Stats 6 :

(1). Bangalore to chennai

$n_1 = 1200$

$x_1 = 452$

$s_1 = 212$

Bangalore to Hosur

$n_2 = 800$

$x_2 = 523$

$s_2 = 185$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} - \frac{s_2^2}{n_2}}}$$

$$t = \frac{(452 - 523) - 0}{\sqrt{\frac{(212)^2}{1200} - \frac{(185)^2}{600}}}$$

$$t = \frac{-71}{\sqrt{37.45 + 42.78}} = \frac{-71}{8.96}$$

$$t = \underline{-7.92}$$

Hypothesis statements:

$H_0 : n_{B \to C} = n_{B \to H}$

$H_1 : n_{B-C} \neq n_{B \to H}$

degree of freedom : $x_1 + x_2 \to n_1 + n_2 - 2$

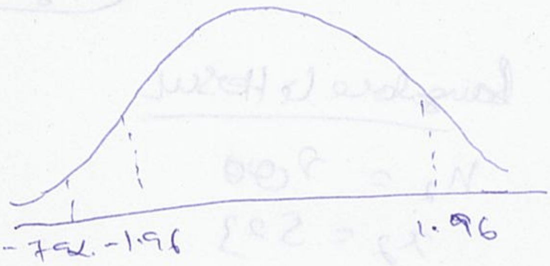
$df = n_1 + n_2 - 2$

$= 1200 + 800 - 2$

$= 1998.$

As the degree of freedom is very high and considering 95% interval

t critical $\approx \underline{\pm 1.96}$



Since $t_{critical} < t$, so rejecting the null hypothesis (H_0). So the number of people travelling from bangalore to chennai is different from number of people travelling from bangalore to Hosur.

(2) Problem statement 2 :-

Hypothesis statement

$x_1 > x_2$ by 45.

$H_0 : \mu_1 - \mu_2 = 45$

$H_1 = \mu_1 - \mu_2 \neq 45.$

Duracell

$n_1 = 100$

$x_1 = 308$

$s_1 = 84.$

Energizer

$n_2 = 100$

$x_2 = 254$

$s_2 = 67$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{(308 - 254) - 45}{\sqrt{\frac{84^2}{100} + \frac{67^2}{100}}} = \frac{9}{\sqrt{70.50 + 44.89}}$$

$= \frac{9}{10.74} = 0.837$

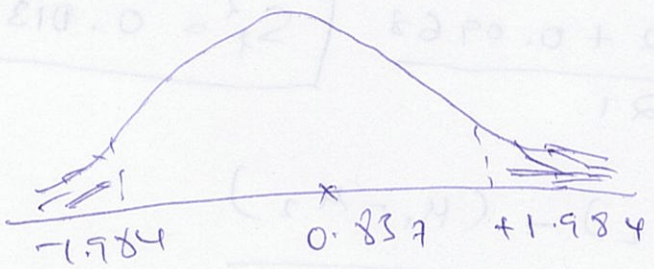
$\boxed{t = 0.837}$

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

$= 100 + 100 - 2$

$df = 198$

$t_{critical} = \pm 1.984 \Big|_{\alpha = 0.05}$



Since t critical > t, unable to reject the null hypothesis.

Problem Statement 3:

Hypotheses statement:

$H_0 : \mu_1 - \mu_2 = 0.$

$H_1 : \mu_1 - \mu_2 \neq 0.$

given data:

Population 1: Price of Sugar = Rs. 27.50

$n_1 = 14.$

$x_1 = 0.317\%.$

$s_1 = 0.12\%.$

Population 2: Price of Sugar = Rs. 20.00.

$s_2 = 0.11\%$

$n_2 = 9$

$x_2 = 0.21\%.$

degree of freedom $df = n_1 + n_2 - 2$

$$df = 14 + 9 - 2$$

$$\boxed{df = 21}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(14-1)(0.12)^2 + (9-1)(0.11)^2}{21}$$

$$= \frac{0.1872 + 0.0968}{21} \qquad \boxed{S_p^2 = 0.0135}$$

$$t = \frac{(x_1 - x_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t = \frac{(0.317 - 0.21) - 0}{0.176 \left(\sqrt{\frac{1}{14} + \frac{1}{9}}\right)}$$

$$= \frac{0.107}{0.126(\sqrt{0.071 + 0.111})} = \frac{0.107}{0.0494}$$

$$t = 2.165$$

$t_{critical}$ at $\alpha = 5\%$ is 2.080. Since $t_{critical} < t$ we are rejecting the null hypothesis.

