة تماماً ، وصيغ التباين التي تعتمد على القيم مباشرة هي :

A. Non- classified data:

$$S^2 = \frac{1}{n-1} \left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i \right)^2}{n} \right]$$

B. classified data:

$$S^2 = \frac{1}{\sum_{i=1}^{k} f_i - 1} \left[\sum_{i=1}^{k} x_i^2 f_i - \frac{\left(\sum_{i=1}^{k} x_i f_i \right)^2}{\sum_{i=1}^{k} f_i} \right]$$

حيث : X_i : مركز الفئة ، f_i : تكرار الفئة

يجب الانتباه بأن هناك فرقاً بين المقدار $\sum_{i=1}^{k} x_i^2 f_i$ والمقدار $\left(\sum_{i=1}^{k} x_i f_i \right)^2$.

Ex5: Calculate the variance and standard deviation of the data mentioned in an example 3, so using the second variance formula (direct values formula).

Sol:

$$S^2 = \frac{1}{n-1} \left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i \right)^2}{n} \right]$$

X_i	7	6	10	9	8	5	5	6	$\Sigma X_i = 56$
X_i^2	49	36	100	81	64	25	25	36	$\Sigma X_i^2 = 416$

$$S^2 = \frac{1}{8-1}\left[416 - \frac{(56)^2}{8}\right]$$

$$= \frac{1}{7}\left[416 - 392\right] = \frac{24}{7} = 3.43$$

$$S = \sqrt{3.43} = 1.85$$

نلاحظ انها نفس النتيجة في المثال السابق.

Ex6: Calculate the variance and standard deviation of the data mentioned in an example 4, so using the second variance formula (direct values formula).

Sol:

$$S^2 = \frac{1}{\sum_{i=1}^{k} f_i - 1}\left[\sum_{i=1}^{k} x_i^2 f_i - \frac{\left(\sum_{i=1}^{k} x_i f_i\right)^2}{\sum_{i=1}^{k} f_i}\right]$$

class	f_i	X_i	$X_i f_i$	X_i^2	$f_i X_i^2$
0 – 10	20	5	100	25	500
10 – 20	80	15	1200	225	18000
20 – 30	50	25	1250	625	31250
30 – 40	40	35	1400	1225	49000
40 – 50	10	45	450	2025	20250
Σ	200		4400		119000

$$S^2 = \frac{1}{200-1}\left[119000 - \frac{(4400)^2}{200}\right]$$

$$= \frac{1}{199}\left[119000 - 96800\right] = 111.56$$

$$S = \sqrt{111.56} = 10.56$$

نلاحظ انها نفس النتيجة التي حصلنا عليها في المثال الرابع.

Properties of Standard Deviation:

- Standard deviation is the most important and widely used measure of dispersion.
- Enter into his calculation all the observed values without neglecting any value.
- It is characterized by its ability to treat algebra.
- It cannot be calculated for open frequency distributions.

Coefficient of variation (معامل الاختلاف):

كل مقاييس التشتت السابقة تعتمد على وحدات القياس، وبالتالي لا يمكن استعمالها لمقارنة تشتت توزيعين مختلفين في وحدات القياس كمقارنة تشتت الأطوال بتشتت الأوزان مثلاً ، ولذلك يجب التعامل مع مقياس نسبي لا يعتمد على الوحدات المستعملة ويسمى هذا المقياس معامل الاختلاف ويحسب كما يلي :

$$\text{Coefficient of variation} = \frac{S}{\bar{x}} * 100$$

$$\text{معامل الاختلاف} = \frac{\text{الانحراف المعياري}}{\text{الوسط الحسابي}} \times 100$$

وبالطبع كلما زاد التشتت أخذ معامل الاختلاف نسبة أكبر .

Ex7: Compare the dispersion of the students' lengths and their weights, using the following data, which represent the lengths and weights 100 students.

Lengths (cm)	Number of students
110 – 120	12
120 – 130	15
130 – 140	25
140 – 150	30
150 – 160	10
160 - 170	8

Weights (Kg)	Number of students
60 – 63	6
63 – 66	16
66 – 69	40
69 – 72	28
72 – 75	10

Sol:

نحسب معامل الاختلاف لكل من التوزيعين ثم نقوم بالمقارنة ، ولحساب معامل الاختلاف لكل توزيع يلزمنا حساب الوسط الحسابي والانحراف المعياري لكل منهما ، وذلك كما يلي :

A. Distribution of lengths:

نوضح العمليات الحسابية التي تلزمنا لحساب الوسط الحسابي والانحراف المعياري في الجدول التالي:

class	f_i	X_i	$X_i f_i$	$(X_i - \bar{x})$	$f_i(X_i - \bar{x})^2$
110 – 120	12	115	1380	-23.5	6627
120 – 130	15	125	1875	-13.5	2733.75
130 – 140	25	135	3375	-3.5	306.25
140 – 150	30	145	4350	6.5	1267.5
150 – 160	10	155	1550	16.5	2722.5
160 - 170	8	165	1320	26.5	5618
Σ	100		13850		19275

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} = \frac{13850}{100} = 138.50$$

$$S^2 = \frac{1}{\sum_{i=1}^{k} f_i - 1}\left[\sum_{i=1}^{k} f_i\,(x_i - \bar{x})^2\right]$$

$$= \frac{1}{100\text{-}1}\ 19275 = 194.7$$

$$S = \sqrt{194.7} = 13.95$$

Coefficient of variation for lengths $= \frac{13.95}{138.5} * 100 = 10.07\%$

B. Distribution of weights:

الجدول التالي يوضح لنا العمليات الحسابية اللازمة للحصول على الوسط الحسابي والانحراف المعياري للاوزان:

class	f_i	X_i	$X_i f_i$	$(X_i - \bar{x})$	$f_i(X_i - \bar{x})^2$
60 – 63	6	61.5	369	-6.6	261.36
63 – 66	16	64.5	1032	-3.6	207.36
66 – 69	40	67.5	2700	-0.6	14.4
69 – 72	28	70.5	1974	2.4	161.28
72 – 75	10	73.5	735	5.4	291.6
Σ	100		6810		936

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i} = 68.1$$

$$S^2 = \frac{1}{\sum_{i=1}^{k} f_i - 1}\left[\sum_{i=1}^{k} f_i (x_i - \bar{x})^2\right]$$

$$= \frac{936}{99} = 9.45$$

$$S = \sqrt{9.45} = 3.07$$

Coefficient of variation for weights $= \frac{3.07}{68.1} * 100 = 4.51\%$

بما أن معامل الاختلاف للأطوال أكبر من معامل الاختلاف للأوزان ، إذن تشتت الأطوال أكبر من تشتت الاوزان.

H.W 2: If we had two sets of data:

set 1: 21, 15, 23, 16, 14, 12, 25.

set 2: 30, 42, 64, 60, 28, 70, 82, 56.

A- For each of these two groups, calculate the following:

Range, Variance, Standard deviation.

B- Compare the dispersion of the two groups using the coefficient of variation.

H.W 3: If you know the coefficient of variation = 25% and that the Variance = 25, that for the following data:

10, 18, 25, 15, X. Calculate the missing value X.

Statistics

Lecture 12

Probability Distribution

التوزيعات الاحتمالية

The benefit of studying probability distributions from the valuable implementation success is in a major field:

1. The possibility of determining or calculating a probability in certain evidence circumstances.
2. Appropriateness of the probability distribution of some of the data collected in the field. (Therefore, it is possible to extract information from this data after making a comparison with these distributions are compatible with it).

1. امكانية تحديد او احتساب احتمال ما في ظروف بيانات معينة.
2. امكانية ملائمة التوزيع الاحتمالي لبعض البيانات التي تم جمعها ميدانيا. وبالتالي امكانية استخراج المعلومات من هذه البيانات بعد عمل مقارنة مع هذه التوزيعات المتلائمة معها.

المتغيرات العشوائية Random variables

إن بعض الكميات تكون قيمتها العددية غير محددة تحديداً نهائياً، وانما تتغير بتغير عوامل عشوائية ومن الامثلة على هذه الكميات اختلاف مقاومات الانضغاط للمكعبات المأخوذة من نفس الخلطة ومعالجتها بنفس الظروف او مقاومات الشد لقضبان حديد التسليح المصنوعة من نفس الفولاذ وبنفس الظروف. إن هذا الاختلاف بالرغم من تساوي فرص الحصول على توافق ظاهرياً هو بسبب ظاهرة التشتت وسببها عوامل عشوائية لم يكن بالإمكان حسابها او التنبؤ بها مسبقاً. لذا يمكن تعريف **الكميات العشوائية**: وهي الكميات التي تأخذ قيماً مختلفة في العمليات أو المحاولات المختلفة وتبقى مختلفة مهما حاولنا جعل ظروف وشروط حدوثها متجانسة تبعاً لتغيرات عشوائية خارجة عن الحسبان. إن اهم ما يجب معلوميته بما يخص المتغيرات العشوائية هو جميع القيم العددية التي يمكن أن تأخذها تلك المتغيرات، بالإضافة الى معرفة الاحتمالات المقابلة لتلك القيم لحالة المتغيرات المدروسة. إن معلومية احتمال القيم الممكنة للكميات العشوائية المدروسة يكفي لحل أي سؤال يرتبط بهذه الكمية العشوائية وتبيان فعاليتها. وتقسم المتغيرات العشوائية الى نوعين:

أ- المتغيرات العشوائية المتقطعة او المنفصلة : ويكون فيها المتغير العشوائي يأخذ أي قيمة من قيم مجموعة متقطعة أو قابلة للعد مثل متغير عدد الطلبة الذكور في احدى المدارس.

ب- المتغيرات العشوائية المتصلة : وهي المتغيرات العشوائية التي من الممكن أن تأخذ أي قيمة داخل مدى (مجال) معين من القيم مثل مقاييس درجة الحرارة أو الوزن أو الطول.

Depending on the type of the random variable, the probability distributions can be divided into two types:

بالاعتماد على انواع المتغير العشوائي ، يمكن تقسيم التوزيعات الاحتمالية الى نوعين:

A. Discrete Probability Distributions ((منفصلة) توزيعات احتمالية متقطعة)

- Binomial Dist. (توزيع ذي الحدين)
- Poisson Dist.(توزيع بواسون)

B. Continuous Probability Distributions (توزيعات احتمالية متصلة)

- Normal Distribution (التوزيع الطبيعي)
- T-Distribution (توزيع تي)
- Chi-Square Distribution (توزيع كاي تربيع)

A. Discrete Probability Distributions ((توزيعات احتمالية متقطعة (منفصلة)

وهي التوزيعات الاحتمالية (دوال كتلة الاحتمال) لمتغيرات عشوائية متقطعة.

تعريف دالة كتلة الاحتمال للمتغير العشوائي المتقطع: إذا كان X متغيراً عشوائياً متقطعاً بحيث إن مجموعة القيم الممكنة له هي

$X(S)=\{X_1, X_2, \ldots\ldots, X_n\}$

فإن دالة كتلة الاحتمال ويرمز لها f(X) تأخذ القيم التالية:

$$f(X) = \begin{Bmatrix} P(X = X_i); & X_i \in X(S) \\ 0 \quad\quad ; & X_i \notin X(S) \end{Bmatrix}$$

ويجب ان تحقق دالة كتلة الاحتمال الشروط التالية:

i. ان قيمة دالة كثافة الاحتمال عند قيمة X ما يجب أن لا تزيد عن 1 ولا تقل عن صفر.
ii. ان مجموع قيم دوال كثافة الاحتمال لكل قيم المتغير X المتوقعة (ضمن القيم الممكنة ل X ان تأخذها أو فضاء العينة) يجب أن يساوي "1".

- **Binomial Dist. (توزيع ذي الحدين)**

وهو من اهم التوزيعات المتقطعة المستعملة في مجالات التطبيقات الهندسية لمتغيرات عشوائية منفصلة. حيث يمكن الحكم على نتائج تجربة ما ب "نعم" أو "لا". ويستخدم عند تحقيق الشروط التالية:

- ان نتيجة أي محاولة لتجربة ما تكون اما النجاح أو الفشل.
- كل محاولة نتيجتها مستقلة عن نتيجة المحاولة الاخرى. (أي لا تؤثر عليها مثل عملية السحب بإرجاع).
- احتمال النجاح في كل محاولة ثابت وليكن (P) وبالتالي فإن احتمال الفشل ثابت أيضاً وهو (q) ويحسب من (q=1-P).
- تجري التجربة بعدد معين من المحاولات أو المرات أو بمعنى آخر (يكون هنالك عدد n من المحاولات).

وعند اجراء تجربة ما وكان عدد المحاولات =n ، عدد النجاحات أو عدد مرات حصول الحادثة =x ، احتمال النجاح للتوزيع = P ، واحتمال الفشل للتوزيع =q فإن الاقتران الاحتمالي لمتغير ذي الحدين ويرمز له P(x;n,p) ويحسب كما يلي:

$$P(x;n,p) = C_x^n . p^x . q^{(n-x)} \qquad ; X=0,1,2,.....n$$

ملاحظة: تم أخذ قيمة الصفر لأنه يمثل احتمال عدم الحصول او احتمال ولا مرة يحصل حادث النجاح.

