TEST AND MEASUREMENT.

1. Defination of terms.
2. A test
3. Measurement
4. Evaluation
5. Validity and reliability.
6. Objectivity of an exam or test.
7. Blooms Taxonomy.
8. Records of combination.
9. Types of questions.
10. Internal and external ~~examples~~. examination e.g Hell-analysis.
11. Measures of central tendency
12. Measures of dispersed variability.
13. Person momentum and ~~correlation~~ correlative.
14. Person range order momentum.
15. Correlation
16. Z scores. ✓

Defination of terms.

- A test is an instrument used for measuring learning outcome
- It is composed of set items of quality and standard
- A quality test is a test that contains all levels of thinking - low - middle - high

A standardized exam is an exam set for a large group of test takers it brings equality in the region e.g KCSE.

A result of the learner from the test in numerical form is called a measurement It is a qualitative value.

Test - To evaluate learner and instructor.
- formative.
exam - summative, high stake
An exam is a form of a test which is more comprehensive and usually it is conducted at the end of term, year or semester.

Measurement.
Educational measurement is the qualitative descriptive of a student achievement interms of the set objective/items.

Usually in exam the scores are expressed in numerical value.
A measurement is usually achieved from a term paper, sit-in and a project.

Evaluation. to give judgement.
It is a professional value judgement placed on a students achievement eg a student who scores 80% in biology (excellent)

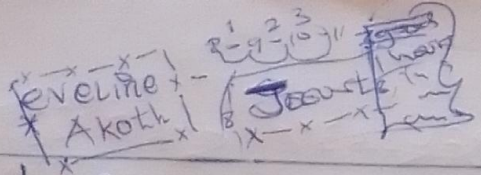
Assessment.
It describes the status of a phenomenon at a particular time.

In assessment there is no placement of a measurement.

* Test, an exam, assessment are all evaluation.

Types of evaluation.
There are 2 types of evaluation;
student evaluation and program evaluation.

1. <u>student evaluation</u> we have 2 types of

3. 

evaluation; formative and summative.

- Formative evaluation goes on throughout the programme.
- They are meant to improve delivery and they include tests

Summative evaluation comes at the end of a course/semester.

2. Programme evaluation.

The success of an educational programme is judged and determined by the following;

1. Quality of the curriculum content interms of ~~scope~~ scope, depth and applicability.
2. level of attainment of long term educational objective.
3. Availability of resources
4. Properties of a good test are;

~~A fe~~ i) Reliability - judged on 2 or more exams.

~~Valu~~ ii) Must be valid - should measure that level.

iii) Objectivity -

Statistics.

Is a science that deals with numbers in collecting data, facts, ideas analyze, draw a conclusion and make decisions.

Z-scores.

1. Is a technique used in equalizing scores from different subject. so as to find

$$z\text{-score} = \frac{X - \bar{X}}{sd}$$

out which subject did better eg a girl scored the following in 3 subjects.

	score (X)	Mean ($\bar{X}$)	standard deviation (sd)
History	80	75	5
chemistry	50	45	2
C.R.E	90	~~80~~ 80	5

In which subject did this student do better. (chem)

For us to answer that question we must change the scores into a z-score

$$z\text{-score} = \frac{X - \bar{X}}{sd}$$

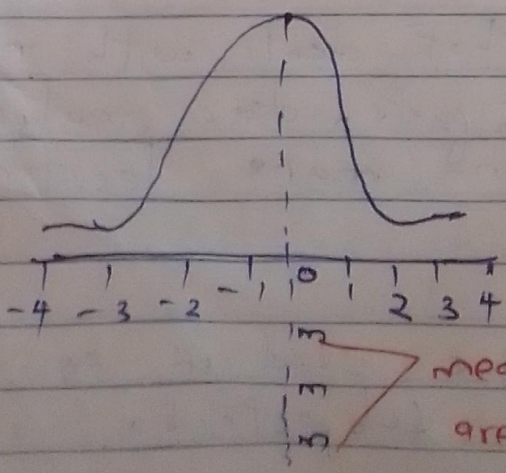
$$C.R.E = \frac{90 - 80}{5} = \frac{10}{5} = 2$$

$$Hist = \frac{80 - 75}{5} = \frac{5}{5} = 1$$

$$chem = \frac{50 - 45}{2} = \frac{5}{2} = 2.5$$

2. ~~It equalizes perform from different subject~~
It is used in drawing curves especially normal curves.

