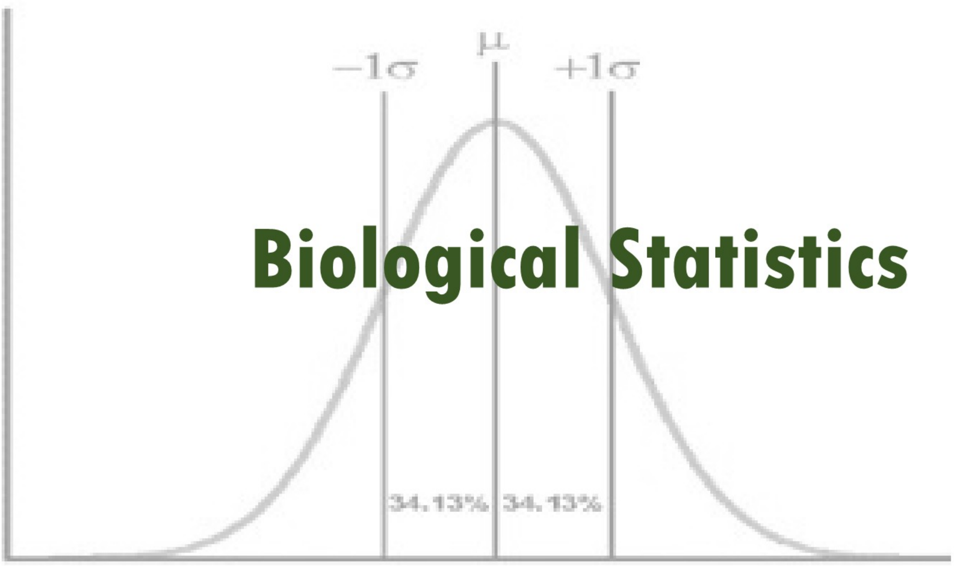
Biological Statistics



Biological statistics ═══════════ *Dr. Labeed Al-Saad*

جدولة البيانات وتمثيلها بيانيا Tabular and graphical presentation ؟

هي الطريقة التي يتم فيها تحويل البيانات الخام Raw data التي جمعت من تجارب الدراسة إلى بيانات مبوبة (معلومات) Grouped data في جداول خاصة تدعى بجداول التوزيع التكراري Frequency distribution tables وهذه الجداول قد تكون بسيطة أو مركبة. والغاية من هذه الجداول هي رصد تكرار المشاهدات أو فئات المشاهدات.

مثال: نريد ان نعرف عدد الذكور و عدد الأناث في العينة أو نريد ان نعرف كم هو عدد الافراد لكل فئة عمرية نحددها نحن؟ فهنا يجري بناء جدول مقسم إلى فئات عمرية محددة أو إلى فئتين (ذكور وأناث) بحيث نأخذ فكرة (معلومة) عن تكرار كل فئة بالعينة.

وهذه الجداول قد تكون بسيطة أو مركبة حسب حاجة الباحث أو الغرض من البحث وفيما يلي شرح مبسط لهذه الجداول:

جدولة البيانات وتمثيلها بيانيا Tabular and graphical presentation ؟

1. الجدول البسيط Simple table: يرتكز هذا الجدول على تمثيل صفة واحدة يجري تقسيمها الى فئات يحدد طولها من قبل الباحث وتحسب التكرارات لكل فئة في حقل خاص ويخصص حقل خاص لمركز الفئة.

مثال: توزيع درجات الطلبة لامتحان مادة الاحصاء البايلوجي في صف معين يحتوي 29 طالب.

مركز الفئة يعني معدل الدرجات في هذه الفئة

فئات درجات الطلبة	تكرار الطلبة	مركز الفئة
25-1	1	13
50-26	3	38
75-51	20	63
100-76	5	88
المجموع	29	

لاحظ بعد ان كانت لدي 29 درجة مختلفة للطلبة اصبحت الصورة بعد تبويبها في جدول تكراري مفهومة واصبح لدينا تصور واضح عن مستويات الطلبة في هذا الصف

فيما تقدم أعلاه ظهرت لدينا بعض المصطلحات الاحصائية التي يجب توضيحها

- الفئة Class: هي عملية تقسيم قيم المتغير إلى مجاميع بحيث ان كل مجموعة تأخذ مدى معين من قيم المتغير ومدى هذه القيم يكون متساوي عادة وتسمى كل مجموعة بالفئة. يعني في مثالنا السابق قسمن متغير الدرجات الذي يبلغ حده الادنى 0 والاعلى 100 درجة إلى اربعة فئات.
- حدود الفئة Class limits: هي الحدين الاعلى الأدنى لكل فئة مثلا من 1 – 25 فحدي الفئة هنا الحد الادنى Lower limit = 1 و الحد الأعلى Upper limit = 25.
- طول الفئة Class width: هو الفرق بين الحدين الادنى أو الحدين الأعلى لفئتين متتاليتين فلو اخذنا اول وثاني فئة في مثالنا السابق يكون طول الفئة (26-1=25) أو (50-25=25) أي ان طول الفئة هو 25.

فيما تقدم أعلاه ظهرت لدينا بعض المصطلحات الاحصائية التي يجب توضيحها

- مركز الفئة Class Mid-point: هو منتصف المدى بين حدي الفئة (الحد الأعلى + الحد الأدنى)/2.
- تكرار الفئة Class frequency: وهو عدد القيم (المشاهدات) التي تظهر ضمن حدود اي فئة ففي مثالنا اعلاه ظهر لدينا ان 20 طالبا كانت درجاتهم ضمن حدود الفئة الثالثة (50-75) اي اننا حصلنا على 20 تكرار ضمن هذه الفئة.

ملاحظة: يتم تحديد عدد الفئات اللازمة لتقسيم العينة اما عن طريق خبرة الباحث وحاجة البحث او بواسطة احد طرق تحديد الفئات مثل طريقة Sturges:

$$No. of\ classes = 1 + (3.3 \times \log n)$$

فيجب ان يصبح عدد الفئات في مثالنا اعلاه بعد تطبيق هذه المعادلة 5.3 أي تقريبا 5 فئات

جدولة البيانات وتمثيلها بيانيا Tabular and graphical presentation ؟

1. الجدول المركب Compounded table: يرتكز هذا الجدول على تمثيل صفتين أو أكثر يجري تقسيمها الى فئات ايضا

مثال: تقسيم طلبة قسم البايلوجي على اساس أطوالهم و اوزانهم.

طول الطالب / وزن الطالب	50-40 كغم	60-50 كغم	70-60 كغم	المجموع
119-110	20	15	5	40
129-120	15	12	8	35
139-130	6	15	14	35
المجموع	41	42	27	110

لاحظ كيف تم ترتيب 220 قيمة في فئات ثنائية الاتجاه (110 للطول و 110 للوزن) في جدول واضح ومفهم أعطى صورة واضحة عن بيانات الدراسة

جدولة البيانات وتمثيلها بيانيا Tabular and graphical presentation ؟

ماذا لو كانت البيانات وصفية وليست كمية؟

مثال: تقسيم طلبة قسم البايلوجي على اساس المرحلة وجنس الطالب.

المرحلة الدراسية / جنس الطالب	الطلبة الذكور	الطلبة الاناث	المجموع
الأولى	18	7	25
الثانية	10	8	18
الثالثة	21	5	26
الرابعة	19	6	25
المجموع	68	26	94

كيفية انشاء جدول تكراري للبيانات

1. حساب الفرق بين أعلى قيمة واقل قيمة لبيانات العينة (رتب البيانات تصاعديا او تنازليا لتسهيل ذلك).
2. تحديد عدد الفئات اما عن طريق خبرة الباحث أو بواسطة احدى طرق تحديد الفئات التي مر ذكرها.
3. حساب طول الفئة بقسمة ناتج الفرق بين اعلى و ادنى قيمة (الخطوة 1) على عدد الفئات (الخطوة 2) وتقريب الناتج لأقرب قيمة صحيحة.
4. كتابة الحدود الدنيا لكل الفئات بحيث نختار الحد الادنى لأول فئة ولتكن اقل قيمة في العينة أو اصغر منها قليلا ثم تكون القيمة الدنيا للفئة الثانية اكبر منها بمقدار طول الفئة وهكذا ثم نكتب الحدود العليا لجميع الفئات ابتداءً من الفئة الاولى التي يكون الحد الأعلى لها اقل بدرجة عن الحد الادنى للفئة التالية والحد الأعلى التالي يكبره بمقدار طول الفئة وهكذا.

كيفية انشاء جدول تكراري للبيانات

مثال: أنشيء جدولا تكراريا لعينة عدد كريات الدم لـ32 شخص اختيرت عشوائيا:

RBC	3.2	2.4	2.5	4.1	2.5	2.1	3.1	2.6	3.7	2.6	2.4
	3.6	3.7	4.11	4.2	3.2	2.9	2.92	2.66	1.8	1.95	1.8
	3.2	3.3	3.1	2.5	2.44	3.6	3.11	2.88	4.1	5.2	

الحل:

1. نحدد أعلى وادنى قيمة: و ونحسب الفرق بينهما : 5.2 - 1.8= 3.4
2. نحدد عدد الفئات حسب المعادلة $\boldsymbol{No.of\ classes = 1 + (3.3 \times \log n)}$= 5.5 اي تقريبا 6.
3. نحدد طول الفئة بقسمة الناتج الاول على الناتج الثاني: 3.4 / 6 = 0.56 تقريبا 0.6

فعليه يكون الجدول كالتالي:

كيفية انشاء جدول تكراري للبيانات

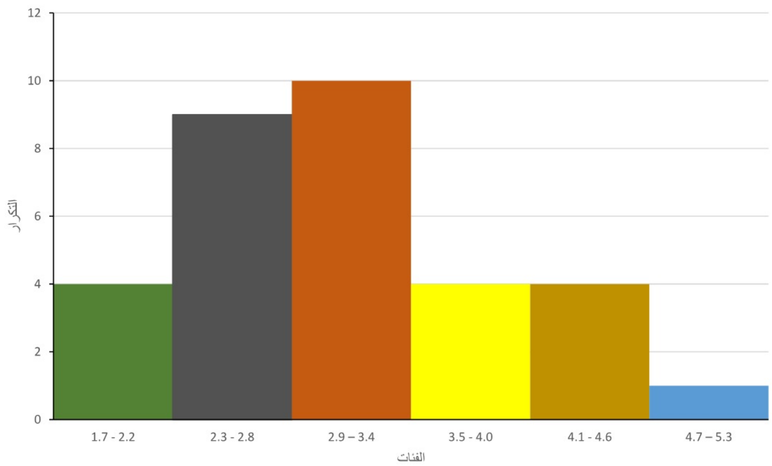
فئات عدد الكريات الحمر	التكرار	مركز الفئة
2.2 - 1.7	4	1.95
2.8 - 2.3	9	2.55
3.4 – 2.9	10	3.15
4.0 - 3.5	4	3.75
4.6 - 4.1	4	4.35
5.3 – 4.7	1	5
المجموع	32	

رسم المدرج والمضلع والمنحنى التكراري

لرسم المدرج التكراري Histogram نتبع الخطوات التالية اعتمادا على الجدول التكراري الذي سبق انشاؤه:

1. نرسم المحورين السيني والصادي.
2. نقسم المحور السيني على اساس الحدود الحقيقية للفئات.
3. نقسم المحور الصادي على اساس التكرارات.
4. نمثل كل فئة برسم مستطيل من بداية الحد الحقيقي للفئة صعودا الى قيمة التكرار ثم عرضيا إلى الد الحقيقي للفئة التالية وقد تترك مسافة بين المستطيلات خصوصا في حالة الفئات الوصفية غير الرقمية.

رسم المدرج التكراري Histogram

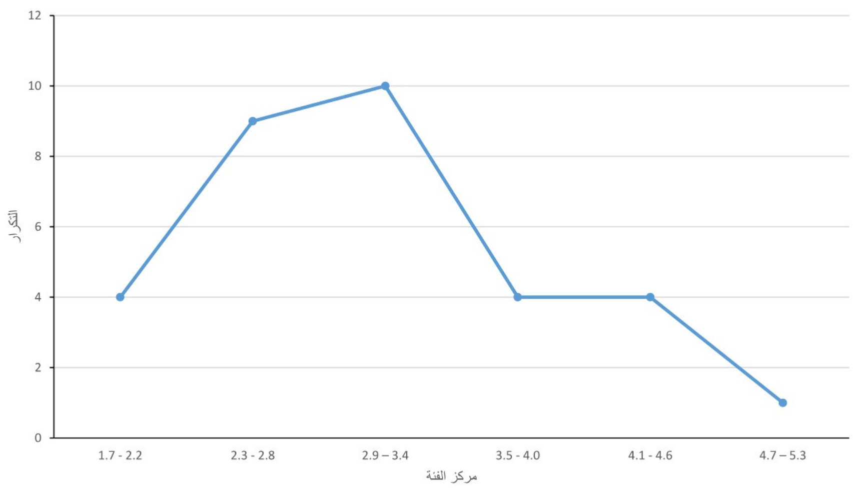


رسم المدرج والمضلع والمنحنى التكراري

لرسم المضلع التكراري Frequency polygon

1. نرسم المحورين السيني والصادي.
2. نقسم المحور السيني على اساس الحدود الحقيقية للفئات.
3. نقسم المحور الصادي على اساس التكرارات.
4. نمثل مركز كل فئة برسم بنقطة تدل على التكرار الخاص بتلك الفئة.
5. نصل بين النقاط بخطوط مستقيمة.

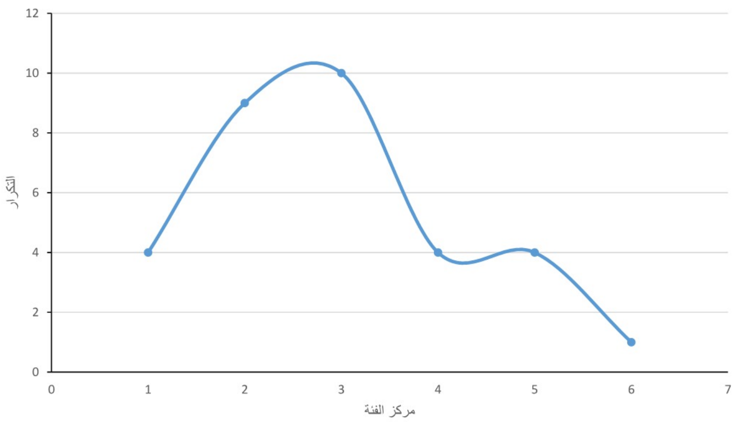
Freequency polygon رسم المضلع التكراري

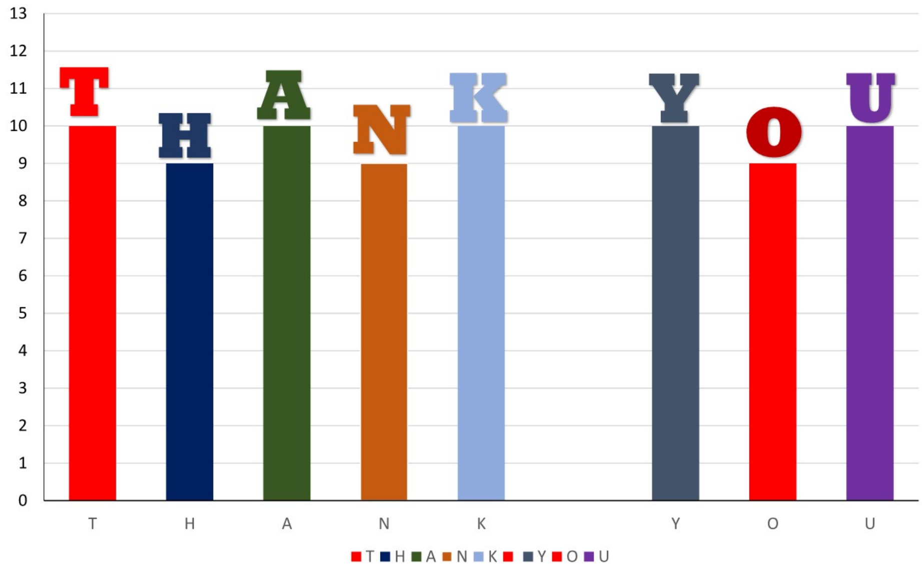


رسم المنحنى التكراري

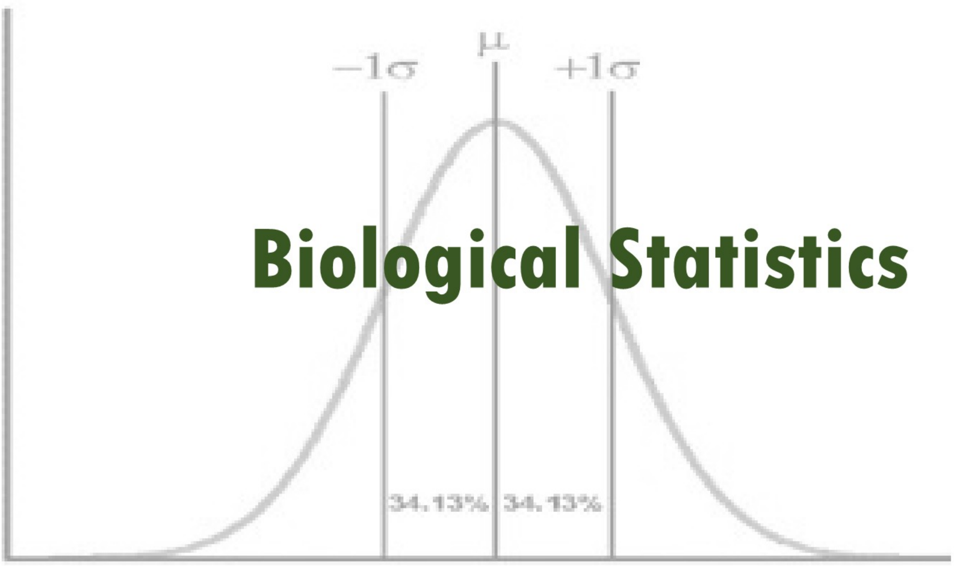
لرسم المنحنى التكراري Frequency curve

نتبع نفس خطوات رسم المضلع غير أننا نصل بين النقاط بخطوط منحنية.





