Notes
مذكرات

Calcualate the value of appropriate test

8.00

عبارة عن value بحسبه من الداتا اللي جمعتها باستعمال eqn

9.00

Z-test → Z-value | ANOVA → f-value
t-test → t-value | chi-square → chi-square

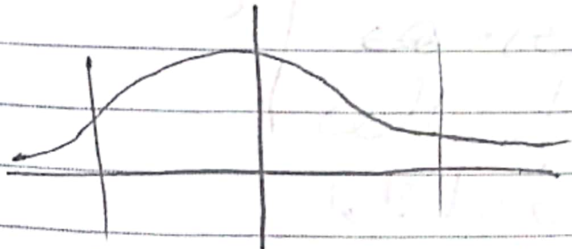
11.00

2 Probability distribution of test statistic

12.00

gives you the probability of occurance
of each value of the test

1.00



2.00

3.00

بعد الحسابات بتخرج لنا قيمة تسمى P-value == α (ألفا)

4.00

وهذه تعبر عن احتمالية حدوث حدث ما أو test ما بالصدفة
ولا تعبر عن شيء حقيقي

5.00

هناك في curve ما يسمى بـ critical value اللي داخلها
areas تعتبر تكون الـ chance
values يعني ممكن كتير الـ test يكون بالصدفة

7.00

كأنه قيمة تفصل بين القيم التي تظهر بالصدفة
وعكسها

December 2022

Notes

ديسمبر ٢٠٢٢

مذكرات

* Make decision

Reject the null

calculated value > tabulated value

or critical region القيمة موجودة في

يعني رفض ال null

هناك اختلاف كبير في الدت

او الملاحظة المشاهدة والقيمة الجدولية

وهذا يعني

level of significant (α)

confidence interval

Accept null

معناه ان الاختلاف المشاهد ليس حقيقي وانه اعطاء كبير يكون حصل بالصدفة

⇒ Z test :

used to compare the Mean of the variable in the sample ($\bar{x}$) to a hypothesized value like population mean

used when there is only one sample or one group data

If the ~~sam~~ standard deviation of the Population is known

$$Z_{calc} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- Z → normal value
- $\bar{x}$ → Mean sample
- μ → Mean Population
- n → number of samples
- σ → standard deviation of the Population

If it's unknown

use $S \rightarrow \sigma$

$$\frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}$$

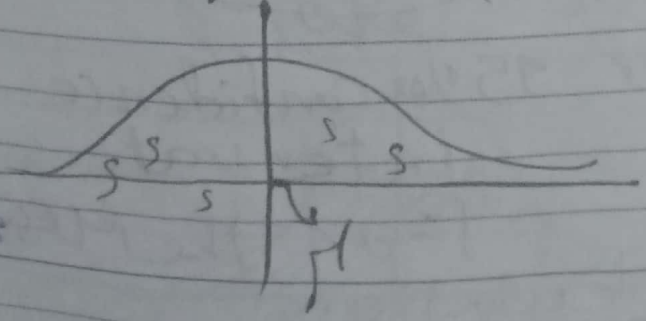
use t-table

Co

Normal dist يكون samples Mean عند توزيع معظم

the Mean of Population (μ) هو نقص التشتت باعتبار

$$\Rightarrow sd = \frac{\sigma_{pop}}{\sqrt{n}}$$



December 2022

Saturday

31

DECEMBER

Sat	Sun	Mon	Tue	Wed	Thu	Fri
31					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

ديسمبر

السبت	الأحد	الاثنين	الثلاثاء	الأربعاء	الخميس	الجمعة
٣١					١	٢
٣	٤	٥	٦	٧	٨	٩
١٠	١١	١٢	١٣	١٤	١٥	١٦
١٧	١٨	١٩	٢٠	٢١	٢٢	٢٣
٢٤	٢٥	٢٦	٢٧	٢٨	٢٩	٣٠

ديسمبر ٢٠٢٢

السبت

٣١

٧ جمادى الآخر ١٤٤٤هـ

٢٢ كيهك ١٧٣٩ق

about 68% are within one standard deviation of the mean and 95% are within two standard deviations of the mean