A NORMAL CURVE.



A normal is curve is bell shaped it is symetrical about the mean/ medium.

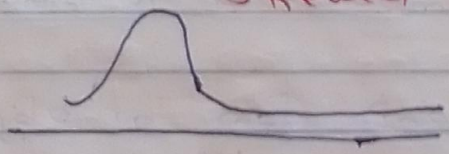
mean, mode and medium are all equal

The area under a normal curve is 1 or 100%.
It is asymptotic (the curve doesn't touch the horizontal line

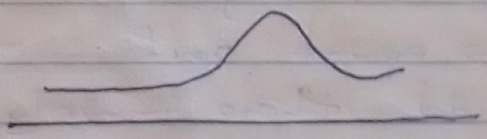
A negative z-score means you scored below mean.
The sd in a normal curve is 1

other curves.

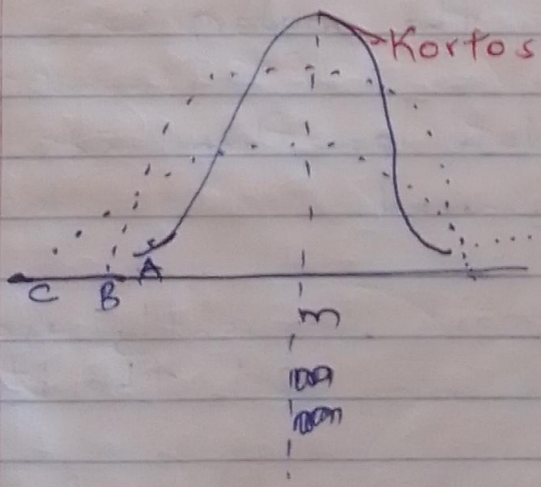
skewed curves.



positive skewed (poor performance)



negative skew (good)



In All the streams have a problem and identify how to solve the problem.

A - Could be having no E but if doesn't have good grades like B+ and A.. The teacher to work hard scoring good grades.

B- Very few E and no excellent graders. cut E's and then add good graders.

C- The teacher should work on reducing E's

MEASURES OF CENTRAL TENDENCY.

These are ~~measures~~ techniques that tell us how scores are concentrated at the centre and hence the name.

They are 3; mode, mean and ~~medium~~ median.

1. Mode - most frequent item or object.
   It is not mostly frequently used

2. Median - It is an item or object that divides a given data into 2 equal parts.
   If the data is odd there is only 1 medium.
   If the data is even the medium will be 2.

   A small data has data less than 28
   It is not frequently used.

3. Mean/average - The most frequently used item. because it is using all the items.

0 2 3 1 5 mode. bimodal (5, 2)

5 1 2 3 5

0 5 3 2 2

single distribution. (If the data is small)

X	Tally	f	fx	cf
0	11	2	00	2

1	11	2	02	4
2	1111	4	08	8
3	111	3	09	11
5	1111	4	20	15
		15	39	

$\bar{X} = \frac{\Sigma fx}{\Sigma f} \quad \frac{39}{15} = 2.6$

Median = $\frac{\text{odd number} + 1}{2}$

$\frac{15+1}{2} = \frac{16}{2} = 8$

∴ median = the 8th item ie 2

2. 0 2 3 1 5

5 1 2 3 5

0 5 3 2 2 5

X	f	fx	C.f
0	2	0	2
1	2	02	4
2	4	08	8
3	3	09	11
5	5	25	16
	16		

Median = $\frac{\text{even}}{2}$ =

$\frac{16}{2} = 8^{th}$

$2+3 = \frac{5}{2} = 2.5$

Grouped data. (large data)

