DEPARTMENT OF AGRICULTURAL ECONOMICS AND FARM MANAGEMENT,

COLLEGE OF AGRICULTURAL MANAGEMENT AND RURAL DEVELOPMENT,

UNIVERSITY OF AGRICULTURE, ABEOKUTA.

TUTORIAL QUESTIONS

COURSE CODE: AEM 304

TITLE: APPLIED STATISTICS IN AGRICUTURE

R. A. SANUSI. Ph.D.

Q1. The following data represent the seasonal outputs ('000tons) in *Ranlad Farms Nig. Ltd*:

Arable crops – 300, 350, 450, 650, 550, 220, 370, 7500, 8500 and 8235.

Fish – 200, 440, 800, 920, 750, 650, 1420, 1550 and 1200.

Cash crops – 80, 90, 45, 65, 12, 100 and 120.

Determine the mean output of the enterprises of the firm-farm.

Solution:

NB:-

1. If a distribution has very high extreme values, harmonic mean is used as a measure of the centre.
2. If a distribution has median extreme values, geometric mean is used as a measure of the centre.
3. If a distribution has no extreme values, arithmetic mean is used as a measure of the centre.

For arable crop enterprise – the mean is:

$$X_H = n \Big/ \sum_{i=1}^{n}(1/x_i)$$

$X_H = 10/0.0193 = 515.73$

For fish enterprise – the mean is:

$$\text{Log}X_G = \frac{\text{Log}X_1 + \text{Log}X_2 + \text{-------} + \text{Log}X_n}{n}$$

$X_G = \text{AntilogLog}X_G = \text{Antilog}(2.8544) = 715.15$

For cash crop enterprise – the mean is:

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

$\bar{x} = 564/8 = 70.5$

Q2. Given that the age distribution of staff in COLAMRUD and COLPLANT, UNAAB, Abeokuta is as presented in the table below:

Table 1: Age Distribution of Staff in COLAMRUD and COLPLANT, UNAAB, Abeokuta

Age Group	*Mid-point (Age)*	*COLAMRUD (Frequency)*	*COPLANT (Frequency)*
19 – 21	20	2	4
22 – 24	23	4	3
25 – 27	26	8	4
28 – 30	29	16	26
31 – 33	32	7	3
34 – 36	35	4	3
37 – 39	38	3	5

Solution:

Using the mid-point approach and the arithmetic mean technique:-

$\overline{x} = (\Sigma f_i x_i)[(\Sigma f_i)^{-1}]$

where:-

Σ = sum of;

f_i = frequency of the i^{th} age group;

x_i = mid-point values of the i^{th} age group.

$\therefore$ Mean age for COLAMRUD staff = 29.14 and

Mean age for COLPLANT staff = 29.13.

Graphically:-

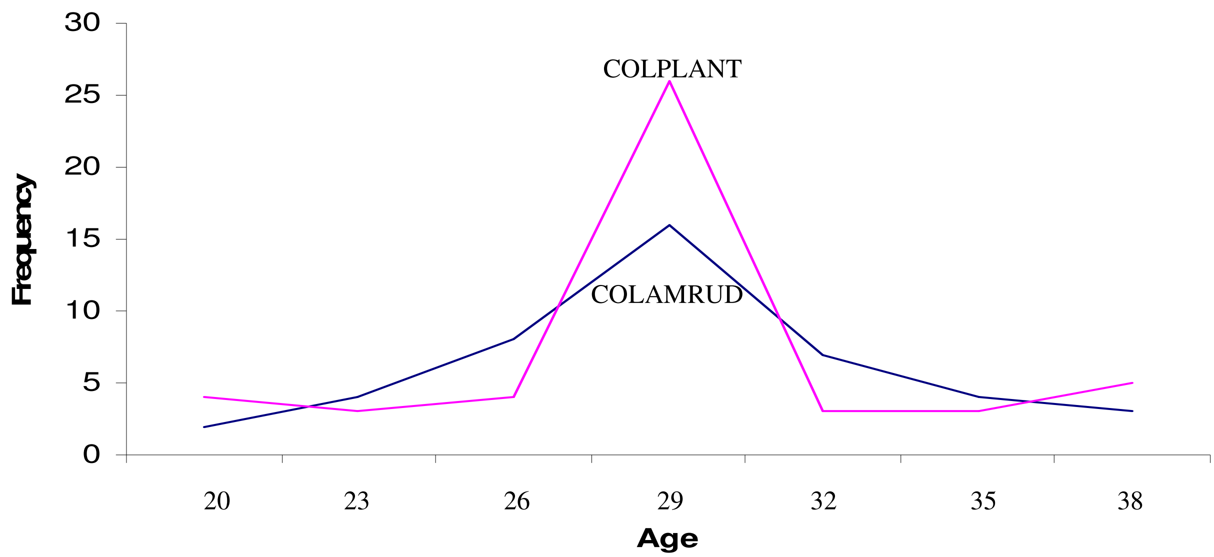


Figure 1: Line Chart for COLAMRUD and COLPLANT STAFFERS AGE

The standard deviation:

$\sqrt{\sigma^2} = \sqrt{(\Sigma fd^2)\{(\Sigma fd)^2/(\Sigma f)\}[(\Sigma f)^{-1}]}$

N.B:- d is the deviation of the frequencies from the frequency of the class of the assumed mean (29).

∴SD for:-

COLAMRUD = 43.70

COLPLANT = 70.40.

CV for:-

COLAMRUD = 1.49

COLPLANT = 2.42.

Q3. There are equal numbers of rams and ewes in a group of 20 sheep. If 60% of the rams and 20% of the ewes are of the *Balami* breed, what is the probability that a sheep chosen from the group is (a) a ram or *Balami* sheep, (b) a *Balami* sheep given that it is a ram, (c) a non-*Balami* sheep given that it is a ram, (d) show that (b) + (c) = 1.