ملاحظة: اذا كان X متغيراً ذا حدين (خاضع لنظرية ذات الحدين) فإن:

i. الوسط او التوقع لقيمة X الممكنة هو $M = n.p$
ii. التباين لقيمة X هو $S^2 = n.p.q$

مثال(1): رُميت زهرة نرد منتظمة 5 مرات ، ما احتمال عدم ظهور الرقم 6 فيها؟ ما احتمال ظهور الرقم 6 ثلاث مرات؟

الحل: ان فضاء العينة لتجربة رمي زهرة نرد واحدة منتظمة لمرة واحدة هو $s = \{1,2,3,4,5,6\}$

- احتمال عدم ظهور الرقم 6 $q(6) = 1 - \frac{1}{6} = \frac{5}{6}$ $\rightarrow$ احتمال ظهور الرقم 6 $P(6) = \frac{1}{6}$

- $P(6) = \frac{1}{6}$, $q(6) = \frac{5}{6}$, n=5 , x=0,1,2,3,4,5

وهي تمثل الصيغة العامة لهذه التجربة $P\left(x;5,\frac{1}{6}\right) = C_x^5.\left(\frac{1}{6}\right)^x.\left(\frac{5}{6}\right)^{(5-x)}$, x=0,1,2,3,4,5

- احتمال عدم ظهور الرقم 6 $P\left(x=0;5,\frac{1}{6}\right) = C_0^5.\left(\frac{1}{6}\right)^0.\left(\frac{5}{6}\right)^{(5-0)} = 1 \times 1 \times \left(\frac{5}{6}\right)^5 = 0.40187$

- احتمال عدم ظهور الرقم 6 ثلاث مرات $P\left(x=3;5,\frac{1}{6}\right) = C_3^5.\left(\frac{1}{6}\right)^3.\left(\frac{5}{6}\right)^{(5-3)} = 10 \times \frac{1}{216} \times \left(\frac{25}{36}\right) = 0.03215$

ولحساب القيمة المتوقعة للمتغير X نحسب الوسط $M = n.p = 5.\frac{1}{6} = \frac{5}{6}$

ولحساب قيمة التباين المتوقعة للمتغير X نحسب التباين $S^2 = n.p.q = 5.\frac{1}{6}.\frac{5}{6} = \frac{25}{36}$

مثال(2): يُراد توزيع عدد من الكراسي الخاصة بالطلبة الأعسرين (مستخدمي اليد اليسرى) في كل قاعة دراسية ، فإذا عُلم ان 10% من جميع الطلبة هم أعسرين فما هو احتمال إن صفاً من 20 طالب: (أ)- لا يحتوي على أي طالب أعسر؟ (ب)- اثنان منهم أعسران؟ (ج)- أكثر من اثنين منهم أعسرين؟

الحل: نفرض ان المتغير X يمثل عدد الطلبة الأعسرين n=20 , x=0,1,2,…….,20

P(L)=0.1 احتمال أن يكون الطالب أعسر ، q(L)=1 - 0.1=0.9 احتمال أن لا يكون الطالب أعسر ،

من المعلومات أعلاه نلاحظ تحقق شروط نظرية ذي الحدين، لذا فإن الصيغة العامة للتجربة هي

$P(x;20,0.1) = C_x^{20}.(0.1)^x.(0.9)^{(20-x)}$, x=0,1,2,…..,20

(أ)- احتمال أن لا يحتوي الصف ولا طالب أعسر $P(x=0;20,0.1) = C_0^{20}.(0.1)^0.(0.9)^{(20-0)} = 1 \times 1 \times 0.12157 = 0.12157$

(ب)- احتمال أن يحتوي الصف طالبين أعسرين $P(x=2;20,0.1) = C_2^{20}.(0.1)^2.(0.9)^{(20-2)} = 190 \times 0.01 \times 0.15009$

=0.28518

- **Poisson Dist.(توزيع بواسون)**

يعتبر توزيع بواسون من التوزيعات الاحتمالية المتقطعة التي تستخدم للحوادث نادرة الحصول أو الوقوع، أو المتغيرات التي تحدث في أزمنة عشوائية معلومة. مثل عدد الأخطاء المطبعية في صفحة كتاب ما، عدد الزلازل السنوية الحاصلة في موقع ما، حدوث فيضان بتصريف معين في نهر ما...الخ.

وتسمى التجارب التي تعطي عدد نجاحات معين في فترة زمنية أو فترة قياس كالطول أو الحجم أو المساحة...الخ بالتجربة البواسونية، وهي تجربة احصائية تمتلك الخصائص التالية:

أ- نتائج التجربة مُعرفة بالنجاح أو الفشل والتجارب مستقلة الواحدة عن الاخرى.

ب- الاحتمال ثابت وذو قيمة صغيرة بشكل عام (تقترب من 0) وإن عدد المحاولات n يكون كبيراً عادةً (يقترب من ∞) ولكن المهم ان مقدار حاصل الضرب بينهما يكون مقدار ثابت يساوي المتوسط أو المعدل (λ) بحيث أن ($\lambda > 0$). $\lambda = n.P$

ومن خصائص توزيع بواسون الاحتمالي:

- المعدل أو المتوسط للتوزيع ($\lambda = n.P$) ، والتباين يساوي (λ)
- الانحراف المعياري يساوي ($\sqrt{\lambda}$)

ان الاقتران الاحتمالي لتوزيع بواسون هو: $P(x,\lambda) = \frac{e^{-\lambda}.\lambda^x}{x!}$, x=0,1,2,.....

<u>حيث ان:</u>

λ: معدل عدد النجاحات (حصول الحدث) ضمن الفترة الزمنية أو فترة القياس (الطول، المساحة، الحجم) وغيرها.

x: عدد النجاحات الفعلي أو عدد مرات القياس المطلوبة وهي تمثل متغير بواسون العشوائي.

e: أساس اللوغاريتم الطبيعي وتساوي تقريباً (e=2.718281828).

مثال(1): معدل عدد حوادث السيارات على طريق صحراوي ما هو 5 (حادث\أسبوع).(أ) ما احتمال عدم حدوث أي حادث على ذلك الطريق في اسبوع معين؟ (ب)ما احتمال حدوث 4 حوادث او أقل في اسبوع معين؟(ج) ما احتمال أن يزيد عدد الحوادث عن حادثين خلال أسبوعين؟

الحل: $P(x,\lambda) = \frac{e^{-\lambda}.\lambda^x}{x!}$, x=0,1,2,..... , λ=5

الصيغة الاساسية لحل المسألة $P(x,\lambda=5) = \frac{e^{-5}.5^x}{x!}$

(أ) عدم حصول أي حادث $P(x=0,\lambda=5) = \frac{e^{-5}.5^0}{0!} = 0.00674$

(ب) حدوث 4 حوادث أو أقل $P(x\leq 4,\lambda=5) = P(0,5)+P(1,5)+P(2,5)+P(3,5)+P(4,5)$

$$= 0.00674 + \frac{e^{-5}.5^1}{1!} + \frac{e^{-5}.5^2}{2!} + \frac{e^{-5}.5^3}{3!} + \frac{e^{-5}.5^4}{4!} = 0.4405$$

(ت) زيادة عدد الحوادث عن 2 خلال إسبوعين (هنا نلاحظ ان معدل حصول الحوادث اختلف بدلاً من الفترة اسبوع الى اسبوعين فهنا ستتغير الصيغة الاساسية للسؤال حسب المعدل الجديد).

$$\lambda = 5(\text{حادث\اسبوع}) \times 2 = 10(\text{حادث\اسبوعين})$$

$P(x,\lambda=10) = \frac{e^{-10}.10^x}{x!}$, $x = 0,1,2,\ldots\ldots\ldots$

$$P(x>2,\lambda=10) = 1-\{P(0,10)+P(1,10)+P(2,10)\} = 1-\left\{\frac{e^{-10}.10^0}{0!}+\frac{e^{-10}.10^1}{1!}+\frac{e^{-10}.10^2}{2!}\right\} = 0.99723$$

Approximation of the binomial distribution of the Poisson distribution

تقريب توزيع ثنائي الحدين لتوزيع بواسون

اذا اعتبرنا n او عدد المحاولات كبيرة بشكل كافٍ بحيث (n≥50) و احتمال النجاح P صغير (P<0.1) بحيث يبقى حاصل ضرب قيمتيهما (n.P) مقدار ثابت وليكن λ، عندها يمكن تقريب توزيع ذي الحدين الى توزيع بواسون بالمقدار (λ=n.P).

مثال (2): في اختبار مُضاف مُحسن للخلطات الخرسانية، اذا افترضنا انه ضار للخلطة باحتمال 0.001 وانه تم استخدامه فعليا في انتاج 2000 خلطة خرسانية، استخدم تقريب بواسون للإجابة على: (أ) ما احتمال ان تتضرر أكثر من خلطتين نتيجة استخدام ذلك المُضاف؟ (ب) ما احتمال تضرر خلطتان على الأقل؟

Solution: n=2000 , P=0.001 (احتمال حصول الضرر)

نلاحظ إن عدد المحاولات كبير جداً (n>50) واحتمال حصول الحدث P أقل من 0.1 لهذا يمكن استخدام تقريب بواسون بذي الحدين أو بمعنى آخر

$\lambda = n.P = 0.001 \times 2000 = 2$

(أ) $P(x > 2, \lambda = 2) = 1 - \{P(0,2) + P(1,2) + P(2,2)\} = 1 - \left\{\frac{e^{-2}.2^0}{0!} + \frac{e^{-2}.2^1}{1!} + \frac{e^{-2}.2^2}{2!}\right\} = 0.32332$

(ب) $P(x \geq 2, \lambda = 2) = 1 - P(x < 2, \lambda = 2) = 1 - \{P(0,2) + P(1,2)\} = 0.59399$

Not:

الفرق الأساسي بين توزيعي ذي الحدين وتوزيع بواسون هو إن n أ و عدد التجارب الكلي في توزيع ذي الحدين يكون معلوماً ولهذا فإن قيم المتغير العشوائي x تكون (x=0,1,2,.....,n) أما في توزيع بواسون فإن n أو عدد التجارب غير معلوم أو مبهم وفي هذه الحالة فإن المتغير العشوائي x يأخذ المدى (x=0,1,2,.....). ولهذا في حالة إحتساب احتمالات أكبر من قيمة ما و لتكن a في توزيع بواسون فإننا نضطر طرح احتمال تلك القيمة والتي اقل منها من (1) وهو الأحتمال الكلي لفضاء العينة أو بمعنى آخر $P(x > a, \lambda) = 1 - P(x \leq a, \lambda)$ وذلك لأننا لا نعرف إلى أي حد سينتهي إليه المتغير العشوائي x.

مثال(2): اذا كان 3% من انتاج أحد معامل الطابوق غير مطابق للمواصفات القياسية وقد تم سحب عينة من الانتاج بشكل عشوائي. ما هو احتمال بأن العينة تحتوي بالضبط على طابوقتين فاشلتين؟ استخدم توزيعين مختلفين للحل ثم قارن النتيجتين؟

الحل: (أ) باستخدام توزيع ذي الحدين (P=0.03 , q=0.97 , n=100)

$$\therefore P(x; 100,0.03) = C_x^{100}.(0.03)^x.(0.97)^{100-x} \qquad x = 0,1,\dots\dots,100$$

$$\therefore P(x = 2; 100,0.03) = \frac{100!}{2! \times (100-2)!} \times (0.03)^2 \times (0.97)^{100-2} = 4950 \times 0.0009 \times 0.0505 = 0.225153$$

(ت) باستخدام توزيع بواسون ($\lambda = n.P = 0.03 \times 100 = 3$)

$$P(x, \lambda = 3) = \frac{e^{-3}.3^x}{x!} \quad , \qquad x = 0,1,2,\dots\dots$$

$$\therefore P(x = 2, \lambda = 3) = \frac{e^{-3}.3^2}{2!} = 0.22404$$

مثال(3): اذا كانت احدى البدالات تتلقى مكالمات هاتفية بمعدل مكالمتين لكل دقيقة. أوجد احتمال أن تتلقى البدالة 3 مكالمات على الأقل؟

Solution: (λ=2 call/min)

$$P(x,2) = \frac{e^{-2}.2^{x}}{x!} \quad , \qquad x = 0,1,2,\dots\dots$$

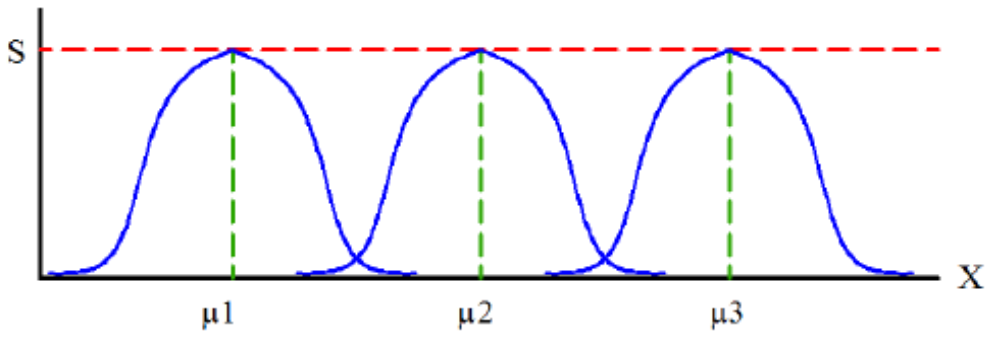
$$P(x \geq 3, \lambda = 2) = 1 - \left\{\frac{e^{-2}.2^{0}}{0!} + \frac{e^{-2}.2^{1}}{1!} + \frac{e^{-2}.2^{2}}{2!}\right\} = 1 - (0.67667) = 0.3233$$

التوزيعات الاحتمالية المتصلة (Continuous Probability Distributions)

1-1- التوزيع الطبيعي (Normal Distribution):

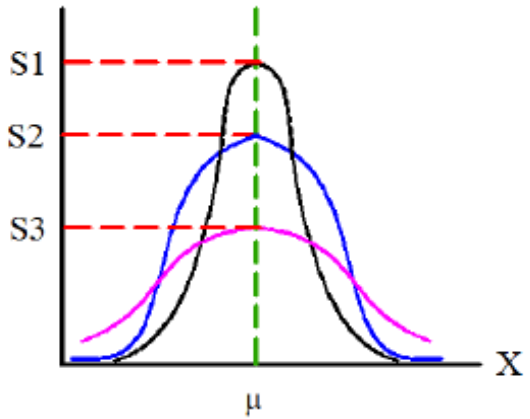
وهو من أشهر التوزيعات الاحتمالية المتصلة وأكثرها استخداماً وتطبيقاً ومن الأمثلة عليه التوزيعات البايومترية (توزيعات الطول والوزن) وتوزيعات اخطاء المشاهدات (الفروق بين القيم الحقيقية والقيم المُشاهدة) وتم تطوير هذا التوزيع من قبل العالم گاوس سنة 1809 ولهذا يطلق عليه في بعض الأحيان اسم (توزيع گاوس).

ان منحني التوزيع الطبيعي الذي يأخذ شكل الجرس هو منحني توزيع نظري لدالة كثافة (كتلة الاحتمال) للمتغير العشوائي الذي يخضع لهذا التوزيع. هنالك ما لانهاية من منحنيات التوزيع الطبيعي التي قد تختلف عن بعضها البعض حسب قيمة كل من الوسط الحسابي (μ) والانحراف المعياري (S)، فمثلاً قد تتفق المنحنيات بالانحراف المعياري لكنها تختلف بالوسط الحسابي كما في الشكل (1) أو بالعكس قد تختلف بالانحراف المعياري وتتفق بالوسط كما في الشكل (2).