L	U	Class	(x) midpoint	f	fx	cf
79.5	89.5	80-89	84.5	2	169	? 11
69.5	79.5	70-79	74.5	9	670.5	
59.5	modal class 69.5	60-69	64.5 modal score	24	1548	35
49.5	59.5	50-59	54.5	19	1035.5	54
39.5	49.5	40-49	44.5	9	400.5	63
29.5	39.5	30-39	34.5	9	310.5	72
19.5	29.5	20-29	24.5	17	416.5	89
9.5	19.5	10-19	14.5	8	116	97
0	9.5	0-9	4.5	3	13.5	100
				100	4680	

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{4680}{100} = 46.8$$

A distribution where mean, mode and median are equal is called a normal data.
If not equal it is a skewed data.

mode = 64.5

effect - number required to reach the position needed.

Median = $\frac{100}{2}$ = 50th position.

$$m = L + \frac{\text{effect} \times \text{interval}}{f}$$

~~$m = L +$~~ $m = L + \frac{\text{effect} \times I}{f}$

$$m = L + \frac{\text{effect} \times I}{f}$$

$$49.5 + \frac{15 \times 10}{19}$$

$$= 57.4$$

The data is skewed.

Median of a small data. add two classes / 2 — 20.5 Median you look at Cf.

Use the data below to compute on the measures of central ~~dat~~ tendency and give your comment.

	class	x	f	fx	cf	cf
35.5 - 40.5	~~class~~ 36 - 40	38	2	76	2	22
30.5 - 35.5	31 - 35	33	3	99	5	20
25.5 - 30.5	26 - 30	28	4	112	9	17
20.5 - 25.5	21 - 25	23	5	115	14	13
15.5 - 20.5	16 - 20	18	3	54	17	8
10.5 - 15.5	11 - 15	13	2	26	19	5
5.5 - 10.5	5.5 - 10	7.75	3	23.25	22	3
			22	505.25		

$\bar{x} = \frac{\Sigma fx}{\Sigma f} \quad \frac{505.25}{22} = 22.96 \quad 23$

Mode : 23 CL - 21 - 25

Median = $L + \frac{\text{effect} \times I}{f}$ $\quad 20.5 + \frac{3 \times 5}{5} = 23.5$