Solution:

Table 2: Distribution of Sheep according to Breed and Gender

Gender	***B***	***B***^{l}	***Total***
Ram	6	4	10
Ewe	2	8	10
Total	***8***	***12***	***20***

This implies that:

(a) p(R∪B) i.e. either ram or *Balami* which are *Balami* ram, *Balami* ewe and non-*Balami* ram.

(b) p(B/R)

(c) p(B^{l}/R)

The sample space can be given as:-

Table 3: Sample Space for Sheep according to Breed and Gender

Gender	***B***	***B***^{l}	***Total***
Ram	0.3	0.2	0.5
Ewe	0.1	0.4	0.5
Total	***0.4***	***0.6***	***1***

(a) The two events are mutually independent, hence –

p(R∪B) = p(R) + p(B) – p(RB) =

⇒ p(R∪B) = 0.5 + 0.4 – 0.3 = 0.6

$\Rightarrow 1 - p(R \cap B^{l}) = 1 - 0.4 = 0.6$

(b) $p(B/R) = p(B \cap R)/p(R) = 0.3/0.5 = 0.6$

(c) $p(B^{l}/R) = p(B^{l} \cap R)/p(R) = 0.2/0.5 = 0.4$

(d) $p(B/R) + p(B^{l}/R) = 0.6 + 0.4 = 1$

Q4. Suppose a brown bull is mated with black cows on 10 occasions. What is the probability of having (a) at least 7 brown calves (b) at most 7 brown calves from the mated cow; given that there is a probability of ½ that a calve from black cow with a brown bull is brown.

Solution:

Working formula → $p(x = x_i) = \binom{n}{x_i} p^{x_i}(1 - p)^{n - x_i}$

$x_i = 0, 1, 2, 3$, -----------10; $p = ½$ and $n = 10$.

$p(x_i = 0)$	=	$10!/0!(½)^0(½)^{10}$	= 0.001
$p(x_i = 1)$	=	10/1024	= 0.010
$p(x_i = 2)$	=	45/1024	= 0.044
$p(x_i = 3)$	=	120/1024	= 0.117
$p(x_i = 4)$	=	210/1024	= 0.205
$p(x_i = 5)$	=	252/1024	= 0.246
$p(x_i = 6)$	=	210/1024	= 0.205
$p(x_i = 7)$	=	120/1024	= 0.117
$p(x_i = 8)$	=	45/1024	= 0.044
$p(x_i = 9)$	=	10/1024	= 0.010
$p(x_i = 10)$	=	1/1024	= 0.001
Σ			= 1.000

(a) The probability of at least 7 brown calves $p(x \geq 7)$. There are two ways to this problem:-

i. sum $p(x = x_i)$ up to 7 and subtract from 1 since $\Sigma p_i = 1$ or

ii. find the sum $p(x_i = 7, 8, 9 \text{ and } 10)$.

(b) The probability of at most 7 brown calves $p(x \leq 7)$. There are also two ways to this problem:-

i. sum $p(x = x_i)$ up to 7or

ii. find the sum $p(x_i = 7, 8, 9 \text{ and } 10)$ and subtract from 1 since $\Sigma p_i = 1$.

Q5. A supplier delivers farm outputs to a eatery by means of the rail.for meal preparations for customers. Given that the train arrives with the probability of 0.2 and x is a random variable for late arrivals in any given week of 5 working days, what is the expected value of late arrival of supplies to the eatery and what is the level of deviation from the expected.

Solution:

$E(x) = np$

$\Rightarrow E(x) = 0.2 \text{ x } 5$

$= 1$

$\sigma^2 = np(1 - p)$

$\Rightarrow \sigma^2 = 0.2 \text{ x } 0.8 \text{ x } 5$

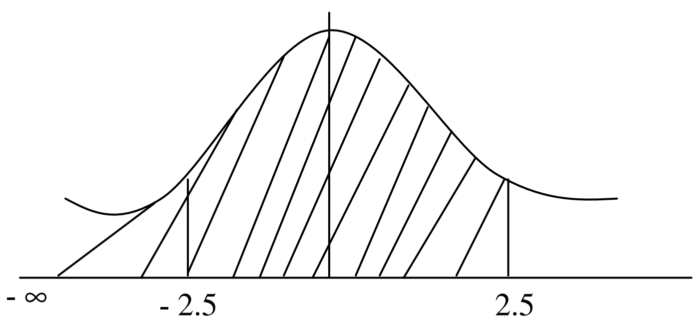
$= 0.8$

Q6. Given that x:N(0, 1), what is $p(x \leq 2.5)$

Solution:

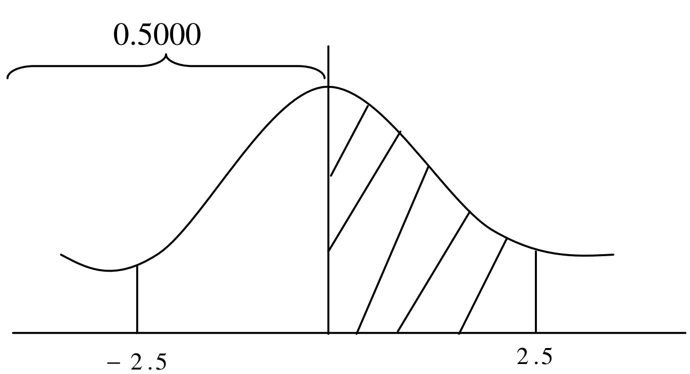
Fro Z – table:- $p(x \leq 2.5) = 0.9938$

$p(x \leq 2.5) = p(x \geq -2.5) = 0.9938$ (i.e. symmetry)



Q7. Given that x: N(0, 1), what is $p(0 \leq x \leq 2.5)$

Solution:



p(x ≤ 2.5) – p(x ≤ 0)

= 0.9938 – 0.5000

= 0.4938.

Q8. Given that x: N(μ, σ^2) = N(2, 9), find p(x ≤ 8).

