A

Short Questions

hypotesis → points towards possible relationships between variables

Null hypothesis → no effect
Alternative hypothesis → an effect

test intelligence | dependent variable
whatever is being ~~a~~ tested is the dependent variable
independent variable → gets manipulated by the experimenter.

Random sample → every member has the same chance of being included.

→ non directional alternative hypothesis → use 2 tailed test.
→ Direction of Prediction uncertain /(determine if there is a difference)
→ directional (one tailed) gives direction eg Group A's productivity is higher than Group B

Degree of freedom is always less than N = formula (N − 1)

manipulation of the independent variable in an experiment is called the treatment factor.

Type 1 error when null hypothesis is rejected when its true.

Test statistic < critical vallue - Do Not reject
" > " " - Reject.

B/

df total = N - 1

→ df total = df group + df error

→ Anova (squared values) is always positive.

→ relationship between nonmetric independent / 2 or more dependent variables ⇒ Multi virate Anova

→ relationship between single nonmetric dependent variable / various metric indepentend variables → multiple discriminant analysis

→ Matched groups N is equal to the number of pairs.

→ In Anova df error = number of cases minus number of groups (N - k)

df = N - k
(done in a table)

- T-test for related samples used when wanting to determine if students that handed in assignments does better in the exams than those that did not
  → the results of the groups have been matched

→ relationship between 2 variables → correlation.
eg relationship between no of days absent and the employee's attitude towards supervisor.

C.

Raw score easier to interpret if you have the Standard deviation.

extreme scores → do not affect → mode
→ median

→ does affect → mean.
→ Range
→ correlation

✓ dependent variable => being measured → Criterion Y
independent variable => manipulated. → predictor — X

measure effected most by extreme score in a distribution.

median → middle score / central value
mode → most frequent occurring score.
mode → best reflection of central tendency
→ not affected by extreme values
mean → average

median
~~mean~~ → not affected extreme values

central tendency → mode - not affected by extreme scores
* → mean → affected
→ median → not affected by extreme scores.

mean → influenced by extreme scores
Range → Affected / influenced by extreme score.

Mode → score that occurs most often

normal distribution = 1 ~~dt~~ standard diviation = 1
↳ mean = 0
↳ 3 standard diviations below → 3 standard diviations above
→ same area between 1 and 2
as -1 and -2

value of the correlation coefficient R can vary between -1 & 1
r =

kuskal wallis → non parametric / one way Anova (F-test)
wilcoxon → non parametric equivalent — t-test matched pair Related Samples
mann-whitney - non parametric " → t-test independent

C/

Descriptive statistics -> organise, summarise + describe data.

Inferential statistics -> technique applied to samples to make inferences about populations / experimental research.

T. test   Non parametric equivalent.

matched pairs tests -> Wilcoxon

Independent samples -> Mann Whitney.

~~K~~

unrelated samples -> Kruskal Wallis

ANOVA equivalent F. test

Related samples -> Wilcoxon

Related groups ->

T- test -> compare 2 estimated population means

Non-parametric -> tests that do not rely on parameter estimation / distributional assumptions

most important difference between correlation + regression is that regression is used to make predictions

null hypothesis = statistical hypothesis

increase in sample size decreases ~~α~~ probability of error II β

Science principal -> can be published + repeated by other researchers / scientists

→ Comment on graphic display
→ Symmetry → modality
→ skewness → kurtosis

Anova - used to test for differences between the means of 2 or more groups.

T- Test - 2 independent samples (means)

R = correlation coefficient → relationship 2 variables
chi-square test → analyse frequency / categorical data.
→ variables

Population → set of events → Parameter ⇒ Population data.
Sample → observed scores → Statistic ⇒ sample data

Alternative H

$H_0 \mu$ = sample mean in Null hypothesis

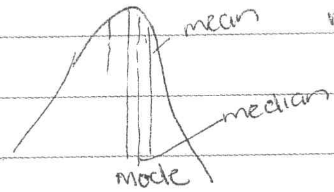
Variances + Standard deviations are always positive.