$\mu = \Sigma$ sample means

2σ ← Margin of error

$\sqrt{n}$

95% of sample means, just of People who will be in this interval

$\left[\mu_{BT} - \frac{2\sigma}{\sqrt{n}}, \mu_{BT} + \frac{2\sigma}{\sqrt{n}}\right)$

$S = \mu$ نقطة المنتصف

Samples Pop

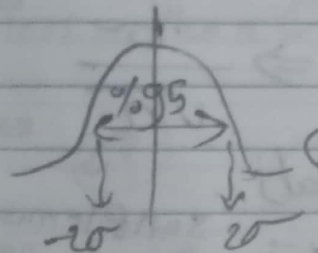
$\bar{X} \rightarrow \frac{2\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{2\sigma}{\sqrt{n}}$

Accepted region or 95% confidence interval for the Mean

with sam calculation

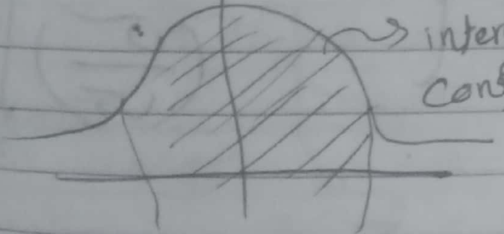
0.95

(للعلم) ←  اهنا كلما كان sample Mean اكثر قربا
M Mean of the Mean يدل هذا على اعتدالية وصالحية
واعتماديه صالحة للصدف



ع [] نذكرها في فترات ثقة Interval ∈

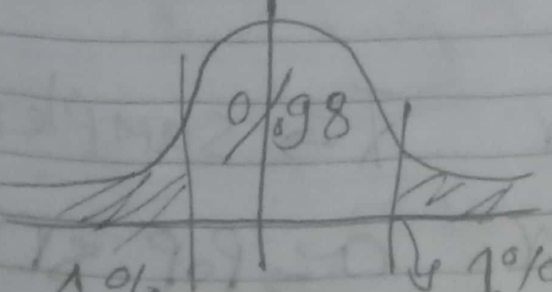
for all, 95% of sample Means fall within 1,96 standard errors from the Population Mean



rule $\bar{X} - 1,96 * \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} +$

rule → The greater our samplesize, the smaller is the confidence intervals

Make 98% confidence intervals



Z ← 2,32

٢٠ كيهك ١٧٢٩ق ٥ جمادى الاخر ١٤٤٤هـ

① calculate Point Estimat or the Population Mean خطوات الحل

② calculate the standard error ⇒ $\frac{\sigma}{\sqrt{n}}$

③ calculate the confidence intervals

for 0.95 ⇒ $\bar{X} + (1.96 * \frac{\sigma}{\sqrt{n}}) < \mu < \bar{X} - (1.96 * \frac{\sigma}{\sqrt{n}})$ (sample number)

⑤ calculate Z values by $\left\{ \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \right\}$

$\sqrt{}$: Population number $\bar{X}$ → sample Mean

n: sample number σ → Pop SD

μ → Pop-Mean S → Sample SD

December 2022 | DECEMBER | ديسمبر | ديسمبر ٢٠٢٢

Wednesday | الاربعاء

28 | ٤ جمادى الآخر ١٤٤٤هـ | ١٩ كيهك ١٧٢٩ق | ٢٨

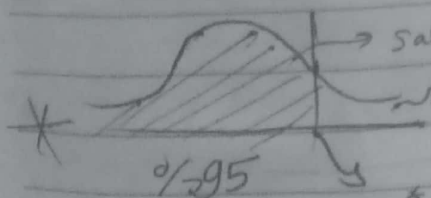
Levels of likelihood

* If the Probability of getting a sample Mean is less than ① ,05 , or 5% ② 1% or ,01