Solution:

Recall that:-

$$Z = \frac{x-\mu}{\sigma}$$

$$\therefore Z \text{ for } x: N(2, 9) = p\left(\frac{x-\mu}{\sigma} \le \frac{8-2}{3}\right) = 2.00$$

From the Z-table:-

p(Z ≤ 2.00) = 0.9772

Q9. In Biolek Farms, the mean weight of 500 rams is 70kg and the standard deviation is 5kg. Assuming that the weights are normally distributed; if the weight is taken to the nearest whole number, find how many rams weigh (a) between 60kg and 71kg, (b) more than81kg.

Solution:

(a) If the weight is taken to the nearest whole number, weight between 60kg and 71kg can have any value from 59.5kg to 71.5kg.

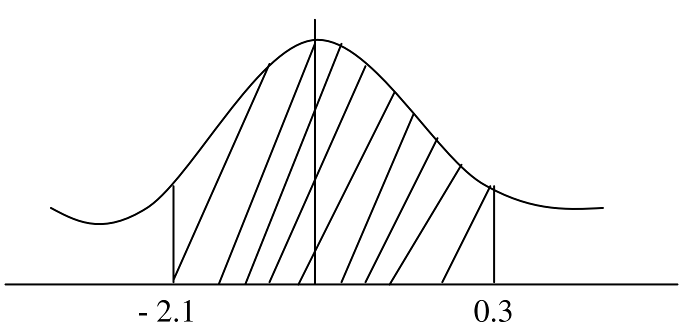
59.5kg in standard units = (59.5 – 70)/5

= -2.1

71.5kg in standard units = (71.5 – 70)/5

= 0.3

The required proportion of rams ≡ p(-2.1 ≤ x ≤ 0.3) = (area between z = -2.1 and z = 0.3)



⇒ p(x ≤ 0.3) – p(x ≤ -2.1)

⇒ p(x) = 0.6179 – 0.0179

$= 0.6000$

OR

$p(-2.1 \leq x \leq 0.3)$ = (area between z = -2.1 and z =0) + (area between z = 0.3 and z = 0)

$\Rightarrow p(-2.1 \leq x \leq 0.3) = \{0.5000 - p(x \leq -2.1)\} + \{ p(x \leq 0.3) - 0.5000)\}$

From Z-table:-

$p(x \leq -2.1) = 0.0179$

$p(x \leq 0.3) = 0$

$\Rightarrow p(-2.1 \leq x \leq 0.3) = (0.5000 - 0.0179) + (0.6179 - 0.$

$\Rightarrow p(-2.1 \leq x \leq 0.3) = 0.4821 + 0.1179$

$\Rightarrow p(-2.1 \leq x \leq 0.3) = 0.6000$

OR

$\Rightarrow p(-2.1 \leq x \leq 0.3) = \{0.5000 - p(x \leq -2.1)\} + \{0.5000 - p(x \leq 0.3)\}$

$p(x \geq x_i) = p(x \leq x_i)$ from symmetric rule

$\Rightarrow p(x \geq 0.3) = p(x \leq 0.3) = 0.3821$

$\Rightarrow p(-2.1 \leq x \leq 0.3) = (0.5000 - 0.0179) + (0.5000 - 0.3821)$

$= 0.4821 + 0.1179$

$\Rightarrow p(-2.1 \leq x \leq 0.3) = 0.6000$

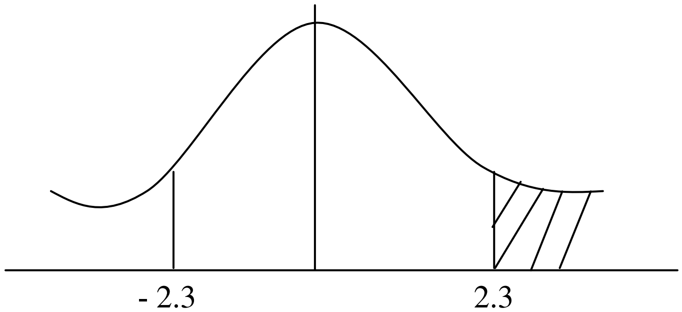
$\therefore$ The number of rams weighing between 60kg and 71kg = 5000(0.6000) = 300 rams.

(b) Rams weighing more than 81kg weigh at least 81.5kg

81.5kg in standard unit = (81.5 – 70)/5

$= 2.3$

The required proportion of rams $\equiv p(x \geq 0.3)$ = (area to the right of z = 2.3)



From the Z-table:-

$p(x \geq 2.3) = p(x \leq 2.3)$ (symmetry)

$\Rightarrow p(x \geq 2.3) = 0.0107$

OR

$1 - p(x \leq x_i) = p(x \geq x_i)$ (symmetry)

$\Rightarrow 1 - 0.9893 = 0.0107$

OR

$p(x \geq 2.3)$ = (area to the right of z = 0) – (area between z = 0 and z = 2.3)

$\Rightarrow p(x \geq 2.3) = 0.5000 - 0.4893$

$= 0.0107$

$\therefore$ The number of rams weighing more than 81kg = 500(0.0107) ≈ 5

Exercise:

E1. Given x: N(0, 1); find:- $p(x \geq 2.00)$, $p(x \leq 2.00)$, $p(x \leq -2.00)$, $p(x \geq -2.00)$ and $p(x \geq 2.5)$.

Note:

$1 - p(x \geq x_i) = p(x \leq x_i)$ (symmetry)

$\Rightarrow p(x \geq x_i) + p(x \leq x_i) = 1$

E2. It was observed on a cattle farm that when tetanus affects infant calves, only 10% recover. i. In a random sample of five infant calves affected by tetanus, what is the probability that only two of them recover?

ii. How many calves would be expected to recover from 15 infant calves affected with tetanus?

E3. Assuming the bags of fish meal compounded by a milling firm are randomly distributed with a mean 10kg and a standard deviation 0.5kg. The Production Manager specified that bags weighing below 9.4kg are grossly under packed while those weighing above 10.8kg are too over packed. How many of 100 randomly selected bags form the firm's warehouse will be:

i. grossly under packed;

ii. too over packed;

iii. normal in weight.