13
12
11
10
9
8
7
6
5
4
3
2
1
0
T H A N K Y O U
T
H
A
N
K
Y
O
U



Biological Statistics

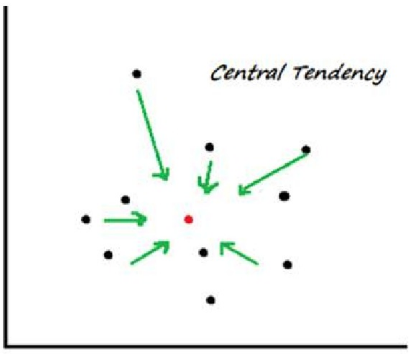
بعض الرموز الاحصائية المهمة

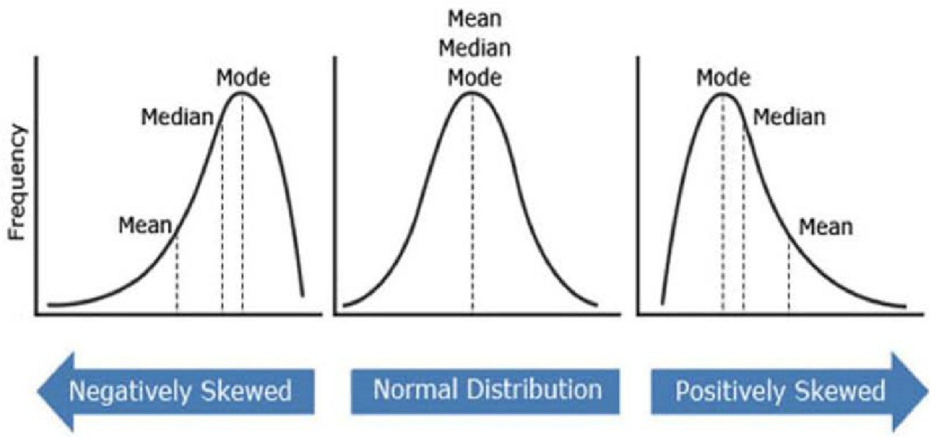
$\sum$	Sum of or summation	n!	N Factorial
$\neq$	Not Equal	α	0.05 statistical level
$>$	Greater than	β	0.01 statistical level
$<$	Less than	$\lvert\ \ \rvert$	Absolute value
μ	Population mean	$\bar{x}$	Sample mean
σ^2 σ	Variance and standard deviation of the population	S^2 S	Variance and standard deviation of the sample
$\hat{Y}$	Y hat: indicate an estimate rather than true	X or Y or Z	A variable
Xi	Individual observation	$\sum Xi$	Total sum of servation
n	Sample size	$(\sum xi)^2$	Square total sum
$\sum Xi^2$	Total sum of square	C.V.	Coefficient of ariation
d	Difference (x1 –x2)	b	Sample regression coefficient
r	Sample correlation	df	Degrees of freedom
SS	Sum of square	MS	Mean square
ns	Not significant	(*) and (**)	Significant and highly significant
H_o	Null hypothesis	H1	Alternative hypothesis

هذه الرموز ستمر علينا اثناء مسيرة الكورس الدراسي لذا فمن المهم ان نطلع عليها ونفهم معانيها

Central tendency measurements مقاييس النزعة المركزية

هي المقاييس التي تتمركز حولها القيم ومن اهمها المتوسط الحسابي (المعدل) والوسيط (القيمة التي تقع في الوسط) والمنوال (القيمة الاكثر تكرارا).





The Arithmetic Mean المتوسط الحسابي

هو عبارة عن معدل قيم العينة = مجموع القيم / عددها

$$\bar{X} = \frac{\sum xi}{n} = \frac{x1 + x2 + x3 + \cdots + xn}{n}$$

حيث ان: ($\bar{X}$ = المتوسط الحسابي) ($\sum xi$ = مجموع القيم) و (n = عددها)

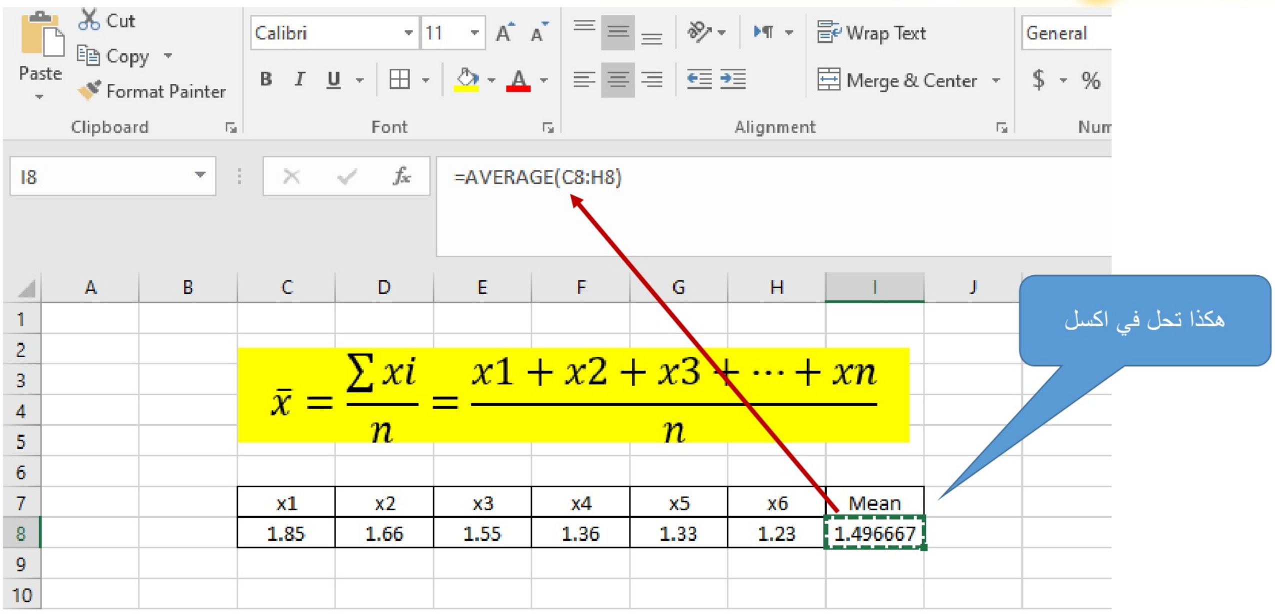
مثال: جد المتوسط الحسابي للقيم التالية:

1.4 1.23 1.33 1.36 1.55 1.66 1.85

الحل: عدد القيم (n) = 7 قيم

$$\bar{X} = \frac{\sum xi}{n} = \frac{1.85 + 1.66 + 1.55 + \cdots + 1.4}{7} = 1.4966$$

The Arithmetic Mean المتوسط الحسابي



خواص المتوسط الحسابي

1. مجموع انحرافات القيم عن متوسطها الحسابي يساوي صفر.

Xi	12	3	5	8	7
$\bar{X}$-xi	7-12 = -5	7-3 = 4	7-5 = 2	7-8 = -1	7-7 = 0

$\bar{X} = 7$

$\sum(\bar{X} - xi)$ = -5 + 4 + 2 -1 +0 = 0

$\sum(\bar{X} - xi)^2$ = 25 + 16 +4 +1 + 0 = 46

2. مجموع مربعات انحرافات القيم عن متوسطها الحسابي هو اقل ما يمكن (يعني لو استبدلنا المتوسط باي قيمة من قيم العينة وقسنا مربعات الانحرافات عنها لكانت اكبر من قيمة مجموع مربعات الانحرافات عن المتوسط).

جرب استبدال قيمة المتوسط باي قيمة من قيم العينة الخمسة وستلاحظ ان مجموع مربعات الانحرافات يكون اكبر من 46

Xi	12	3	5	8	7
$\bar{X}$-xi	7-12 = -5	7-3 = 4	7-5 = 2	7-8 = -1	7-7 = 0
$(\bar{X} - xi)^2$	25	16	4	1	0

$\bar{X} = 7$

$\sum(\bar{X} - xi)^2$ = 25 + 16 + 4 +1 + 0 = 46

خواص المتوسط الحسابي

3. اذا اضيف أو طرح عدد ثابت الى جميع القيم فان قيمة المتوسط تتغير بمقدار القيمة الثابتة المضافة أو المطروحة.

4. اذا ضربت كل القيم بعدد ثابت فان المتوسط الحسابي الجديد = المتوسط القديم × العدد الثابت.

تأكد من النقطتين اعلاه بنفسك

5. الوسط الحسابي لمجموع قيم متغيرين = مجموع الوسطين الحسابيين.

	xi	yi	xi+yi
	10	13.16	23.16
	12	4	16
	3	12	15
	5	7	12
Mean	7.5	9.04	16.54

لاحظ معدل المجموع = مجموع المعدلين

المتوسط الحسابي الموزون The Weighing Mean

هو حاصل قسمة (مجموع (حاصل ضرب القراءات في اوزانها)) على مجموع الاوزان.

$$\overline{WX} = \frac{\sum wixi}{\sum wi} = \frac{(w1 * x1) + (w2 * x2) + (w3 * x3) + \cdots + (wn * xn)}{w1 + w2 + w3 + \cdots + wn}$$

حيث ان: ($\overline{WX}$ = المتوسط الحسابي الموزون) ($\sum wixi$ = مجموع القيم × اوزانها) و (wi = وزن القيمة i)

ماذا نقصد بالاوزان ؟؟؟؟؟

الوزن هو تعبير عن اهمية القيمة فمثلا لو كانت لديك 10 دنانير عراقية و 10 دولارات فبرغم ان القيمة العددية هي 10 الا ان القيمة الموزونة للـ10 دولارات تكون 10 x 1300= 13000 دينارا فالـ1300 هو وزن كل دولار بينما وزن كل دينار هو 1 فقط. وكمثال اخر درجات المرحلة الأولى في الكلية وزنها 10% في المعدل النهائي بينما درجات المرحلة الرابعة وزنها 40% اي ان قيمة الدرجة في المرحلة الرابعة =0.4 من درجة المعدل النهائي بينما الدرجة في المرحلة الاولى=0.1 من درجة المعدل النهائي.

The Weighing Mean المتوسط الحسابي الموزون

مثال: إذا كانت معدلات احد الطلبة لأربعة مراحل دراسية هي (60، 70، 80، 75%) على التوالي وكان وزن معدل المرحلة الأولى 10% والمرحلة الثانية 20% والمرحلة الثالثة 30% والمرحلة الرابعة 40% فكم سيكون المعدل النهائي للطالب؟

الحل:

المرحلة	معدل الطالب % (xi)	وزن المعدل % (wi)	المعدل الموزون wi × xi
الاولى	60	10	600
الثانية	70	20	1400
الثالثة	80	30	2400
الرابعة	75	40	3000
المجموع		100	7400

$$\overline{WX} = \frac{\sum wixi}{\sum wi} = \frac{600 + 1400 + 2400 + 3000}{100} = 74$$

The Weighing Mean المتوسط الحسابي الموزون

I10 =SUM(H10:H13)/SUM(G10:G13)

$$\overline{WX} = \frac{\sum wixi}{\sum wi} = \frac{(w1 * x1) + (w2 * x2) + (w3 * x3) + \cdots + (wn * xn)}{w1 + w2 + w3 + \cdots + wn}$$

المرحلة	معدل الطالب % (xi)	وزن المعدل % (wi)	المعدل الموزون Wi × xi	المتوسط الموزون WX
الاولى	60	10	600	74
الثانية	70	20	1400	
الثالثة	80	30	2400	
الرابعة	75	40	3000	

هكذا تحل في اكسل

=SUM(H10:H13)/SUM(G10:G13)

=F10*G10

The Median الوسيط

هي القيمة التي تقع في الوسط بعد ترتيب قيم العينة تصاعديا أو تنازليا.

❖ اذا كان عدد القيم فردي فان ترتيب الوسيط = عدد القيم + 1 مقسوما على 2.

❖ اذا كان عدد القيم زوجي فان قيمة الوسيط = عدد القيم مقسوما على 2 ثم ناخذ هذه القيمة والتي بعدها ونستخرج معدلهما والذي يمثل قيمة الوسيط.

في حالة عدد القيم فردي

3، 2، 4، 7، 5، 10، 9

⬇

2، 3، 4، 5، 7، 9، 10

ترتيب الوسيط = (7+1)/2 = 4 اي القيمة الرابعة (5)