شـكل (1)

منحنيـــات توزيـــع طبيعـــي لها نفـــس الانحـراف المعيـــاري مع اختـــلاف الوســـط الحســـابي



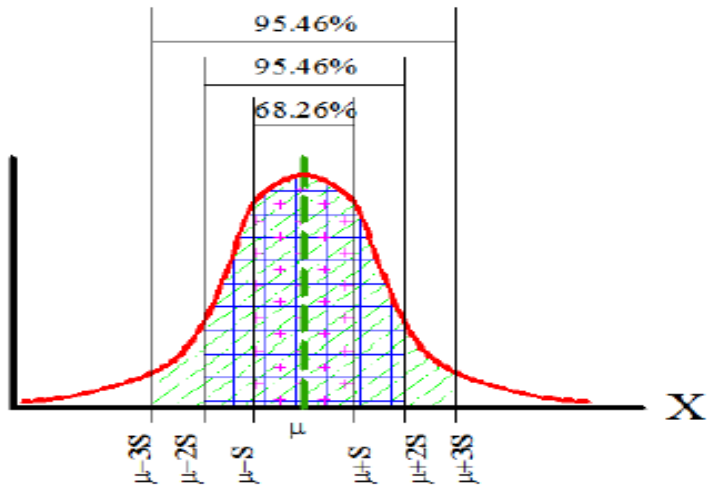
شـكل (2)

منحنيـــات توزيـــع طبيعـــي لها نفـــس الوســـط الحســـابي مع اختـــلاف الانحـــراف المعيـــاري

خصائص التوزيع الطبيعي:

- منحني التوزيع الطبيعي متماثل حول الوسط μ (إلتوائه يساوي صفراً) ، وشكله بشكل الجرس.
- له قيمة منوالية (منوال) واحد ينطبق على الوسط μ ويقترب طرفاه من قيمتي (-∞ ، ∞) دون أن تقطع المحور الافقي.
- قيمة الانحراف المعياري تحدد عُرض المنحني الطبيعي، مثلاً قيمة انحراف معياري أعلى تنتج منحني أعرض وأكثر تسطحاً.
- الأحتمالات للمتغير الطبيعي العشوائي تُعطى بالمساحة تحت منحني التوزيع الطبيعي. والمساحة الكلية تحت منحني التوزيع الطبيعي للفترة (-∞≤x≤+∞) تساوي (1) و تتوزع مناصفةً 0.5 للنصف الأيمن و 0.5 للنصف الأيسر.
- قيمة الاحتمال عند قيمة مفردة محددة للمتغير العشوائي الطبيعي x تساوي (صفر) ولأي قيمة وذلك لان المساحة تحت المنحني والمحددة بقيمة واحدة تكون عبارة عن خط مستقيم ولهذا فلحساب الاحتمال يجب ان تحدد فترة ما للمتغير العشوائي ومن خلالها تحسب المساحة تحت المنحني ضمن هذه الفترة والتي سوف تساوي قيمة الاحتمال ضمن تلك الفترة.

$P(x=0) = 0$, $P(x=1) = 0$, $P(x=100) = 0$, $P(0<x<a) = ?$, $P(x>a) = ?$



المساحة تحت المنحني الطبيعي تتوزع كما يلي (ثابتة لجميع منحنيات التوزيع الطبيعي)

1- 68.26% من المساحة تحت المنحني تكون ضمن حدود (μ-S≤x≤μ+S)
2- 95.46% من المساحة تحت المنحني تكون ضمن حدود (μ-2S≤x≤μ+2S)
3- 99.74% من المساحة تحت المنحني تكون ضمن حدود (μ-3S≤x≤μ+3S)
4- 95% من المساحة تحت المنحني تكون ضمن حدود (μ-1.96S≤x≤μ+1.96S)
5- 99% من المساحة تحت المنحني تكون ضمن حدود (μ-2.58S≤x≤μ+2.58S)

- يُقال للمتغير العشوائي المتصل x بأنه موزع طبيعياً اذا كانت دالة كثافة (كتلة) الاحتمال له

مُعطاة بالصيغة :

$$F(x) = \frac{1}{S\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{S}\right)^2}$$, للفترة (-∞≤x≤+∞)

ويرمز لهذا التوزيع بالرمز التالي $N(\mu,S^2)$

التوزيع الطبيعي المعيارى (القياسي):

بسبب ان شكل منحني التوزيع الطبيعي يعتمد على قيمتي الوسط μ والانحراف المعياري S فلهذا هنالك مالانهاية من المنحنيات الخاصة بذلك النوع من التوزيع وبالتالي فان قيم الاحتمالات المعتمدة على قيمة المساحة المحصورة تحت هذه المنحنيات سوف تتغير تبعاً لتلك القيم وبالتالي فلأغراض المقارنة بين قيم الاحتمالات للتوزيعات الطبيعية تم إقتراح توزيع طبيعي معياري موحد يمكنه التعبير عن جميع منحنيات التوزيعات الطبيعية: وهو التوزيع الطبيعي الذي وسطه (μ=0) وانحرافه المعياري (S=1) وله دالة كثافة (كتلة) احتمال معرفة كما يلي:

$$F(z) = \frac{1}{1\times\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-0}{1}\right)^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} \quad , \quad \text{للفترة } (-\infty \le z \le +\infty)$$

هنا لاستخراج المساحة تحت منحني التوزيع الطبيعي المعياري لاستخراج الاحتمال ضمن الفترة المطلوبة والتي سوف تمثل حدود التكامل باستخدام صيغة التكامل التالية:

$$P(-\infty \le z \le +\infty) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} . dz$$

أما لحساب الاحتمال لفترة ما فيتم بتبديل حدود التكامل أعلاه بقيم تلك الفترة فنستخرج قيمة المساحة تحت المنحني ضمن تلك الفترة والتي تساوي قيمة الاحتمال لتلك الفترة.

ملاحظة: لتسهيل الحسابات والعمل بهذا النوع من التوزيع يتم استبدال التكامل بجداول خاصة بالتوزيع الطبيعي المعياري واعتماداً على قيمة المتغير الطبيعي العشوائي (z) ، وعلى قيم الحدود للفترة المطلوب حساب الاحتمال لها.

ملاحظة: للتحويل من حالة التوزيع الطبيعي الى حالة التوزيع الطبيعي المعياري نقوم بتحويل المتغير العشوائي المعياري x الى المتغير العشوائي z من خلال العلاقة التالية:

$$z = \frac{x-\mu}{s}$$

مثال (1): إذا كان x متغير عشوائي طبيعي خاضع للتوزيع N(3,4) فما هي قيمة الاحتمال عند وقوع x بين القيمتين 3 و5 ؟

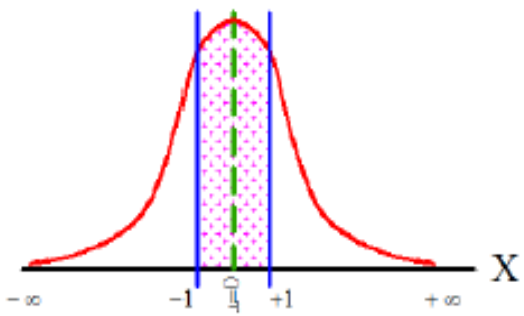
Solution: $z = \frac{x-\mu}{S}$, $z_1 = \frac{3-3}{\sqrt{4}} = 0$, $z_2 = \frac{5-3}{\sqrt{4}} = 1$ وبتطبيق هذه الحدود تصبح المسألة → P(0≤z≤1)=?

ومن الجداول الخاصة بالتوزيع الطبيعي المعياري → P(0≤z≤1)=0.3413

مثال (2): إذا كان المتغير العشوائي الطبيعي x خاضع للتوزيع $N(\mu, S^2)$ ، إحسب كلٍ من $P(\mu-S\leq x \leq \mu+S)$ ، $P(\mu-2S\leq x \leq \mu+2S)$ و $P(\mu-3S\leq x \leq \mu+3S)$ ؟

Solution:

(1) For $P(\mu-S\leq x \leq \mu+S)$, , $z_1 = \frac{(\mu-S)-\mu}{S} = -1$, $z_2 = \frac{(\mu+S)-\mu}{S} = 1$, $\rightarrow P(-1 \leq z \leq 1) = 2 \times P(0 \leq z \leq 1) = 2 \times 0.3413 = 0.6826 = 68.26\%$ (باستخدام الجداول)



(2) و (3) H.W

ملاحظات لاستخدام الجداول:

1. في حالة وقوع قيمة المتغير العشوائي بين قيمتين في الجدول فيتم استخدام التقريب الخطي وتشابه المثلثات لإيجاد قيمة الاحتمال الصحيح.
2. يجب قبل استخدام الجدول ملاحظة الرسم الموجود في أعلى الصفحة وملاحظة المساحة المظللة تحت المنحني فهي تمثل حدود التكامل التي على غرارها أو وفقاً لها تم استخراج قيم الجدول.
3. في حالة كانت قيمة المتغير z خارج قيم الجدول أكبر أو أقل فيتم أعتماد طريقة التكامل لاستخراج المساحة تحت المنحني وبالتالي قيمة الاحتمال المطلوبة.

مثال: للتوزيع الطبيعي المعياري $Z\sim N(0,1)$ أوجد قيمة K ، اذا كانت:

1) $P(Z \leq K) = 0.5$
2) $P(Z \geq K) = 0.2451$
3) $P(Z \leq K) = 0.8289$
4) $P(-K \leq Z \leq K) = 0.901$
5) $P(0 \leq Z \leq K) = 0.492$
6) $P(Z > K) = 0.0548$
7) $P(|Z| < K) = 0.901$
8) $P(|Z| > K) = 0.099$
9) $P(-1.24 < Z < K) = 0.8$

الحل:

1. $P(Z \leq K) = 0.5$:

نلاحظ هنا بخلاف المثال السابق، القيمة المعطاة هي قيمة المساحة تحت المنحنى بينما المجهول هو إيجاد القيمة على المحور الأفقي، وبما أن جدول Z متماثل حول الصفر والمساحة الكلية تساوي الواحد فهذا يعني أن المساحة يمين الصفر تساوي 0.5 ويسار الصفر تساوي 0.5 ، أي أن قيمة K تساوي صفر.

2. $P(Z \geq K) = 0.2451$

نلاحظ أن قيمة المساحة أصغر من 0.5، وعلامة الاحتمال أكبر من ، أي أن قيمة K ستكون موجبة :

$$P(Z \geq K) = 0.2451 \rightarrow 1 - P(Z < K) = 0.2451 \rightarrow P(Z < K) = 0.7549$$

ومن خلال الجدول نجد أنه عند المساحة 0.7549 قيمة K تساوي 0.69 .

3. $P(Z \leq K) = 0.8289$

نلاحظ أن قيمة المساحة أكبر من 0.5، وعلامة الاحتمال أصغر من ، أي أن قيمة K ستكون موجبة ومن خلال الجدول مباشرة نجد أن قيمة K عند المساحة 0.8289 تساوي 0.95 .

4. $P(-K \leq Z \leq K) = 0.901$

$$P(-K \leq Z \leq K) = P(-K < Z < 0) + P(0 < Z < K)$$

من خلال خاصية التماثل في جدول التوزيع الطبيعي المعياري، نجد أن:

$$P(-K < Z < 0) = P(0 < Z < K)$$

وبذلك يكون لكل منهما نصف المساحة أي أن :

$$P(0 \leq Z \leq K) = 0.4505$$

$$P(0 \leq Z \leq K) = P(Z \leq K) - P(Z \leq 0) = 0.4505$$

$$P(Z \leq K) = P(Z \leq 0) + 0.4505 = 0.5 + 0.4505 = 0.9505$$

من خلال جدول التوزيع وعند المساحة 0.9505 نجد أن K تساوي 1.65

5. $P(0 \leq Z \leq K) = 0.492$

بنفس الطريقة السابقة: $P(0 \leq Z \leq K) = P(Z \leq K) - P(Z \leq 0) = 0.492$

$$P(Z \leq K) = 0.492 + 0.5 = 0.992$$

من خلال الجدول ، نجد أن قيمة K تساوي 2.41 .

6. $P(Z > K) = 0.0548$

$$P(Z > K) = 1 - P(Z < K) = 0.0548 \rightarrow P(Z < K) = 1 - 0.0548 = 0.9452$$

من خلال الجدول نجد أن قيمة K تساوي 1.6 .

7. $P(|Z| < K) = 0.901$

لاحظ أن علامة الاحتمال أصغر من وبالتالي نستطيع تحويل القيمة المطلقة الى:

$$P(-K \leq Z \leq K) = 0.901$$

وكما في الفقرة 4 تكون قيمة 1.65 K .

8. $P(|Z| > K) = 0.099$: لاحظ أن علامة الاحتمال أكبر من أي عكس الفقرة السابقة ويتم التعامل معها كالتالي:

$$P(|Z| > K) = P(Z > K) + P(Z < -K) = 1 - P(-K \leq Z \leq K) = 0.099$$

$$\rightarrow P(-K \leq Z \leq K) = 1 - 0.099 = 0.901$$

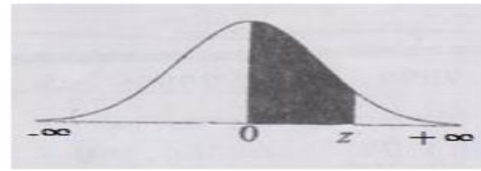
ومن خلال الفقرة السابقة تكون قيمة K تساوي 1.65

9. $P(-1.24 < Z < K) = 0.8$:

$$P(Z < K) - P(Z < -1.24) = 0.8 \rightarrow P(Z < K) = 0.8 + 0.1075 = 0.9075$$

ومن خلال الجدول وعند المساحة 0.9075 نجد أن قيمة K تساوي تقريبا 1.325 .