③ ,001 or ,1%

that's consider unlikely to occur

these level called α levels



Probability of (%95) = 1,65

$\bar{z}_{critical}$ = 1,65

* Z critical of (%1) is 2,33

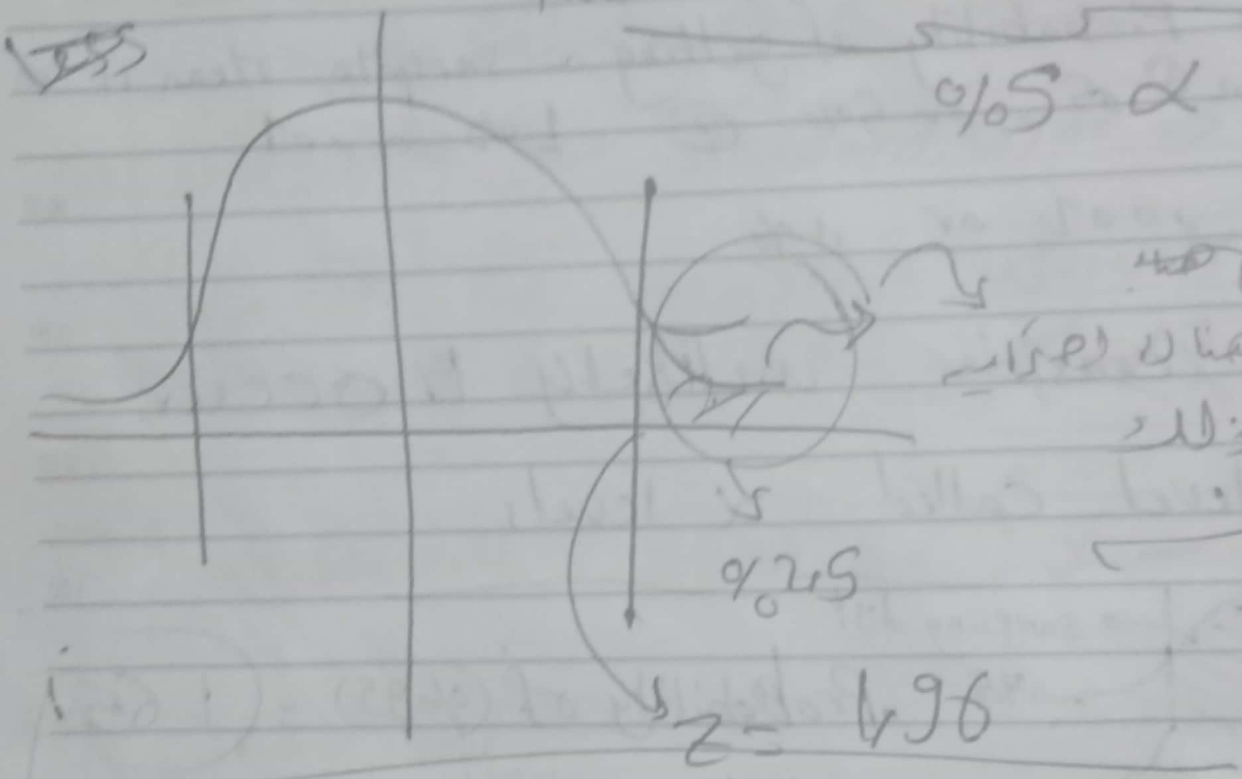
Z critical of (,1%) 3,08

for any hypothesis

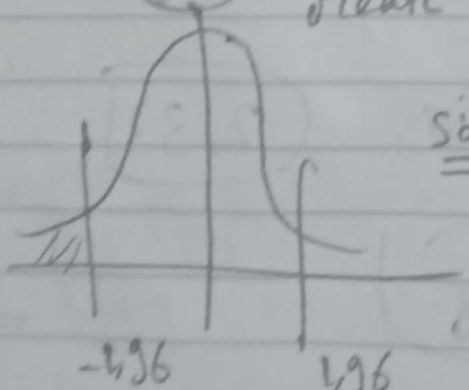
$$z = \frac{\bar{x} - \mu}{\sigma(\sqrt{n})}$$

Enter the interval

Two-Tailed Critical Values



ex) What can we say about the sample Mean $\bar{x}$ = 7,13 [Z-score = -2,59] ?