في حالة عدد القيم زوجي

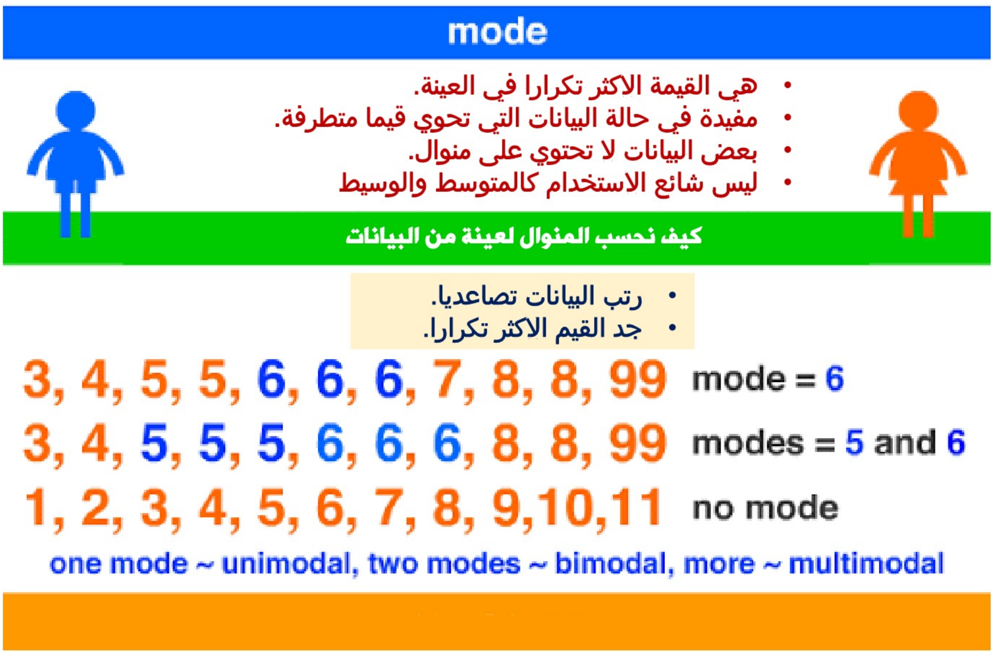
13، 19، 3، 6، 4، 10

⬇

3، 4، 6، 10، 13، 19

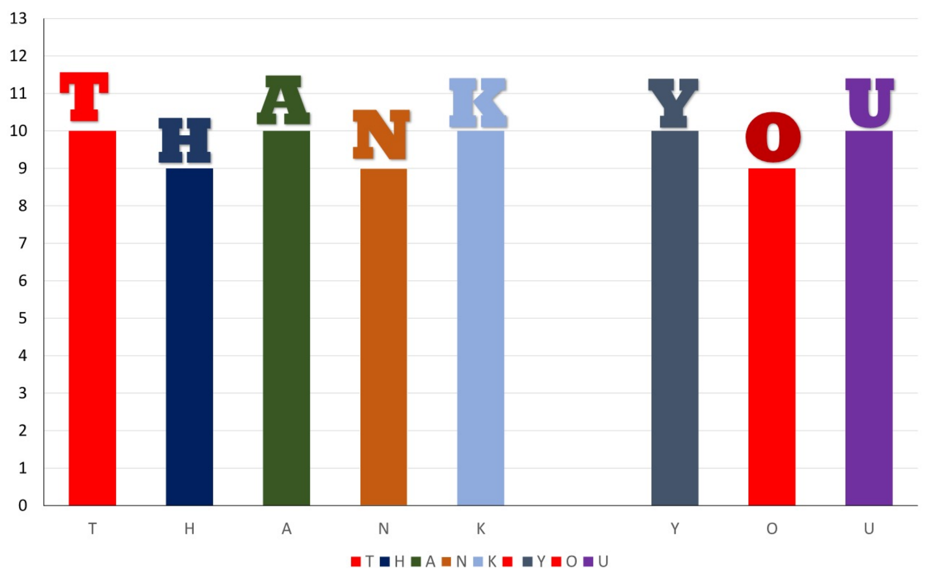
قيمة الوسيط =(6+10)/2 = 8

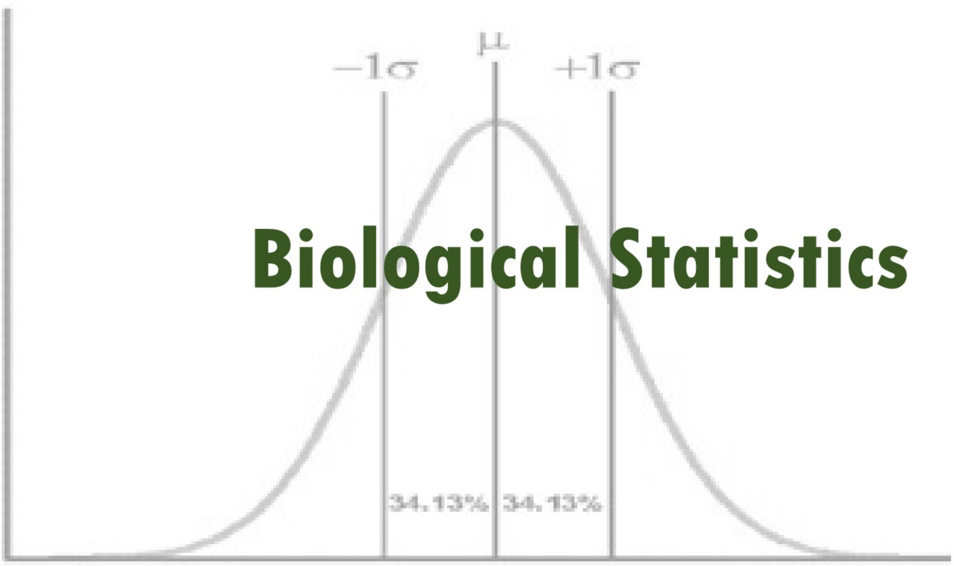
The Mode المنوال




MEAN
The "mean" is the "average". To find the mean, you add up all the numbers and then divide by the number of numbers.
TO FIND THE MEAN FOR THIS SET OF NUMBERS: 13, 18, 13, 14, 13, 16, 14, 21, 13
average the set of numbers:
(13 + 18 + 13 + 14 + 13 + 16 + 14 + 21 + 13) ÷ 9 = 15
Note that the mean isn't a value from the original list. This is a common result.
DO NOT assume that the mean will be one of the original numbers.
MEDIAN
The "median" is the "middle" value in the list of numbers. To find the median, your numbers have to be listed in numerical order, so you may have sort the list first.
FOR AN ODD NUMBER OF VALUES: 1,5,2,8,7
Sort the numbers 1, 2, 5, 7, 8
FOR AN EVEN NUMBER OF VALUES: 1,5,2,10,8,7
Sort the numbers: 1, 2, 5,7, 8, 10.
TAKE THE AVERAGE OF THE TWO MEAN NUMBERS: (5+7)/2 = 6
MODE
The "mode" is the value that occurs most often. If no number is repeated, then there is no mode for the list.
TO FIND THE MODE FOR THIS SET OF NUMBERS: 13, 18, 13, 14, 13, 16, 14, 21, 13
Sort the numbers: 13, 13, 13, 13, 14, 14, 16, 18, 21

الخلاصة


T
H
A
N
K
Y
O
U
13
12
11
10
9
8
7
6
5
4
3
2
1
0
T H A N K Y O U



Biological Statistics

Dispersion of Variation measurements مقاييس التشتت أو الاختلاف

هي المقاييس التي تحاول ايجاد قيمة تدل على مدى تشتت أو تباعد قيم المتغير (المشاهدات) عن وسطها الحسابي لإعطاء صورة عن مدى تجانس القيم (اي هل هناك تباين فيما بينها أم هي متجانسة)

ومن أهم هذه المقاييس:

1. المدى Range.
2. التباين Variance.
3. الانحراف القياسي Standard deviation.
4. الخطأ القياسي Standard error

توضيح:

لو كان لدينا مجموعتين:

X = 17 16 15 18 19 17

Y = 5 11 20 50 13 7

فان المتوسط الحسابي للـ X = 17

وكذلك المتوسط الحسابي للـ Y = 17

ولكن المجموعة Y قيمها اكثر تشتتا من قيم المجموعة X

المدى (Range (R

هو الفرق بين أعلى وادنى قيمة في المشاهدات.

كيف نحسبه ؟ !

1. نرتب القيم تصاعديا أو تنازليا.
2. نختار أعلى قيمة .Max وأدنى قيمة .Min
3. نطبق المعادلة التالية: $R = Max. - Min.$

مثال: جد المدى لقيم المتغيرات التالية:

X = 2 5 7 6 10 6 7

Y = 8 13 9 16 8 10 9

الحل: نرتب القيم تصاعديا

X = 2 5 6 6 7 7 10 R= 10 – 2 = 8

Y = 8 8 9 9 10 13 16 R= 16 – 8 = 8

المدى معيار غير دقيق لأنه يعتمد على اكبر
واقل قيمة ففي المثال نلاحظ ان العينتين
لديهما نفس المدى لكن لو حسبنا قيمة التباين
لكانت قيم الثانية متباينة اكثر من الاولى

Variance (S ² or σ ²) التباين

من أكثر مقاييس التشتت استخداما ويمكن أن يعرف على انه معدل مربعات انحرافات القيم عن متوسطها الحسابي.

كيف نحسبه ؟ !

هناك طريقتين لحسابه:

طريقة انحرافات القيم	طريقة تربيع القيم (أكثر تفضيلا)
$$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$	$$S^2 = \frac{\sum X_i{}^2 - \frac{\sum(X_i)^2}{n}}{n-1}$$

مثال: احسب التباين للعينة التالية:

X = 22 21 19 18 23 20 17

نحسبه بطريقة انحرافات القيم

1. نحسب المتوسط الحسابي:

$$\bar{X}_i = \frac{22+21+18+23+20+17}{7} = 20$$

2. نشكل جدول الانحرافات ومربعات الانحرافات وكالتالي:

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
22	22-20 = 2	4
21	21-20 = 1	1
19	19-20 = -1	1
18	18 – 20 = -2	4
23	23 – 20 = 3	9
20	20 – 20 = 0	0
17	17 – 20 = -3	9
140	$\sum X_i - \bar{X}$= 0	$\sum(X_i - \bar{X})^2$= 28

ثم

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

$$S^2 = \frac{28}{7-1} = 4.67$$

مثال: احسب التباين للعينة التالية:

X = 22 21 19 18 23 20 17

نحسبه بطريقة تربيع القيم

نشكل جدول القيم مربعاتها:

X_i	${X_i}^2$
22	464
21	441
19	361
18	324
23	529
20	400
17	289
$\sum X_i$ =140	$\sum {X_i}^2$= 2828

ثم

$$S^2 = \frac{\sum {X_i}^2 - \frac{\sum (X_i)^2}{n}}{n-1}$$

$$S^2 = \frac{2828 - \frac{(140)^2}{7}}{7-1} = 4.67$$

Standard deviation (S) الانحراف القياسي

وهو الجذر التربيعي للتباين وهو من المقاييس الشائعة الاستخدام ايضا بل يستخدم اكثر من التباين نفسه وسبب استخدامه اننا لو رعنا الانحرفات في التباين فيجب علينا ان نربع الوحدات فمثلا:

سم يصبح سم^2 فتفهم على انها مساحة وليست تباين طول فعند جذر قيمة مربع الانحراف فان الوحدات ستعود لطبيعتها. بناءً على ذلك فان النحراف القياسي يكون:

طريقة انحرافات القيم

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

طريقة تربيع القيم (أكثر تفضيلا)

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

عليه فالانحراف القياسي لمثالنا السابق يكون

مثال: احسب الانحراف القياسي للعينة التالية:

X = 22 21 19 18 23 20 17

طريقة انحرافات القيم

$$S = \sqrt{\frac{28}{7-1}} = 2.16$$

طريقة تربيع القيم (أكثر تفضيلا)

$$S = \sqrt{\frac{2828 - \frac{(140)^2}{7}}{7-1}} = 2.16$$

الخطأ القياسي Standard Error (SE)

وهو عبارة عن مقياس يخبرنا ما مدى دقة الوسط الحسابي للعينة أي هل ان هذا الوسط هو فعلا تقدير حقيقي لوسط المجتمع أم لا. وكلما قلت قيمة الخطأ القياسي كلما زادت دقة تقدير الوسط الحسابي واعطانا ذلك انطباعا عن دقة حسابه ومدى تجانس قيم العينة.

كيف نحسبه ؟ !

يحسب ببساطة بقسمة الانحراف القياسي على جذر عدد قيم العينة:

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

δ= standard deviation

n= number of samples

توضيح:
ان S هي نفسها σ غير أن S تستخدم للعينة و σ تستخدم للمجتمع أما من حيث الحساب فهي نفس الشيء

عليه فالخطأ القياسي لمثالنا السابق يكون

مثال: احسب الخطأ القياسي للعينة التالية:

X = 22 21 19 18 23 20 17

طريقة تربيع القيم (أكثر تفضيلا)

$$S = \sqrt{\frac{2828 - \frac{(140)^2}{7}}{7-1}} = 2.16$$

طريقة انحرافات القيم

$$S = \sqrt{\frac{28}{7-1}} = 2.16$$

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

$$SE = \frac{2.16}{\sqrt{7}} = 0.816$$

حساب التباين والانحراف القياسي والخطأ القياسي للبيانات المبوبة (التي تحتوي على فئات وتكرارات)

يقصد بهذه البيانات البيانات التي تكون على شكل جداول تكرارية يعني ان القيم غير جاهزة وانما مقسمة الى فئات وكل فئة تحتوي على تكرار مثلا: (10-20) هناك 5 قيم لكننا في هذه الحالة لا نعلم بالضبط كم تساوي كل قيمة بالضبط لأنها مبهمة ولا نعرف عنها شيء سوى انها ضمن حدود الفئة. عموما المعادلة العامة لهذا النوع من التباين هي:

معادلة تباين البينات المبوبة

$$S^2 = \sum FiX_i{}^2 - \frac{\sum(FiX_i)^2}{\sum Fi}$$

حيث أن:

X_i: مركز الفئة = (الحد الاعلى +الحد الانى)/2

Fi: قيمة التكرار

حساب التباين والانحراف القياسي والخطأ القياسي للبيانات المبوبة (اي : التي تحتوي على فئات وتكرارات)

مثال: احسب التباين والانحراف القياسي والخطأ القياسي للبيانات المبوبة التالية:

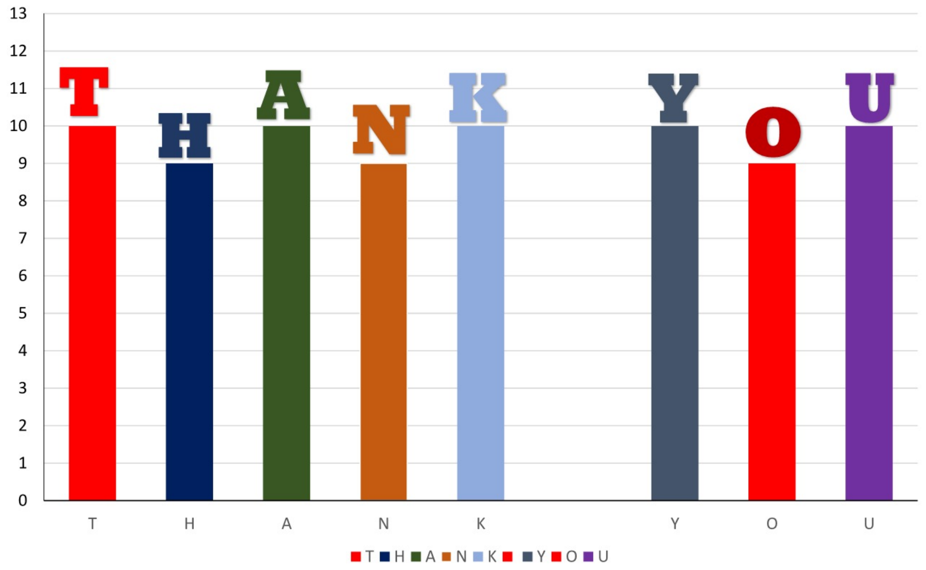
مدخلات السؤال		خطوات الحل			
الفئات	Fi	X_i	FiX_i	X_i^2	FiX_i^2
60-62	5	(60+62)/2= 61	(5 ×61) = 305	3721	(5 × 3721) = 18605
63-65	18	(63+65)/2= 64	(18 ×64) = 1152	4096	(18 ×4096) = 73728
66-68	42	(66+68)/2= 67	(42 ×67) = 2814	4489	(42 × 2814) = 188538
69-71	27	(69+71)/2= 70	(27 ×70) = 1890	4900	(27 × 4900) = 132300
72-74	8	(72+74)/2= 73	(8 ×73) = 584	5329	(8 × 5329) = 42632
	$\sum Fi = 100$		$\sum F_iX_i = 6745$		$\sum FiX_i^2 = 455803$

$$S^2 = \sum FiX_i^2 - \frac{\sum(FiX_i)^2}{\sum Fi}$$

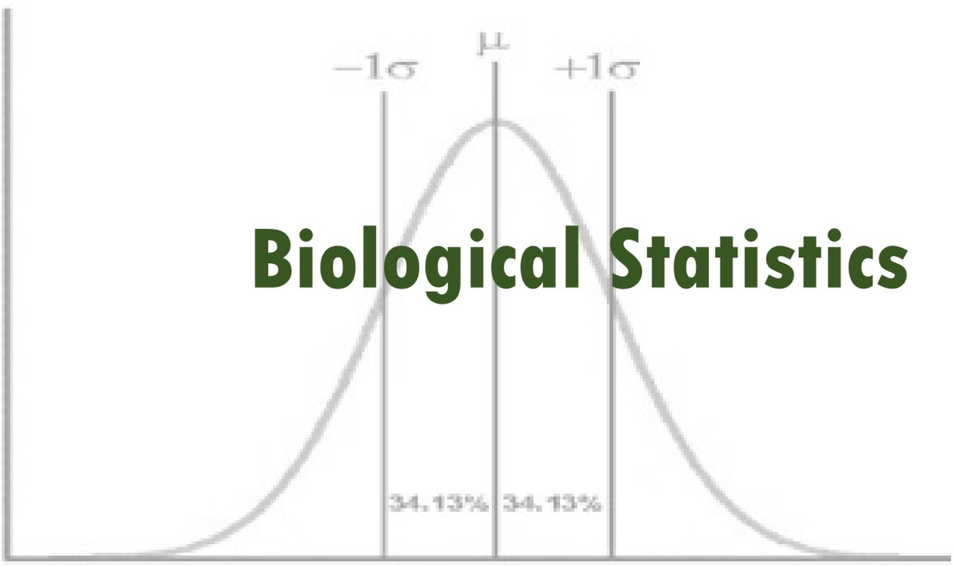
$$S^2 = 455803 - \frac{(6745)^2}{100} = \cancel{8.61}$$

$$S = \sqrt{8.61} = \cancel{2.9}$$

$$SE = \frac{2.9}{\sqrt{100}} = \cancel{0.29}$$



T
H
A
N
K
Y
O
U
13
12
11
10
9
8
7
6
5
4
3
2
1
0
T
H
A
N
K
Y
O
U
T H A N K Y O U



Biological Statistics

العينة Sample

ما هي العينة؟

العينة جزء ممثل عن المجتمع .