Standard Normal Distribution Table



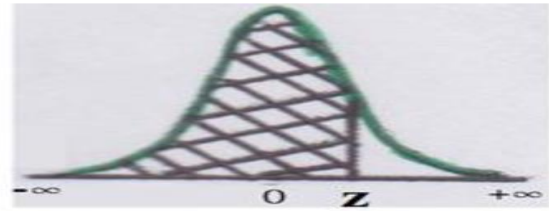
لاحظ هذا الجدول يمكن استخدامه فقط للفترة المُظللة ، وقبل استخدام أي جدول يجب الانتباه الى الفترة المظللة بالرسم المرفق لأن الأرقام تكون فقط لها أعتماداً على حدود التكامل.

اتجاه استخدام قيم z الواقعة ضمن المساحات الأقل من **0.1**

اتجاه استخدام قيم z الواقعة ضمن المساحات الأكبر من **0.1**

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998

Standard Normal Distribution Table (2)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Good luck

Statistics

Lecture 13

Probability Distribution

التوزيعات الاحتمالية

The benefit of studying probability distributions from the valuable implementation success is in a major field:

1. The possibility of determining or calculating a probability in certain evidence circumstances.
2. Appropriateness of the probability distribution of some of the data collected in the field. (Therefore, it is possible to extract information from this data after making a comparison with these distributions are compatible with it).

1. امكانية تحديد او احتساب احتمال ما في ظروف بيانات معينة.
2. امكانية ملائمة التوزيع الاحتمالي لبعض البيانات التي تم جمعها ميدانيا. وبالتالي امكانية استخراج المعلومات من هذه البيانات بعد عمل مقارنة مع هذه التوزيعات المتلائمة معها.

المتغيرات العشوائية Random variables

إن بعض الكميات تكون قيمتها العددية غير محددة تحديداً نهائياً، وانما تتغير بتغير عوامل عشوائية ومن الامثلة على هذه الكميات اختلاف مقاومات الانضغاط للمكعبات المأخوذة من نفس الخلطة ومعالجتها بنفس الظروف او مقاومات الشد لقضبان حديد التسليح المصنوعة من نفس الفولاذ وبنفس الظروف. إن هذا الاختلاف بالرغم من تساوي فرص الحصول على توافق ظاهرياً هو بسبب ظاهرة التشتت وسببها عوامل عشوائية لم يكن بالإمكان حسابها او التنبؤ بها مسبقاً. لذا يمكن تعريف **الكميات العشوائية**: وهي الكميات التي تأخذ قيماً مختلفة في العمليات أو المحاولات المختلفة وتبقى مختلفة مهما حاولنا جعل ظروف وشروط حدوثها متجانسة تبعاً لتغيرات عشوائية خارجة عن الحسبان. إن اهم ما يجب معلوميته بما يخص المتغيرات العشوائية هو جميع القيم العددية التي يمكن أن تأخذها تلك المتغيرات، بالإضافة الى معرفة الاحتمالات المقابلة لتلك القيم لحالة المتغيرات المدروسة. إن معلومية احتمال القيم الممكنة للكميات العشوائية المدروسة يكفي لحل أي سؤال يرتبط بهذه الكمية العشوائية وتبيان فعاليتها. وتقسم المتغيرات العشوائية الى نوعين:

أ- المتغيرات العشوائية المتقطعة او المنفصلة : ويكون فيها المتغير العشوائي يأخذ أي قيمة من قيم مجموعة متقطعة أو قابلة للعد مثل متغير عدد الطلبة الذكور في احدى المدارس.

ب- المتغيرات العشوائية المتصلة : وهي المتغيرات العشوائية التي من الممكن أن تأخذ أي قيمة داخل مدى (مجال) معين من القيم مثل مقاييس درجة الحرارة أو الوزن أو الطول.

Depending on the type of the random variable, the probability distributions can be divided into two types:

بالاعتماد على انواع المتغير العشوائي ، يمكن تقسيم التوزيعات الاحتمالية الى نوعين:

A. Discrete Probability Distributions ((منفصلة) توزيعات احتمالية متقطعة)

- Binomial Dist. (توزيع ذي الحدين)
- Poisson Dist.(توزيع بواسون)

B. Continuous Probability Distributions (توزيعات احتمالية متصلة)

- Normal Distribution (التوزيع الطبيعي)
- T-Distribution (توزيع تي)
- Chi-Square Distribution (توزيع كاي تربيع)

A. Discrete Probability Distributions ((توزيعات احتمالية متقطعة (منفصلة)

وهي التوزيعات الاحتمالية (دوال كتلة الاحتمال) لمتغيرات عشوائية متقطعة.

تعريف دالة كتلة الاحتمال للمتغير العشوائي المتقطع: إذا كان X متغيراً عشوائياً متقطعاً بحيث إن مجموعة القيم الممكنة له هي

$X(S)=\{X_1, X_2, \ldots\ldots, X_n\}$

فإن دالة كتلة الاحتمال ويرمز لها f(X) تأخذ القيم التالية:

$$f(X) = \begin{Bmatrix} P(X = X_i); & X_i \in X(S) \\ 0 \quad\quad ; & X_i \notin X(S) \end{Bmatrix}$$

ويجب ان تحقق دالة كتلة الاحتمال الشروط التالية:

i. ان قيمة دالة كثافة الاحتمال عند قيمة X ما يجب أن لا تزيد عن 1 ولا تقل عن صفر.
ii. ان مجموع قيم دوال كثافة الاحتمال لكل قيم المتغير X المتوقعة (ضمن القيم الممكنة ل X ان تأخذها أو فضاء العينة) يجب أن يساوي "1".

- **Binomial Dist. (توزيع ذي الحدين)**

وهو من اهم التوزيعات المتقطعة المستعملة في مجالات التطبيقات الهندسية لمتغيرات عشوائية منفصلة. حيث يمكن الحكم على نتائج تجربة ما ب "نعم" أو "لا". ويستخدم عند تحقيق الشروط التالية:

- ان نتيجة أي محاولة لتجربة ما تكون اما النجاح أو الفشل.
- كل محاولة نتيجتها مستقلة عن نتيجة المحاولة الاخرى. (أي لا تؤثر عليها مثل عملية السحب بإرجاع).
- احتمال النجاح في كل محاولة ثابت وليكن (P) وبالتالي فإن احتمال الفشل ثابت أيضاً وهو (q) ويحسب من (q=1-P).
- تجري التجربة بعدد معين من المحاولات أو المرات أو بمعنى آخر (يكون هنالك عدد n من المحاولات).

وعند اجراء تجربة ما وكان عدد المحاولات =n ، عدد النجاحات أو عدد مرات حصول الحادثة =x ، احتمال النجاح للتوزيع = P ، واحتمال الفشل للتوزيع =q فإن الاقتران الاحتمالي لمتغير ذي الحدين ويرمز له P(x;n,p) ويحسب كما يلي:

$$P(x;n,p) = C_x^n . p^x . q^{(n-x)} \qquad ; X=0,1,2,.....n$$

ملاحظة: تم أخذ قيمة الصفر لأنه يمثل احتمال عدم الحصول او احتمال ولا مرة يحصل حادث النجاح.

ملاحظة: اذا كان X متغيراً ذا حدين (خاضع لنظرية ذات الحدين) فإن:

i. الوسط او التوقع لقيمة X الممكنة هو $M = n.p$
ii. التباين لقيمة X هو $S^2 = n.p.q$

مثال(1): رُميت زهرة نرد منتظمة 5 مرات ، ما احتمال عدم ظهور الرقم 6 فيها؟ ما احتمال ظهور الرقم 6 ثلاث مرات؟

الحل: ان فضاء العينة لتجربة رمي زهرة نرد واحدة منتظمة لمرة واحدة هو $s = \{1,2,3,4,5,6\}$

- احتمال عدم ظهور الرقم 6 $q(6) = 1 - \frac{1}{6} = \frac{5}{6}$ $\rightarrow$ احتمال ظهور الرقم 6 $P(6) = \frac{1}{6}$

- $P(6) = \frac{1}{6}$, $q(6) = \frac{5}{6}$, n=5 , x=0,1,2,3,4,5

وهي تمثل الصيغة العامة لهذه التجربة $P\left(x;5,\frac{1}{6}\right) = C_x^5.\left(\frac{1}{6}\right)^x.\left(\frac{5}{6}\right)^{(5-x)}$, x=0,1,2,3,4,5

- احتمال عدم ظهور الرقم 6 $P\left(x=0;5,\frac{1}{6}\right) = C_0^5.\left(\frac{1}{6}\right)^0.\left(\frac{5}{6}\right)^{(5-0)} = 1 \times 1 \times \left(\frac{5}{6}\right)^5 = 0.40187$

- احتمال عدم ظهور الرقم 6 ثلاث مرات $P\left(x=3;5,\frac{1}{6}\right) = C_3^5.\left(\frac{1}{6}\right)^3.\left(\frac{5}{6}\right)^{(5-3)} = 10 \times \frac{1}{216} \times \left(\frac{25}{36}\right) = 0.03215$

ولحساب القيمة المتوقعة للمتغير X نحسب الوسط $M = n.p = 5.\frac{1}{6} = \frac{5}{6}$

ولحساب قيمة التباين المتوقعة للمتغير X نحسب التباين $S^2 = n.p.q = 5.\frac{1}{6}.\frac{5}{6} = \frac{25}{36}$

مثال(2): يُراد توزيع عدد من الكراسي الخاصة بالطلبة الأعسرين (مستخدمي اليد اليسرى) في كل قاعة دراسية ، فإذا عُلم ان 10% من جميع الطلبة هم أعسرين فما هو احتمال إن صفاً من 20 طالب: (أ)- لا يحتوي على أي طالب أعسر؟ (ب)- اثنان منهم أعسران؟ (ج)- أكثر من اثنين منهم أعسرين؟

الحل: نفرض ان المتغير X يمثل عدد الطلبة الأعسرين n=20 , x=0,1,2,…….,20

P(L)=0.1 احتمال أن يكون الطالب أعسر ، q(L)=1 - 0.1=0.9 احتمال أن لا يكون الطالب أعسر ،

من المعلومات أعلاه نلاحظ تحقق شروط نظرية ذي الحدين، لذا فإن الصيغة العامة للتجربة هي

$P(x;20,0.1) = C_x^{20}.(0.1)^x.(0.9)^{(20-x)}$, x=0,1,2,…..,20

(أ)- احتمال أن لا يحتوي الصف ولا طالب أعسر $P(x=0;20,0.1) = C_0^{20}.(0.1)^0.(0.9)^{(20-0)} = 1 \times 1 \times 0.12157 = 0.12157$

(ب)- احتمال أن يحتوي الصف طالبين أعسرين $P(x=2;20,0.1) = C_2^{20}.(0.1)^2.(0.9)^{(20-2)} = 190 \times 0.01 \times 0.15009$

=0.28518

- **Poisson Dist.(توزيع بواسون)**

يعتبر توزيع بواسون من التوزيعات الاحتمالية المتقطعة التي تستخدم للحوادث نادرة الحصول أو الوقوع، أو المتغيرات التي تحدث في أزمنة عشوائية معلومة. مثل عدد الأخطاء المطبعية في صفحة كتاب ما، عدد الزلازل السنوية الحاصلة في موقع ما، حدوث فيضان بتصريف معين في نهر ما...الخ.

وتسمى التجارب التي تعطي عدد نجاحات معين في فترة زمنية أو فترة قياس كالطول أو الحجم أو المساحة...الخ بالتجربة البواسونية، وهي تجربة احصائية تمتلك الخصائص التالية:

أ- نتائج التجربة مُعرفة بالنجاح أو الفشل والتجارب مستقلة الواحدة عن الاخرى.

ب- الاحتمال ثابت وذو قيمة صغيرة بشكل عام (تقترب من 0) وإن عدد المحاولات n يكون كبيراً عادةً (يقترب من ∞) ولكن المهم ان مقدار حاصل الضرب بينهما يكون مقدار ثابت يساوي المتوسط أو المعدل (λ) بحيث أن ($\lambda > 0$). $\lambda = n.P$

ومن خصائص توزيع بواسون الاحتمالي:

- المعدل أو المتوسط للتوزيع ($\lambda = n.P$) ، والتباين يساوي (λ)
- الانحراف المعياري يساوي ($\sqrt{\lambda}$)

ان الاقتران الاحتمالي لتوزيع بواسون هو: $P(x,\lambda) = \frac{e^{-\lambda}.\lambda^x}{x!}$, x=0,1,2,.....

<u>حيث ان:</u>

λ: معدل عدد النجاحات (حصول الحدث) ضمن الفترة الزمنية أو فترة القياس (الطول، المساحة، الحجم) وغيرها.

x: عدد النجاحات الفعلي أو عدد مرات القياس المطلوبة وهي تمثل متغير بواسون العشوائي.

e: أساس اللوغاريتم الطبيعي وتساوي تقريباً (e=2.718281828).

مثال(1): معدل عدد حوادث السيارات على طريق صحراوي ما هو 5 (حادث\أسبوع).(أ) ما احتمال عدم حدوث أي حادث على ذلك الطريق في اسبوع معين؟ (ب)ما احتمال حدوث 4 حوادث او أقل في اسبوع معين؟(ج) ما احتمال أن يزيد عدد الحوادث عن حادثين خلال أسبوعين؟

الحل: $P(x,\lambda) = \frac{e^{-\lambda}.\lambda^x}{x!}$, x=0,1,2,..... , λ=5

الصيغة الاساسية لحل المسألة $P(x,\lambda=5) = \frac{e^{-5}.5^x}{x!}$

(أ) عدم حصول أي حادث $P(x=0,\lambda=5) = \frac{e^{-5}.5^0}{0!} = 0.00674$

(ب) حدوث 4 حوادث أو أقل $P(x\leq 4,\lambda=5) = P(0,5)+P(1,5)+P(2,5)+P(3,5)+P(4,5)$

$$= 0.00674 + \frac{e^{-5}.5^1}{1!} + \frac{e^{-5}.5^2}{2!} + \frac{e^{-5}.5^3}{3!} + \frac{e^{-5}.5^4}{4!} = 0.4405$$

(ت) زيادة عدد الحوادث عن 2 خلال إسبوعين (هنا نلاحظ ان معدل حصول الحوادث اختلف بدلاً من الفترة اسبوع الى اسبوعين فهنا ستتغير الصيغة الاساسية للسؤال حسب المعدل الجديد).

$$\lambda = 5(\text{حادث\اسبوع}) \times 2 = 10(\text{حادث\اسبوعين})$$

$P(x,\lambda=10) = \frac{e^{-10}.10^x}{x!}$, $x = 0,1,2,\ldots\ldots\ldots$

$$P(x>2,\lambda=10) = 1-\{P(0,10)+P(1,10)+P(2,10)\} = 1-\left\{\frac{e^{-10}.10^0}{0!}+\frac{e^{-10}.10^1}{1!}+\frac{e^{-10}.10^2}{2!}\right\} = 0.99723$$

Approximation of the binomial distribution of the Poisson distribution

تقريب توزيع ثنائي الحدين لتوزيع بواسون

اذا اعتبرنا n او عدد المحاولات كبيرة بشكل كافٍ بحيث (n≥50) و احتمال النجاح P صغير (P<0.1) بحيث يبقى حاصل ضرب قيمتيهما (n.P) مقدار ثابت وليكن λ، عندها يمكن تقريب توزيع ذي الحدين الى توزيع بواسون بالمقدار (λ=n.P).