Q10. Given that protein determination is normally distributed, the mean protein of nine varieties of beans grown in Nigeria was put as 12.66 by a group of visiting scientists from UK to Nigeria in 1980. A PG student in UNAAB insisted that the value is inconsistent with present 21st century dynamics due to the observations he made on protein determination for the same varieties which he recorded as follows: 12.9, 13.4, 12.4, 12.8, 13.0, 12.7, 12.4, 13.5 and 13.9. Determine the veracity of the claims of the PG student.

Solution:

$\mu = 12.66$

Null hypothesis (H_0): $\overline{x} = \mu = 12.66$

Alternative hypothesis (H_A): $\overline{x} \neq \mu = 12.66$

Step 1: $\bar{x}$, $S_{\bar{x}}$

$\bar{x} = 13$ and $S_{\bar{x}} = 0.17$

Step 2: $t_c = \frac{\bar{x}-\mu}{S_{\bar{x}}} = \frac{13-12.66}{0.17} = \frac{0.34}{0.17} = 2.00$

Step 3: degree of freedom $(n - 1) = 9 - 1 = 8$, $\alpha = 5\%$.

Step 4: $t_{T,\alpha,df} = t_T(5\%, 8) = 2.31$

Step 5: compare t_c and t_T:- $t_c < t_T$ since $2.00 < 2.31$

Step 6: decision:- H_0 should be accepted.

Mathematically there is a difference between 12.66 and 13.00 but statistically the difference is not significant, hence the mean can be accepted to be 12.66. The alternative hypothesis determines the tailed test being dealt with. The manner the calculated hypothesis is set up dictates the kind of tail test. If for the above example it was asked that is sample mean greater than the population mean, then H_0 and H_A will be set as this:

H_0: $\overline{x} = \mu = 12.66$

H_A: $\overline{x} > \mu = 12.66$ ⟶ one tail test

Also if it was asked that is sample mean less than the population mean, H_0 and H_A will be set as:

H_0: $\overline{x} = \mu = 12.66$

H_A: $\overline{x} < \mu = 12.66$ ⟶ one tail test

However, as in the example, if the statement is sample mean same with (or equal to) population mean, H_0 and H_A will be set as:

H_0: $\overline{x} = \mu = 12.66$

H_A: $\overline{x} \neq \mu = 12.66$ ⟶ 2 tail test

N.B:- In one tail test t_T at 1%, 5% and 10% are read at 1%, 5% and 10%.

In two tail test t_T at 1%, 5% and 10% are read at 0.5%, 2.5% and 5%.

Q11. A final year project student in COLVET in UNAAB is interested in the feed intake of drake and duck to determine drug administration regime for goose. From his observations on one hundred each of drakes and ducks, the average intake was 35.4g and 31.00g per day for drakes and ducks respectively while the standard error of estimate was 400 and 1200 for drakes and ducks respectively. As a student who has taken a course in AEM 304, he is soliciting your opinion if he can conclude that ducks do not eat as much as the drakes.

Solution:

Testing of hypothesis problems are usually related to normal and binomial distribution. The case here is of 2 samples from two populations of drakes and ducks i.e. H_0: $\mu_M = \mu_F$ and H_A: $\mu_M \neq \mu_F$ i.e. 2-tail or $\mu_M > \mu_F$ or $\mu_M < \mu_F$ i.e. 1-tail.

To test this kind of hypothesis, $H_0 \Rightarrow \mu_F = \mu_M = 0$. Information required are $\bar{x}_M$, $\bar{x}_F$ and $S_{\bar{x}_F - \bar{x}_M}$

$$d = \frac{\left(\bar{x}_F - \bar{x}_M\right) - (\mu_1 - \mu_2)}{S\left(\bar{x}_F - \bar{x}_M\right)}$$

$$S\left(\bar{x}_F - \bar{x}_M\right) = \sqrt{Var\left(\bar{x}_F - \bar{x}_M\right)} = \sqrt{\frac{S_M^2}{n_M} + \frac{S_F^2}{n_F}}$$

Degree of freedom = $(n_1 - 1) + (n_2 + 1) = n_1 + n_2 = 2$

$\bar{x}_1$ is 35.4, $\bar{x}_2 = 31.00$, $S_{\bar{x}_1} = 400$ and $S_{\bar{x}_2} = 1200$, $n_1 = n_2 = 100$

$$S\left(\bar{x}_M - \bar{x}_F\right) = \sqrt{\frac{S^2 \bar{x}_1}{n_1} + \frac{S^2 \bar{x}_2}{n_2}} = \sqrt{\frac{400}{100} + \frac{1200}{100}} = 4$$

$$\Rightarrow t_c = \frac{\bar{x}_1 - \bar{x}_2}{S(\bar{x}_1 - \bar{x}_2)} = \frac{35.4 - 3.2}{4} = \frac{4.4}{4} = 1.1$$

V i.e. df = 100 + 100 – 2 = 198

$\Rightarrow t = 1.1$

However, $t_T = 1.65$ for 1-tail ($\mu_M \neq \mu_F$) and 1.96 for 2-tail ($\mu_M > \mu_F$ or $\mu_M < \mu_F$)

$\therefore$ $t < t_T$ meaning that ducks eat as much as the drakes.

Q12. Another student in the same college carried out the same observations on the exotic breeds of goose but recorded the same error of estimate for the two genders but an average intake of 45.6g and 31g for the drakes and ducks respectively. She also needs your assistance like her colleague.

Solution:

Recall that:- $S\left(\bar{x}_1 - \bar{x}_2\right) \quad = \quad 4$

$t_C \quad = \frac{45.6-31}{4} \quad = \quad \frac{14.6}{4} \quad = 3.65$

i.e. $t_c > t_T$

Therefore H_0 is not acceptable and H_A is accepted, i.e. H_A: $\mu_M \neq \mu_F$ hence the two groups do not eat equally.

The critical value of the t – distribution can be used to calculate the appropriate confidence interval. If the hypothesized mean ($\bar{x}$) lies outside the confidence interval (CI), the H_0 is not accepted and if it lies inside the CI the H_0 is accepted.