ما هي شروطها؟

1. العشوائية: ان تكون العينة غير متحيزة.
2. متوزعة طبيعيا: عدد التوقعات الصحيحة يساوي تقريبا عدد التوقعات الخاطئة.
3. أن تكون كل مشاهدة مستقلة عن الاخرى.

حساب حجم العينة الافتراضي Sample size

لماذا نحسب حجم العينة الافتراضي؟

ج: لكي تكون العينة ممثلة للمجتمع بحيث ان الفرق بين متوسط العينة ومتوسط المجتمع الحقيقي يكون اقل مايمكن (غير معنوي) وبذلك فان كل ما يترتب على حجم العينة من عمليات احصائية ستكون نتائجها دقيقة وتعطي قراءة واضحة عن المجتمع وبالتالي فان اي قرارات ستتخذ ستكون صائبة.

حساب حجم العينة الافتراضي Sample size

كيف نحسب حجم العينة الافتراضي؟

حيث أننا يجب أن نوفر المتطلبات التالية لحل المعادلة:

1. مستوى الثقة Confidence level وهو على الاغلب = 0.95%
2. احتمالية نسبة تواجد حاملي الصفة قيد البحث ($\hat{p}$) في المجتمع والحالة الطبيعية إذا كان المجتمع متوزع طبيعيا فانه من المفروض ان يكون نصفهم يمتلك الصفة أي انها = 0.5
3. هامش الخطأ Error margin: بما اننا اخترنا مستوى ثقة 95% اي نسبة الخطأ 5% فهامش الخطأ سيكون اقل بقليل = 0.04
4. حجم المجتمع الكلي (N): وليكن 5000 فرد
5. مستوى المعنوية مقسوما على 2: 0.05 / 2 = 0.025
6. نستخدم دالة NORM.S.INV() في اكسل لحساب قيمة Z من خلال وضع قيمة مستوى المعنوية / 2 مطروحة من 1 بين قوسي الدالة:-

= NORM.S.INV(1-0.025) = 1.96

7. نقوم بحل المعادلة.

$$Sample\ size = \frac{\dfrac{Z^2 \times \hat{p} \times (1-\hat{p})}{e^2}}{1 + \left(\dfrac{Z^2 \times \hat{p} \times (1-\hat{p})}{e^2 \times N}\right)}$$

$$Sample\ size = \frac{\dfrac{1.96^2 \times 0.5 \times 0.5}{0.04^2}}{1 + \left(\dfrac{1.96^2 \times 0.5 \times 0.5}{0.04^2 \times 5000}\right)} = 535.91$$

ملاحظة: يمكن الحصول على قيمة Z باستخدام جداول احصائية خاصة تدعى بجدول قيم Z متوفرة على الانترنيت أو من خلال برامج Z calculators على الانترنيت ايضا

Confidence intervals حدود الثقة

قبل شرح حدود الثقة لابد ان نعرف مفهوم يدعى مستوى الثقة Confidence level

هو عبارة عن نسبة مئوية (95% مثلا) وهي تعني انك لو أعدت اجراء نفس التجربة أو المسح الميداني مرات ومرات فانك 95% ستحصل على نتائج مطابقة لقراءات المجتمع أي بمعنى آخر احصائك سيكون صحيح تماما ويعكس ما يحدث في المجتمع فعلا.

ما هي حدود الثقة؟

هي مقياس لمدى وثوقية طريقة جمع العينات. حيث انها تحدد القيم العليا والدنيا التي يجب ان تقع بينها قيم الصفة المدروسة في مجتمع معين ضمن مستوى الثقة المحدد للعينة مهما كررت مرات اخذ العينات.

Confidence intervals حدود الثقة

مثال مبسط:

لو أردت أن آخذ عينة من طلبة كلية العلوم لدراسة متوسط طول الطلبة عند مستوى ثقة 95% وكانت حدود الثقة (150-180) وكانت هذه العينة مطابقة لشروط العينة الصحيحة فاني لو اعدت اخذ عينة من هذا المجتمع مرات ومرات فان قيم الطول التي سأقيسها 95% ستقع ضمن مدى حدود الثقة. واذا لم يحدث ذلك فهذا يعني انني لم أكن موفقا في اختيار العينة وان اي عمليات احصاء ستجرى عليها لا تمتلك المصداقية وان اي قرارات ستتخذ بناءً على النتائج المستحصلة ستكون خاطئة تماما وغير موثوقة.

Confidence intervals حدود الثقة

متى نستخدم حدود الثقة؟

عندما لا نمتلك معلومات عن متوسط المجتمع. في هذه الحالة لا نستطيع ان نقارن مدى اقتراب او ابتعاد متوسط العينة عن متوسط المجتمع وبالتالي لا نستطيع ان نعرف هل ان عينتنا ممثلة للمجتمع ام لا؟

كيفية حساب حدود الثقة Confidence intervals

يتم حساب حدود الثقة حسب المعادلة التالية:

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

CI= Confidence interval.
$\bar{x}$ = Sample mean.
z = Confidence level value.
s= Sample standard deviation
n= Sample size

$$z \frac{s}{\sqrt{n}}$$

هذه تسمى حدود الخطأ Error margins وعند اضافتها أو طرحها من معدل العينة $\bar{x}$ نحصل على الحدود الدنيا والعليا للثقة

$$S = \sqrt{\frac{\sum_{i=1}^{n}(Xi - \bar{X})^2}{n-1}}$$

هذا هو الخطأ القياسي وهو جذر التباين كما تعلمنا سابقا

Z = قيمة t عند درجات حرية = n-1
ومستوى معنوية = 1- confidence level

كيفية حساب حدود الثقة Confidence intervals

مثال: أخذت عينة لأطوال 10 طلبة كعينة ممثلة لمجتمع طلبة قسم علوم الحياة لتحديد معدل الطول العام لطلبة القسم. المطلوب تحديد حدود الثقة ضمن مستوى ثقة 95% للتعرف على مدى وثوقية العينة المختارة ومطابقة قياساتها لقياسات مجتمع الطلبة في القسم.

أطوال الطلبة: 155 176 180 175 160 155 150 166 179 162

الحل:

$$CI = \bar{x} \pm z \frac{s}{\sqrt{n}}$$

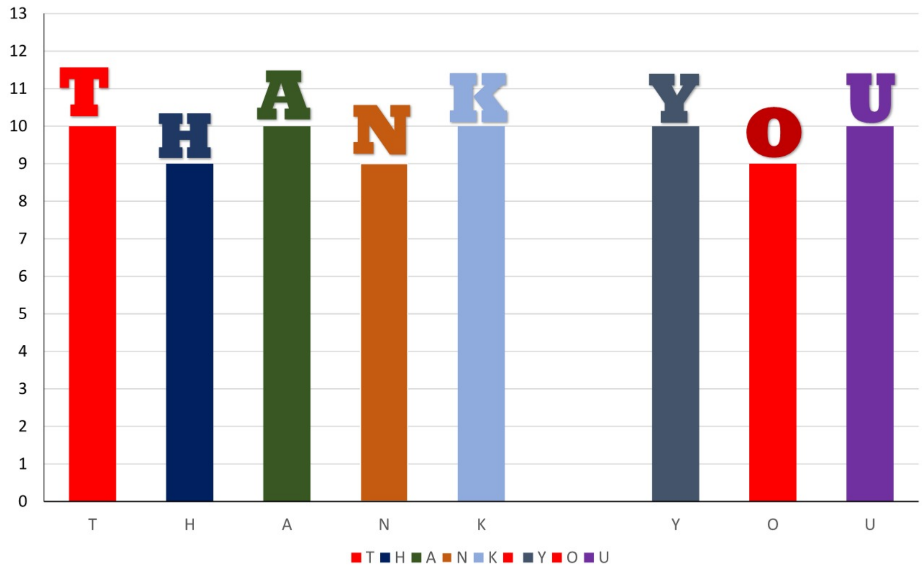
Xi	150	155	160	175	180	176	155	166	179	162
$(Xi - \bar{X})^2$	249.64	116.64	33.64	84.64	201.64	104.04	116.64	0.04	174.24	14.44

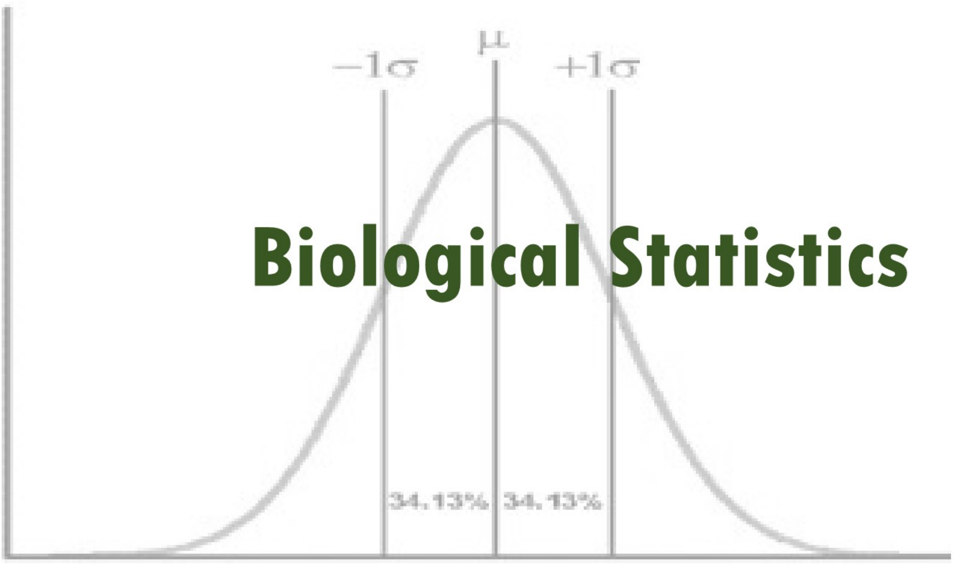
$$\bar{X} = \frac{\sum Xi}{n} = \frac{1658}{10} = 165.8$$

$$S = \sqrt{\frac{\sum_{i=1}^{n}(Xi - \bar{X})^2}{n-1}} = \sqrt{\frac{1095.6}{10-1}} = 11.033$$

$$CI = 165.8 \pm 11.033$$

$$CI = (154.766 - 176.833)$$


13
12
11
10
9
8
7
6
5
4
3
2
1
0
T
H
A
N
K
Y
O
U
T H A N K Y O U



Biological Statistics

Coefficient of variance (CV) معامل الاختلاف

- هو أحد مقاييس التباين النسبي وهو نسبة الانحراف القياسي إلى المتوسط الحسابي.
- يستخدم هذا المقياس عادة لمقارنة تباين متغيرين مختلفين بالحجم أو الوحدات.
- يتم حسابه باستخدام المعادلة التالية:

$$CV = \frac{\sigma}{\mu}$$ للمجتمع

OR

$$CV = \frac{S}{\bar{X}}$$ للعينة

حيث أن:

σ أو S: هو الانحراف القياسي.

μ أو $\bar{X}$: هو المتوسط الحسابي.

Coefficient of variance (C.V.) معامل الاختلاف

مثال: لدينا عينتين من الطلبة لصفين مختلفين X و Y ونريد مقارنة التباين في مستويات الطلبة لمادة الأحياء لكل صف حيث كانت درجات الطلبة كالآتي:

35	90	65	25	90	30	88	70	50	Xi
641.61	880.31	21.81	1248.21	880.31	919.91	765.63	93.51	106.71	$(X_i - \bar{X})^2$
60	55	77	82	85	67	75	70	65	**Yi**
4994.25	4994.25	4994.25	4994.25	4994.25	4994.25	4994.25	4994.25	4994.25	$(Y_i - \bar{Y})^2$

$$\bar{X} = \frac{543}{9} = 60.33$$

$$\bar{Y} = \frac{636}{9} = 70.67$$

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

$$S = \sqrt{\frac{5558}{8}} = 26.36$$

$$CV = \frac{S}{\bar{X}} = \frac{26.36}{60.33} = 0.44$$

$$S = \sqrt{\frac{\sum(Y_i - \bar{Y})^2}{n-1}} \quad S = \sqrt{\frac{798}{8}} = 9.99$$

$$CV = \frac{S}{\bar{Y}} = \frac{9.99}{70.67} = 0.14$$

بما أن معامل الاختلاف للعينة X أكبر منه للعينة Y اذا نستنتج ان التباين في مستويات الطلبة في العينة X أعلى منه للعينة Y.

Skewness and Kurtosis measurements مقاييس الالتواء والتفلطح

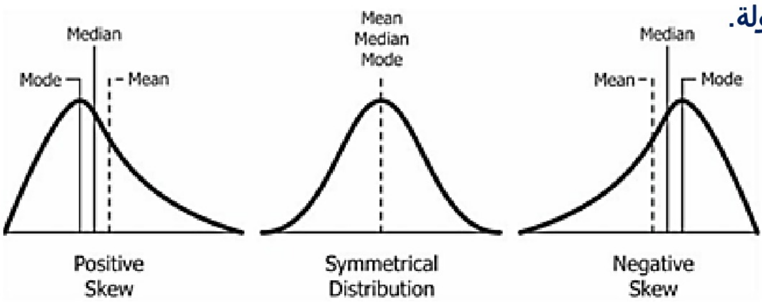
- الإلتواء Skewness: هو مقياس لمدى تناظر جانبي منحنى توزيع العينة اي مدى تطابقه مع منحنى التوزيع الطبيعي (الناقوس) أو بشكل اكثر دقة هو تحديد مدى اختلاف شكل منحنى توزيع العينة عن منحنى التوزيع الطبيعي.

- مقدار الالتواء في منحنى التوزيع الطبيعي = صفر.

- عندما تكون قيمة الالتواء بين 0.5 و – 0.5 تكون مقبولة.

- عندما تكون قيمة الالتواء بين – 0.5 و – 1 فان الالتواء يكون سالبا (Negatively skewed) وعندما يكون بين 0.5 و 1 فان الالتواء يكون موجبا (Positively skewed).

- اذا كانت قيمة الالتواء أقل من – 1 فسيكون التواء سالب عالي واذا كانت اعلى من 1 فسيكون التواء موجب عالي.



https://codeburst.io/2-important-statistics-terms-you-need-to-know-in-data-science-skewness-and-kurtosis-388fef94eeaa

مقاييس الالتواء والتفلطح Skewness and Kurtosis measurements

كيف نحسب قيمة الالتواء؟

يمكن حسابه باستخدام ال معادلة التالية:ـ

حيث ان:

Xi: قيمة المشاهدة

$\bar{X}$: متوسط العينة.

σ: الانحراف القياسي.

$$Skewness = \frac{\sum_{i=1}^{n}(Xi - \bar{X})^3}{n \times \sigma^3}$$

https://www.wallstreetmojo.com/skewness/

Skewness and Kurtosis measurements مقاييس الالتواء والتفلطح

مثال: افترض أن لدينا المشاهدات التالية والتي تمثل عينة عشوائية اخذت من مجتمع معين ونريد ان نختبر هل ان هذه البيانات متوزعة طبيعيا ام لا؟

Xi	$(X_i - \bar{X})^2$	$(Xi - \bar{X})^3$
12	400	-8000
13	361	-6859
54	484	10648
56	576	13824
25	49	-343
160	$\sum(X_i - \bar{X})^2$= 1870	$(Xi - \bar{X})^3$= 9270

$$\bar{X} = \sum \frac{Xi}{n} = \frac{160}{5} = 32$$

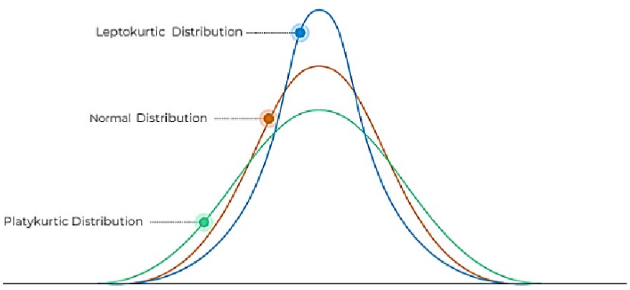
$$S = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{1870}{4}} = 21.62$$

$$Skewness = \frac{\sum_{i=1}^{n}(Xi - \bar{X})^3}{n \times \sigma^3} = \frac{9270}{5 \times 10108.17} = 0.183$$

بما ان قيمة الالتواء اكبر من صفر، إذًا منحنى التوزيع لهذه العينة موجب الالتواء. وبما انها تقع بين 0.5 و – 0.5 فهي مقبولة ويمكن اعتبار البيانات متوزعة طبيعيا.