Sol
1) That's Mean It's unlikely to have gotten a Mean engagment score of 7,13

2) ~~If Mean~~ there is evidence that Katie's singing Made Students less engaged

3) A Mean engagement score of 7,13 is significant at $p < .05$

Monday

26

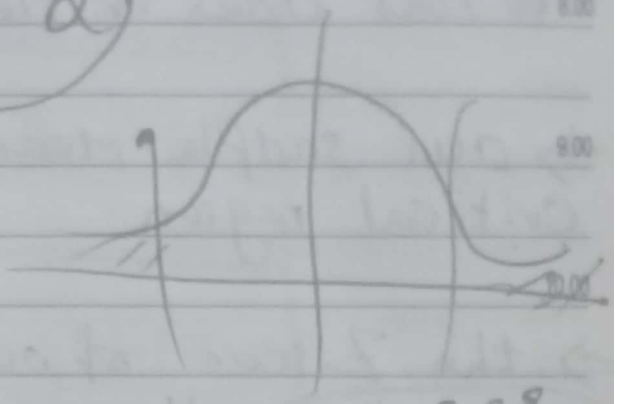
الأثنين

٢٦

٢ جمادى الآخر ١٤٤٤هـ ١٧ كيهك ١٧٢٩ق

② for 1% α

$Z = \pm 2,58$



③ for 0,1% = 0,001 -2,58 2,58

$Z = \pm 3,27$

one-tail	Two-tailed
$\alpha = 0,05$ $Z = 1,65$	$\alpha = 0,05$ $Z = \pm 1,96$
$\alpha = 0,01$ $Z = 2,32$	$\alpha = 0,01$ $Z = \pm 2,57$
$\alpha = 0,001$ $Z = 3,08$	$\alpha = 0,001$ $Z = \pm 3,27$

null hypothesis: H_0

الفرضية الصفرية $\mu_1 = \mu_2$

there is no significant difference

alternative hypothesis H_1

$\mu_1 < \mu_2$ $\mu_1 > \mu_2$

$\mu_1 \neq \mu_2$

there is difference

Hint:- we can't Prove that H_0 is true. we can only obtain evidence to reject the null hypothesis

Sunday
25
١٦ كيهك ١٧٢٩ ق
١ جمادى الآخر ١٤٤٤ هـ

What does it Mean to reject the null hypr?

⇒ our sample Means fall (within) outside the critical region ✗

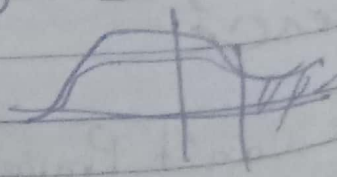
⇒ the Z-score of our sample Mean is (greater than) less than the Z-critical value ✗

⇒ the Probability of obtaining the sample Mean is less than [redacted] the alpha level

Significant difference : Means the Sample Statistic we test is very different than what we would expect for all other samples we could have gotten from the Population

$$\left(\bar{z} = \frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right)$$

reject the null ⇒ lie in the critical region



	Reject Ho	Retain Ho
Ho True	wrong disci Type I error	True discion
Ho false	Correct discion	Wrong discion Type II error

Problem set « estimations
~ α the range of confidence interval

TMP Problems

For understand (Udacity Problem set 8

Numerous studies have demonstrate that listening to music while studing can improve Memory.

To demonstrate this phenomenon a researcher obtains a sample of 36 college student and give them a standardized Memory test while they listen to background Music. Under normal circumstances (without any music) the scores on this test form a normal distribution with μ = 25 and σ = 6. The sample mean is 28. Use a two-tailed hypothesis test.

ember 2022

الأربعاء

lednesday

٢١

21

٢٧ جمادى الأول ١٤٤٤هـ

١٢ كيهك ١٧٢٩ق

Z-test works when we know σ

Samples

How different a sample Mean is from a population.

2- How different two sample Means are from each other ~~dependent~~ ~~independent~~

Sample Mean $\sqrt{\frac{\Sigma(Xi-X)^2}{n-1}}$

When n increase the

① t-dis approaches a normal distribution

② the t-distribution act skinner tails

③

Tuesday
20

١١ كيهك ١٧٢٩ق
٢٦ جمادى الأول ١٤٤٤هـ

ديسمبر ٢٠٢٢
الثلاثاء
٢٠

* degree of freedom
→ it's the number of pieces of information that can be freely varied without violating any given restrictions.

whil (n-1) degree of freedom increases, the t-dist is better approximates the normal dist

ex) 1

You have a sample of size 30. what is the t-critical values for two tailed test with $\alpha = 0.05$?