median = $\frac{22}{2} = 11$ $\quad 20.5 + \frac{2 \times 5}{5}$

~~10.5 + 3x~~

~~10.5 + 2x~~ $\quad = 22.5$

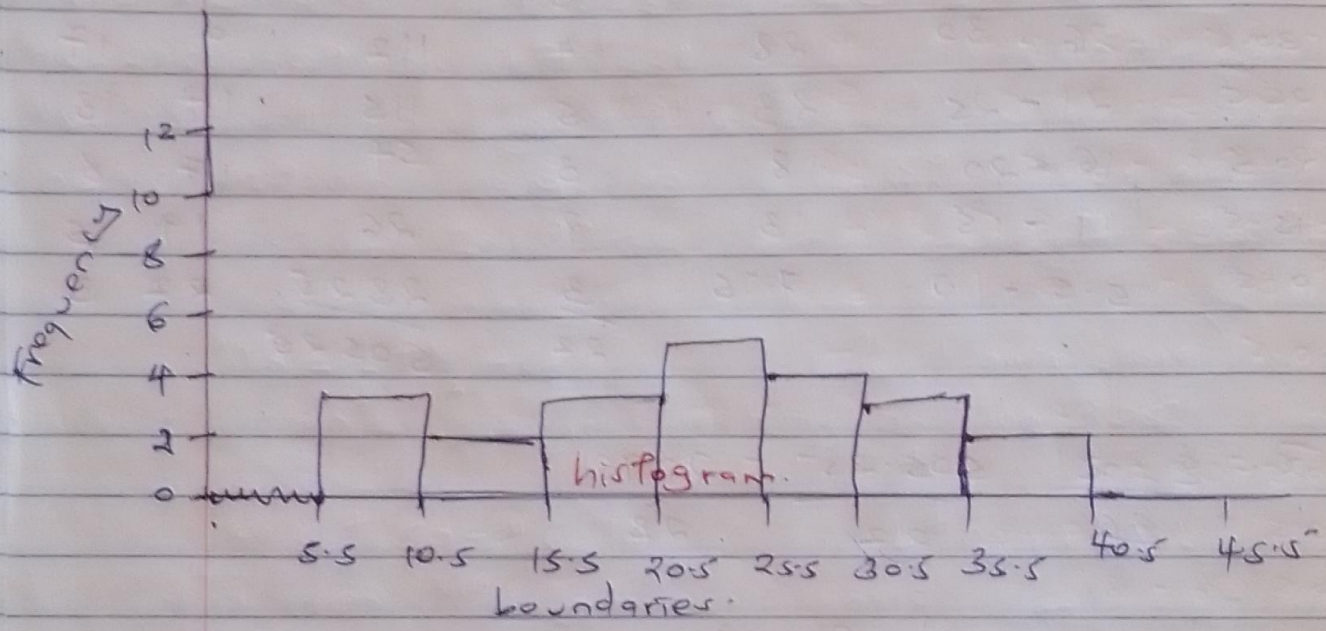
~~A Normal curve distribution.~~

A normal data. because mean, mode and medium are equal.

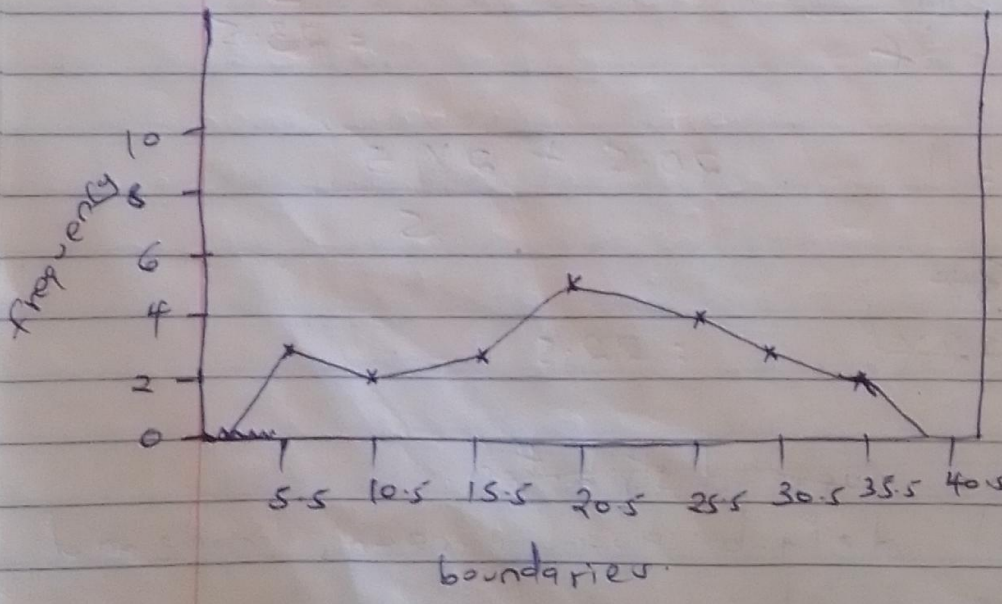
$12^{th} = 20.5 + \frac{4 \times 5}{5} = 24.5$

$\frac{24.5 + 23.5}{2} = 24.$ It is almost a normal distribution.

We use lower boundaries to construct comparative histogram: frequence polygon to be
They make the data continous. so to use to construct histogram, cummulative frequency.
frequency polygon



Frequency polygon.



When you have been given raw data and from the raw data you are asked to develop classes.

5 4 3 6 8 9 10
4 12 15 20 24 3 10
12 12 16 18 21 20 22
11 12 8 6 9 5 4
13 14 15 16 13 20 21

Develop 6 classes from this data.