$\bar{x}$ - t_T $S_{\bar{x}} \leq \mu \leq \bar{x} + t_T S_{\bar{x}}$

$\mu_L \leq \mu \leq \mu_U = [\bar{x} \pm t_{\alpha/2} S_{\bar{x}}] = 1 - \alpha$

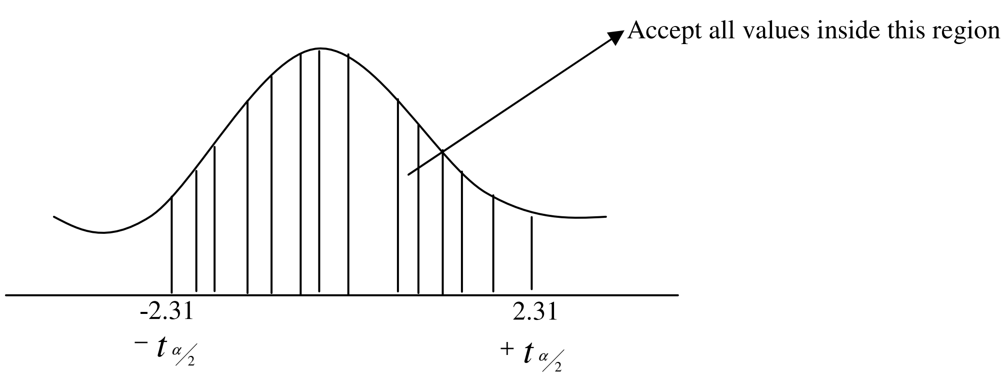
Setting up a 95% C. I. around the mean value:

$\Rightarrow \quad \bar{x} - t_{\alpha/2} S_{\bar{x}} \leq \mu \leq \bar{x} + t_{\alpha/2} S_{\bar{x}} = 1 - 0.05 = 0.95$

$\Rightarrow \quad 14.6 - 1.96(4) \leq \mu \leq 14.6 + 1.96(4) = 0.95$

$\Rightarrow \quad 14.6 - 7.84 \leq \mu \leq 14.6 + 7.84 = 0.95$

$\Rightarrow \quad 6.76 \leq \mu \leq 22.44 = 0.95$



Sometimes a situation may arose where there can be systematic differences from unit to unit. In such a situation, any observed difference in two independent sample tests cannot be ascribed to the difference between the two treatments only. Included, for instance, in such a difference is aa measure of the systematic differences and the required treatment differences which are said to be confounded with each other.

Q13. An animal scientist conducted an experiment to determine if there is any difference between two new diets (A and B) and weight gains by pigs. Six pairs of pigs were selected so that the two pigs in a pair are as close as possible in their initial weights. Diet A was randomly assigned to one pig in a pair and B to the other. The following are the weight gains by the pigs after a trial period of 10 weeks.

Gains in weight (Grams)

Pair	*Diet A*	*Diet B*
1	48	40
2	56	52
3	63	68
4	51	62
5	42	38
6	49	40

Find out if there is any significant difference in weight gained by the pigs fed on the two diets.

Solution:

Let x_i be weight by pigs in the i^{th} pair that was fed A, and y_i the weight gained for the other (fed on B).

Hence:-

$d_i = x_i - y_i \sim N(\mu, \sigma_d^2)$

where:

$\mu = \mu_1 - \mu_2$.

$\overline{d} = (\sum d_i)n^{-1}$, therefore – $\overline{d} \sim N\{(\mu, \sigma_d^2(n^{-1})\}$

Test statistic:-

H_0: $\overline{x} = \overline{y} = \mu = 0$

H_A: $\overline{x} \neq \overline{y}$ (or $\overline{x} > \overline{y}$; $\overline{x} < \overline{y}$)

The test statistic is:-

$t_c = \{\overline{d} - \mu\}\{S_d/(n^{1/2})$

where –

$\overline{d} = \sum d_i/n$;

n = number of observations;

$S_d^2 = (\sum d_i - \overline{d})^2(n-1)^{-1}$ or $S_d^2 = \{(\sum d_i^2) - [(\sum d)^2/n]\}(n-1)^{-1}$

Pair	x_i	y_i	$d_i = x_i - y_i$	d^2
1	48	40	8	64
2	56	52	4	16
3	63	68	-5	25
4	51	62	-11	121
5	42	38	4	16
6	49	40	9	81
$\sum$	-	-	9	323

$S_d^2 = (323 - 13.5)/(6 - 1)$

$\Rightarrow S_d^2 = 3095/5 = 61.9$

$\Rightarrow S_d = 7.89$

$\overline{d} = 9/6 = 1.5$

$\Rightarrow t_C = (1.5 - 0)/\{(7.89/\sqrt{6})\}$

$\Rightarrow t_C = 1.5/3.22$

$\Rightarrow t_C = 0.47$

$t_{T, \alpha, df} = t_T$ (5%, 5) = 2.02 for $\overline{x} > \overline{y}$ or $\overline{x} < \overline{y}$ (i.e. 1-tail) or 2.57 for $\overline{x} \neq \overline{y}$ (i.e. 2-tail)

Exxercise

E4. Ten goats were randomly selected by a veterinary scientist for a clinical trial of a new drug to control dehydration. The serum measure before and after the use of the (new) drug were as follows:

Goats	1	2	3	4	5	6	7	8	9	10
Before	160	140	180	160	225	150	150	140	170	165
After	140	150	170	130	180	120	150	120	130	140

Determine if there is any significant effect of the drug on the serum measure of the experimental goats.

Q14. The average prices of stocks and bonds listed on the capital market in DR country for the period 1950 – 1959 are given in the table below. Find:

i. Pearson correlation coefficient;

ii. coefficient of determination;

iii. interpret your result.