Sol:- You can search at $(n-1, \frac{\alpha}{2})$ in table

because two-tailed

$$t \text{ value} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

Sunday 18 الأحد ١٨

٩ كيهك ١٧٣٩ق — ٢٤ جمادى الأول ١٤٤٤هـ

when we use large sample size (n) the t-dist goes from wider to skinner

Dependent: t-test for paired samples

~~for last t for~~ ~~α = .05~~ critical

Dependent samples

① Repeated Measures design

دي حالة الفرق بين الأخطاء اللي ممكن أرتكبها لو استخدمت الأكبر من النوع أ والأخطاء — استخدام الأكبر نوع ب

② longitudinal design

نفس العينة نفس الكلام لكن في وقتين مختلفين

H_0: $time_1$ & $time_2$ null will $time_1 - time_2$

③ Pretest Posttest

H_0: μ

ملاحظات

December 2022
Saturday
17

٨ كيهك ١٧٣٩ق
٢٣ جمادى الأول ١٤٤٤هـ
السبت ١٧

Types of effect size Measurs

1) Difference Measures
→ Mean difference → SD differences
Like Cohen's d ⇒ standardized difference
Cohen's d = SD units

2) correlation Measures
r^2 - Proportion (%) of variation in one variable that is related to another variable

Statistical significance

Means:-
1) rejected the null hypothesis.
also
2) results are not likely due to chance

☎ when you see word "statistical significant"
All it Means is that the results are Probably not due to Chance

December 2022

Friday

16

DECEMBER

Sat	Sun	Mon	Tue	Wed	Thu	Fri
31					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

ديسمبر

الجمعة	الخميس	الأربعاء	الثلاثاء	الاثنين	الأحد	السبت
٢	١					٣١
٩	٨	٧	٦	٥	٤	٣
١٦	١٥	١٤	١٣	١٢	١١	١٠
٢٣	٢٢	٢١	٢٠	١٩	١٨	١٧
٣٠	٢٩	٢٨	٢٧	٢٦	٢٥	٢٤

دیسمبر ٢٠٢٢

الجمعة

١٦

٢٢ جمادى الأول ١٤٤٤هـ

٧ كيهك ١٧٣٩ق

r^2 ⇒ coefficient of determination

$\alpha \longrightarrow 1$

$r^2 = t^2 / (t^2 + df)$

↙

not critical

8.00

9.00

10.00

Results sections

11.00

1. Descriptive Statistics (Mean, SD)

12.00

↓↓

it's the heart of our study

1.00

We can report those in 3 ways

2.00

in text - in graphs - in tables

3.00

② inferential statistics (hypothesis test - confidence intervals

4.00

~~②~~ test statistics (the value of t

5.00

* the degree of freedom * P value

6.00

+ direction of the test (one tail - two tail)

7.00

Don't forget the alph level α

APA style :-

منكرات

$t(df) = X,XX$, $P >=< .,XX$, direction

$t(24) = -2,5$, $P < ,05$, one tailed

summary :→ We need CI, CI level (α)
lower limit, upper limit

ex :→ 95% CI = (4, 6)

Formulas

df = n − 1 , standard error (SE) = $\frac{s}{\sqrt{n}}$

$t = \frac{\bar{x} - \mu}{SE}$ " " Margin of error = $(t_{critical} * SE)$

(CI) $\bar{x} \pm$ Margin of error

cohen's d = $\frac{\bar{x} - \mu}{s}$

$r^2 = \frac{t^2}{t^2 + df}$

not-critical

منكرات

December 2022 | Wednesday | 14 | ديسمبر ٢٠٢٢ | الاربعاء | ١٤

٢٠ جمادى الأول ١٤٤٤هـ | ٥ كيهك ١٧٣٩ق

* When P-value of 5% or lower is often considered to be statistically significant

* Statistical significance is a determination about the null hypothesis which suggests that the results are due to chance alone.

A data set provides statistical significance when the P-value is sufficiently small

* When P-value is large, then the results in the data are explainable by chance alone and the data are deemed consistent with null hypothesis "Accept null"

* When the P-value is small "typically 5% or less" the results are not easily explained by chance only and the data are deemed inconsistence with the null hypothesis. So the null hypothesis is rejected

statistical significance

* alternative hypothesis tells us what we're hoping to find.