Skewness and Kurtosis measurements مقاييس الالتواء والتفلطح

- التفلطح Kurtosis: هذا المقياس في الواقع يهتم بنهايتي الناقوس (مدى اتساعه) وليس بمقدار انخفاضه او ارتفاعه حيث انه يقيس مدى تواجد القيم المتطرفة (أعلى او ادنى من المتوسط بكثير) في العينة لأن هذه القيم تجعل قيمة المتوسط غير ممثلة للعينة. اي انه مقياس لمدى تطرف قيم العينة.



- عندما تكون قيمة التفلطح = 3 يسمى Mesokurtic وهي الحالة القياسية لمنحنى التوزيع الطبيعي.
- اذا كانت قيمة التفلطح أكبر من 3 يسمى Leptokurtic اي ان القيم المتطرفة في العينة اقل مما في العينة المتوزعة طبيعيا.
- إذا كانت قيمة التفلطح اقل من 3 يسمى Platykurtic ومعناه ان العينة تحتوي على قيم شديدة التطرف مقارنة بالعينة المتوزعة طبيعيا.
- قيمة التفلطح بين -2 و +2 تعد مقبولة حسب George & Mallery (2010)

George, D., & Mallery, M. (2010). SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update (10a ed.) Boston: Pearson.

مقاييس الالتواء والتفلطح Skewness and Kurtosis measurements

كيف نحسب قيمة التفلطح؟

يمكن حسابه باستخدام ال معادلة التالية:ـ

$$Kurtosis = \frac{\sum_{i=1}^{n}(Xi - \bar{X})^4}{n \times \sigma^4}$$

https://analystprep.com/cfa-level-1-exam/quantitative-methods/kurtosis-and-skewness-types-of-distributions/

حيث ان:

Xi: قيمة المشاهدة

$\bar{X}$: متوسط العينة.

σ: الانحراف القياسي.

وحيث ان قيمة تفلطح منحنى التوزيع الطبيعي هي 3 فيتم طرح 3 من ناتج المعادلة اعلاه للحصول على القيمة الفعلية للتفلطح عندما يكون تفلطح المنحنى الطبيعي = صفرا

Skewness and Kurtosis measurements مقاييس الالتواء والتفلطح

مثال: لو أخذنا نفس بيانات العينة في المثال السابق واردنا ان نعرف نوع التفلطح في منحنى توزيعها الطبيعي؟

Xi	$(X_i - \bar{X})^2$	$(Xi - \bar{X})^4$
12	400	160000
13	361	130321
54	484	234256
56	576	331776
25	49	2401
160	$\sum(X_i - \bar{X})^2 = 1870$	$(Xi - \bar{X})^4 = 858754$

$$\bar{X} = \sum \frac{Xi}{n} = \frac{160}{5} = 32$$

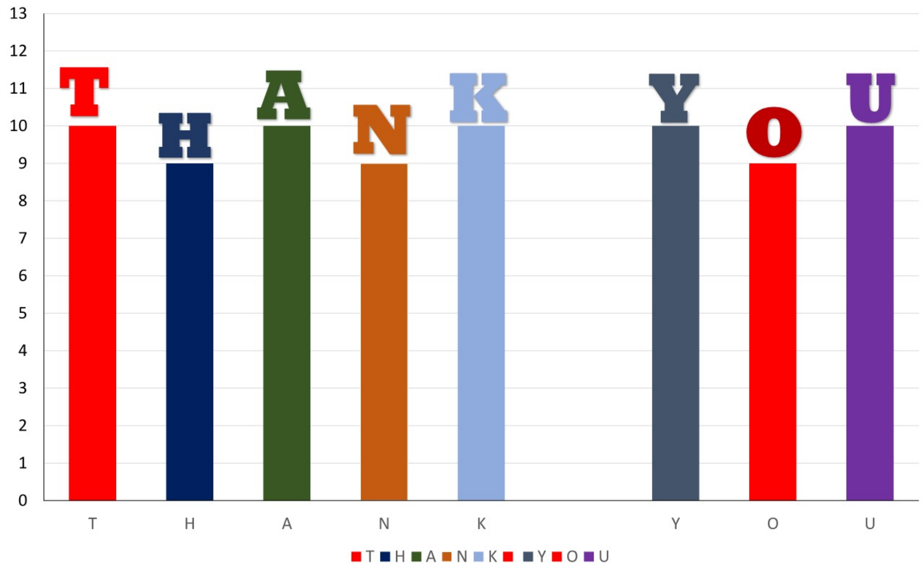
$$S = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{1870}{4}} = 21.62$$

$$Kurtosis = \frac{\sum_{i=1}^{n}(Xi - \bar{X})^4}{n \times \sigma^4} = \frac{858754}{5 \times 218556.3} = 0.785$$

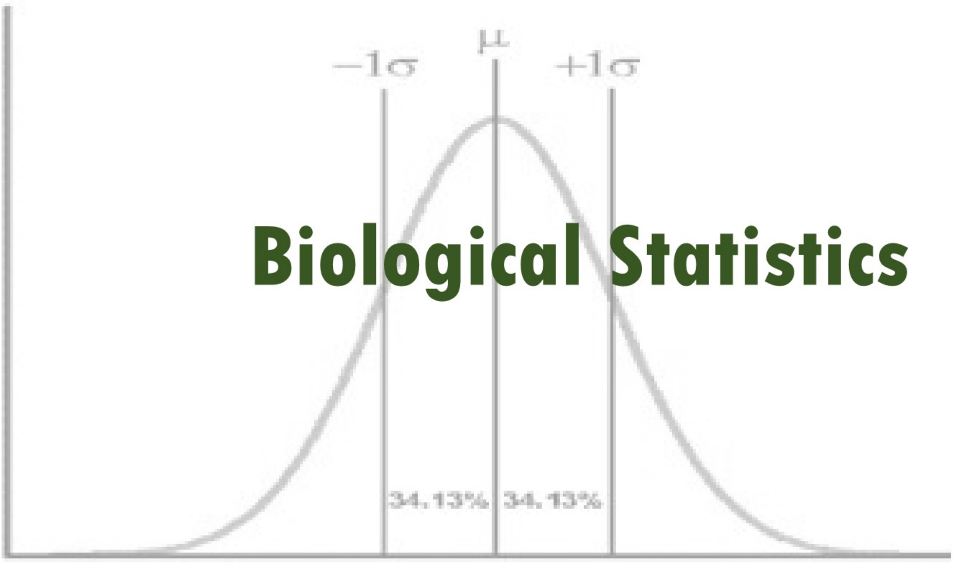
بعد ذلك نطرح القيمة القياسية لتفلطح منحنى التوزيع الطبيعي من القيمة اعلاه لنستخرج فائض التفلطح او التفلطح الفعلي:

$$Excess\ kurtosis = 0.785 - 3 = -2.213$$

وحيث أن قيمة التفلطح سالبة إذًا تفلطح هذه العينة من النوع Platykutrosis. وبما ان القيمة اقل من -2 فان البيانات تعد غير متوزعة طبيعيا.


13
12
11
10
9
8
7
6
5
4
3
2
1
0
T
H
A
N
K
Y
O
U
T H A N K Y O U

Biological Statistics



اختبار التوزيع الطبيعي للبيانات Normality test

- هو اختبار احصائي مهم جدا للتأكد من كون مشاهدات العينة (القياسات) متوزعة طبيعيا ام لا؟ ويجب ان يجرى قبل تحليل البيانات احصائيا لأن تحليل البيانات الغير متوزعة طبيعيا بطرق الاحصاء الشائعة سوف يؤدي إلى اعطاء نتائج غير صحيحة وبالتالي اتخاذ قرارات خاطئة كون معظم طرق المقارنات الاحصائية كاختبار t و اختبارات ANOVA وبعض حالات مربع كاي Chi^2 تختبر قبول أو عدم قبول فرضية العدم والتي تتبع التوزيع الطبيعي اصلا فاذا كانت البيانات غير متوزعة طبيعياً فلا يمكن استخدامها لاختبار الفرضية.
- يعني على اساس نتيجة هذا الاختبار نحدد بالضبط نوع التحليل الاحصائي المناسب لبياناتنا كون الاختبارات المشار اليها اعلاه تعتمد على اساس ان البيانات متوزعة طبيعيا.
- في حالة كونها غير متوزعة طبيعيا فبالامكان معالجتها باحد طرق التحويل الرياضي و من ثم تحليلها.

كيف نجري هذا الاختبار ؟

- هناك اكثر من طريقة لإجراء هذا الاختبار وسنتطرق إلى الاكثر شيوعا:
- طريقة الرسوم البيانية: ابسط هذه الطرق هي مقارنة المدرج التكراري للعينة مع منحنى التوزيع الطبيعي غير انها غير دقيقة. وهناك طرق اكثر دقة مثل منحنيات الـQ-Q (Quantile-quantile) ومنحنيات التوزيع الطبيعي والتي تقارن عادة بمنحنى التوزيع الطبيعي القياسي.
- طريقة Shapiro – Wilk: وهي من أدق طرق اختبار التوزيع الطبيعي وقد قدمها Samuel Sanford و Martin Wilk عام 1963 حيث يخبرنا هذا الاختبار هل ان العينة قد اتت من مجتمع متوزع طبيعيا أم لا. وسنناقش هذه الطريقة بشيء من التفصيل ونعتمدها في الاختبار.

كيف نجري هذا الاختبار ؟

ويجرى هذا الاختبار حسب المعادلة التالية:ـ

$$W = \frac{\left(\sum_{i=1}^{n} aiX_{(i)}\right)^2}{\sum_{i=1}^{n}(Xi-\bar{X})^2}$$

حيث ان:

$X_{(i)}$: (X_n-X_1)، ...، $(X_{n/2+1} - X_{(n/2)})$ في حالة n: عدد زوجي

و (X_n-X_1)، ...، $(X_{n/2+1} - X_{(n/2)-1})$ في حالة n: عدد فردي

ai: ثابت Shapiro – Wilk (يتم الحصول عليه من جداول خاصة)

$\bar{X}$: المتوسط الحسابي.

هناك الكثير من البرامجيات يمكنها ان تقوم باجراء هذا الاختبار ومنها SPSS

حيث يتم رفض فرضية العدم اذا كانت قيمة W المحسوبة أكبر من Wa الجدولية أي أن قيمة P أقل من α (0.05). (رفض فرضية العدم يعني أن البيانات غير متوزعة طبيعيا)

كيف نجري هذا الاختبار ؟

مثال: أخذت عينة عشوائية مكونة من 12 شخص لدراسة متوسط اعمار مجتمع معين. كيف تتأكد من ان العينة متوزعة طبيعيا؟

الحل:

1. نقوم بترتيب البيانات تصاعديا.
2. نحسب مربعات انحراف القيم عن متوسطها.

Age	Sorted	$(Xi - \bar{X})^2$
65	35	765.444
61	45	312.111
63	55	58.778
86	58	21.778
70	61	2.778
55	63	0.111
74	65	5.444
35	68	28.444
72	70	53.778
68	72	87.111
45	74	128.444
58	86	544.444
$\sum(Xi - \bar{X})^2=$		2008.667

كيف نجري هذا الاختبار ؟

3. نقوم باستخراج قيم ثابت Shapiro – Wilk (ai) من الجداول الخاصة للعينة المكونة من 12 فرد (n=12) حيث يلاحظ هنا انها 6 قيم.

n =	2	3	4	5	6	7	8	9	10	11	12	13	14
a1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739	0.5601	0.5475	0.5359	0.5251
a2			0.1677	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291	0.3315	0.3325	0.3325	0.3318
a3					0.0875	0.1401	0.1743	0.1976	0.2141	0.2260	0.2347	0.2412	0.2460
a4							0.0561	0.0947	0.1224	0.1429	0.1586	0.1707	0.1802
a5									0.0399	0.0695	0.0922	0.1099	0.1240
a6											0.0303	0.0539	0.0727
a7													0.0240

كيف نجري هذا الاختبار ؟

4. نقوم بحساب قيم الـ $X_{(i)}$ ومن ثم نحسب $\left(\sum ai * X_{(i)}\right)^2$

ai	ai value	X(i)	Dif	a*dif
a1	0.5475	x12-x1=	51	27.9225
a2	0.3325	x11-x2=	29	9.6425
a3	0.2347	x10-x3=	17	3.9899
a4	0.1586	x9-x4=	12	1.9032
a5	0.0922	x8-x5=	7	0.6454
a6	0.0803	x7-x6=	2	0.1606
			$\left(\sum aiX_{(i)}\right)^2$	**~~44.2641~~**

$$W = \frac{\left(\sum_{i=1}^{n} aiX_{(i)}\right)^2}{\sum_{i=1}^{n}(Xi-\bar{X})^2}$$

$$W = \frac{\cancel{44.2641}}{2008.667} = 0.9754$$

5. نقارن قيمة W المحسوبة بقيمة W الجدولية من جداول قيم W ونستخرج ايضا قيمة P من نفس الجداول.

n \ P	0.01	0.02	0.05	0.1	0.5	0.9	0.95	0.98	0.99
3	0.753	0.756	0.767	0.789	0.959	0.998	0.999	1.000	1.000
4	0.687	0.707	0.748	0.792	0.935	0.987	0.992	0.996	0.997
5	0.686	0.715	0.762	0.806	0.927	0.979	0.986	0.991	0.993
6	0.713	0.743	0.788	0.826	0.927	0.974	0.981	0.986	0.989
7	0.730	0.760	0.803	0.838	0.928	0.972	0.979	0.985	0.988
8	0.749	0.778	0.818	0.851	0.932	0.972	0.978	0.984	0.987
9	0.764	0.791	0.829	0.859	0.935	0.972	0.978	0.984	0.986
10	0.781	0.806	0.842	0.869	0.938	0.972	0.978	0.983	0.986
11	0.792	0.817	0.850	0.876	0.940	0.973	0.979	0.984	0.986
12	0.805	0.828	0.859	0.883	0.943	0.973	0.979	0.984	0.986
13	0.814	0.837	0.866	0.889	0.945	0.974	0.979	0.984	0.986
14	0.825	0.846	0.874	0.895	0.947	0.975	0.980	0.984	0.986
15	0.835	0.855	0.881	0.901	0.950	0.975	0.980	0.984	0.987
16	0.844	0.863	0.887	0.906	0.952	0.976	0.981	0.985	0.987
17	0.851	0.869	0.892	0.910	0.954	0.977	0.981	0.985	0.987
18	0.858	0.874	0.897	0.914	0.956	0.978	0.982	0.986	0.988
19	0.863	0.879	0.901	0.917	0.957	0.978	0.982	0.986	0.988
20	0.868	0.884	0.905	0.920	0.959	0.979	0.983	0.986	0.988
21	0.873	0.888	0.908	0.923	0.960	0.980	0.983	0.987	0.989
22	0.878	0.892	0.911	0.926	0.961	0.980	0.984	0.987	0.989
23	0.881	0.895	0.914	0.928	0.962	0.981	0.984	0.987	0.989

كيف نجري هذا الاختبار ؟

$W = 0.9754$

$N=12$

لاحظ انها وقعت بين القيمتين 0.973 و 0.979 اذا قيمة Wα= 0.976 وهي أكبر من قيمة W المحسوبة.

كذلك قيمة P وقعت بين 0.9 و 0.95 اي ان قيمة P= 0.925 وهي أكبر من 0.05 وعليه نقبل فرضية العدم ونرفض البديلة أي ان البيانات متوزعة طبيعيا.

معالجة البيانات الغير متوزعة طبيعيا

ماذا نفعل اذا كانت البيانات غير متوزعة طبيعيا ونريد أن نحللها احصائيا؟

هذا النوع من البيانات يجب ان يعالج قبل الذهاب للتحليل الاحصائي ويعاد اختبار التوزيع الطبيعي لها من جديد للتأكد من سلامتها قبل تحليلها احصائيا باستخدام طرق التحليل التي تعتمد على التوزيع الطبيعي أما اذا لم تنجح طرق المعالجة في ذلك فيجب استخدام طرق اختبار لا معلمية Non-parametric tests مثل اختبار Chi^2 لمقارنة عينتين أو Kruscal – Wallis للمقارنات المتعددة. ومن ابسط المعالجات التي تستخدم لتصحيح البيانات الغير متوزعة طبيعيا:

1. تحويل قيم العينة إلى لوغارتم القيم ومن ثم تحليلها.
2. تحويل القيم إلى الجذر التربيعي للقيم ومن ثم تحليلها.