مثال (2): في اختبار مُضاف مُحسن للخلطات الخرسانية، اذا افترضنا انه ضار للخلطة باحتمال 0.001 وانه تم استخدامه فعليا في انتاج 2000 خلطة خرسانية، استخدم تقريب بواسون للإجابة على: (أ) ما احتمال ان تتضرر أكثر من خلطتين نتيجة استخدام ذلك المُضاف؟ (ب) ما احتمال تضرر خلطتان على الأقل؟

Solution: n=2000 , P=0.001 (احتمال حصول الضرر)

نلاحظ إن عدد المحاولات كبير جداً (n>50) واحتمال حصول الحدث P أقل من 0.1 لهذا يمكن استخدام تقريب بواسون بذي الحدين أو بمعنى آخر

$\lambda = n.P = 0.001 \times 2000 = 2$

(أ) $P(x > 2, \lambda = 2) = 1 - \{P(0,2) + P(1,2) + P(2,2)\} = 1 - \left\{\frac{e^{-2}.2^0}{0!} + \frac{e^{-2}.2^1}{1!} + \frac{e^{-2}.2^2}{2!}\right\} = 0.32332$

(ب) $P(x \geq 2, \lambda = 2) = 1 - P(x < 2, \lambda = 2) = 1 - \{P(0,2) + P(1,2)\} = 0.59399$

Not:

الفرق الأساسي بين توزيعي ذي الحدين وتوزيع بواسون هو إن n أ و عدد التجارب الكلي في توزيع ذي الحدين يكون معلوماً ولهذا فإن قيم المتغير العشوائي x تكون (x=0,1,2,.....,n) أما في توزيع بواسون فإن n أو عدد التجارب غير معلوم أو مبهم وفي هذه الحالة فإن المتغير العشوائي x يأخذ المدى (x=0,1,2,.....). ولهذا في حالة إحتساب احتمالات أكبر من قيمة ما و لتكن a في توزيع بواسون فإننا نضطر طرح احتمال تلك القيمة والتي اقل منها من (1) وهو الأحتمال الكلي لفضاء العينة أو بمعنى آخر $P(x > a, \lambda) = 1 - P(x \leq a, \lambda)$ وذلك لأننا لا نعرف إلى أي حد سينتهي إليه المتغير العشوائي x.

مثال(2): اذا كان 3% من انتاج أحد معامل الطابوق غير مطابق للمواصفات القياسية وقد تم سحب عينة من الانتاج بشكل عشوائي. ما هو احتمال بأن العينة تحتوي بالضبط على طابوقتين فاشلتين؟ استخدم توزيعين مختلفين للحل ثم قارن النتيجتين؟

الحل: (أ) باستخدام توزيع ذي الحدين (P=0.03 , q=0.97 , n=100)

$$\therefore P(x; 100,0.03) = C_x^{100}.(0.03)^x.(0.97)^{100-x} \qquad x = 0,1,\dots\dots,100$$

$$\therefore P(x = 2; 100,0.03) = \frac{100!}{2! \times (100-2)!} \times (0.03)^2 \times (0.97)^{100-2} = 4950 \times 0.0009 \times 0.0505 = 0.225153$$

(ت) باستخدام توزيع بواسون ($\lambda = n.P = 0.03 \times 100 = 3$)

$$P(x, \lambda = 3) = \frac{e^{-3}.3^x}{x!} \quad , \qquad x = 0,1,2,\dots\dots$$

$$\therefore P(x = 2, \lambda = 3) = \frac{e^{-3}.3^2}{2!} = 0.22404$$

مثال(3): اذا كانت احدى البدالات تتلقى مكالمات هاتفية بمعدل مكالمتين لكل دقيقة. أوجد احتمال أن تتلقى البدالة 3 مكالمات على الأقل؟

Solution: (λ=2 call/min)

$$P(x,2) = \frac{e^{-2}.2^{x}}{x!} \quad , \qquad x = 0,1,2,\dots\dots$$

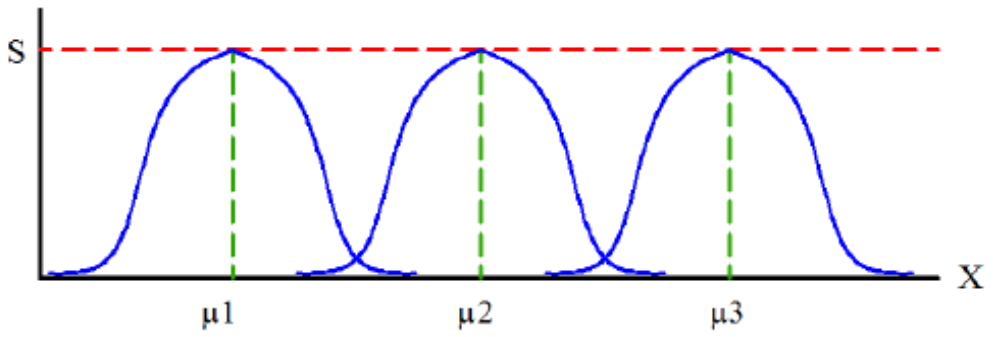
$$P(x \geq 3, \lambda = 2) = 1 - \left\{\frac{e^{-2}.2^{0}}{0!} + \frac{e^{-2}.2^{1}}{1!} + \frac{e^{-2}.2^{2}}{2!}\right\} = 1 - (0.67667) = 0.3233$$

التوزيعات الاحتمالية المتصلة (Continuous Probability Distributions)

1-1- التوزيع الطبيعي (Normal Distribution):

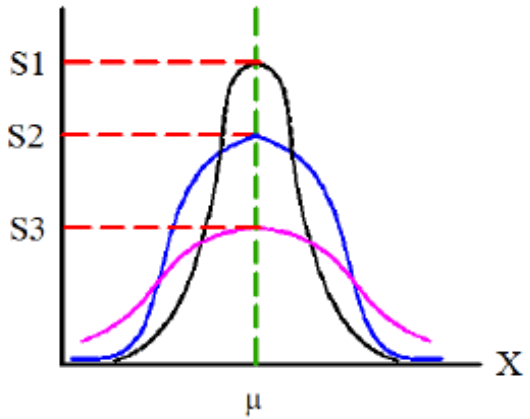
وهو من أشهر التوزيعات الاحتمالية المتصلة وأكثرها استخداماً وتطبيقاً ومن الأمثلة عليه التوزيعات البايومترية (توزيعات الطول والوزن) وتوزيعات اخطاء المشاهدات (الفروق بين القيم الحقيقية والقيم المُشاهدة) وتم تطوير هذا التوزيع من قبل العالم گاوس سنة 1809 ولهذا يطلق عليه في بعض الأحيان اسم (توزيع گاوس).

ان منحني التوزيع الطبيعي الذي يأخذ شكل الجرس هو منحني توزيع نظري لدالة كثافة (كتلة الاحتمال) للمتغير العشوائي الذي يخضع لهذا التوزيع. هنالك ما لانهاية من منحنيات التوزيع الطبيعي التي قد تختلف عن بعضها البعض حسب قيمة كل من الوسط الحسابي (μ) والانحراف المعياري (S)، فمثلاً قد تتفق المنحنيات بالانحراف المعياري لكنها تختلف بالوسط الحسابي كما في الشكل (1) أو بالعكس قد تختلف بالانحراف المعياري وتتفق بالوسط كما في الشكل (2).



شـكل (1)

منحنيـــات توزيـــع طبيعـــي لها نفـــس الانحـراف المعيـــاري مع اختـــلاف الوســـط الحســـابي



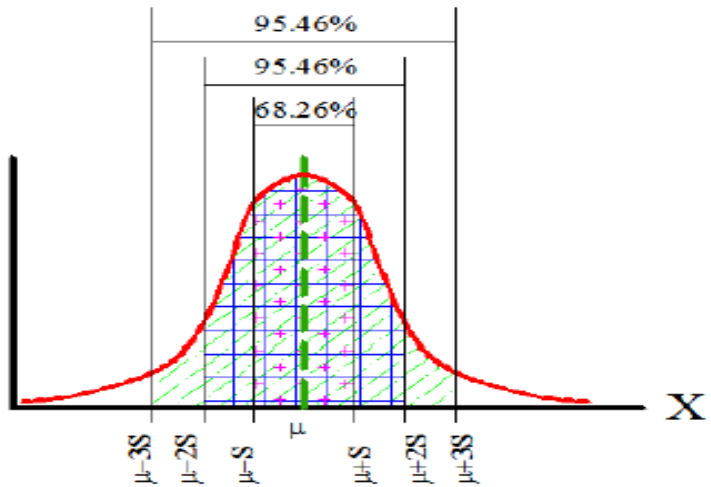
شـكل (2)

منحنيـــات توزيـــع طبيعـــي لها نفـــس الوســـط الحســـابي مع اختـــلاف الانحـــراف المعيـــاري

خصائص التوزيع الطبيعي:

- منحني التوزيع الطبيعي متماثل حول الوسط μ (إلتوائه يساوي صفراً) ، وشكله بشكل الجرس.
- له قيمة منوالية (منوال) واحد ينطبق على الوسط μ ويقترب طرفاه من قيمتي (-∞ ، ∞) دون أن تقطع المحور الافقي.
- قيمة الانحراف المعياري تحدد عُرض المنحني الطبيعي، مثلاً قيمة انحراف معياري أعلى تنتج منحني أعرض وأكثر تسطحاً.
- الأحتمالات للمتغير الطبيعي العشوائي تُعطى بالمساحة تحت منحني التوزيع الطبيعي. والمساحة الكلية تحت منحني التوزيع الطبيعي للفترة (-∞≤x≤+∞) تساوي (1) و تتوزع مناصفةً 0.5 للنصف الأيمن و 0.5 للنصف الأيسر.
- قيمة الاحتمال عند قيمة مفردة محددة للمتغير العشوائي الطبيعي x تساوي (صفر) ولأي قيمة وذلك لان المساحة تحت المنحني والمحددة بقيمة واحدة تكون عبارة عن خط مستقيم ولهذا فلحساب الاحتمال يجب ان تحدد فترة ما للمتغير العشوائي ومن خلالها تحسب المساحة تحت المنحني ضمن هذه الفترة والتي سوف تساوي قيمة الاحتمال ضمن تلك الفترة.

$P(x=0) = 0$, $P(x=1) = 0$, $P(x=100) = 0$, $P(0<x<a) = ?$, $P(x>a) = ?$



المساحة تحت المنحني الطبيعي تتوزع كما يلي (ثابتة لجميع منحنيات التوزيع الطبيعي)

1- 68.26% من المساحة تحت المنحني تكون ضمن حدود (μ-S≤x≤μ+S)
2- 95.46% من المساحة تحت المنحني تكون ضمن حدود (μ-2S≤x≤μ+2S)
3- 99.74% من المساحة تحت المنحني تكون ضمن حدود (μ-3S≤x≤μ+3S)
4- 95% من المساحة تحت المنحني تكون ضمن حدود (μ-1.96S≤x≤μ+1.96S)
5- 99% من المساحة تحت المنحني تكون ضمن حدود (μ-2.58S≤x≤μ+2.58S)

- يُقال للمتغير العشوائي المتصل x بأنه موزع طبيعياً اذا كانت دالة كثافة (كتلة) الاحتمال له

مُعطاة بالصيغة :

$$F(x) = \frac{1}{S\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{S}\right)^2}$$, للفترة (-∞≤x≤+∞)

ويرمز لهذا التوزيع بالرمز التالي $N(\mu,S^2)$

التوزيع الطبيعي المعيارى (القياسي):

بسبب ان شكل منحني التوزيع الطبيعي يعتمد على قيمتي الوسط μ والانحراف المعياري S فلهذا هنالك مالانهاية من المنحنيات الخاصة بذلك النوع من التوزيع وبالتالي فان قيم الاحتمالات المعتمدة على قيمة المساحة المحصورة تحت هذه المنحنيات سوف تتغير تبعاً لتلك القيم وبالتالي فلأغراض المقارنة بين قيم الاحتمالات للتوزيعات الطبيعية تم إقتراح توزيع طبيعي معياري موحد يمكنه التعبير عن جميع منحنيات التوزيعات الطبيعية: وهو التوزيع الطبيعي الذي وسطه (μ=0) وانحرافه المعياري (S=1) وله دالة كثافة (كتلة) احتمال معرفة كما يلي:

$$F(z) = \frac{1}{1\times\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-0}{1}\right)^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} \quad , \quad \text{للفترة } (-\infty \le z \le +\infty)$$

هنا لاستخراج المساحة تحت منحني التوزيع الطبيعي المعياري لاستخراج الاحتمال ضمن الفترة المطلوبة والتي سوف تمثل حدود التكامل باستخدام صيغة التكامل التالية:

$$P(-\infty \le z \le +\infty) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} . dz$$

أما لحساب الاحتمال لفترة ما فيتم بتبديل حدود التكامل أعلاه بقيم تلك الفترة فنستخرج قيمة المساحة تحت المنحني ضمن تلك الفترة والتي تساوي قيمة الاحتمال لتلك الفترة.

ملاحظة: لتسهيل الحسابات والعمل بهذا النوع من التوزيع يتم استبدال التكامل بجداول خاصة بالتوزيع الطبيعي المعياري واعتماداً على قيمة المتغير الطبيعي العشوائي (z) ، وعلى قيم الحدود للفترة المطلوب حساب الاحتمال لها.