* the standard error tells us that we expect sample ean to differ from the true population mean y $/a, on average.

explain Meaning of P-value

assume t = -1,711

calc

P =

-1,711

the P-value is the area less than -2,15

it just a proportion scores falls beyond a t-value

so P < 0,05

December 2022

Monday

12

DECEMBER

ديسمبر ٢٠٢٢

الأثنين

١٢

١٨ جمادى الأول ١٤٤٤هـ

٢ كيهك ١٧٣٩ق

Advantages of dependent samples

→ Repeated Measures ←

* Two conditions * longitudinal * Pre & Post - test

① it controls individual differences

② use fewer subjects ③ cost-effective

④ less time-consuming

⑤ less expensive

disadvantages

* carry-over effects

* the second measurement can be affected by first treatment

* order may influence results

Independent samples

Using experimental test

or observational

Very imp note

$N(\mu_1, \sigma_1) - N(\mu_2, \sigma_2) = N(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

ملكرات

So $\sqrt{S_1^2+S_2^2} > S_1$

8.00

9.00
So standard error $= \frac{S_t}{\sqrt{n}} = \sqrt{\frac{S_1^2+S_2^2}{n}} = \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$

10.00

$H_0: \mu_1 - \mu_2 = 0$

11.00

$H_a: \mu_1 - \mu_2 \neq 0$

12.00

1.00
and Because of two n_1, n_2 sample size

2.00
degree of freedom (df) $= n_1 + n_2 - 2$

3.00
$(n_1 - 1) + (n_2 - 1)$

4.00

5.00

6.00

7.00

ملكرات

December 2022

Thursday

8

DECEMBER

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

ديسمبر

						١
٢	٣	٤	٥	٦	٧	٨
٩	١٠	١١	١٢	١٣	١٤	١٥
١٦	١٧	١٨	١٩	٢٠	٢١	٢٢
٢٣	٢٤	٢٥	٢٦	٢٧	٢٨	٢٩
٣٠	٣١					

٢٩ هاتور ١٧٢٩ ق

١٤ جمادى الأول ١٤٤٤هـ

ديسمبر ٢٠٢٢

الخميس

* the smaller the distance between sample Means, the less likely Population Means will differ Significantly

* the greater the variability of each individual sample, the less likely Population Means will differ significantly

So

we will study

analysis of variance (ANOVA)

Any and How much samples we need to test ANOVA Make it in one test

f ratio : it's ratio between group variability and within group variability

and represent How much group means differ

$$F = \frac{n\sum (\bar{X}_K - \bar{X}_G)/(K-1)}{\sum (\bar{X}_i - \bar{X}_K)^2 / df}$$

n, sample size in each group

ملكررات

$$df = n_1 + n_2 + n_3 - K$$

$$= N - K$$

(N total no. of values from all values in samples)

Wednesday
7

DECEMBER

ديسمبر ٢٠٢٢
الاربعاء
٧

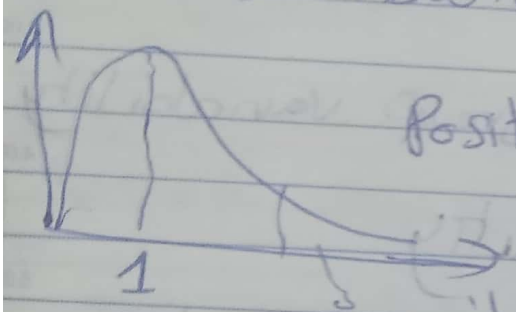
٢٨ هاتور ٦١٧٢٩ق
١٣ جمادى الأول ١٤٤٤هـ

$F = \frac{SS_{between}}{df_{between}} \Big/ \frac{SS_{within}}{df_{within}} \leq \frac{MS_{between}}{MS_{within}}$

$df_{between} + df_{within} = N - 1$

f-distribution is not symMetrical

Positively-skewed



always

$df_{between} = (\text{no. groups} - 1)$

$df_{within} = (n_1 - 1) + (n_2 - 1) + n_3 - 1$

How to calce the $\bar{X}$ grand ?