Problem statement 4:-

Given data:-

Population 1

$n_1 = 15$

$X_1 =$ Rs. 6598.

$S_1 =$ Rs. 844.

Population 2

$n_2 = 19.$

$X_2 = 6870$

$S_2 = 669$

Hypothesis statement.

$H_0 : \mu_1 = \mu_2$ (no increase in sales)

$H_1 : \mu_1 \neq \mu_2$ (increase in sales).

$$t = \frac{(x_1 - x_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p = \sqrt{\frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}}$$

$$S_p = \sqrt{\frac{(15-1)(844)^2 + (12-1)(669)^2}{15+12-2}}$$

$S_p = 771.903.$

$$t = \frac{(6568 - 6870) - 0}{771.903\sqrt{\frac{1}{15} + \frac{1}{12}}}$$

$$\boxed{t = -0.909}$$

degree of freedom $df = n_1 + n_2 - 2$

$= 15 + 12 - 2$

$= 25.$

Considering $\alpha = 10\%$. it is a 2 tailed test $\alpha/2$

$$t_c = \pm 1.476$$

$t_c < t$ rejecting the null hypothesis.

Problem statement 5:-

Given Data:-

Population 1

$n_1 = 1000$

$x_1 = 53.$

$P_1 = 0.53$

Population 2

$n_2 = 100.$

$x_2 = 43$

$P_2 = 0.53.$

Hypothesis statement

$H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0.$

$$t = \frac{(P_1 - P_2) - (\mu_1 - \mu_2)}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\hat{P} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\hat{P} = \frac{53 + 43}{1000 + 100} = \frac{96}{1100} = 0.087.$$

$$t = \frac{0.53 - 0.43 - 0}{\sqrt{0.0087\ (1 - 0.087)\left(\frac{1}{1000} + \frac{1}{100}\right)}}$$

$$= \frac{0.1}{\sqrt{0.087\ (1 - 0.087)\left(\frac{1}{1000} + \frac{1}{100}\right)}}$$

$$= \frac{0.1}{\sqrt{0.087 \times 0.913 \times 0.011}} = \frac{0.1}{0.03}$$

$$\boxed{t = 3.33}$$

t critical at $\alpha = 10\%$.

t critical = 1.645.

Since tcritical < t, we are rejecting the null hypotheses.

Since the value of the test statistics is above the critical point even at a 10% level of significance, we may conclude that there is a statistically significant difference between bank shares of car loans in 1980 and 1995.

Problem Statement 6:

Given data:

Population 1:

$n_1 = 300.$

$x_1 = 120$

$p_1 = 0.40.$

Population 2:

$n_2 = 700$

$x_2 = 140$

$p_2 = 0.20$

Hypotheses Statement:

$H_0 : \mu_1 - \mu_2 \leq 10\%.\ (0.10)$

$H_1 = \mu_1 - \mu_2 > 10\%.\ (0.10)$

$$t = \frac{(p_1 - p_2) - (\mu_1 - \mu_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

8

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{120 + 140}{300 + 700} \quad \hat{p} = \frac{260}{1000} = 0.26$$

$$t = \frac{(0.40 - 0.20) - 0.10}{\sqrt{0.26\,(1 - 0.26)\left(\frac{1}{300} + \frac{1}{700}\right)}}$$

$$= \frac{0.20 - 0.10}{\sqrt{0.26 \times 0.74 \times 0.0047}}$$

$$= \frac{0.10}{0.030} = 3.33$$

Zcritical at $\alpha = 1\%$

$z_c = 3.09$.

Since $z_c < z$, H_0 may be rejected.

Since the value of the test statistics is above the critical point even for a level of significance as small as 0.001 the null hypotheses may be rejected. and we may conclude that the proportion of customers buying at least \$2500 of travelers checks at least 10% higher than sweepstakes users.