ملاحظة: للتحويل من حالة التوزيع الطبيعي الى حالة التوزيع الطبيعي المعياري نقوم بتحويل المتغير العشوائي المعياري x الى المتغير العشوائي z من خلال العلاقة التالية:

$$z = \frac{x-\mu}{s}$$

مثال (1): إذا كان x متغير عشوائي طبيعي خاضع للتوزيع N(3,4) فما هي قيمة الاحتمال عند وقوع x بين القيمتين 3 و5 ؟

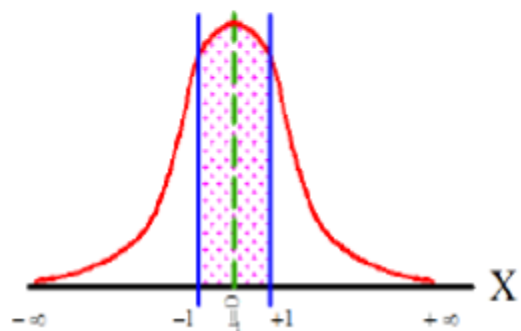
Solution: $z = \frac{x-\mu}{s}$, $z_1 = \frac{3-3}{\sqrt{4}} = 0$, $z_2 = \frac{5-3}{\sqrt{4}} = 1$ وبتطبيق هذه الحدود تصبح المسألة → P(0≤z≤1)=?

ومن الجداول الخاصة بالتوزيع الطبيعي المعياري → P(0≤z≤1)=0.3413

مثال (2): إذا كان المتغير العشوائي الطبيعي x خاضع للتوزيع $N(\mu, S^2)$ ، إحسب كلٍ من $P(\mu-S\leq x\leq\mu+S)$ ، $P(\mu-2S\leq x\leq\mu+2S)$ و $P(\mu-3S\leq x\leq\mu+3S)$ ؟

Solution:

(1) For $P(\mu-S\leq x\leq\mu+S)$, , $z_1 = \frac{(\mu-s)-\mu}{s} = -1$, $z_2 = \frac{(\mu+s)-\mu}{s} = 1$, $\rightarrow P(-1 \leq z \leq 1) = 2 \times P(0 \leq z \leq 1) = 2 \times 0.3413 = 0.6826 = 68.26\%$ (باستخدام الجداول)



Statistics

Lecture 6

Measures of central tendency

مقاييس النزعة المركزية

The basic idea behind measures of central tendency is to represent a large set of data with a single value. This value is usually in the middle of the data, that is, in its center, and it is the value that other values tend to and clustered around Thus, this property that most data possesses is called the central tendency property, and I term measures That measures the average value around which the cluster gets, measures of central tendency or averages.

The average value is taken as a representative of the whole group, on the basis that it is a non-extreme value, but rather a value that is gathered around it Most of the values are, and therefore, first than others in representing data. Knowing the middle value of the data is useful for us Study the characteristics of the frequency distribution and compare the different repetitive distributions for the same phenomenon.

الفكرة الأساسية التي تعتمد عليها مقاييس النزعة المركزية هي تمثيل مجموعة كبيرة من البيانات بقيمة واحدة ، وهذه القيمة عادة تكون في وسط البيانات أي في مركزها وهي القيمة التي تميل إليها بقية القيم وتتجمع حولها، وبالتالي سميت هذه الخاصية التي تتمتع بها معظم البيانات بخاصية النزعة المركزية ، وأطلق على المقاييس التي تقيس هذه القيمة المتوسطة التي يحصل حولها التجمع، مقاييس النزعة المركزية أو المتوسطات.وتؤخذ قيمة المتوسط كممثل للمجموعة كلها ، على أساس أنها قيمة غير متطرفة بل هي قيمة تتجمع حولها أغلبية القيم ، وبالتالي هي أولى من غيرها في تمثيل البيانات. ومعرفة القيمة الوسطى للبيانات تفيدنا في دراسة خصائص التوزيع التكراري والمقارنة بين التوزيعات التكرارية المختلفة لنفس الظاهرة.

One of the main averages that we are going to study the following:

1. Mean:

The mean is defined as the value that represents the center of gravity of the data, i.e. The anchor point then the equilibrium occurs, if we have a set of data and match it with weights of equal weight on a stepped board, it will balance if suspended or fixed from its center of gravity and center of gravity this is what represents the mean value of these values.

$$\text{Mean} = \frac{Sum\ of\ values}{Number\ of\ values}$$

We denote the mean of the sample by the symbol $(\overline{X})$, and we denote the mean of the community by the symbol μ .

عرف الوسط الحسابي بأنه القيمة التي تمثل مركز ثقل البيانات أي نقطة الارتكاز التي يحصل عندها التوازن ، فإذا كان لدينا مجموعة من البيانات ومثلناها بأثقال متساوية الوزن على لوح مدرج ، فسنجد أن هذا اللوح سيتزن إذا علق أو ثبت من مركز ثقله ومركز الثقل هذا هو الذي يمثل قيمة الوسط الحسابي لهذه القيم.

In the following, we will show how to apply the mean formula in the case of ungrouped data, i.e. Not shown in the frequency distributions tables, and in the case of grouped data, i.e. shown in Frequency distributions tables.

1. Calculate the arithmetic mean of unclassified data :

$$\overline{X} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

($\overline{X}$) : الوسط الحسابي

x : رمز المتغيير

n : عدد القيم

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

Ex1: The following data represent 6 scores in the statistics subject, (4, 6, 3, 7, 9, 1) calculate the arithmetic mean

of these scores.

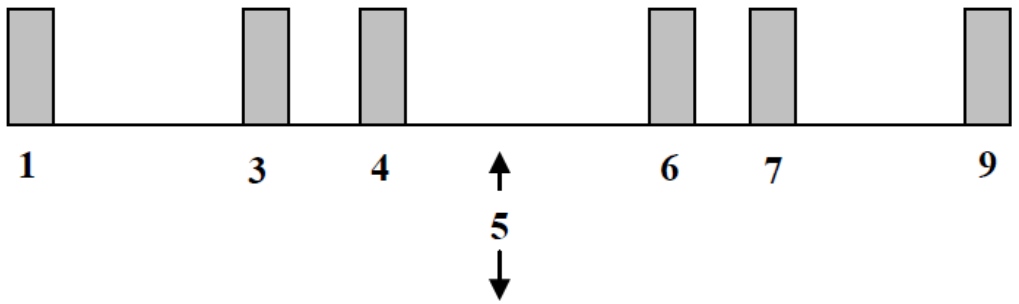
Sol:

$x_1 = 4$, $x_2 = 6$, $x_3 = 3$, ……. $x_6 = 1$.

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

$$= \frac{4+6+3+7+9+1}{6} = 5$$

This means that the center of gravity, i.e., the focal point at which the balance occurs, is degree 5 this is as shown in the figure.



Ex2: The following data shows the value of imports (in millions of Iraqi dinars) for a port in years from 1976 to 1980, according to bulletins issued by the Bureau of Statistics and Census.

the years	1976	1977	1978	1979	1980
Import value	381	292	293	333	391

Calculate the arithmetic mean of the value of this port's imports in these five years.

Sol:

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

$$= \frac{381+272+293+333+391}{5} = \frac{1670}{5} = 334$$

2. Calculation of the arithmetic mean of the classified data:

A. If the data is classified in a frequency distribution table to represent each category of the table only one value, in this case in order to calculate the arithmetic mean, we perform the following steps:

1. نحسب المجموع الكلي للقيم التابعة لكل فئة بضرب قيمة x_i التي تمثلها الفئة في تكرار الفئة f_i ، اي نحسب $(x_i\ f_i)$ لكل الفئات.
2. نحسب المجموع الكلي لجميع القيم وذلك بجمع المجاميع الخاصة بكل الفئات والتي تحصلنا عليها في الخطوة السابقة، أي نحسب:

$$\left(\sum_{i=1}^{k} x_i f_i\right)$$

حيث k = عدد الفترات أو عدد القيم المختلفة.

3. نحسب العدد الكلي للقيم المشاهدة وهو يساوي المجموع الكلي للتكرارات $\left(\sum_{i=1}^{k} f_i\right)$.
4. نحسب الوسط الحسابي بقسمة المجموع الكلي للقيم $\left(\sum_{i=1}^{k} x_i f_i\right)$ على العدد الكلي للقيم $\left(\sum_{i=1}^{k} f_i\right)$.

B. If the data is grouped in a repeating table so that each of the table categories is more representative Of a single value**.**

إذا كانت البيانات مبوبة في جدول تكراري بحيث كل فئة من فئات الجدول تمثل أكثر من قيمة واحدة ، ففي هذه الحالة لا نستطيع معرفة القيم المشاهدة التابعة لكل فئة ، والذي نعرفه هو عددها فقط والمتمثل في تكرار الفئة ، ولحساب المجموع الكلي للقيم التابعة لكل فئة سنفترض أن القيم المشاهدة موزعة حول مركز الفئة داخل كل فئة من فئات الجدول توزيعاً عادلاً ، وبالتالي تكون قيمة مركز الفئة هي القيمة الافتراضية لجميع القيم داخل الفئة.

ولكي نحسب الوسط الحسابي في هذه الحالة نجري الخطوات التالية :

1. نحسب مركز كل فئة ونرمز للمركز بالرمز x_i.
2. نحسب المجموع الكلي للقيم التابعة لكل فئة بضرب القيمة x_i (حيث x_i مركز الفئة) في تكرار الفئة f_i ، اي نحسب $(x_i\ f_i)$ لكل الفئات.
3. نحسب المجموع الكلي لجميع القيم وذلك بجمع المجاميع الخاصة بكل الفئات والتي تحصلنا عليها في الخطوة السابقة، أي نحسب $(\sum_{i=1}^{k} x_i f_i)$
4. نحسب العدد الكلي للقيم المشاهدة وهو يساوي المجموع الكلي للتكرارات $(\sum_{i=1}^{k} f_i)$.
5. نحسب الوسط الحسابي بقسمة المجموع الكلي للقيم $(\sum_{i=1}^{k} x_i f_i)$ على العدد الكلي للقيم $(\sum_{i=1}^{k} f_i)$.

وهكذا تكون صيغة المتوسط الحسابي للبيانات المبوبة كما يلي :

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}$$

حيث :

x : القيمة التي تمثلها الفئة (عندما الفئة تمثل قيمة واحدة فقط)،

وهي عبارة عن مركز الفئة (عندما الفئة تمثل اكثر من قيمة واحدة).

f: تكرار الفئة.

Ex1: The following data represent the results of the compression test for fifty concrete cubes in units (MPa).

(29 30 32 39 38 38 44 28 33 31 35 37 42 49 25 34 30 31 35 37 40 26 32 33 33 39 44 45 26 31 34 31 36 36 41 27 30 31 32 37 40 39 25 34 31 30 38 35 43 48).

Calculate the mean for the two cases:

1. Unclassified data.
2. Data classified in a frequency table with categories.

Sol:

1. $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{29+30+32+\cdots+48}{50} = \frac{1744}{50} = 34.88$

2.

Class limits حدود الفئات	Centers of classes مراكز الفئات	Frequencies التكرارات	The arithmetic mean of the actual values within the classes
25 - 29	27	7	26.571
30 - 34	32	19	31.737
35 - 39	37	14	37.071
40 - 44	42	7	42
45 - 49	47	3	47.333

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}$$

$$\overline{X} = \frac{27*7+32*19+37*14+42*7+47*3}{50} = 35$$

- نلاحظ هنا الفرق بين النتيجتين وسبب واضح وهو وجود تقريب في حالة تبويب البيانات لاننا استخدمنا بديل عن البيانات الواقعية وهو مراكز الفئات بدلا من الفئات نفسها.

Ex2: The data shown in the following frequency table represent the scores of 50 students:

the scores	Number of students
30 – 39	4
40 – 49	6
50 – 59	8
60 – 69	12
70 – 79	9
80 – 89	7
90 – 99	4

Calculate the mean of these scores.

Sol:

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}$$

the scores	Number of students f_i	Centers of classes مراكز الفئات x_i	$x_i f_i$
30 – 39	4	34.5	138
40 – 49	6	44.5	267
50 – 59	8	54.5	436
60 – 69	12	64.5	774
70 – 79	9	74.5	670.5
80 – 89	7	84.5	591.5
90 – 99	4	94.5	378
SUM=	50		3255

$$\overline{X} = \frac{3255}{50} = 65.1$$

Statistics

Lecture 7

Properties of the arithmetic mean:

ان لقيمة الوسط الحسابي صفات تؤهله بأن يكون ممثلا عن القيم للبيانات الاخرى المحسوب منها. من مزايا الوسط الحسابي انه ياخذ بنظر الاعتبار كل القيم ومن عيوبه انه يتأثر بالقيم الشاذة و المتطرفة و يصعب حسابه في حالة البيانات الوصفية و في حالة الجداول التكرارية المفتوحة.

ومن اهم خصائصه:

1. أهم مقاييس النزعة المركزية وأكثرها استعمالاً نظراً لسهولة حسابه وإمكانية التعامل معه رياضياً ، ولذلك له أهمية قصوى في التحليل الإحصائي إذ إنه يدخل في حساب كثير من المقاييس الأخرى.
2. يأخذ بعين الاعتبار جميع القيم المشاهدة ، ونستطيع أن نحصل على مجموع القيم المشاهدة إذا عرفنا قيمة الوسط الحسابي كما يلي :

   مجموع القيم = الوسط الحسابي × عدد القيم
3. الوسط الحسابي هو قيمة نظرية وليس بالضروري أن تكون واحدة من القيم التي يمكن أن يأخذها المتغير محل الدراسة.
4. يتأثر بوجود قيم متطرفة في البيانات وينحاز لها ، فمثلاً إذا تبرع خمسة عشر فرداً لعمل خيري بالمبالغ التالية :

10 , 15 , 15 , 20 , 15 , 15 , 12 , 20 , 10 , 10 , 18 , 20 , 10 , 10 , 1000 .

نلاحظ أن الوسط الحسابي للتبرعات كبير ولا يمثل ما دفعه معظم الأفراد ، فهنا انحاز الوسط الحسابي للقيمة المتطرفة 1000 ، وبالتالي يعدُّ الوسط الحسابي في هذه الحالة مقياساً مضللاً . ولكن لو استبعدنا القيمة المتطرفة فسنلاحظ أن الوسط الحسابي سيكون واقعياً .