* mice starved / mice fed → Running speed you should be able to compute correlation between them

Frequency polygon ⇒ Line chart → both this + ~~bp~~ histogram is drawn from frequency information ∴ should provide same info if based ~~on~~ the same data

* Raw score you need → mean μ
→ Standard deviation σ
→ no of cases N

possitively skewed distribution

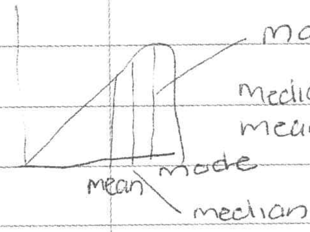


mode -> highest.

median ->

mean -> smallest

negatively.



mode -> highes

median middle

mean ~~highest~~ small

correlation ->effected by extreme/outlying points

-> misleading if the population are not homogenous

generalization of sample data => population => statistical inference

regression -> make prediction.

average score for ~~exp~~ example of a class = descriptive statistic

Categorical data = frequency data

sample => observations drawn from a population

histogram -> extreme score -> highest part

data in a frequency table/graph is called a frequency distribution distribution

nominal scales -> used for stat. processing => way to measure

Descriptive statistics -> describe a set of data

inference statistics -> generalisation from sample to population

central values = descriptive of nature.

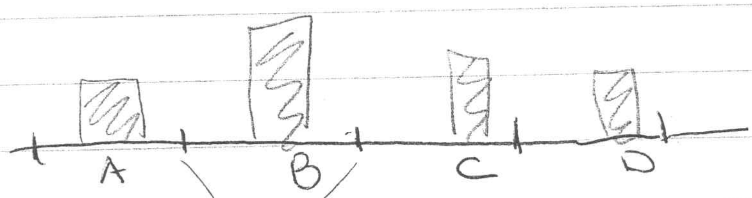
hypoteois => research hypotheois want to research

① Frequency distr. table

	f	fcummulative	%	cum. %
A } write in order it was given / or highest / lower	↓	A	A/N	% +
-		A+B	B/N	A% + B%
-		A+B+C	C/N	ABC % together
-		↓	D/N	Total
-				
	N = *sum.		100%	

Graph

NOTE: gaps between graphs



Note between Markers

Real limits → eg 93 - 98

Real limit 92.5 → 98.5

If an interval size is given. calculate

Midpoint → halfway mark of the class interval

Real limit → class interval → .5 (or if given .9) .3 etc.

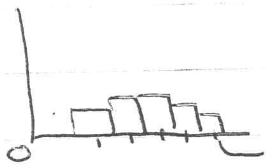
below the low level +

.5 above top level.

DRAW the histogram with midpoints calculated above

(2)

Histogram

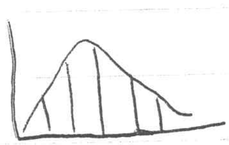


classmidpoint
no gaps between classes / gap bars
could have blank space in beginning

kurtosis -> peakness / skewness.

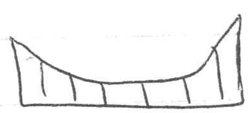
 — Positive  — negative.

unimodal symetrical



peak in middle

Symetrical Bimodal



peak on sides

Symetrical -> center most points in midde

Peaked - leptokurtic

normal mesokurtic

Flat. platykuric

∴ peaked - leptokurtic
normal - mesokurtic.
flat -> platykuric.

(3)

pesentile rank

% below = cum. % below class in which you are calculating for.

Score of P = RLL =

Find the cumm % where the score will fall in, that is what you use

interval % = % f of the group you are workink with

Use .2 desimals. 0,43
2,53
7,41 etc.

Always do ∴ + write a short summary
eg ∴ Score at 25th persentile is - - -

Mode = number that occurs the most

median = middle = sort data in order write out
use median location $\frac{N+1}{2}$

∴ median = that position → look at the sorted data + find it.

(4)

Mean = $\bar{X} = \frac{\Sigma x}{N}$

Variance = $sx^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N-1}$

calculate x^2 in a table / calculate y^2 and xy (x times y)

Standard deviation =

$sx = \sqrt{s^2x}$

$sx = \sqrt{\text{variance}}$

Scatter diagram → name points A, B
1, 2 etc. as
per question.

- Round to 2 decimals

correlation. 0 = no correlation
1 → strong positive
-1 → strong negative
-0,25 → weak negative
0.25 → weak positive.

Correlation does not imply that one factor causes the other only implies an association.