Problem statement 7 :-

A die is thrown 132 times.

Number turned up 1, 2, 3, 4, 5, 6.

frequency 16, 20, 25, 14, 29, 28.

Hypothesis Statement:

H_0: die is unbiased.

H_1 : die is biased.

	observation	Expected frequency ($n \times p$)
$x_1 = 1$	16	22
$x = 2$	20	22
$x = 3$	25	22
$x = 4$	14	22
$x = 5$	29	22
$x = 6$	28	22

$$\chi^2 = \Sigma \left[\frac{(o - E_i)^2}{E_i} \right]$$

$$= \frac{1}{22} \left[(16-22)^2 + (20-22)^2 + (25-22)^2 + (14-22)^2 + (29-22)^2 + (28-22)^2 \right]$$

$$= \frac{1}{22} \left[36 + 4 + 9 + 64 + 49 + 36 \right]$$

$$\boxed{\chi^2 = 9}$$

Degree of freedom $= n - 1 = 6 - 1 = \underline{5}$.

From chisquare table Critical value $\boxed{\chi_c^2 = 11.07}$

$\chi_c^2 > \chi^2$

Since the test statistical critical value is greater that test value unable to reject the null hypothesis. So the dice is unbiased.

Problem statement 8r

No of defects	observed frequency
0	32
1	15
2	9
3	4

Hypothesis statement :-

H_0: Printed circuit boards follow Poison distribution

H_1: Printed circuit boards does not follow Poison distribution

$n = 60$.

$$\mu = \frac{0 \times 32 + 1 \times 15 + 2 \times 9 + 3 \times 4}{60} = 0.75$$

Poison distribution

$$P(x) = \frac{e^{-\mu} \mu^x}{x!}$$

$$P(0) = \frac{e^{-0.75} \, 0.75^0}{0!} = 0.472$$

$$P(1) = \frac{e^{-0.75} (0.75)^1}{1!} = 0.354.$$

$$P(2) = \frac{e^{-0.75}(0.75)^2}{2!} = \underline{0.132}$$

$$P(3) = \frac{e^{-0.75}(0.75)^3}{3!} = 0.\underline{\underline{0}}41$$

frequency $= P_i \times n$ expected frequency.

$E_0 = 0.472 \times 60 = 28.32 = 28.$

$E_1 = 0.354 \times 60 = 21.24 = 21$

$E_2 = 0.132 \times 60 = 7.92 \approx 7$

$E_3 = 0.041 \times 60 = 2.46 = 2$

$$\chi^2 = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

$$\chi^2 = \frac{(32-28)^2}{28} + \frac{(15-21)^2}{21} + \frac{(9-7)^2}{7} + \frac{(4-2)^2}{2}$$

$$= \frac{16}{28} + \frac{36}{21} + \frac{4}{7} + \frac{4}{2}$$

$$= 0.571 + 1.714 + 0.571 + 2$$

$$\boxed{\chi^2 = 4.856}$$

$df = 4 - 1 = 3$ $\alpha = 5\%$.

$\chi^2_{critical} = 7.82$

$\boxed{\chi^2_c > \chi^2}$ So failed to reject the null hypotheses.

It indicates that printed circuit board follow Poisson distribution

Problem statement 9:-

	Men	women	Total
Voted	2792	3591	6383
Not voted	1486	2131	3617
	4278	5722	10000

Hypotheses Statement :-

H_0 : Gender and Voting are independent.

H_1 : Gender and Voting are dependent.

$$\text{men voted} = \frac{6383 \times 4278}{10000} = 2730$$

$$\text{women voted} = \frac{5722 \times 6383}{10000} = 3652.$$

$$\text{men not voted} = \frac{4278 \times 3617}{10000} = 1547.$$

$$\text{women not voted} = \frac{5722 \times 3617}{10000} = 2069.$$

$$\chi^2 = \Sigma \left[\frac{(O - E_i)^2}{E_i} \right]$$

$$(ii) = \frac{(2792-2730)^2}{2730} + \frac{(3591-3652)^2}{3652} + \frac{(1486-1547)^2}{1547} + \frac{(2131-2069)^2}{2069}$$

$$X^2 = 6.689$$

Degree freedom $df = (n-1)(n-1)$
$= (2-1)(2-1)$
$= 1$

From chi square table
for 95 % → $X^2_{critical} = 3.84$ ($\alpha = 0.05$)

$$\boxed{X^2_{critical} < X^2}$$

So, rejecting the null hypothesis which indicates that Gender and voting are independent.