تساؤلات مهمة

ما هو الاحصاء ؟

الاحصاء هو طريقة رياضية للحصول على معلومات مهمة اعتمادا على بيانات (قياسات) عشوائية جمعت من مجتمع معين (مجتمع بشري أو مجتمع نباتي أو مجتمع حيواني أو مجتمع ميكروبي ...ألخ).

ما فائدة الاحصاء الوصفي Descriptive statistic ؟ ولماذا اتعبنا انفسنا بدراسته ؟

الاحصاء الوصفي يساعدنا على تبسيط فهم عدد هائل من البيانات بطريقة عرض مبسطة تمكننا من فهم حقيقة الصفة أو الصفات المدروسة لأفراد مجتمع معين. بكلام آخر يعطينا صورة واضحة عن الصفة المدروسة في ذلك المجتمع. فلو أخذنا صفة الطول كمثال فسوف يعطينا الاحصاء الوصفي صورة واضحة عن كون افراد المجتمع طوال القامة عموما أم قصار القامة أم متوسطي الطول؟

تساؤلات مهمة

ما فائدة مقاييس النزعة المركزية Central tendency measurements ؟

هي عبارة عن نشرة موجزة عن مجموعة من القيم (المشاهدت أو القياسات) من خلال قيمة معينة ممثلة عنها تقع في المركز وتتوزع حولها بقية القيم.

ما فائدة المتوسط الحسابي Mean ؟

هو تعبير عن معدل قيم العينة (القيمة المركزية). اي معدل قيمة الصفة المدروسة مثل معدل الطول أو

ما فائدة الوسيط Median ؟

بما انه تعبير عن القيمة التي تقع في وسط قيم العينة بالضبط فإنه يكون تعبير اصدق من المتوسط الحسابي في حالة وجود قيم متطرفة في العينة (قيم صغيرة جدا وقيم كبيرة جداً).

تساؤلات مهمة

ما فائدة دراسة مقاييس التشتت ؟

لقياس مدى تجانس توزيع القيم في مجتمع معين. أي مدى ابتعادها او اقترابها من متوسطها الحسابي. فهذه المقاييس تستخدم لتدقيق مدى مطابقة مقاييس النزعة المركزية للواقع ومن خلالها نعرف مقدار دقة مقاييس التمركز في رسم صورة الصفة في العينة قيد الدراسة.

ما فائدة قياس التباين Variance ؟

للتعرف على مدى انحراف القيم عن متوسطاتها اي مدى تشتتها وعدم تجانسها. كلما قلت قيمة التباين دل ذلك على تجانس قياسات الصفة المدروسة في المجتمع واصبحت العينة المأخوذة ممثلة بشكل اكثر وثوقية لذلك المجتمع وبالتالي فان متوسطها سيكون مماثل لمتوسط المجتمع واي قرار سيتخذ على اساس نتائج ذلك التحليل سيكون صائباً.

تساؤلات مهمة

ما فائدة قياس معامل الاختلاف Coefficient of variance ؟

في حالة وجود عينتين مختلفتين في الحجم (واحدة صغيرة والأخرى كبيرة) ولكن معدلهما متشابه فإن الانحراف القياسي لا يعطينا صورة عن فرق التغاير بينهما فنظطر إلى قياس معامل الاختلاف (النحراف القياسي مقسوما على المتوسط) لمعرفة ايهما تمتلك قيما اكثر تغايرا.

ما فائدة قياس المدى Range ؟

للتعرف على مقدار التطرف في قيم العينة. فاذا كانت قيمته عالية دل ذلك على مدى عدم تجانس القيم وابتعادها عن متوسطاتها وبالتالي فان قيم المتوسطات تعطي تصور خاطئ عن وضع العينة وعلينا ان نجري اختبارت التدقيق الاخرى كالتباين أو الانحراف القياسي للتأكد من دقة الطريقة التي جمعت بها العينة ومدى مطابقتها للشروط.

تساؤلات مهمة

ما فائدة قياس التفلطح Kurtosis ؟

للتعرف على مدى زيادة تشتت القيم عن مستوى االتوزيع الطبيعي القياسي للمجتمع. كلما زادت او قلت قيمة التفلطح عن الصفر دل ذلك على زيادة التشتت وابتعد توزيع المجتمع عن التوزيع الطبيعي القياسي.

ما فائدة قياس الالتواء Skewness ؟

للتعرف على مقدار التشوه أو الانحياز الموجب أو السالب لتوزيع افراد مجتمع معين مقارنة بالتوزيع الطبيعي القياسي. زيادة أو نقصان القيمة عن الصفر يدل على ابتعاد توزيع المجتمع عن التوزيع الطبيعي القياسي.

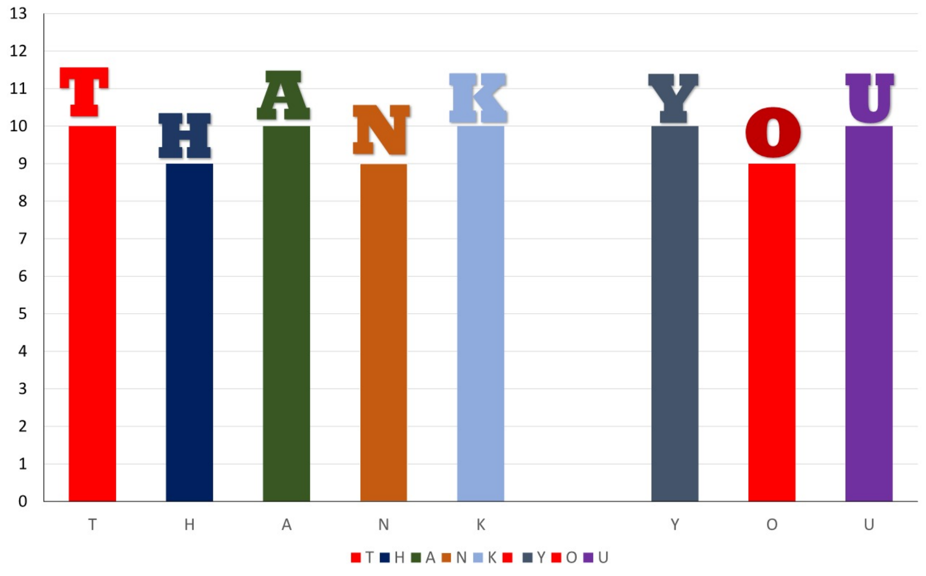

T
H
A
N
K
Y
O
U
13
12
11
10
9
8
7
6
5
4
3
2
1
0
T H A N K Y O U
T H A N K Y O U

Table D4. Coefficients a_i for the Shapiro-Wilk test for normality.
(Source: Shapiro and Wilk, 1965)

i \ n	2	3	4	5	6	7	8	9	10
1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.5888	0.5739
2	-	0.0000	0.1677	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291
3	-	-	-	0.0000	0.0875	0.1401	0.1743	0.1976	0.2141
4	-	-	-	-	-	0.0000	0.0561	0.0947	0.1224
5	-	-	-	-	-	-	-	0.0000	0.0399

i \ n	11	12	13	14	15	16	17	18	19	20
1	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734
2	0.3315	0.3325	0.3325	0.3318	0.3306	0.3290	0.3273	0.3253	0.3232	0.3211
3	0.2260	0.2347	0.2412	0.2460	0.2495	0.2521	0.2540	0.2553	0.2561	0.2565
4	0.1429	0.1586	0.1707	0.1802	0.1878	0.1939	0.1988	0.2027	0.2059	0.2085
5	0.0695	0.0922	0.1099	0.1240	0.1353	0.1447	0.1524	0.1587	0.1641	0.1686
6	0.0000	0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334
7	-	-	0.0000	0.0240	0.0433	0.0593	0.0725	0.0837	0.0932	0.1013
8	-	-	-	-	0.0000	0.0196	0.0359	0.0496	0.0612	0.0711
9	-	-	-	-	-	-	0.0000	0.0163	0.0303	0.0422
10	-	-	-	-	-	-	-	-	0.0000	0.0140

i \ n	21	22	23	24	25	26	27	28	29	30
1	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254
2	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043	0.3018	0.2992	0.2968	0.2944
3	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533	0.2522	0.2510	0.2499	0.2487
4	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151	0.2152	0.2151	0.2150	0.2148
5	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836	0.1848	0.1857	0.1864	0.1870
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630
7	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316	0.1346	0.1372	0.1395	0.1415
8	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089	0.1128	0.1162	0.1192	0.1219
9	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876	0.0923	0.0965	0.1002	0.1036
10	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672	0.0728	0.0778	0.0822	0.0862
11	0.0000	0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650	0.0697
12	-	-	0.0000	0.0107	0.0200	0.0284	0.0358	0.0424	0.0483	0.0537
13	-	-	-	-	0.0000	0.0094	0.0178	0.0253	0.0320	0.0381
14	-	-	-	-	-	-	0.0000	0.0084	0.0159	0.0227
15	-	-	-	-	-	-	-	-	0.0000	0.0076

Table D4. (continued)

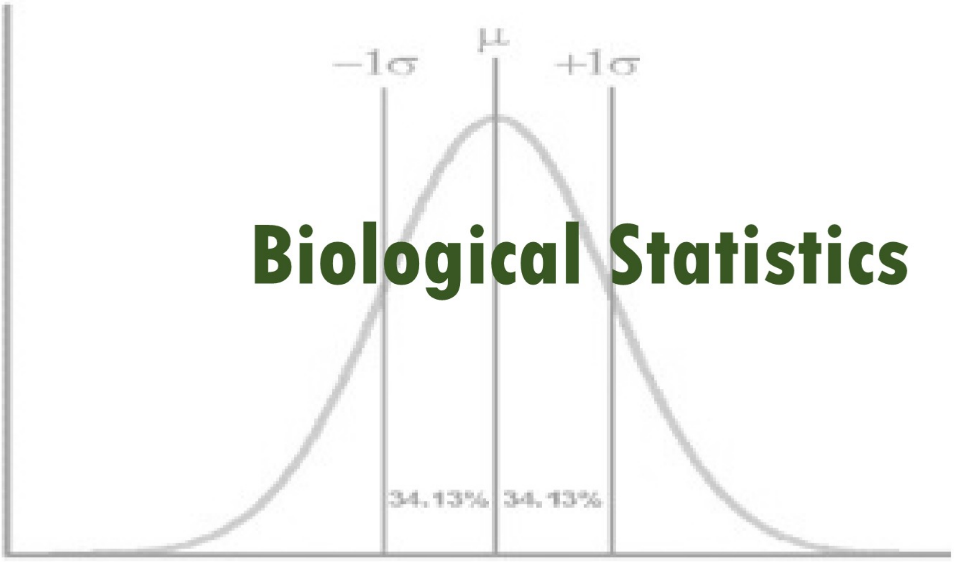
i \ n	31	32	33	34	35	36	37	38	39	40
1	0.4220	0.4188	0.4156	0.4127	0.4096	0.4068	0.4040	0.4015	0.3989	0.3964
2	0.2921	0.2898	0.2876	0.2854	0.2834	0.2813	0.2794	0.2774	0.2755	0.2737
3	0.2475	0.2462	0.2451	0.2439	0.2427	0.2415	0.2403	0.2391	0.2380	0.2368
4	0.2145	0.2141	0.2137	0.2132	0.2127	0.2121	0.2116	0.2110	0.2104	0.2098
5	0.1874	0.1878	0.1880	0.1882	0.1883	0.1883	0.1863	0.1881	0.1880	0.1878
6	0.1641	0.1651	0.1660	0.1667	0.1673	0.1678	0.1683	0.1686	0.1689	0.1691
7	0.1433	0.1449	0.1463	0.1475	0.1487	0.1496	0.1505	0.1513	0.1520	0.1526
8	0.1243	0.1265	0.1284	0.1301	0.1317	0.1331	0.1344	0.1356	0.1366	0.1376
9	0.1066	0.1093	0.1116	0.1140	0.1160	0.1179	0.1196	0.1211	0.1225	0.1237
10	0.0699	0.0931	0.0961	0.0988	0.1013	0.1036	0.1056	0.1075	0.1092	0.1108
11	0.0739	0.0777	0.0812	0.0844	0.0873	0.0900	0.0924	0.0947	0.0967	0.0986
12	0.0585	0.0629	0.0669	0.0706	0.0739	0.0770	0.0798	0.0824	0.0848	0.0870
13	0.0435	0.0485	0.0530	0.0572	0.0610	0.0645	0.0677	0.0706	0.0733	0.0759
14	0.0289	0.0344	0.0395	0.0441	0.0484	0.0523	0.0559	0.0592	0.0622	0.0651
15	0.0144	0.0206	0.0262	0.0314	0.0361	0.0404	0.0444	0.0481	0.0515	0.0546
16	0.0000	0.0068	0.0131	0.0187	0.0239	0.0287	0.0331	0.0372	0.0409	0.0444
17	-	-	0.0000	0.0062	0.0119	0.0172	0.0220	0.0264	0.0305	0.0343
18	-	-	-	-	0.0000	0.0057	0.0110	0.0158	0.0203	0.0244
19	-	-	-	-	-	-	0.0000	0.0053	0.0101	0.0146
20	-	-	-	-	-	-	-	-	0.0000	0.0049

i \ n	41	42	43	44	45	46	47	48	49	50
1	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3806	0.3789	0.3770	0.3751
2	0.2719	0.2701	0.2684	0.2667	0.2651	0.2635	0.2620	0.2604	0.2589	0.2574
3	0.2357	0.2345	0.2334	0.2323	0.2313	0.2302	0.2291	0.2281	0.2271	0.2260
4	0.2091	0.2085	0.2078	0.2072	0.2065	0.2058	0.2052	0.2045	0.2038	0.2032
5	0.1876	0.1874	0.1871	0.1868	0.1865	0.1862	0.1859	0.1855	0.1851	0.1847
6	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
7	0.1531	0.1535	0.1539	0.1542	0.1545	0.1548	0.1550	0.1551	0.1553	0.1554
8	0.1384	0.1392	0.1398	0.1405	0.1410	0.1415	0.1420	0.1423	0.1427	0.1430
9	0.1249	0.1259	0.1269	0.1278	0.1286	0.1293	0.1300	0.1306	0.1312	0.1317
10	0.1123	0.1136	0.1149	0.1160	0.1170	0.1180	0.1189	0.1197	0.1205	0.1212
11	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
12	0.0891	0.0909	0.0927	0.0943	0.0959	0.0972	0.0986	0.0998	0.1010	0.1020
13	0.0782	0.0804	0.0824	0.0842	0.0860	0.0876	0.0892	0.0906	0.0919	0.0932
14	0.0677	0.0701	0.0724	0.0745	0.0765	0.0783	0.0801	0.0817	0.0832	0.0846
15	0.0575	0.0602	0.0628	0.0651	0.0673	0.0694	0.0713	0.0731	0.0748	0.0764
16	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
17	0.0379	0.0411	0.0442	0.0471	0.0497	0.0522	0.0546	0.0568	0.0588	0.0608
18	0.0283	0.0318	0.0352	0.0383	0.0412	0.0439	0.0465	0.0489	0.0511	0.0532
19	0.0186	0.0227	0.0263	0.0296	0.0328	0.0357	0.0385	0.0411	0.0436	0.0459
20	0.0094	0.0136	0.0175	0.0211	0.0245	0.0277	0.0307	0.0335	0.0361	0.0386
21	0.0000	0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
22	-	-	0.0000	0.0042	0.0081	0.0118	0.0153	0.0185	0.0215	0.0244
23	-	-	-	-	0.0000	0.0039	0.0076	0.0111	0.0143	0.0174
24	-	-	-	-	-	-	0.0000	0.0037	0.0071	0.0104
25	-	-	-	-	-	-	-	-	0.0000	0.0035