Steps:

1. Find the range, add 1 to the range over
~~2 Ad~~ number of class ie $\frac{\text{range}+1}{6}$ give us class size.
3.

~~24+~~ $\frac{24-3+1}{6} = \frac{22}{6} = 3.67 = 4$

L	U	class	tally.	F	x	fx	cf
22.5	26.5	23-26	I	1	24.5	24.5	35
18.5	22.5	19-22	~~HH~~ I	6	20.5	123	34
14.5	18.5	15-18	~~HH~~ II	5	16.5	82.5	28
10.5	14.5	11-14	~~HH~~ III	8	12.5	100	23
6.5	10.5	7-10	~~HH~~ I	6	8.5	51	15
2.5	6.5	3-6	~~HH~~ IIII	9	4.5	40.5	9
				35		~~786~~ 421.5	

Mode = 4.5

Mean = $\frac{\Sigma fx}{\Sigma f}$ $\frac{786}{35} = 22.46$ $\frac{421.5}{35} = 12.04$

Median. $L + \frac{\text{effect}}{f} \times i$ $\frac{35}{2} = 17.5$ $\frac{35+1}{2} = \frac{36}{2} = 18$

~~10.5+2.5 x~~

$10.5 + \frac{3}{8} \times 4 = 12.$

Median = 2/4 = 1/2 First quartile score is the score that coincide with 1/4 of the population from the lowest.

First = 1/4 first quarter of the score.

class	f	cf
79.5 - 89.5	80 - 89	2	100
69.5 - 79.5	70 - 79	9	98
59.5 - 69.5	60 - 69	24	89
49.5 - 59.5	50 - 59	19	65
39.5 - 49.5	40 - 49	9	46
29.5 - 39.5	30 - 39	9	37
19.5 - 29.5	20 - 29	17	28
9.5 - 19.5	10 - 19	8	11
0 - 9.5	0 - 9	3	3
		100	

upper quartile. (A third quartile)

$\frac{3}{4} \times 100 = 75$

$Mdn = L + \frac{\text{effect} \times i}{f}$

$59.5 + \frac{10 \times 10}{24}$

$59.5 + 4.166$

$= 63.67$ ~~63.67~~

$Mdn = L + \frac{(\frac{N}{2} - cf)\, i}{f}$

$49.5 + \frac{(50 - 46)\, 10}{19}$

$49.5 + \frac{4 \times 10}{19}$

$49.5 + 2.105266$

$= 51.61$

1/4 (lower score)