Problem statement 10:-

Given data:

$n = 100$ (Sample voters).

observation for 4 candidates

Higgins	Reardon	white	Charlton
41	19	24	16

$\alpha = 0.05 \rightarrow 5\%$

$df = 3$

$X^2 = 14.96$

Hypothesis statement

H_0: All candidates are Equally popular.

H_1: All candidates are not Equally popular.

Expected value $E_i = \frac{100}{4} = 25$ (Equal probability)

$$\chi^2 = \Sigma\left(\frac{(O_i - E_i)^2}{E_i}\right)$$

$$= \frac{1}{25}\left[(41-25)^2 + (19-25)^2 + (24-25)^2 + (16-25)^2\right]$$

$$= \frac{1}{25}\left[256 + 36 + 1 + 81\right]$$

$$\boxed{\chi^2 = 14.96}$$

$df = 3$ $\chi^2_{critical} = 7.81$ at $\alpha = 0.05$

$\chi^2_{critical} < \chi^2$ so rejecting the null hypothesis. All the candidates are not Equally popular

Problem Statement 10:

	A	B	C	
5 - 6	18	22	20	60
7 - 8	2	28	40	70
9 - 10	20	10	40	70
	40	60	100	200

$\chi^2 = 29.1$, $df = 4$, $\alpha = 0.05$

Hypothesis statement:

H_0: No relationship between age and photograph preferences.

H_1: Relationship is there between age and photograph preferences.

Age group (5-6) for A = $\frac{40 \times 60}{200} = 12.$

Age group (5-6) for B = $\frac{60 \times 60}{200} = 18$

Age group (5-6) for C = $\frac{60 \times 100}{200} = 30.$

Age group (7-8) for A = $\frac{40 \times 70}{200} = 14$

Age group (7-8) for B = $\frac{60 \times 70}{200} = 21$

Age group (7-8) for C = $\frac{100 \times 70}{200} = 35$

Age group (9-10) for A = $\frac{40 \times 70}{200} = 14.$

Age group (9-10) for B = $\frac{60 \times 70}{200} = 21.$

Age group (9-10) for C = $\frac{100 \times 70}{200} = 35.$

$$X^2 = \Sigma \left(\frac{(O_i - E_i)^2}{E_i} \right)$$

$$= \frac{(18-12)^2}{12} + \frac{(22-18)^2}{18} + \frac{(20-30)^2}{30} + \frac{(2-34)^2}{14} + \frac{(28-21)^2}{21} +$$

$$\frac{(40-35)^2}{35} + \frac{(20-14)^2}{14} + \frac{(10-21)^2}{21} + \frac{(40-35)^2}{35}$$

$$\boxed{X^2 = 29.6.}$$

degree of freedom = (n-1)(n-1)
= (3-1)(3-1)

$\boxed{df = 4}$

$\boxed{\chi^2_{Critical} = 9.49}$

$\boxed{\chi^2_c < \chi^2}$

Since Critical value is less than the χ value. rejecting the null hypothesis.
It indicates that there is a significant relationship between age and photograph preferences.

Problem Statement 12 :-

	Support	No Support	
Conform	18	40	58
Not Conform	32	10	42
	50	50	100

$\chi^2 = 19.87$, $df = 1$, $\alpha = 0.05$

Hypothesis Statement :-

H_0 : No Significant difference between the support and no support condition.

H_1 : Significant difference between the support and no support condition

Conform Support $= \frac{50 \times 58}{100} = 29$.