Table D5. Quantiles of the Shapiro-Wilk test for normality (values of W such that 100p% of the distribution of W is less than W_p). (Source: Shapiro and Wilk, 1965)

n	$W_{0.01}$	$W_{0.02}$	$W_{0.05}$	$W_{0.10}$	$W_{0.50}$
3	0.753	0.756	0.767	0.789	0.959
4	0.687	0.707	0.748	0.792	0.935
5	0.686	0.715	0.762	0.806	0.927
6	0.713	0.743	0.788	0.826	0.927
7	0.730	0.760	0.803	0.838	0.928
8	0.749	0.778	0.818	0.851	0.932
9	0.764	0.791	0.829	0.859	0.935
10	0.781	0.806	0.842	0.869	0.938
11	0.792	0.817	0.850	0.876	0.940
12	0.805	0.828	0.859	0.883	0.943
13	0.814	0.837	0.866	0.889	0.945
14	0.825	0.846	0.874	0.895	0.947
15	0.835	0.855	0.881	0.901	0.950
16	0.844	0.863	0.887	0.906	0.952
17	0.851	0.869	0.892	0.910	0.954
18	0.858	0.874	0.897	0.914	0.956
19	0.863	0.879	0.901	0.917	0.957
20	0.868	0.884	0.905	0.920	0.959
21	0.873	0.888	0.908	0.923	0.960
22	0.878	0.892	0.911	0.926	0.961
23	0.881	0.895	0.914	0.928	0.962
24	0.884	0.898	0.916	0.930	0.963
25	0.886	0.901	0.918	0.931	0.964
26	0.891	0.904	0.920	0.933	0.965
27	0.894	0.906	0.923	0.935	0.965
28	0.896	0.908	0.924	0.936	0.966
29	0.898	0.910	0.926	0.937	0.966
30	0.900	0.912	0.927	0.939	0.967
31	0.902	0.914	0.929	0.940	0.967
32	0.904	0.915	0.930	0.941	0.968
33	0.906	0.917	0.931	0.942	0.968
34	0.908	0.919	0.933	0.943	0.969
35	0.910	0.920	0.934	0.944	0.969
36	0.912	0.922	0.935	0.945	0.970
37	0.914	0.924	0.936	0.946	0.970
38	0.916	0.925	0.938	0.947	0.971
39	0.917	0.927	0.939	0.948	0.971
40	0.919	0.928	0.940	0.949	0.972
41	0.920	0.929	0.941	0.950	0.972
42	0.922	0.930	0.942	0.951	0.972
43	0.923	0.932	0.943	0.951	0.973
44	0.924	0.933	0.944	0.952	0.973
45	0.926	0.934	0.945	0.953	0.973
46	0.927	0.935	0.945	0.953	0.974
47	0.928	0.936	0.946	0.954	0.974
48	0.929	0.937	0.947	0.954	0.974
49	0.929	0.937	0.947	0.955	0.974
50	0.930	0.938	0.947	0.955	0.974

Biological Statistics



مراجعة مبسطة لما تطرقنا له سابقاً

لكي لا يكون لديك تشتت أو التباس حول كثرة الطرق والاختبارات التي تعلمناها في المحاضرات السابقة وما ستتعلمه لاحقا سنوجز هنا التسلسل المنطقي لإجراء التحليل الاحصائي وسبب تعلمنا كل هذه الطرق وكيف نتصرف وفق منهجية علمية صحيحة لتحليل البيانات واستخلاص النتائج من أجل تفسيرها واتخاذ القرارات على اساس علمي صحيح يضمن لنا حل المشكلة قيد الدراسة.

كل ما تعلمناه لغاية هذه المحاضرة يتعلق بالاحصاء الوصفي للبيانات الاحصائية والذي يشمل استعراض بياناتنا وفحصها ومن ثم تحضيرها للاختبارات الاحصائية وهذا واقعا يمثل نصف التحليل الاحصائي تقريبا.

أما النصف الثاني فسيشمل دراسة العلاقات أو اجراء المقارنات وتحديد كونها معنوية احصائيا ام لا؟

مراجعة مبسطة لما تطرقنا له سابقاً – خطوات العمل – الشريحة الاولى

تحديد المشكلة التي نريد دراستها: مثلا
[كفاءة عقار معين في علاج مرض معين]

تحديد الكثافة السكانية للمجتمع قيد الدراسة
[مثلا مدينة فيها اربع ملايين نسمة]

حساب حجم العينة اللازم لتمثيل المجتمع
[لمثالنا هذا سيكون 385 مشاهدة تقريبا]
معناه ان النتائج التي سنحصل عليها من تحليل 350 مشاهدة عشوائية جمعت من المجتمع اعلاه ستعطينا نتائج ممثلة له، أي كاننا حللنا بيانات 4 ملايين فرد

جمع البيانات
[تأكد من كون بيانات العينة عشوائية وغير منحازة]

تنظيم البيانات في جداول مبوبة
[تحويلها من بيانات غير مفهومة إلى معلومات]

استعراض بيانات العينة احصائيا
[مقاييس التمركز والتشتت والالتواء والتفلطح]

مراجعة مبسطة لما تطرقنا له سابقاً – خطوات العمل – الشريحة الثانية

استعراض بيانات العينة احصائيا

مقاييس التمركز
[تحدد القيم التي تتمركز حولها المشاهدات]

- المتوسط الحسابي Mean
  [هو معدل القيم]
- الوسيط Median
  [القيمة التي تقع وسط المشاهدات]
- المنوال Mode
  [القيمة الاكثر تكرار]

مقاييس التشتت
[تصف لنا مدى التباين بين المشاهدات وتشتتها]

- التباين Variance
  [متوسط مربع انحراف القيم عن متوسطها]
- الانحراف القياسي Standard deviation
  [جذر التباين: لتلافي مشكلة تربيع الوحدات]
- الخطأ القياسي Standard error
  [يقيس مقدار تجانس قيم العينة]
- المدى Range
  [يقيس مقدار التطرف في قيم العينة ويحدد اعلى واقل قيمة]
- معامل الاختلاف CV
  [يقيس التباين النسبي ويعطي صورة ادق عن تباين قيم العينة]

مقاييس التفلطح والالتواء
[في الشريحة التالية]

مراجعة مبسطة لما تطرقنا له سابقاً – خطوات العمل - الشريحة الثالثة

استعراض بيانات العينة احصائيا

مقاييس التفلطح والالتواء
[تحدد شكل منحنى توزيع قيم العينة]

الالتواء Skewness
[يحدد مدى تماثل جانبي منحنى توزيع قيم العينة]

التفلطح Kurtosis
[يحدد مدى اتساع او انكماش منحنى توزيع قيم العينة]

اختبار كون البيانات متوزعة طبيعيا Normality test

مثل اختبار Shapiro-wilk أو باستخدام المدرج التكراري
[هل أن قيم العينة متوزعة طبيعيا أم لا؟]

نعم:
[نبقى البيانات على حالها ونستمر بالتحليل]

لا:
[أما ان اعالج البيانات بحيث تتوزع طبيعيا او انتقل الى الاختبارات التي لا تحتاج إلى توزيع طبيعي]

انتهى استعراض و فحص البيانات وتهيئتها وسننتقل للاختبارات الاحصائية [الشريحة الرابعة]

مراجعة مبسطة لما تطرقنا له سابقاً – خطوات العمل - الشريحة الرابعة

اجراء الاختبارات الاحصائية Statistical tests

اختبار العلاقات Relationship tests
[لتحديد نوع العلاقة بين متغيرين]

- الارتباط Correlation
  [يحدد معنوية العلاقة بين متغيرين مستقلين]
- الانحدار Regression
  [يحدد معنوية العلاقة بين متغير أو متغيرات مستقلة ومتغير تابع]

مقارنة بيانات
[تحديد معنوية الفروق لاتخاذ قرار بقبول فرضية العدم أو رفضها]

- T-test
  [تصف معنوية الفرق بين قيم عينة واحدة او عينتين مستقلتين أو مرتبطتين على ان تكون القيم متوزعة طبيعيا ولا تزيد عن 30 قيمة]
- مربع كاي Chi^2
  [يصف معنوية الفروق بين القيم المشاهدة والقيم المتوقعة ولا يهتم بكون القيم متوزعة طبيعيا ام لا ويستخدم لمقارنة متغيرين مستقلين]
- تحليل التباين ANOVA
  [يحدد معنوية الفرق بين متوسطات ثلاث متغيرات فما فوق ويشترط ان تتوزع قيم المتغيرات طبيعيا]

مراجعة مبسطة لما تطرقنا له سابقاً – خطوات العمل - الشريحة الخامسة

تفسير النتائج واتخاذ القرارات

- Relationship tests اختبار العلاقات
  [وجود أم عدم وجود علاقة معنوية]
- Hypothesis tests اختبارات الفرضيات
  [قبول فرضية العدم أو رفضها]

جدول مبسط يساعدك في اختيار نوع الاختبار الاحصائي بعد ان هيأت بياناتك للتحليل

Choosing the best statistical test

- **Q1: What type of data do you have?**
- **Q2: How many samples do you have?**
- **Q3: What is the test supposed to do?**

Q3→	Compare the Data		Seek Relationships
Q2 ↓ Q1→	Categorical Data	Quantitative Data	
One Sample	1 sample proportion	1 sample t	
Two Samples	2 sample proportions	2 sample t	
Two Samples Special		2 sample t Paired t	Correlation/ Regression
Three or more samples		One-Way ANOVA	

دراسة العلاقة بين المتغيرات – معامل الارتباط Correlation

يستخدم معامل الارتباط لتقدير العلاقة الخطية بين متغيرين واتجاه هذه العلاقة. وتتراوح قيمة معامل الارتباط بين +1 و -1 حيث تعني الاشارة الموجبة أن العلاقة طردية بينما تشير الاشارة السالبة إلى ان العلاقة عكسية وكلما اقتربت القيمة من الـ1 بغض النظر عن الاشارة دل ذلك على قوة العلاقة.

أنواع معامل الارتباط Correlation

هناك ثلاثة انواع من معامل الارتباط الخطي:

1. معامل ارتباط بيرسون Pearson: وهو من المقاييس المعلمية Parametric اي التي تستخدم مع البيانات الكمية (البيانات بشكل ارقام حقيقية).

دراسة العلاقة بين المتغيرات – معامل الارتباط Correlation coefficient

أنواع معامل الارتباط Correlation coefficient

2. معامل ارتباط Spearman’s rho: من المقاييس اللامعلمية Non-parametric ويستخدم مع البيانات التي تكون على شكل رتب (اي تهتم برتبة القيمة بين القيم وليس مقدارها، مثلا بدل درجات الطلاب نهتم بمن هو الاول ومن الثاني ..الخ بغض النظر عن قيمة الدرجة).
3. معامل ارتباط Kendall’s tau: ويستخدم مع المقاييس اللامعلمية ايضا Non-parametric.

Pearson correlation coefficient كيفية حساب معامل الارتباط

يمكن حساب (معامل ارتباط بيرسون) باستخدام احدى المعادلتين التاليتين:

$$r = \frac{\sum_{i=1}^{n}[(xi-\bar{x})\times(yi-\bar{y})]}{\sqrt{\sum_{i=1}^{n}(xi-\bar{x})^2 \times \sum_{i=1}^{n}(yi-\bar{y})^2}}$$

OR

$$r = \frac{\sum_{i=1}^{n} xi \times yi - \frac{\sum_{i=}^{n} xi \times \sum_{i=1}^{n} yi}{n}}{\sqrt{\left(\sum_{i=1}^{n} xi^2 - \frac{(\sum_{i=1}^{n} xi)^2}{n}\right) \times \left(\sum_{i=1}^{n} yi^2 - \frac{\left(\sum_{i=1}^{n} yi\right)^2}{n}\right)}}$$

يعني هو ببساطة عبارة عن:

$$r = \frac{Variance\ xy}{Variance\ x\ * Variance\ y} \qquad r = \frac{S^2 x * y}{S^2 x\ \times S^2 y}$$

$$|t| = r \times \sqrt{\frac{n-2}{1-r^2}}$$

ويستخدم اختبار T للتعرف على معنوية الارتباط فاذا كانت T المحسوبة اكبر من T الجدولية عند درجات حرية *df* = n-2 فالارتباط معنوي.

حيث أن:

r: معامل الارتباط

xi: قيم المتغير X

yi: قيم المتغير Y

$\bar{x}$: متوسط قيم المتغير X

$\bar{y}$: متوسط قيم المتغير Y

n: عدد القيم.

مثال: البيانات التالية تمثل قيم المشاهدات لمتغيرين عشوائيين مستقلين. بين ما هي احتمالية ارتباط المتغيرين من عدمه.

X-value	1.24	1.34	1.39	1.41	1.64	1.44	1.48	1.51	1.54	1.54	1.54	1.62
Y-value	1.30	1.50	1.70	1.50	1.44	1.47	1.60	1.60	1.80	1.50	1.70	1.90

الحل:

xi	yi	$(xi-\bar{x})$	$(xi-\bar{x})^2$	$(yi-\bar{y})$	$(yi-\bar{y})^2$	$(xi-\bar{x})\times(yi-\bar{y})$
1.24	1.30	-0.23	0.0529	-0.28	0.08	0.06
1.34	1.50	-0.13	0.02	-0.08	0.01	0.01
1.39	1.70	-0.08	0.01	0.12	0.01	-0.01
1.41	1.50	-0.06	0.00	-0.08	0.01	0.00
1.64	1.44	0.17	0.03	-0.14	0.02	-0.02
1.44	1.47	-0.03	0.00	-0.11	0.01	0.00
1.48	1.60	0.01	0.00	0.02	0.00	0.00
1.51	1.60	0.04	0.00	0.02	0.00	0.00
1.54	1.80	0.07	0.00	0.22	0.05	0.02
1.54	1.50	0.07	0.00	-0.08	0.01	-0.01
1.54	1.70	0.07	0.00	0.12	0.01	0.01
1.62	1.90	0.15	0.02	0.32	0.10	0.05
Sum: 17.69	19.01	0.0500	0.1485	0.050	0.309700	0.116700

مثال: البيانات التالية تمثل قيم المشاهدات لمتغيرين عشوائيين مستقلين. بين ما هي احتمالية ارتباط المتغيرين من عدمه.

الحل:

$$r = \frac{\sum_{i=1}^{n}[\,(xi-\bar{x})\times(yi-\bar{y})\,]}{\sqrt{\sum_{i=1}^{n}(xi-\bar{x})^2\times\sum_{i=1}^{n}\ (yi-\bar{y})^2}}$$

$$r = \frac{0.116700}{\sqrt{0.148500\ \times 0.309700}} = 0.544$$

$$|t| = r\times\sqrt{\frac{n-2}{1-r^2}} \quad = 0.544\times\sqrt{\frac{12-2}{1-0.544^2}} = 2.050$$

ثم نستخرج قيمة t الجدولية عند درجات حرية 10 ومستوى احتمال 0.05 والتي تساوي 2.228

لاحظ ان المتغيرين مرتبطين ارتباطا موجبا ولكن غير معنوي أي ان هناك علاقة طردية بينهما لكن غير معنوية

Spearman's rho correlation coefficient كيفية حساب معامل الارتباط

يمكن حساب (معامل ارتباط سبيرمان) باستخدام المعادلة التالية:

$$r_s = 1 - \frac{6 \times \sum d^2}{n(n^2 - 1)}$$

حيث أن:

r_s: معامل ارتباط سبيرمان

d^2: مربع الفرق بين الرتبتين المتقابلتين (rank xi – rank yi)

n: عدد القيم.

xi	Xi rank	yi	Yi rank	di	di²
25	3	80	2	1	1
15	4	77	3	1	1
30	2	35	4	-2	4
50	1	90	1	0	0
					$\sum d^2 = 6$

$$r_s = 1 - \frac{6 \times 6}{4 \times 15} = 0.6$$

كيفية حساب معامل الارتباط Kendall's tau correlation coefficient

يمكن حساب (معامل ارتباط كندال) باستخدام المعادلة التالية:

$$\tau = \frac{n_c - n_d}{n(n-1)/2}$$

C تعني أن y الزوج الايمن اكبر من y الزوج الايسر و x الزوج الايمن أكبر من x الزوج الايسر. ونفس الشيء في حالة اصغر يعني بنفس الاتجاه. أما اذا اختلفا فانها تصبح D.

حيث أن:

τ: معامل ارتباط كندال تاو

n_c: عدد الازواج المتوافقة (C) Concordant اي قيمة x من الثاني اكبر/اصغر من قيمة x من الاول وقيمة y من الثاني أكبر/اصغر من قيمة y من الاول.