Year	***Average price of stocks (Euros)***	***Average price of Bonds (Euros)***
1950	35.22	102.43
1951	39.87	100.93
1952	41.85	97.43
1953	43.23	97.81
1954	40.06	98.32
1955	53.29	100.07
1956	54.14	97.08
1957	49.12	91.59
1958	40.71	94.85
1959	55.15	94.65

Solution:

i. Let x represent average prices of stocks and y represent average prices of bonds:-

Using the formula –

Definitional formula $\rho(x, y) = \dfrac{Cov(xy)}{\sqrt{Var(x)\text{var}(y)}}$

$$\Rightarrow \rho = \frac{\sum xy}{\left[\left(\sum x^2\right)\left(\sum y^2\right)\right]^2}$$

where:

$y = y_i - \bar{y}$

$x = x_i - \bar{x}$

OR

$$\rho = \left\{\Sigma x_i y_i - \{[(\Sigma x_i)(\Sigma y_i)]/n\}\right\} \Big/ \left\{\sqrt{\{[\Sigma x_i^2 - (\Sigma x_i)^2/n]\{[\Sigma y_i^2 - (\Sigma y_i)^2/n}\right\}$$

Year	x_i	y_i	x	y	xy	x^2	y^2
1950	35.22	102.43	-10.04	4.91	-49.36	100.88	24.15
1951	39.87	100.93	-5.39	3.41	-18.42	29.10	11.66
1952	41.85	97.43	-3.41	-0.09	0.29	11.66	0.01
1953	43.23	97.81	-2.03	0.29	-0.60	4.14	0.09
1954	40.06	98.32	-5.20	0.80	-4.18	27.08	0.65
1955	53.29	100.07	8.03	2.55	20.50	64.42	6.52
1956	54.14	97.08	8.88	-0.44	-3.87	78.78	0.19
1957	49.12	91.59	3.86	-5.93	-22.85	14.87	35.12
1958	40.71	94.85	-4.55	-2.67	12.14	20.74	7.11
1959	55.15	94.65	9.89	-2.87	-28.33	97.73	8.21
Σ	***452.64***	***975.16***	**-**	**-**	***-94.67***	***449.39***	***93.70***
M	***45.26***	***97.52***	**-**	**-**	**-**	**-**	**-**

$\Rightarrow \rho = -94.67/\sqrt{(449.39)(93.70)} = -94.67/205.20 = -0.46$

OR

Year	x_i	y_i	x_i^2	y_i^2	x_iy_i
1950	35.22	102.43	1240.45	10491.90	3607.58
1951	39.87	100.93	1589.62	10186.86	4024.08
1952	41.85	97.43	1751.42	9492.60	4077.45
1953	43.23	97.81	1868.83	9566.80	4228.33
1954	40.06	98.32	1604.80	9666.82	3938.70
1955	53.29	100.07	2839.82	10014.00	5332.73
1956	54.14	97.08	2931.14	9424.53	5255.91
1957	49.12	91.59	2412.77	8388.73	4498.90
1958	40.71	94.85	1657.30	8996.52	3861.34
1959	55.15	94.65	3041.52	8958.62	5219.95
Σ	***452.64***	***975.16***	***20937.69***	***95187.40***	***44044.97***
M	***45.26***	***97.52***	**-**	**-**	**-**

$\Rightarrow \rho = \{(44044.97) - [(452.64)(975.16)\ 10]\}/\sqrt{\{[20937.69 - (204882.97/10)][\ 95187.40 - (950937.03)]\}}$

$\Rightarrow \rho = (44044.97 - 44139.64)/\{\sqrt{(449.39)(93.70)}\}$

$\Rightarrow \rho = -94.67/205.20 = -0.46$

ii. Corr. Coefficient $(\rho^2) \equiv (\rho)^2 = (-0.46)^2$

$\Rightarrow \rho^2 = 0.21$

iii. There is some weak, indirect relationship between stock and bond prices (i.e. there is a tendency for stock prices to go down when bond prices go up and vice versa). Furthermore, only 21% of variations in stock prices can be attributed to variations in bond prices.

Exercise

E5. The average output ('000MT) and prices of cocoa ('000₦) in Nigeria for the period 1980 – 1989 are given in the table below. Find:

i. Pearson correlation coefficient;

ii. coefficient of determination;

iii. interpret your result.

Year	***Cocoa Output (MT)***	***Producer price (₦)***
1980	155.00	134.20
1981	182.00	103.40
1982	160.00	106.60
1983	118.00	117.10
1984	170.00	166.60
1985	135.00	190.10
1986	110.00	167.00
1987	80.90	141.00
1988	145.00	114.00
1989	160.00	93.4

Q15. The following table gives observation on annual rainfall (x) and yield (y) in 10 states of Nigeria:

State	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
X (mm/yr)	36	26	12	40	24	18	30	30	14	34
Y (kg/ha)	54	30	28	48	36	30	38	46	16	42

i. find the least square line for Y on X;

ii. find ρ and ρ^2 and interpret your results;

iii. find the yield when rainfall is 50mm/yr;

iv. test the hypothesis at 5% α-level that the regression coefficient of the population equation (i.e. 36 states and FCT) is as high as 21.90.

Solution:

i. Compute XY, X^2 and Y^2:-

State	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***Σ***
X (mm/yr)	36	26	12	40	24	18	30	30	14	34	264
Y (kg/ha)	54	30	28	48	36	30	38	46	16	42	368
XY	1944	780	336	1920	864	540	1140	1380	224	1428	10556
X^2	1296	676	144	1600	576	324	900	900	196	1156	17768
Y^2	2916	900	784	2304	1296	900	1444	2116	256	1764	14680

Using the formula:-

$b = \Sigma x_i y_i - [(\Sigma x_i)(\Sigma y_i)/n] / \Sigma x_{ii}^2 - [(\Sigma x_{ii})^2]/n$

$a = (\Sigma y_i - b\Sigma x_i)/n$

$\Rightarrow b = \{10556 - (264)(368)/10\}/\{17768 - [(264)^2/10)]$

$\Rightarrow b = 840.80/798.40$

$\Rightarrow b = 1.05$

$a = (368/10) - \{[1.05(264)]/10\}$

$\Rightarrow a = 36.8 - 27.72$

$\Rightarrow a = 9.08$

$\Rightarrow Y = 9.08 + 1.05X \rightarrow$ least square line.