(5)

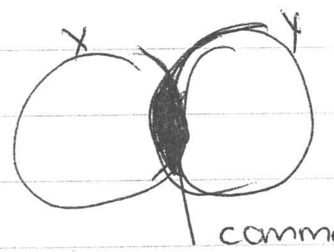
persentage of common variance + diagram

$r = xx$

$r^2 = (xx)^2$

$= xxx \times 100$

$=$ % common variance

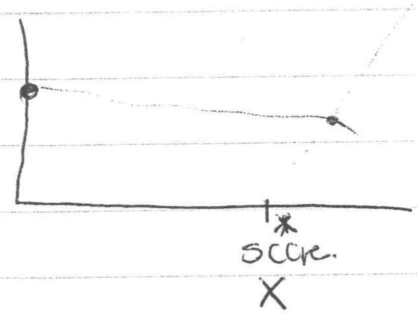


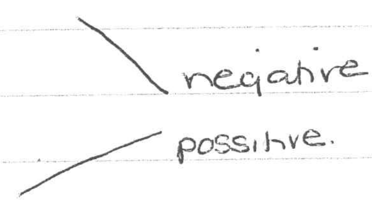
X = predictor variable

$$b = \text{slope} = b = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma X^2 - (\Sigma X)^2}$$

predict a score. $\hat{Y} = bx + a$ (where x is given

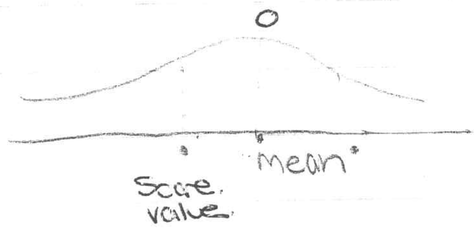
Regression line





6 when doing z table - Standard NORMAL distribution

~~also~~ ALWAYS do Sketch



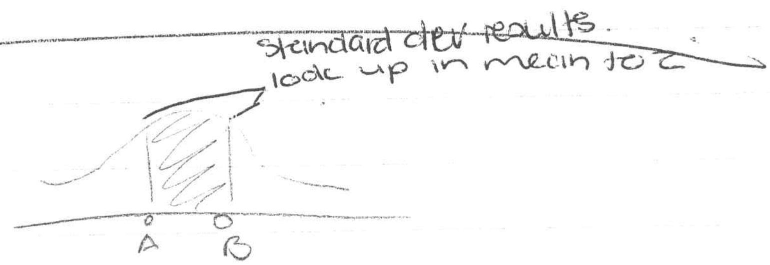
color in the section → look at bigger → smaller ←

~~no~~ number of cases with a raw score use z calculated + look up mean to z proportions for both

∴ Add together

× N = no of people etc.

draw



∴ $(x^1 + x^2) \times N$

2 common fixed limits = z values.

95% = 1.96

99 = 2.58

(7) T. test -> normal distribution
Calculate D + D^2 For Groups. (Same Sample amount N)
D = A - B = D D × D = D^2

Mean of difference scores = $\sum D / N$

one tailed test $\alpha = 0{,}01$ look up in ~~score~~ tables
make sure you look at the correct table
1 tailed + 2 tailed tables / look at Right line.

Rejecting / not rejecting Ho (Null hypothesis)

(chi f, t) test statistic > critical value reject Ho. (Degree of freedom)
" " < " do not reject "

N amounts differ for 2 group.
Null Hypothesis Ho = $H_A = AB$
Alter. " $H_1 = H_A \neq H_B$

Calculate X_A X^2_A $\bar{X}_A = \sum X_A / N(A)$
X_B X^2_B $\bar{X}_B = \sum X_B / N(B)$

only round final answer
write tailed lookups as $t_{0.05}$ / 0.1 etc = value position ——> value lookup

(8)

Reject

Take 100% → Less value of lookup 0,5 , 01

0,01 = 99% sure

0,05 = 95% sure

etc.

for A 99%/95% certainty that there is a significant difference

---

3 or more groups different methods tested/used.

Calculate X^2 for each method + Σ of each

use SS total

group

error

N = total of all participants

n = total no in each group

for SS group remember to do $\frac{\Sigma X \text{ (group)}}{N}$ total participants

Do Table + fill in from calculations above

Source	df	SS	MS	F
Group				
Error				
Total				

9

look up the critical value in table

use df (group →
error ↓)

to determine value - make sure you are

on the right table

eg:

$F_{0,01}: (2,27) = 3,35$

$F_{0,05}: (2,15) = 6,36$

∴ Null Hypothesis = Fcalc F (lookup)

= > Reject

< Do not reject

- Data Set - Chi Test.