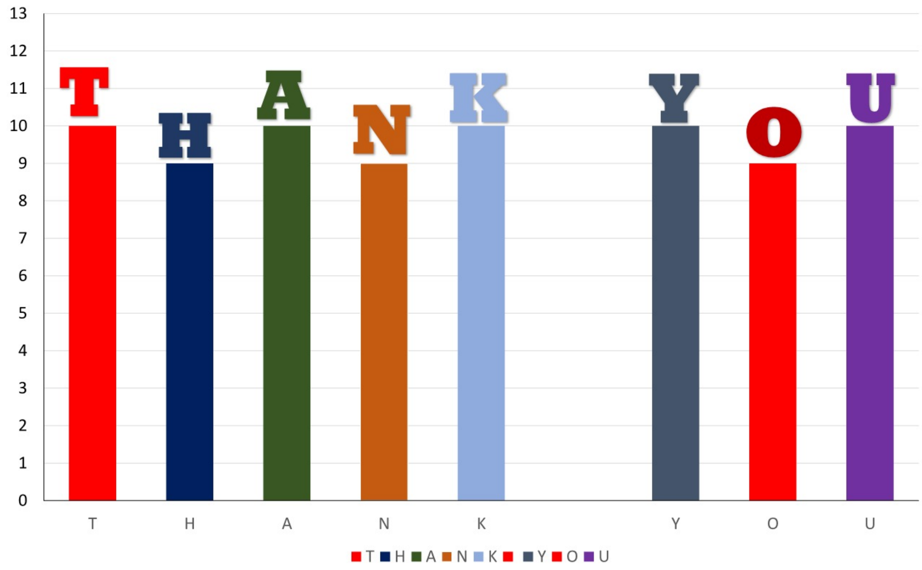
n_d: عدد الازواج الغير المتوافقة (D) Discordant اي التي لا ينطبق عليها الشرط أعلاه

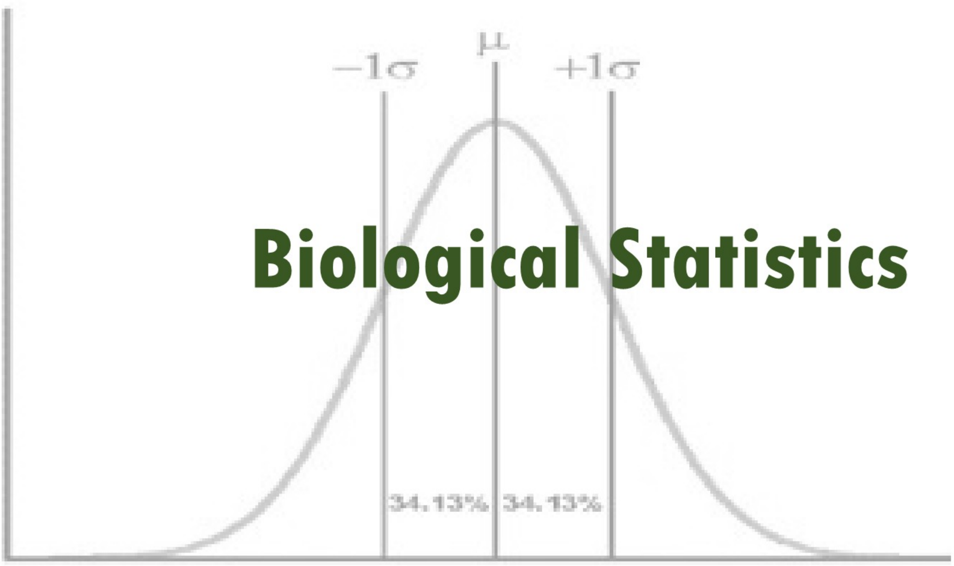
n: عدد القيم.

xi	yi	Xi rank	Yi rank	Pair 1		Pair 2		C/D
15	77	4	3	4	3	3	2	C
25	80	3	2	4	3	2	4	D
30	35	2	4	4	3	1	1	C
50	90	1	1	3	2	2	4	D
				3	2	1	1	C
				2	4	1	1	C

$$\tau = \frac{4-2}{6} = 0.33$$

- يتم ترتيب قيم X تصاعديا وتستخراج رتبها.
- تستخرج رتب Y وتوضع كل رتبة امام ما يقابلها من X.
- تجرى مقارنات متعددة بين كل زوج وباقي الازواج لتحديد عدد الازواج المتوافقة Concordant (n_c) وغير المتوافقة Discordant (n_d)


13
12
11
10
9
8
7
6
5
4
3
2
1
0
T
H
A
N
K
Y
O
U
T H A N K Y O U



Biological Statistics

اختبار T للمقارنة بين متوسطين T-test

وهو من الاختبارات المعلمية Parametric test ويستخدم لمقارنة متوسطين (عينتين) ومعرفة مدى معنوية الاختلاف بينهما، أي هل ان الختلاف ناجم عن الصدفة أم لا ويستخدم عندما يكون حجم العينة صغير (30 مشاهدة أو أقل).

شروط الاختبار

- توزيع بيانات العينة يجب ان يتبع التوزيع الطبيعي Normal distribution.
- يجب أن تكون العينة عشوائية أي غير منحازة.

أنواع اختبار T

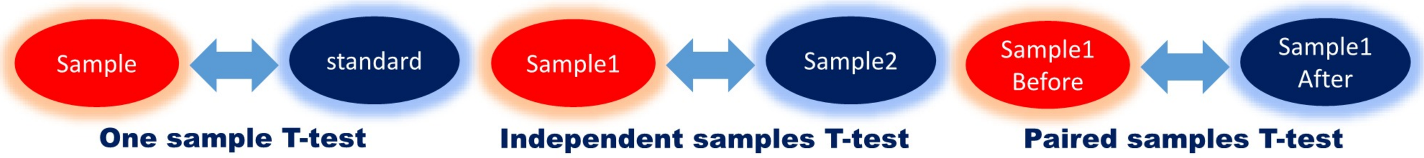
هناك ثلاثة أنواع من هذا الاختبار:

1. اختبار T لعينة واحدة One sample T-test: يستخدم لمقارنة متوسط عينة مع متوسط قياسي معروف وقياس معنوية الفرق بينهما مثل قياس معدل ضربات القلب/دقيقة لعينة من الطلبة ومقارنته مع المعدل الطبيعي (72 ضربة / دقيقة) لقياس مدى اختلاف معدل العينة عن المعدل القياسي وهل ان هذا الاختلاف معنوي ام لا؟ حيث ان الموسط القياسي المعروف هنا هو الـ72.
2. اختبار T لعينتين مستقلتين Independent sample T-test: يستخدم لمقارنة متوسطي عينتين مستقلتين مثل قياس معدل ضربات القلب/دقيقة لعينة من الطلبة من المجموعة A مع عينة من الطلبة من المجموعة B.

أنواع اختبار T

هناك ثلاثة أنواع من هذا الاختبار:

3. اختبار T لعينتين مرتبطتين Paired samples T-test: يستخدم لمقارنة متوسطي عينتين مرتبطتين مثل مقارنة متوسط عدد ضربات القلب/دقيقة لعينة من الطلبة قبل اداء تمرين رياضي معين مع متوسط عدد ضربات القلب/دقيقة لنفس العينة بعد أداء التمرين الرياضي.



كيفية اجراء اختبارات T ؟

اختبار T لعينة واحدة One sample T-test:

يتم إجراء هذا الاختبار من خلال حساب قيمة T المحسوبة ومقارنتها بقيمة T الجدولية عند درجات حرية = n-1 ومستوى معنوية 0.05 فاذا كانت قيمة T المحسوبة أكبر أو تساوي قيمة T الجدولية فإن الفرق بين متوسط العينة والمتوسط القياسي معنوي واذا كانت اقل فالفرق غير معنوي:

$$|t| = \frac{\bar{X} - \mu_0}{S_x/\sqrt{n}}$$

حيث أن:

$|t|$: قيمة T المحسوبة وتكون مطلقة لأن الاشارة لا تؤثر على النتيجة.

$\bar{X}$: متوسط العينة.

μ_0 : المتوسط القياسي للمجتمع.

$S_x/\sqrt{n}$: الخطأ القياسي للعينة.

مثال : لو أخذنا قياس عدد ضربات القلب/دقيقة لعينة من الاشخاص واردنا معرفة مدى اختلافها عن المعدل الطبيعي لضربات القلب (72 ضربة/دقيقة) لمعرفة هل أن الاختلاف في عدد ضربات القلب طبيعي أم معنوي غير خاضع للصدفة.

Xi	70	77	90	100	85	80	72	65	90	88	76	101	95	87

الحل : بداية علينا ان نتأكد من كون البيانات متوزعة طبيعيا أم لا باستخدام طريقة Shapiro-wilk؟

$$W = \frac{\left(\sum_{i=1}^{n} aiX_{(i)}\right)^2}{\sum_{i=1}^{n}(Xi - \bar{X})^2}$$

$$W = \frac{39^2}{1594} = 0.967$$

Xi	*sorted xi*	*(Xi-mean)^2*	*ai*	*X(i)*	*ai*x(i)*
70	65	361	0.5251	36	18.9036
77	70	196	0.3316	30	9.948
90	72	144	0.246	23	5.658
100	76	64	0.1802	14	2.5228
85	77	49	0.124	13	1.612
80	80	16	0.0727	8	0.5816
72	85	1	0.024	2	0.048
65	87	9			
90	88	16			
88	90	36			
76	90	36			
101	95	121			
95	100	256			
87	101	289			
Sum		1594			39.274

حيث أن قيمة W الجدولية = 0.953 أي اصغر من المحسوبة وان قيمة P = 0.7 أي أكبر من 0.05 أي أن العينة متوزعة طبيعيا. عليه يمكننا الانتقال إلى اختبار T

الحل :

Xi	70	77	90	100	85	80	72	65	90	88	76	101	95	87	1176
Xi^2	4900	5929	8100	10000	7225	6400	5184	4225	8100	7744	5776	10201	9025	7569	100378

$$S^2 = \frac{\sum {X_i}^2 - \frac{\sum (X_i)^2}{n}}{n-1} = \frac{100378 - \frac{(1176)^2}{14}}{14-1} = 122.615$$

$$S = \sqrt{S^2} = \sqrt{122.615} = 11.037$$

$$|t| = \frac{\bar{X} - \mu_0}{S_x/\sqrt{n}} = \frac{84-72}{11.037/\sqrt{14}} = 4.054$$

بما أن قيمة T الجدولية = 2.144 وأن T المحسوبة أكبر منها بكثير، إذا متوسط عدد ضربات قلب الاشخاص العينة (84) يختلف اختلافا معنويا عن المتوسط القياسي (72 ضربة/دقيقة).

كيفية اجراء اختبارات T ؟

اختبار T لعينتين مستقلتين Independent samples T-test:

يتم إجراء هذا الاختبار من خلال حساب قيمة T المحسوبة ومقارنتها بقيمة T الجدولية عند درجات حرية = n1+ n2 -2 ومستوى معنوية 0.05 فاذا كانت قيمة T المحسوبة أكبر أو تساوي قيمة T الجدولية فإن الفرق بين متوسط العينة والمتوسط القياسي معنوي واذا كانت اقل فالفرق غير معنوي:

$$|t| = \frac{\bar{X} - \bar{Y}}{\sqrt{(S_x^2/n_1) + (S_y^2/n_2)}}$$

حيث أن:

$\bar{X}$: متوسط العينة X.

$\bar{Y}$: المتوسط القياسي للمجتمع.

S_x^2: تباين العينة X.

S_y^2: تباين العينة Y.

n_1: عدد مشاهدات العينة X.

n_2: عدد مشاهدات العينة Y.

مثال: لو أردنا اختبار معنوية الفرق بين متوسطي عدد ضربات القلب بين العينة X والعينة Y واللاتي كانت بياناتهما كالتالي وعلى اعتبار انهما متوزعتين طبيعيا:

Xi	70.00	77.00	90.00	100.00	85.00	80.00	72.00	65.00	90.00	88.00	76.00	101.00	95.00	87.00
Yi	82.00	80.00	100.00	102.00	77.00	85.00	72.00	70.00	85.00	72.00	77.00	75.00	80.00	75.00

الحل:

	Xi	*Yi*	$\bar{X}-\bar{Y}$	S_x^2/n_1	S_y^2/n_2
	70.00	82.00	3.14	8.76	6.72
	77.00	80.00			
	90.00	100.00			
	100.00	102.00			
	85.00	77.00			
	80.00	85.00			
	72.00	72.00			
	65.00	70.00			
	90.00	85.00			
	88.00	72.00			
	76.00	77.00			
	101.00	75.00			
	95.00	80.00			
	87.00	75.00			
Mean	84.00	80.86			

$$|t| = \frac{\bar{X}-\bar{Y}}{\sqrt{(S_x^2/n_1)+(S_y^2/n_2)}} = \frac{84-80.86}{\sqrt{8.76+6.72}} = 0.798$$

$$df = n_1 + n_2 - 2 = 14 + 14 - 2 = 26$$

نستخرج قيمة T الجدولية عند درجات حرية = 26 من جداول قيمة T والتي تساوي 2.06 وحيث أن قيمة T المحسوبة (0.798) اقل من قيمة T الجدولية فهذا يعني عدم وجود فروق معنوية بين متوسطي العينتين X و Y وعليه نقبل فرضية العدم ونرفض البديلة.

كيفية اجراء اختبارات T ؟

اختبار T لعينتين مرتبطتين Paired samples T-test:

يتم إجراء هذا الاختبار أيضا من خلال حساب قيمة T المحسوبة ومقارنتها بقيمة T الجدولية عند درجات حرية = n-1 ومستوى معنوية 0.05 فاذا كانت قيمة T المحسوبة أكبر أو تساوي قيمة T الجدولية فإن الفرق بين متوسط العينة والمتوسط القياسي معنوي واذا كانت اقل فالفرق غير معنوي:

حيث أن:

$\bar{d}$: متوسط الفروقات بين القرائتين المرتبطتين.

S_d: الانحراف القياسي للفروقات

$$|t| = \frac{\bar{d}}{S_d/\sqrt{n}}$$

مثال: لو فرضنا أن لدينا مجموعة من الرياضيين واردنا اختبار تأثير تمرين الهرولة لمدة ربع ساعة على متوسط عدد ضربات القلب قبل وبعد التمرين حيث تمثل X1 عدد ضربات القلب قبل التمرين و X2 عدد ضربات القلب بعد التمرين. على اعتبار ان البيانات متوزعة طبيعيا.

X1i	70.00	77.00	90.00	100.00	85.00	80.00	72.00	65.00	90.00	88.00	76.00	101.00	95.00	87.00
X2i	100.00	102.00	100.00	102.00	110.00	105.00	101.00	95.00	105.00	103.00	100.00	106.00	100.00	105.00

الحل:

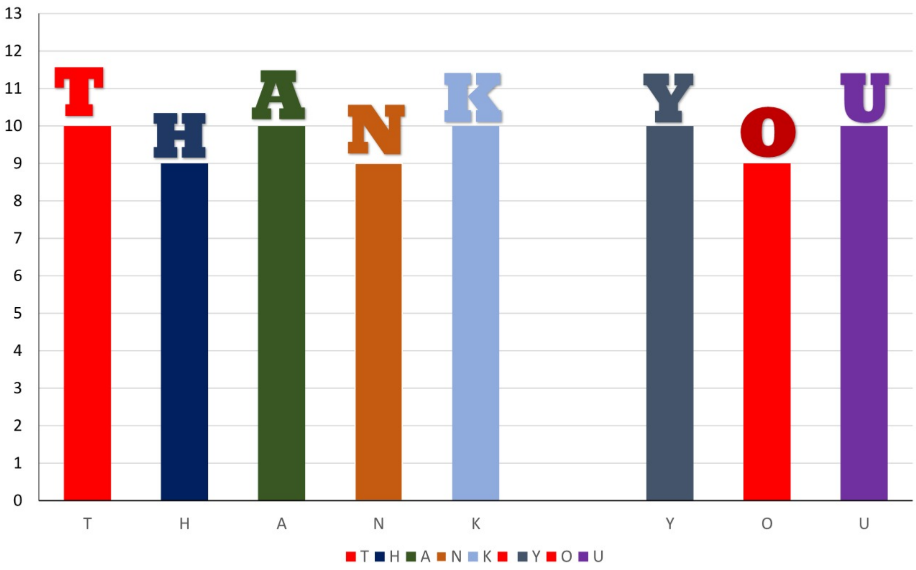
X1i	70.00	77.00	90.00	100.00	85.00	80.00	72.00	65.00	90.00	88.00	76.00	101.00	95.00	87.00
X2i	100.00	102.00	100.00	102.00	110.00	105.00	101.00	95.00	105.00	103.00	100.00	106.00	100.00	105.00
di=|X1-X2|	30.00	25.00	10.00	2.00	25.00	25.00	29.00	30.00	15.00	15.00	24.00	5.00	5.00	18.00

$$\bar{d} = \frac{\sum di}{n} = \frac{258}{14} = 18.43$$

$$S^2 = \frac{\sum {d_i}^2 - \frac{\sum (d_i)^2}{n}}{n-1} = 9.866$$

$$|t| = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{18.43}{9.866/\sqrt{14}} = 6.988$$

وحيث أن قيمة T المحسوبة (6.988) أكبر من قيمة T الجدولية (2.160) اذا فالفرق بين متوسط ضربات القلب قبل وبعد التمرين معنوي وعليع نقبل فرضية العدم ونرفض البديلة



T
H
A
N
K
Y
O
U
13
12
11
10
9
8
7
6
5
4
3
2
1
0
T H A N K Y O U