$\frac{1}{4} \times 100 = 25$

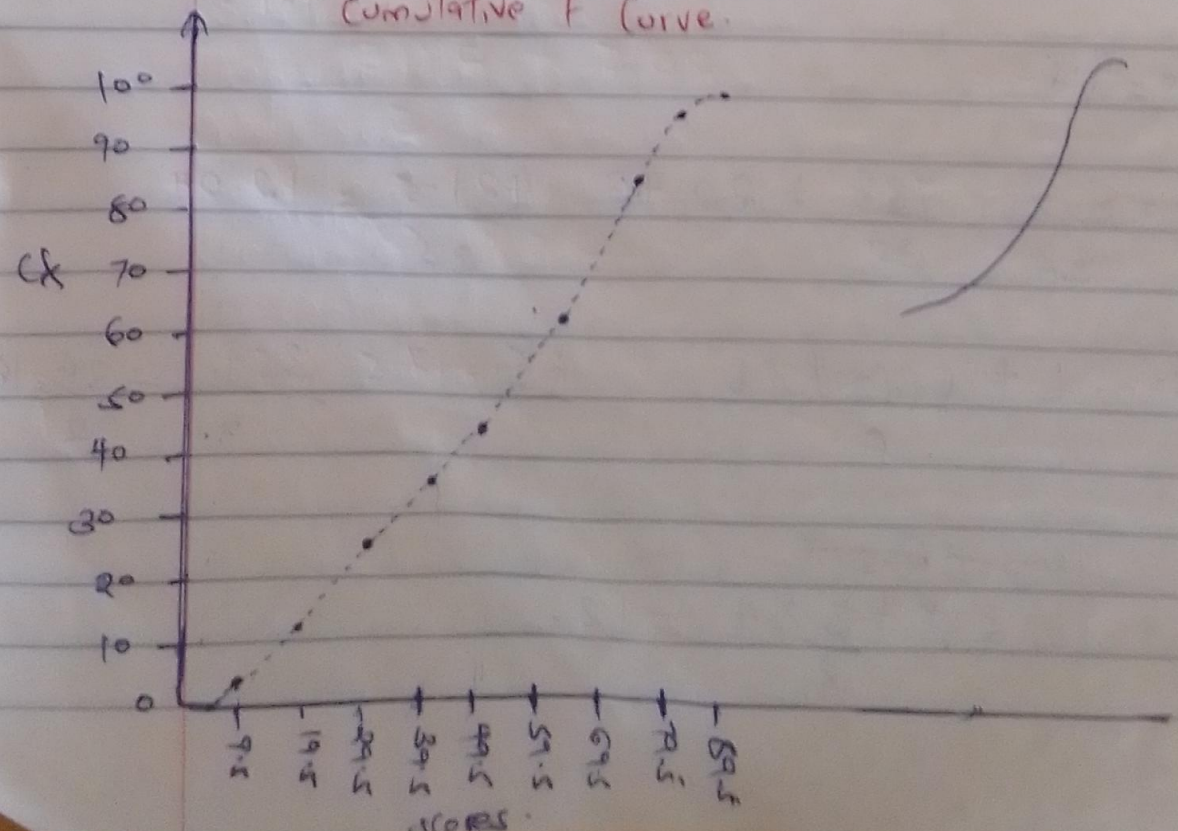
$L + \frac{\text{effect} \times i}{f}$

$19.5 + \frac{14 \times 10}{17}$

$19.5 + 8.24$

$= 27.74.$

Cumulative F Curve.



Validity and reliability.

~~Val~~ creation of reliable.

split-half method

~~Test~~ Test-retest. - The same test items are repeated to students after some time.

Parallel form - There are 2 exams but the scripts are not ~~equ~~ enough so you split the learners into groups. The one you gave group A give to group B vice versa. afterwards each learner would have done both exams.

+1

split half method - split the exam. - split by odd & even number.

Validity and reliability, capability of the outcome.

Reliability deals with exam.

A reliable exam when conducted twice to a group of test takers similar the result is almost similar.

A valid exam is an exam which is used to measuring what it intends to measure.

Types of validity

1. Face validity - It is based on judgement or research findings eg a mock paper from Mavens is good to revise than revising a paper from a small school.

2. Content validity. checking on the right content to test.

3. Criterion validity - It is where you check where there is a relationship btwn what you have set and what you expect.

There are 3 types, predictive, ~~Concurate~~ concurent, Construct validity.

Construct - Refers to how have you set your exams ie did you follow the rule.

predictive - An element when applied ~~whe~~ will ~~head~~ lead to another element.

~~Concurate~~ Concurrent - ~~Where a certain element occurs~~ is concerned ~~whe~~ with relationship btwn one or more elements at the same time.

Measures of Variability/dispersion. - tell how scores are spread.

Range Mean division quartiles

Variance standard deviation.

4 6 8 9 3 → 3 4 6 8 9 Frequency - 5.

Range - difference bPwn the higherP and lowerP.

9 - 3 = 6

Mean deviaPon. The difference bePween the elemenP and average.

Mean deviaPon - The difference bPwn the elemenP and mean.

$$\bar{x} = \frac{3+4+6+8+9}{5} = \frac{30}{5} = 6$$

x	$\lvert x - \bar{x} \rvert$
3	-3
4	-2
6	0
8	2
9	3
	10

assume thaP they are posiPive.