$\overline{X} = \frac{1200}{15} = 80$

5. لا يمكن حسابه في حالة البيانات النوعية (الوصفية) .

6. لا يمكن إيجاده من الرسم (أي بيانياً) .

7. لا يمكن حسابه في حالة جداول التوزيعات التكرارية المفتوحة ؛ لأننا لا نستطيع حساب مراكز الفئات المفتوحة.

8. مجموع انحرافات القيم عن وسطها الحسابي يساوي صفراً ، حيث انحراف القيمة عن الوسط الحسابي المقصود به القيمة مطروحاً منها الوسط الحسابي ، وبالتالي فهذه الخاصية تعني أن:

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

For example:

اذا كانت لدينا البيانات التالية: 10, 11, 5, 8, 6 فأن الوسط الحسابي لهذه البيانات هو:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{6+8+5+11+10}{5} = 8$$

فسنجد أن مجموع انحرافات هذه القيم عن وسطها الحسابي يساوي صفراً ، كما هو مبين فيما يلي :

القيمة (x_i)	الانحرافات ($x_i - \bar{x}$)
6	6 – 8 = -2
8	8 – 8 = 0
5	5 – 8 = -3
11	11– 8 = 3
10	10 – 8 =2
المجموع	صفر

9. مجموع مربعات انحرافات القيم عن وسطها الحسابي أقل ما يمكن.

2. The Median (الوسيط):

The median is the value in the middle of the data after arranging it in ascending or descending order (M).

الوسيط هو القيمة الموجودة في منتصف البيانات بعد ترتيبها تصاعدياً أو تنازلياً ، أي هوالقيمة الوسطى التي يكون نصف البيانات أقل منها والنصف الآخر أكبر منها.

The median is calculated as follows:

A. <u>Non-classified data:</u>

يحسب الوسيط في حالة البيانات غير المبوبة باتّباع الخطوات التالية :

أ – نرتب القيم المشاهدة تصاعدياً ، أي نبدأ بأصغر قيمة ثم نرتب ما بعدها الأكبر فالأكبر .

ب – نحدد ترتيب الوسيط بين القيم وذلك كما يلي :

ترتيب الوسيط هو $\frac{n+1}{2}$ حيث n عدد القيم.

Median arrangement = $\frac{n+1}{2}$

ج – نحدد قيمة الوسيط وتكون هي القيمة التي ترتيبها $\frac{n+1}{2}$ عندما يكون عدد القيم n عددا فرديا، اما اذا كانت قيمة n عددا زوجيا سيقع الترتيب بين قيمتين ويكون الوسيط هومعدل هاتين القيمتين.

Ex1: Find the median value of the following data:

13, 17, 15, 18, 20, 19, 16.

Sol:

نرتب القيم ترتيبا تصاعديا:

13, 15, 16, 17, 18, 19, 20

n = 7 (عدد فردي)

Median arrangement = $\frac{n+1}{2} = \frac{7+1}{2} = 4$

أي أن قيمة الوسيط هي القيمة الرابعة في البيانات بعد ترتيب البيانات تصاعدياً وبالتالي فإن :

الوسيط = 17

Ex2: Find the median value of the following data:

3, 8, 9, 6, 5, 11.

Sol:

نرتب القيم ترتيبا تصاعديا:

3, 5, 6, 8, 9, 11.

n = 6 (عدد زوجي)

Median arrangement = $\frac{n+1}{2}$ = $\frac{6+1}{2}$ = 3.5

قيمة الوسيط تقع بين القيمتين الثالثة والرابعة ، إذاً الوسيط يساوي المعدل لهاتين القيمتين أي:

M= $\frac{6+8}{2}$ = 7

B. classified data:

لحساب الوسيط في حالة البيانات المعروضة في جداول تكرارية بحيث تكون فئات الجدول مرتبة ترتيباً تصاعدياً أو تنازلياً، نتبع الخطوات التالية:

1. نحدد ترتيب الوسيط ويساوي في البيانات المبوبة

$$\text{Median arrangement} = \frac{\sum_{i=1}^{k} f_i}{2}$$

2. نحدد الفئة الوسيطية وهي الفئة التي تحتوي على الوسيط ويتم تحديدها بالاستعانة بالتكرارالمتجمع الصاعد ، فتكون الفئة الوسيطة هي أول فئة يكون تكرارها المتجمع الصاعد أكبر من أو يساوي ترتيب الوسيط .

3. نقوم بالتعويض في القانون التالي للحصول على قيمة الوسيط :

$$m = L + \left(\frac{\frac{\sum_{i=1}^{k} f_i}{2} - F}{f_m} \right) \times c$$

حيث:

L: الحد الأدنى للفئة الوسيطية.

F: التكرار المتجمع الصاعد للفئة السابقة للفئة الوسيطية .

f_m: تكرار الفئة الوسيطية .

c: طول الفئة الوسيطية .

والقاعدة الرياضية التي يعتمد عليها قانون الوسيط هو افتراض أن القيم التابعة لكل فئة موزعة حول مركزها توزيعاً عادلاً ، وبالتالي سيكون هناك تناسب بين بعد الوسيط عن الحد الأدنى للفئة الوسيطية والذي نرمز له بالرمز (L) وطول الفئة الوسيطية (C) وبين عدد القيم السابقة للوسيط داخل الفئة الوسطية (ترتيب الوسيط F) والعدد الكلى للقيم المشاهدة التابعة للفئة الوسيطية (f_m)، أي أن هناك تناسباً بين المسافات والتكرارات .

Ex3: Find the median value of the following data:

Class limits	f_i
50 – 54	10
54 – 58	30
58 – 62	90
62 – 66	60
66 – 70	20

Sol:

Median arrangement = $\frac{\sum_{i=1}^{k} f_i}{2} = \frac{210}{2} = 105$

نحدد الفئة الوسيطية أي نحدد الفئة التي تحتوي على الوسيط وهو القيمة التي ترتيبها 105 ويتم ذلك بالاستعانة بالتكرار المتجمع الصاعد الموضح في الجدول التالي:

Class limits	f_i	The up word combined frequency
50 – 54	10	10
54 – 58	30	40
58 – 62	90	130
62 – 66	60	190
66 – 70	20	210

الفئة الوسطية

حيث إن الفئة الثانية تكرارها المتجمع يساوي 40 ، والفئة الثالثة تكرارها المتجمع يساوي 130 ، فيعني ذلك أن التسعين قيمة الموجودة في الفئة الثالثة ترتيباتها تبدأ من الترتيب 41 الى الترتيب 130 ، وحيث إن ترتيب الوسيط يساوي 105 ، إذن يجب أن يكون ضمن قيم الفئة الثالثة ، وبالتالي فإن الفئة الثالثة هي الفئة الوسيطية.

وبعد تحديد الفئة الوسيطية نطبق قانون الوسيط ، حيث :

$f_m = 90 \qquad C=4 \qquad L=58 \qquad F=40$

$$m = L + \left(\frac{\frac{\sum_{i=1}^{k} f_j}{2} - F}{f_m} \right) \times c$$

$$= 58 + \frac{(105 - 40)}{90} \times 4$$

$$= 58 + 2.89 = 60.89$$

ويعني ذلك ان 50% الفئات تكون اقل من 60.89 والنسبة المتبقية تكون اكثر من 60.89 .

Statistics

Lecture 8

Measures of dispersion

In statistics, the measures of dispersion help to interpret the variability of data i.e. to know how much homogenous or heterogeneous the data is. In simple terms, it shows how squeezed or scattered the variable is.

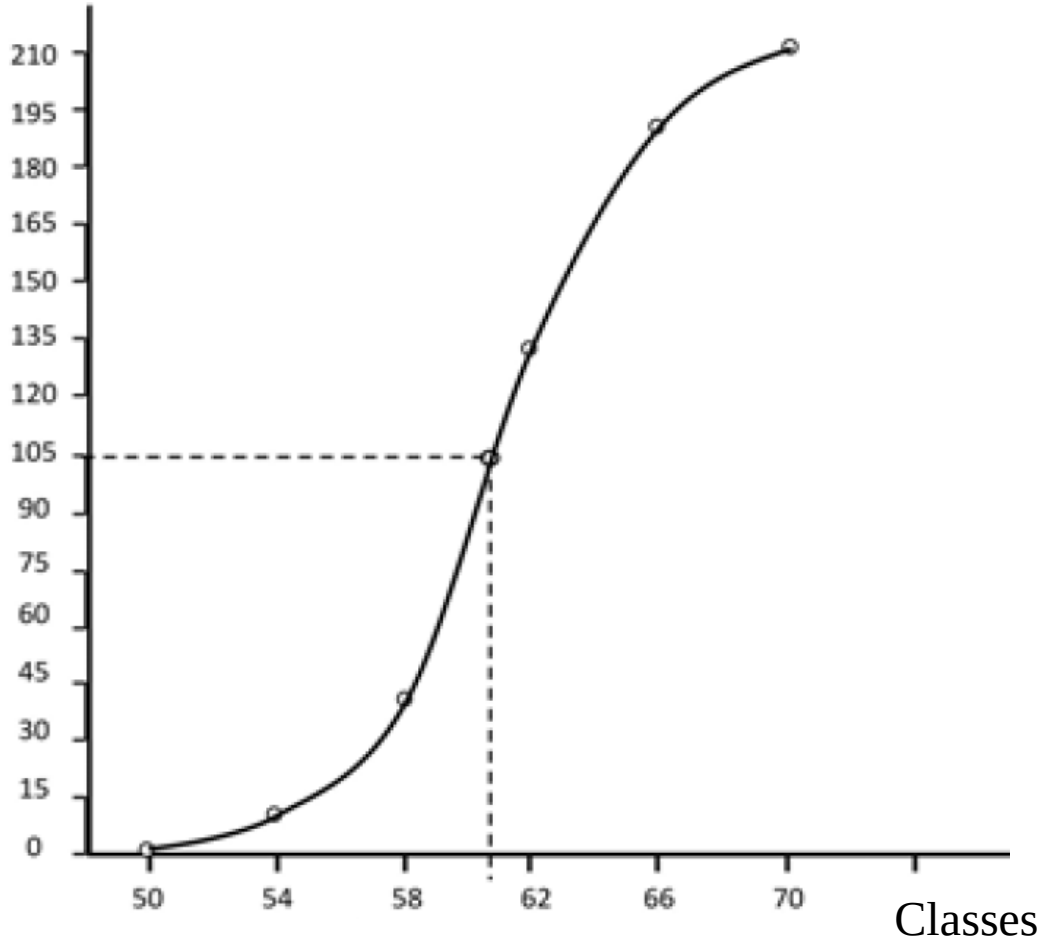
في الإحصاء ، تساعد مقاييس التشتت في تفسير تباين البيانات ، أي معرفة مقدار البيانات المتجانسة أو غير المتجانسة. بعبارات بسيطة ، يوضح مدى ضغط أو تشتيت المتغير.

Sol:

نكوّن أولاً جدول التكرار المتجمع الصاعد أو جدول التكرار المتجمع الهابط لهذه البيانات وبعد رسم المنحنى المتجمع وتحديد ترتيب الوسيط الذي يساوي 105 على المحور الرأسي ، نرسم من هذه النقطة خطاً أفقياً يوازي محور السينات حتى يلاقى المنحنى المتجمع ومن نقطة التلاقي نسقط عموداً ليلاقي محور السينات فنحصل على قيمة الوسيط.

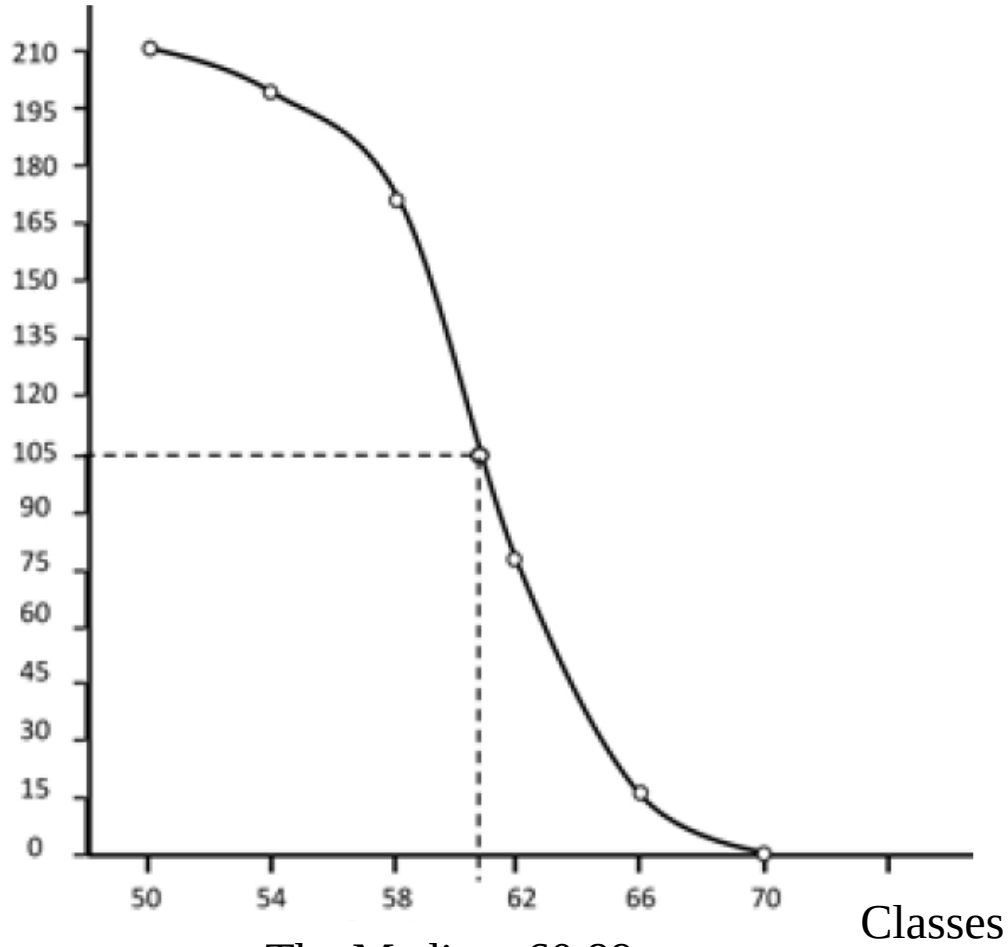
Classes	The up word combined frequency	Classes	The down word combined frequency
Less than 50	0	More than 50	210
Less than 54	10	More than 50	200
Less than 58	40	More than 50	170
Less than 62	130	More than 50	80
Less than 66	190	More than 50	20
Less than 70	210	More than 50	0

The up word combined frequency



The Median=60.89

The down word combined frequency



The Median=60.89

Median properties:

- سهل التعريف وسهل الحساب .
- يعتمد على القيمة الوسطى فقط إذا كانت (n) فردية وعلى القيمتين الوسيطيتين إذا كانت زوجية، ويهمل بقية القيم .
- لا يتأثر بوجود قيم متطرفة .