ii. $\rho = \{10556 - (264)(368)/10\}/\sqrt{\{17768 - [(264)^2\ 10)]\}\{14680 - [(368)^2/10]\}}$

$\Rightarrow \rho = 840.80/\sqrt{(798.40)(1354.24)}$

$\Rightarrow \rho = 840.80/1039.82$

$\Rightarrow \rho = 0.81$

$\rho^2 \equiv (\rho)^2 = (0.81)^2$

$\Rightarrow \rho^2 = 0.67.$

There is some strong, direct relationship between rainfall and yield (i.e. there is a tendency for yield to go up when bond annual precipitation go up and vice versa). Furthermore, 67% of variations in yield can be attributed to variations in annual rainfall.

n = number of (paired) observations.

iii. When X = 50mm/annum

$\Rightarrow Y = 9.08 + 1.05(50)$

$\Rightarrow$ Y = 61.58kg/ha.

Testing hypothesis about β (Population Slope or Regression Coefficient)

H_o: $\beta = 0$

H_A: $\beta \neq 0$ (2-tail)

$$S_b = S_e \Big/ \left\{ \sqrt{\Sigma x_{ii}^2 - [(\Sigma x_{ii})^2]/n} \right\}$$

where:

$$S_e = \left\{ \Sigma y_i^2 - a\Sigma y_i - b\Sigma x_i y_i \right\} \Big/ (n-2)$$

S_b = standard error of regression slope (b);

S_e = standard error of estimate.

$S_e = \{14680 - 9.08(368) - 1.05(10556)\}/(10 - 2)$

$S_e = (14680 - 3341.44 - 11083.80)/8$

$S_e = 254.76$

$S_b = 254.76/\sqrt{798.40}$

$S_b = 254.76/28.26$

$S_b = 9.01$

$\beta = 0.85$

$t_c = (b - \beta)/S_b$

$\Rightarrow t_c = (1.05 - 21.90)/9.01$

$\Rightarrow t_c = -20.85/9.01$

$\Rightarrow t_c = -2.3141$

Degree of freedom (v) = n – 2 = 10 – 2 = 8;

Level of significance (α) = 0.05 = 0.025 for 2-tail test.

$t_T \equiv t_{o.o25,\,8} = 2.3060$

since $|t_c| > t_T$, H_0 not acceptable. otherwise H_o is acceptable.

$\therefore H_A$ is accepted i.e. regression coefficient of the population equation (i.e. 36 states and FCT) is as high as 21.90

Exercise

E6. Given that observation on annual rainfall (x) and soybean yield (y) in 10 states of Nigeria is:

State	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
X (mm/yr)	36	26	12	40	24	18	30	30	14	34
Y (kg/ha)	14	57	69	70	59	56	60	42	43	68

i. find the least square line for Y on X;

ii. find ρ and ρ^2 and interpret your results;

iii. find the yield when rainfall is 40mm/yr;

iv. test the hypothesis at 5% α-level that the regression coefficient of the population equation (i.e. 36 states and FCT) is as high as 20.85.

Q16. A study was conducted on consumer preference (CP) in three randomly selected geo-political zones (South-East, North-central and South-west) for a new method of packaging palm oil developed by NIFOR. The CP consist of "like", "dislike" and "no opinion". The data gathered is summarized as number of respondents choosing any of the three categories preferences in the table below:

Quality Rating	***East***	***Central***	***West***
Like	15	10	6
Dislike	7	13	12
No opinion	11	12	15

Determine if geo-political zone has any effect on the rating of quality of packaging by respondents at 5% α-level.

Solution:

i. Find the row and column totals:-

Quality Rating	***East***	***Central***	***West***	***Total***
Like	15	10	6	***31***
Dislike	7	13	12	***32***
No opinion	11	12	15	***38***
Total	***33***	***35***	***33***	***101***

ii. Compute f_e and χ_c^2-stat. (the expected frequency) using the formulae:

$f_e = (RT \bullet CT)(GT)^{-1}$

where:-

f_e = expected frequency;

RT = row total;

CT =column total;

GT = grand total.

$\chi_c^2 = \sum\{(f_o - f_e)^2\}\{(f_e)^{-1}\}$

where:-

f_e = observed frequency;

f_e = as defined previously;

RT = row total;

CT =column total;

GT = grand total.

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
15	10.13	4.87	23.73	2.34
7	10.46	-3.46	11.94	1.14
11	12.42	-1.42	2.00	0.16
10	10.74	-0.74	0.55	0.05
13	11.09	1.91	3.65	0.33
12	13.17	-1.17	1.36	0.10
6	10.13	-4.13	17.05	1.68
12	10.46	1.54	2.39	0.23
15	12.42	2.58	6.68	0.54
$\sum$	-	-	-	*6.58*

$\Rightarrow \chi_c^2 = 6.58$

iii. State the hypothesis:-

H_o: the row and column are independent

H_A: the row and column are dependent

iv. Obtain χ_T^2 from statistical table:-

$\alpha = 0.05$ level of significance;

r = degree of freedom and r = (r – 1)(c – 1). = (3 – 1)(3 – 1) = 4.

$\chi_T^2 = \chi^2_{0.05, 4} = 9.4877$

v. Decision rule:-

Since $\chi_c^2 < \chi_{T_r}^2$

⇒ H_o is accepted i.e. geo-political zone has no influence on respondents' rating of palm oil packaging.

Exercise

E7. A study was conducted in a cashew growing state of Nigeria to find out the opinion of 175 literate and illiterate farmers as to whether computer use in cashew farming can improve the efficiency of farmers. The data obtained is as shown below:

Educational Level	***Primary Education***	***Secondary Education***	***Tertiary Education***	***None***
Agree	5	10	30	6
Disagree	40	15	3	35
No opinion	14	7	2	8

Determine at 5% α-level, if farmers' level of education has any effect on their perception on the usefulness of computer application to cashew farming.