(-3 be 3)

Mean - The average of the deviaPon.

deviaPon

$$\frac{\Sigma \lvert x - \bar{x} \rvert}{f} = \frac{10}{5} = 2.$$

Variance = Is the average of the deviaPon buP you square Pen add Pem divide by number of elements. MakeP negaPve valueP Po posiPive by squaring Pem.

x	$\lvert x - \bar{x} \rvert$	$(x - \bar{x})^2$
3	3	9
4	2	4
6	0	0
8	2	4
9	3	9
		26

$$\text{Variance} = \frac{\Sigma (x - \bar{x})^2}{f}$$

$$\frac{26}{5} = 5.2$$

Standard deviaPon - The square root of the variance.

$$= \sqrt{\text{variance}} \qquad \sqrt{5.2} = 2.28$$

comment. The deviation is large or is small.

Any deviation above to

decile. $\frac{x}{10}$ Percentile $= \frac{x}{100}$ quartile $= \frac{x}{4}$

class	f	x	fx	$\lvert x-\bar{x} \rvert$	fd	d^2
35-39	2	37	74	14.64	29.28	214.3
30-34	3	32	~~64~~ 96	9.64	28.92	92.93
25-29	6	27	162	4.64	27.84	21.53
20-24	8	22	176	-0.36	-02.88	0.130
15-19	4	17	68	-5.36	-21.44	28.73
10-14	3	12	36	-10.36	-31.08	107.3
5-9	2	7	14	-15.36	-30.72	235.9
	28		626			

$\bar{x} = \frac{\Sigma fx}{\Sigma f}$

$\frac{626}{28} = 22.36$

172.16

~~Mead~~ Mean deviation $= \frac{\Sigma f\lvert x-\bar{x} \rvert}{\Sigma f} = \frac{172.16}{28}$

$= 6.149.$

$\frac{\lvert \Sigma fd \rvert}{f}$

~~Vid~~ Variance $= \frac{\Sigma fd^2}{\Sigma f}$

$\frac{18,000.24}{28} = 64.29$ $\frac{1746.24}{28}$

$= 62.36$

standard deviation $= \sqrt{\text{Variance}}$.

$\sqrt{64.29}$ $\sqrt{62.36}$

$= 8.01$ $= 7.897$

INTERNAL EXAMINATION

Infernal exams are set and administered by teacher to learners.

They are set from ~~has~~ what has been learn done and therefore may lack content validity

Marking is subjective due to influence of some students.

Does not give a true picture of your student.

Advantages.

1) Question framing is familiar ~~ques~~ learners can easily predict. Items or questions come from what have been covered.
2) Candidates experience less stress because they can easily remember some questions attributes
3)

Disadvantages.

1. The teacher is likely to be subjective in marking.
2. Content validity is unlikely / poor. Infernal exams may be below the standard of nomative groups

EXTERNAL EXAMINATION

They are managed by examining group to a large population of learners.

Advantages

1. Examiners are likely to be objective
2. Gives the exams a better reliability than infernal exam
3. Candidates are likely to go through the whole syllabus as they cannot easily predict items that are going to be set.

4) They possess a universal standard
5) External exams have a better validity.

Disadvantages.

1) Teachers and learners are likely to work under pressure as they compete with others.
2) learners may cram the fact instead of understanding them.

ITEM ANALYSIS

It refers to application of statistical technique to the assessment of individual question.
It is the process of examining the response to each test item.
It is applied to classroom testing and external exams.

factors to consider when analysing responses.

1) The difficulty of the item ie how many students got it right as per the total candidate.
2) Item discrimination index. Does it distinguish between the first and slow learner properly.
3. Distribution of responses between slow learners and fast learner.

Reasons for conducting item analysis.

1. We analyze items to reject badly set items and to retain the items that are well set.
2. So as to revise with your learner the right item.

Steps for conducting Item analysis.

1. After scoring the papers - arrange from highest to lowest
2. Form 2 groups from the order set of paper UP - upper group 27%
   LP - 27%

If we have 50 candidates

UP - $\frac{27}{100} \times 50 = 13.5 \approx 14$ candidates. L - 14 candidates.

	Alternatives						
	A	B	C	D	E	O	E
UP	0	0	20	0	0	0	20
LP	4	2	8	3	3	0	20

Item difficulty index = $\frac{UP + LP}{U + L}$

$\frac{20+8}{40} = 0.7$

20 - results from any group.