Ex2: Find the median value of the following data:

(2 , 7 , 6 , 4 , 10 , 1 , 9).

Sol:

1 , 2 , 4 , 6 , 7 , 9 , 10 نرتب الاعداد ترتيبا تصاعديا

n = 7

Median arrangement = $\frac{n+1}{2} = \frac{7+1}{2} = 4$

The Median = 6

لو استبدلنا بدل 10 رقم اخر يعتبر متطرف مثلا 80 فأن ترتيب الوسيط يبقى نفسه وهوالرقم الرابع وقيمته نفسها 6 .

- يمكن إيجاد الوسيط للبيانات النوعية **(الوصفية)** بشرط إمكانية ترتيبها تصاعدياً أو تنازلياً ويكون عددها فردياً .

فمثلاً إذا كان لدينا البيانات النوعية التالية والتي تمثل تقديرات 9 طلبة في مادة الإحصاء :

مقبول، جيد، ضعيف، ممتاز، جيد جداً، مقبول، مقبول، جيد، مقبول الترتيب التصاعدي للبيانات:

ضعيف،مقبول،مقبول، مقبول،مقبول،جيد،جيد، جيد جداً، ممتاز

$$\text{Median arrangement} = \frac{9+1}{2} = 5$$

إذن الوسيط هو التقدير الخامس بعد ترتيب البيانات تصاعدياً ،أي أن الوسيط هو تقدير مقبول.

- يمكن حسابه من جداول التوزيعات التكرارية المفتوحة.
- يمكن إيجاد الوسيط بيانياً.

The Mode (M_o):

المنوال: هو القيمة أو الصفة الأكثر شيوعاً في البيانات ، أي القيمة أو الصفة التي لها أكبر تكرار ، أي التي تتكرر أكثر من غيرها من القيم أو الصفات .

The mode is calculated as follows:

A. Non-classified data:

في هذه الحالة لا توجد أي عمليات حسابية لإيجاد المنوال ، كل ما يتطلبه إيجاد المنوال هو معرفة القيمة التي تتكرر أكثر من غيرها.

Ex2: Find the mode value of the following data:

(11 , 12 , 9 , 10 , 9 , 11 , 13 , 9)

Sol:

The mode of these values is equal 9 ,because it is the value that is repeated the most.

إذا كان في البيانات منوال واحد فتسمى بيانات وحيدة المنوال ، وإذا وجد في البيانات منوالان فتسمى بيانات ثنائية المنوال ، وإذا وجد في البيانات أكثر من منوالين فتسمى بيانات عديدة المنوالات ، وأحياناً لا توجد في البيانات قيمة أو صفة تتكرر أكثر من غيرها من القيم فتسمى بيانات عديمة المنوال .

Ex3: Find the mode for the data reported in each of the following groups:

Group 1: 18 , 14 , 15 , 19 , 20 , 24.

Group 2: 12 , 10 , 12 , 16 , 10 , 15 , 14.

Group 3: 12 , 12 , 15 , 13 , 15 , 13.

Group 4: 12 , 15 , 19 , 13 , 16 , 14 , 15 , 15.

Group 5: 10 , 9 , 15 , 12 , 5 , 11 , 5 , 10 , 12.

Sol:

For Group 1: There is no mode (بيانات عديمة المنوال)

For Group 2: 12 , 10 (بيانات ثنائية المنوال)

For Group 3: 12 , 13 , 15 (بيانات عديدة المنوال)

For Group 4: 15 (بيانات أحادية المنوال)

For Group 5: 5 , 12 , 10 (بيانات عديدة المنوال)

B. Classified data:

لحساب المنوال لبيانات مبوبة في جداول تكرارية منتظمة أي فئاتها متساوية الطول نتبع الخطوات التالية :

- نحدد أولاً الفئة التي تحتوي المنوال ويطلق عليها الفئة المنوالية وهي الفئة المقابلة لأكبرتكرار.
- نحسب قيمة المنوال باستخدام القانون التالي :

$$M = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

حيث:

L: الحد الأدنى للفئة المنوالية (أو الحد الأدنى الحقيقي في حالة البيانات التي تمثل متغيراً منفصلا)

Δ_1 = تكرار الفئة المنوالية – تكرار الفئة السابقة لها.

Δ_2 = تكرار الفئة المنوالية – تكرار الفئة اللاحقة لها.

C: طول الفئة المنوالية .

ملاحظة : إذا كان جدول التوزيع التكراري غير منتظم يجب تعديل تكراراته قبل تطبيق خطوات إيجاد المنوال التي أشرنا إليها .

Ex4: Find the mode value of the following data:

Class limits	f_i
50 – 54	10
54 – 58	30
58 – 62	90
62 – 66	60
66 – 70	20
Σ	210

Sol:

الفئة المنوالية هي الفئة (58 - 62) لأنها هي الفئة التي لها أكبر تكرار ونجد أن:

L=58 C=4 Δ_1 = 90 - 30 = 60 Δ_2 = 90 – 60 = 30

$$M = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

$$M = 58 + \frac{60}{60+30} * 4 = 60.67$$

Determine the Mode Graphically:

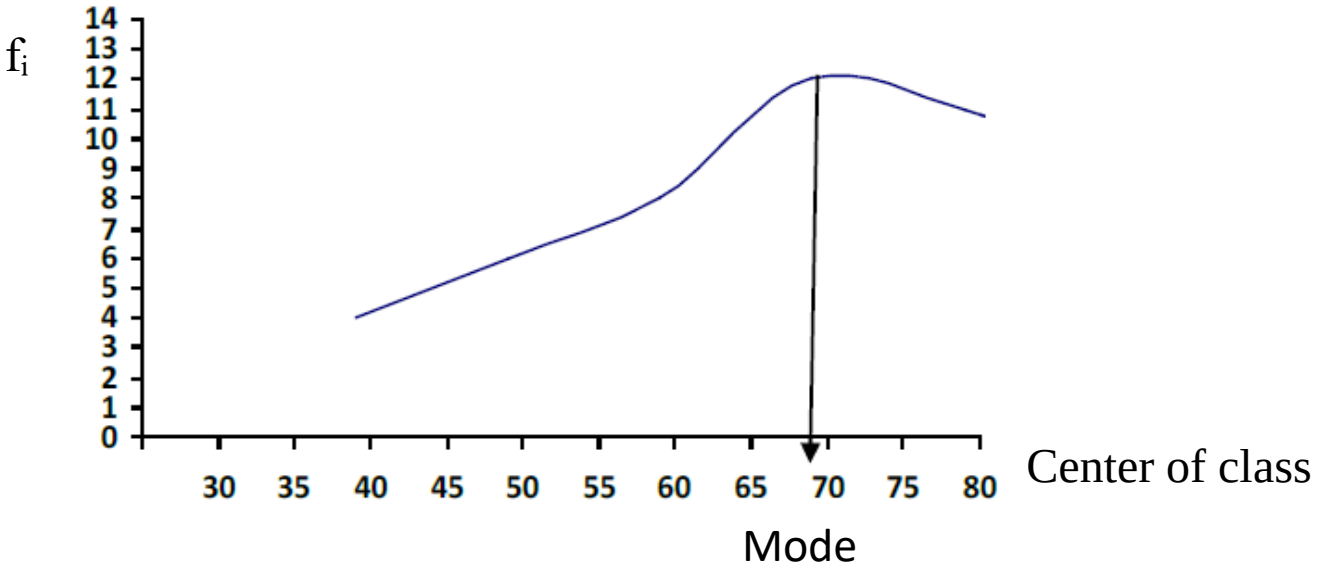
يمكن تحديد قيمة المنوال بيانياً باستخدام المنحنى التكراري أو المضلع فتكون قيمة المنوال هي القيمة المقابلة لقمة المنحنى ، لأن القمة تمثل أكبر تكرار ، وحيث إن المنحنى يكون ممهداً باليد فغالباً ستكون قيمة المنوال التي نتحصل عليها بهذه الطريقة غير دقيقة ، كما يمكن تحديد قيمة المنوال باستخدام المدرج التكراري.

Ex5: Determine the mode Graphically of the following data:

Class limits	f_i
30 – 39	4
40 – 49	6
50 – 59	8
60 – 69	12
70 – 79	11
80 – 89	9

Sol:

Class limits	f_i	Center of class	Point (x , y)
30 – 39	4	34.5	(43.5 , 4)
40 – 49	6	44.5	(44.5 , 6)
50 – 59	8	54.5	(54.5 , 8)
60 – 69	12	64.5	(64.5 , 12)
70 – 79	11	74.5	(74.5 , 11)
80 – 89	9	84.5	(84.5 , 9)



Mode properties:

- أسهل مقاييس النزعة المركزية في حسابه .
- لا يتأثر بوجود قيم متطرفة .
- يمكن حسابه في حالة التوزيعات التكرارية المفتوحة بشرط ألا تكون الفئة المفتوحة هي للفئة المنوالية.
- يمكن إيجاد المنوال للبيانات النوعية.
- ليس له معنى إذا كانت البيانات قليلة العدد وقد لا يوجد أصلاً ، أما في حالة البيانات كثيرة العدد فله معنى معقول وله أهمية كبيرة وخاصة في عملية التسويق ، فمثلاً شركات تسويق الأحذية في مدينة ما لا تهتم بالوسط الحسابي أو بالوسيط بل تهتم بالمقياس الأكثر شيوعاً وهو المنوال .
- يمكن إيجاد المنوال بيانياً .
- قد لا يكون للبيانات منوالٌ وقد تحتوي على منوالين أو أكثر .
- يتأثر كثيراً بطريقة اختيار الفئات التكرارية للتوزيع ، فإذا غيرنا تقسيم الفئات لنفس التوزيع فيحدث تغيراً في التكرارات، وفي الغالب يحدث تغيراً في موقع الفئة المنوالية ، ولذلك نحصل على قيم مختلفة للمنوال .

H.W:

The following data shows the number of teachers in 25 primary schools in a city:

(30 , 24 , 23 , 27 , 28 , 31 , 21 , 25 , 18 , 29 , 29 , 20 , 27 , 28 , 26 , 26 , 24 , 23 , 27 , 28 , 22 , 26 , 26 , 28 , 30)

1. Create a frequency distribution table for this data, using 5 classes of equal length.
2. Find the up word combined frequency for this data.
3. Find the down word combined frequency for this data.
4. Find the arithmetic mean.
5. Find the median value for this data.
6. Determine the Median Graphically for this data.
7. Find the mode value for this data.
8. Determine the mode Graphically for this data.

Statistics

Lecture 9

- **H.W:**

The following data shows the number of teachers in 25 primary schools in a city:

(30 , 24 , 23 , 27 , 28 , 31 , 21 , 25 , 18 , 29 , 29 , 20 , 27 , 28 , 26 , 26 , 24 , 23 , 27 , 28 , 22 , 26 , 26 , 28 , 30)

1.Create a frequency distribution table for this data, using 5 classes of equal length.

2.Find the up word combined frequency for this data.

3.Find the down word combined frequency for this data.

4.Find the arithmetic mean.

5.Find the median value for this data.

6.Determine the Median Graphically for this data.

7.Find the mode value for this data.

8.Determine the mode Graphically for this data.

Sol:

1.

Number of classes = 5.

Rang = 31 – 18 = 13

Class length = $\frac{13}{5}$ = 2.6 = 3

The minimum limit for the first class =18

The actual minimum for the first class = 18 – 0.5 = 17.5

The actual upper limit for the first class = 17.5 + 3 = 20.5

The actual upper limit for the first class = 20.5 – 0.5 = 20

$X_i = \frac{18+20}{2}$ = 19

Table:

Class limits	Actual limits for classes	Centers of classes	Frequencies
18 – 20	17.5 – 20.5	19	2
21 – 23	20.5 – 23.5	22	4
24 – 26	23.5 – 26.5	25	7
27 – 29	26.5 – 29.5	28	9
30 - 32	29.5 – 32.5	31	3
		Σ	25

2.

Actual limits for classes	Frequencies	The up word combined frequency
Less than 17.5	0	0 فئة غير موجودة
17.5 – 20.5	2	2
20.5 – 23.5	4	6
23.5 – 26.5	7	13
26.5 – 29.5	9	22
29.5 – 32.5	3	25
		The total number of data (n) = 25

3.

Actual limits for classes	Frequencies	The down word combined frequency
17.5 – 20.5	2	25
20.5 – 23.5	4	23
23.5 – 26.5	7	19
26.5 – 29.5	9	12
29.5 – 32.5	3	3
More than 32.5	0	0

4.

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i f_i}{\sum_{i=1}^{k} f_i}$$

$$\overline{X} = \frac{19*2+22*4+25*7+28*9+31*3}{25} = 25.84$$

5.

Median arrangement = $\frac{\sum_{i=1}^{k} f_i}{2} = \frac{25}{2} = 12.5$

$$m = L + \left(\frac{\sum_{i=1}^{k} f_i}{2} - F \right) \times c$$

$$f_m$$

L = 24

F = 6

f_m = 7

C = 3

$$\frac{\sum_{i=1}^{k} f_i}{2} = 12.5$$

m = 26.78

6.

نرسم رسما بيانيا ونجد الوسيط حسب الطريقة المشروحة

7.

$$M = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times C$$

L = 27

$\Delta_1 = 9 - 7 = 2$

$\Delta_2 = 9 - 3 = 6$

C=3

M= 27.75

8.

نرسم رسما بيانيا ونجد الوسيط حسب الطريقة المشروحة