assume that $\bar{X}_1 = 13$ for sample 1, $\bar{X}_2 = 48$ for sample 2, $\bar{X}_3 = 45$ for sample 3

$$\bar{X}_{Grand} = \frac{13 + 48 + 45}{3} = 35,33$$

ex 2

you have 3 samples

Ahmed	Omar	Sara
15	39	1
12	45	2
14	49	3
11	1	4

$df_{between} = 3 - 1 = 2$

$df_{within} = (4-1) + (4-1) + (4-1) = 9$

3

$df_1 = 4-1$ $df_2 = 4-1$

مذكرات

How to know which sample is differ from other samples?

We use ~~an~~ a new test called:-

《Multiple comparison tests》

One of them is called: "Tukey's HSD"

→ Honestly significant different

هذا الاختبار يعلمنا من اجراء اختبارات ومقارنة ثنائية أي $\bar{x}_1$ & $\bar{x}_2$ ، $\bar{x}_3$ & $\bar{x}_1$ وهكذا

Margin of error $= z^* \cdot \frac{\sigma}{\sqrt{n}}$

or $= t^* \cdot \frac{s}{\sqrt{n}}$

with Tukey's HSD we have q^*

$= q^* \sqrt{\frac{MS_{within}}{n}} = q^* \cdot \frac{S}{\ldots}$

Studentized Range statistics

the HSD value which we calculate from eqn:-

Tukey's HSD $= q^{*}\sqrt{\frac{MS_{within}}{n}}$

Means that if any two samples have an absolute difference greater than 2,51 , this difference is consider honestly significant

So $IF|\bar{X}_1 - \bar{X}_2| > =$ HSD value

this test Made only IF all sample size is the same

Cohen's d for knowing the effect of sample size

$$\frac{\bar{X}_i - \bar{X}_j}{\sqrt{MS_{within}}}$$

another effect size Measure for one way ANOVA is caled eta squere η^2

η^2 similar to r^2 in t-test

* this effect size tells us the Proportion of total variation that is due to between group differences

$$\eta^2 = \frac{SS_{between}}{SS_{between} + SS_{within}}$$

Reporting results of ANOVA

$F(2,6) = 27$ $P < ,05$; $\eta^2 = 90\%$

ANOVA assumptions

① normaly distributed

② homogeneity of variance

③ Independence of observation

منكرات

t test for test the true correlations

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

راجع درس ال correlation

رقم 17 كورس

intro to infere statistical

Sat	Sun	Mon	Tue	Wed	Thu	Fri
31					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

السبت	الأحد	الاثنين	الثلاثاء	الأربعاء	الخميس	الجمعة
٣١					١	٢
٣	٤	٥	٦	٧	٨	٩
١٠	١١	١٢	١٣	١٤	١٥	١٦
١٧	١٨	١٩	٢٠	٢١	٢٢	٢٣
٢٤	٢٥	٢٦	٢٧	٢٨	٢٩	٣٠

(Measurement scale)

tutorial topics:-
① what measurement scale is?
② four types of measurements scales
nominal - ordinal - Interval - Ratio

definition

* Measurement involves the assignment of a value
* Measurement scales define the nature of values

some scales is word measurement
like gender {Male - female}
others use numbers

Properities of measurement scales:
① identity: each value on the measurement has unique meaning
② Magnitude: values has an ordered relationship to one another
③ Equal interval

⇒ scale units along the scale are equal to one another

متكررات
↓
some are smaler and some are larger

for ex:-

the difference between 1,2 == the diff between 19,20 "درجات الحرارة - القوة وهكذا"

4* A minimum value of zero: the scale has a true zero point that's mean there is a case that there is no value exist.

no value < 0 صفر

absolute zero, equal interval الفرق الرئيسي بين
هو هنا الصفر لا يوجد اقل من الصفر
في absolute zero

X^2 test used for categorical data

df = no. categories − 1

مهم جدا lesson Kai square ان 20 درس

expected values حساب

Cramer's V $(\Phi_c) = \sqrt{\dfrac{X^2}{n(k-1)}}$

k ه الأصغر بين عدد ال row, column