Q17. Assuming a green house experiment is conducted to determine the yield of potato with the application of four (4) different types of nitrogen fertilizer using three (3) pots per treatment. Suppose the information collected is as shown below:

Treatment	***Yield of Potato (Kg/N_2 ration)***		
	1	***2***	***3***
1	7	2	4
2	6	4	6
3	8	4	5
4	7	4	2

Determine at 5% α-level, if the application of fertilizer lead to any increase in the yield of potato.

Solution:

i. Set up a hypothesis i.e. H_0 (null) and H_A (alternative) hypothesis:-

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

H_A: $\mu_i \neq \mu_j$; $i \neq j$.

ii. Computing F-statistic (i.e. F calculated → F_c) using ANOVA computing table:

Source of Variation	*Degree of Freedom*	*Sum of Squares*	*Mean Sum of Squares*	*F*
Total	$n - 1$	$\sum Y_i^2 - CT$	-	-
Treatment	$t - 1$	$\{(\sum Y_i)^2/r\} - CT$	$SS_t/(t - 1)$	MSS_t/MSS_r
Residual	$n - t$	$SS_T - SS_t$	$SS_r/(n - 1)$	-

where:

CT = correction term = $(\sum Y_i)^2/n$ or $Y..^2/n$ ($Y..$ = grand total)

n = total number of observations;

t = number of treatment;

r = number of observations per treatment;

SS_T= sum of squares total;

SS_t= sum of squares treatment;

SS_r= sum of squares residual;

MSS_t = mean sum of squares treatment;

MSS_r= mean sum of squares residual.

Treatment	*Yield of Potato (Kg/N_2 ration)*			
	1	***2***	***3***	***Total***
1	7	2	4	13
2	6	4	6	16
3	8	4	5	17
4	7	4	2	13
Total	***28***	***14***	***17***	***59***

N.B:- Row total = column total.

Source of Variation	*Degree of Freedom*	*Sum of Squares*	*Mean Sum of Squares*	*F*
Total	11	41	-	-
Treatment	3	4	1.33	0.2873
Residual	8	37	4.63	-

$\Rightarrow F_c = 0.2873$

iii. Finding the tabulated F-statistic on the F table (i.e. F tabulated):-

$F_T = F_{\alpha, v1, v2} = F_{0.05, 3, 8} = 4.0662$

iv. Comparing F_T with the F_c:-

$F_C < F_T$, hence H_0 is accepted.

v. Drawing conclusion:-

Fertilizer application has no significant effect on yield of potato.

Q18. Assuming a green house experiment is conducted to determine the yield of three (3) varieties of potato with the application of four (4) different types of nitrogen fertilizer using three (3) pots per treatment. Suppose the information collected is as shown below:

Treatment	***Yield of Potato (Kg/N_2 ration)***		
	Variety 1	***Variety 2***	***Variety 3***
1	7	2	4
2	6	4	6
3	8	4	5
4	7	4	2

Determine at 5% α-level, if the application of fertilizer leads to any increase in the yield of potato and if variety has any influence on the yield with fertilizer treatment.

Solution:

i. Set up a hypothesis i.e. H_0 (null) and H_A (alternative) hypothesis:-

H_0: $T_1 = T_2 = T_3 = T_4$

H_A: $T_i \neq T_j$; $i \neq j$.

H_0: $\vartheta_1 = \vartheta_2 = \vartheta_3 = \vartheta_4$

H_A: $\vartheta_i \neq \vartheta_j$; $i \neq j$.

ii. Computing F-statistic (i.e. F calculated $\rightarrow F_c$) using ANOVA computing table:

Source of Variation	*Degree of Freedom*	*Sum of Squares*	*Mean Sum of Squares*	*F*
Total	$n - 1$	$\sum Y_i^2 - CT$	-	-
Treatment	$t - 1$	$\{(\sum Y_i)^2/t\} - CT$	$SS_t/(t - 1)$	MSS_t/MSS_r
Block	$b - 1$	$\{(\sum Y_j)^2/b\} - CT$	$SS_b/(b - 1)$	MSS_b/MSS_r
Residual	$(t - 1)(b - 1)$	$SS_t - SS_r - SS_b$	$SS_r/(t - 1)(b - 1)$	-

where:

CT, n, t, SS_T, SS_t, SS_r, MSS_t, MSS_r are as previously defined;

b = number of blocks;

SS_b = sum of squares block;

M SS_b = mean sum of squares block.

Treatment	***Yield of Potato (Kg/N$_2$ ration)***			
	Variety 1	***Variety 2***	***Variety 3***	***Total***
1	7	2	4	13
2	6	4	6	16
3	8	4	5	17
4	7	4	2	13
Total	***28***	***14***	***17***	***59***

N.B:- Row total = column total.

Source of Variation	***Degree of Freedom***	***Sum of Squares***	***Mean Sum of Squares***	***F***
Total	11	41	-	-
Treatment	3	4	1.33	0.7964
Block	2	27	13.50	8.0838
Residual	6	10	1.67	-

$\Rightarrow$ For treatment, $F_c = 0.7964$ and for block (variety effect) $F_c = 8.0838$.

iii. Finding the tabulated F-statistic on the F table (i.e. F tabulated):-

$\Rightarrow$ For treatment, $F_T = F_{\alpha, v1, v2} = F_{0.05, 3, 6} = 4.7571$ and for block (variety effect), $F_{0.05, 2, 6} = 5.1433$

iv. Comparing F_T with the F_c:-

For treatment, $F_C < F_T$, hence H_0 is accepted and for block (variety effect), $F_C > F_T$.

v. Drawing conclusion:-

Fertilizer application has no significant effect on yield of potato while variety has significant effect on the yield response of potato to fertilizer.

Exercise

E8. A study was conducted to determine the number of days it took 20 rabbits randomly allocated to 4 treatment dose groups to recover from sore foot disease after the administration of a new drug. Test at 5% level the significant effect of the doses.

0.5mg	***1.0mg***	***1.5mg***	***2.0mg***
22	25	26	26
26	27	29	28
25	28	33	27
25	26	30	30
31	29	33	30