Conform Non-Support $= \frac{58 \times 50}{100} = 29$.

not conform support = $\frac{50 \times 42}{100} = 21$

not conform support = $\frac{50 \times 42}{100} = 21.$

Chi-square distribution

$$\chi^2 = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

$$\chi^2 = \frac{(18-29)^2}{29} + \frac{(40-29)^2}{29} + \frac{(32-21)^2}{21} + \frac{(10-21)^2}{21}$$

$$= \frac{121}{29} + \frac{121}{29} + \frac{121}{21} + \frac{121}{21}$$

$$\boxed{\chi^2 = 19.868}$$

degree freedom df = $(2-1)(2-1)$
$= 1$

$$\boxed{\chi^2_{critical} = 3.84 \big|_{\alpha = 0.05}}$$

$$\chi^2_{critical} < \chi^2$$

So rejecting the null hypothesis.

This indicates that there is a significant difference between the support and no support.

Problem statement 13:

	Short	Tall	
Leader	12	32	44
follower	22	14	36
unclassifiable	9	6	15
	43	52	95

Leader - Short $= \frac{43 \times 44}{95} = 19.92$

Leader - tall $= \frac{52 \times 44}{95} = 24.08$

follower - Short $= \frac{43 \times 36}{95} = 16.29$

Follower - tall $= \frac{52 \times 36}{95} = 19.70$

unclassifiable - Short $= \frac{43 \times 15}{95} = 6.79$

unclassifiable - tall $= \frac{52 \times 15}{95} = 8.21$

degree of freedom $= (n_1 - 1)(n_2 - 1) = (2-1)(3-1)$ $\boxed{df = 2}$

Since we have 3×2 table and (then df = 2) we will work out Chi-Square incorporating ~~x~~ Correction for Continuity

$$\chi^2 = \Sigma\left[\frac{(|O - E| - 0.5)^2}{E}\right]$$

$$\chi^2 = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

$$= \frac{(12-19.92)^2}{19.92} + \frac{(32-24.08)^2}{24.08} + \frac{(22-16.29)^2}{16.29} + \frac{(14-19.71)^2}{19.71}$$

19

$$+ \frac{(9-6.79)^2}{6.79} + \frac{(6-8.21)^2}{8.21}$$

$$\boxed{\chi^2 = 10.712}$$

$$\boxed{\chi^2_{critical} = 5.991}$$

10.712 is greater than χ^2 at 0.01. Significance level. Hence there is a relationship between heights and leadership qualities.

Problem statement 14:

	Married.	Widowed/d/separated.	Never married.
Employed.	679	103	114
Unemployed.	63	10	20
Not in labor force	42	18	25

$$\text{Employed and married} = \frac{784 \times 896}{1074} = 654.063.$$

$$\text{Employed \& widowed, divorced} = \frac{131 \times 896}{1074} = 109.288$$

$$\text{Employed \& never married} = \frac{159 \times 896}{1074} = 132.648.$$

$$\text{unemployed \& married} = \frac{784 \times 93}{1074} = 67.888.$$

$$\text{unemployed \& divorced} = \frac{131 \times 93}{1074} = 11.343.$$

$$\text{unemployed \& never married} = \frac{159 \times 93}{1074} = 13.768.$$

Not in labor & married = $\frac{784 \times 85}{1074} = 62.048$. $\boxed{20}$

Not in labor & widowed / divorced = $\frac{131 \times 85}{1074} = 10.367$

Not in labor & never married = $\frac{159 \times 85}{1074} = 12.583$.

$$\chi^2 = \Sigma\left[\frac{(O_i - E_i)^2}{E_i}\right]$$

$$= \frac{(679 - 654.063)^2}{654.063} + \frac{(103 - 109.288)^2}{109.288} + \frac{(114 - 132.648)^2}{132.648}$$

$$+ \frac{(63 - 67.888)^2}{67.888} + \frac{(10 - 11.343)^2}{11.343} + \frac{(20 - 13.763)^2}{13.763}$$

$$+ \frac{(42 - 62.048)^2}{62.048} + \frac{(18 - 10.367)^2}{10.367} + \frac{(25 - 12.583)^2}{12.583}$$

$$\boxed{\chi^2 = 31.616}$$

$df = (n-1)(n-1) = (3-1)(3-1) = 4.$

$\alpha = 0.05 \quad \chi_c^2 = 9.49.$

$\chi_c^2 < \chi^2$ so rejecting the null hypothesis. it indicates that many different marital status have different distribution of labour force status.