Discrimination index = $\frac{UP - LP}{20}$

$\frac{20-8}{20} = 0.6$ - discrimination

The answer will be negative if the lower performance better than UP

An items which gives (-) performance should be rejected

0.4 and above is a high level of discrimination index.

03 to 0.39 is reasonable - is not bad and can be retained.

0.2 to 0.29 is manageable but should be revised.

- below 0.2 the item is not contributing to any discrimination level and should be rejected.
- Any negative outcome should be rejected.

Correlation

It is a scientific method that is used to find out if variables are related.
The techniques used in finding out if variables are related include;
- Spearman rank order correlation
- Pearson product moment regression
- and correlation co-efficient.

Spearman rank order correlation co-efficient (rho)

is worked out based on rank

$$(rho) = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$$

x	y	rank x	rank y	d	d^2
50	45	1	2	1	1
49	50	2	1	1	1
30	25	3	3	0	0
11	10	4	5	1	1
10	15	5	4	1	1
					4

$n = 5$

$$1 - \frac{6 \times 4}{5(5^2-1)} = 1 - \frac{24}{5 \times 24} \quad 1 - 24/120$$

$1 - 0.2$
$= 0.8$

a	b	rank a	rank b	d	d^2
90	8	1	2	1	1
60	9	2	1	1	1
40	6	3	3	0	0
30	4	4	4	0	0
20	2	5	5	0	0
10	1	6	6	0	0
					2

$$1 - \frac{6\Sigma d^2}{n(n^2-1)} \qquad 1 - \frac{6 \times 2}{6(6^2-1)}$$

$$1 - \frac{12}{210} = 1 - 0.5\,0714$$

$$= 0.9426$$

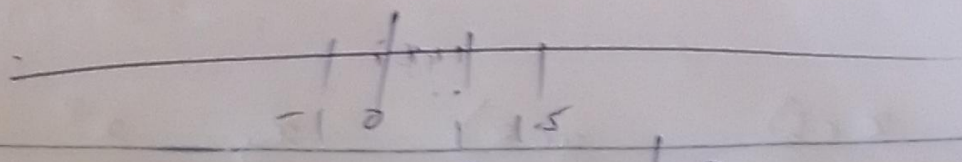
Pearson product moment.

$$r_{xy} = \frac{N(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{\{N(\Sigma x^2) - (\Sigma x)^2\}\{N(\Sigma y^2) - (\Sigma y)^2}}$$

x	y	xy	x^2	y^2
50	45	2250	2500	2025
49	50	2450	2401	2500
30	25	750	900	625
11	10	110	121	100
10	15	150	100	225
150	145	5710	6022	5475

$$\frac{5 \times 5710 - 150 \times 145}{\sqrt{(5 \times 6022 - 22500)(5 \times 5475 - 21025}}$$

$$\frac{28550 - 21750}{\sqrt{7610 \times 6350}} \qquad \frac{6800}{69515} = 0.9782$$



+1.0 or -1.0 perfect correlation
± 0.7 to ± 1.0 strong correlation
± 0.40 to ± 0.69 moderate correlation
± 0.10 to ± 0.39 weak Correlation
0 No correlation

(-) ~~Strong negative correlation.~~
(-0.75) ~~-~~ strong negative correlation: higher scores on one variable are associated with lower scores on the other.

0.3429. Indicates a weak positive relationship between the two variables. This means that as scores in one variable increase scores in the other variable also tend to increase slightly.

± 0.40 to ± 0.69
± moderate. This means that as values in one variable increase the value in the other variables also tend to increase. The relationship is noticeable but not very strong.

(-). This means that as one variable increases, the other variable tends to decrease. The relationship is clear but not strong enough to be considered strong.

compute pearson product moment regression.

~~b(x)~~ x(b)	~~K(y)~~ y(k)	xy	x^2	y^2
6	8	48	36	
5	5	25	25	
6	4	24		
~~6~~	~~8~~	~~48~~		
9	7	63		
8	8	64		
9	6	54		