Yes/No

2 Groups

Do Table example

		Yes	No	Total
Male	A	10	6	16
Female	B	8	10	18
Total		18	16	34

Calculate E first $E_{11} = \frac{R_1 C_1}{N}$

Do for each value Yes/No Male + Female ∴ 4 in this case

10

Do this by using the totals/

Enter them next to Amounts in table

Eg.

	Yes	No	Total
Male	(A) 10 (8.47)	6 (AA) (7.53)	16
Female	(B) 8 (9.53)	(AB) 10 (8.47)	18
	18	16	34

18 X 16 (A) / 34   16 X 16 (AA) / 34

18 X 18 (B) / 34   16 X 18 (AB) / 34

Now calculate $x^2 = \frac{\Sigma(O - E)^2}{E} + \frac{\Sigma(O - E_1)^2}{E_1} + $ etc.

O = Answer given

eg 10, 6, 8, 10

in this case 4

Remember only the top portion $(\ )^2$ not the bottom part

Use. 2 decimals only. Show all calculations

Critical value → tables (z table) A.1.7) look for the x^2

calculate df = (R-1)(c-1) = ~~GROUPS~~ ANSWERS in this case

2 groups / answers

=(2-1)(2-1) = 1

(11)

NB

x^2 > critical value. - Reject
x^2 < critical value Do NOT reject

Non-parametric equivalent

F. test - Anova. - Kruskal Wallis

T. matched pairs / related groups - t. test Wilcoxon
T - Independent → Mann - Whitney.

Regression ⇒
equation $\hat{Y} = bx + a$
$\hat{Y}$ = predicted Y value on y axis
b = slope
a = Y intercept point
x = predictor variables - value on x axis

used to make predictions
- can never pass through all values unless correlation is perfect
→ minimizes the squared ~~differences~~ distances of the observed data points from the line

---

Variance → indicate the degree to which scores are dispersed
S_x^2 or different from each other.
→ average of squared distances of individual scores from mean of distribution

---

(1) Frequency graph : → Draw [illegible] A B

- Write same as given
- calculate top & bottom

- When doing a histogram calculate · Real limits
  - Midpoint
  - interval

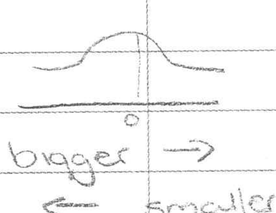
Work with 2 decimals unless from a table

find at 25% or 65% or ...
* find group the % would most likely fall in

PR = Percentile Rank of this score that you wish to calculate

= % below is cummulative % below
interval % is class - % f that we are working in

When doing raw scores check for Bigger or Smaller + draw the graph



use z table to lookup scores

When doing scores in between look up the mean
add together X by N

(2)

Test > Critical value −

Test < critical + Topic? Don't Reject

Hypothesis − Alternative

→ N is the same amount

$H_1 = U_A > U_B$

If N is not same sample amount

$H_1 = U_A \neq U_B$

→ If N is not the same use Pooled variance

Sp

→ you can only use t of f answers in the regecrion

Must **do** t-tests

T. test normally distributed

$t = \frac{\bar{D} - 0 \rightarrow \text{zero?}}{\frac{SD}{\sqrt{N}}}$

Pooled variance − N's differ

$(N_1 - 1) S_1^2 + (N_2 - 1) S_2^2) \rightarrow$ Standard deviation → NOT $(S^2)^2$

3 Groups Remember

Do X^2 for each

Σ for each

$\bar{X}$ for each + $\bar{X}$ for Group

N = total

n = each group (same)

$\text{group}(\bar{X}_j - \bar{X})^2$

$= (\text{indiv. } \bar{X} - \text{group } x)^2$

(3)

K = no of groups

Do Summary table

Source	df	SS	MS	F
Group				
error				

only df + SS has total

see the formulas

df error = k - amount of groups not (k-1)

Critical value

df as always

'Make sure you are on the right table

df group → top line nummurator

error ↓ side → denominator

Chi-Table lookup X^2 scores

df = (r-1)(c-1) → no of answers

↳ no of groups