Problem - 3.

a) Covariance between advertising spend & sales revenue.

X = Advertising spend

Mean (X) = $\dfrac{5+7+6+8+5+9+7+10+6+8}{10}$

$= \dfrac{71}{10} = 7.1$

Y = Sales Revenue

Mean (Y) = $\dfrac{52+65+60+70+55+75+68+80+58+72}{10}$

$= \dfrac{655}{10} = 65.5$

$$Cov(x,y) = \frac{\Sigma[(x_i - \mu x)(Y_i - \mu y)]}{N-1}$$

$$= \frac{\begin{aligned}[&(5-7.1)(52-65.5) + (7-7.1)(65-65.5) + \\ &(6-7.1)(60-65.5) + (8-7.1)(70-65.5) + \\ &(5-7.1)(55-65.5) + (9-7.1)(75-65.5) + \\ &(7-7.1)(68-65.5) + (10-7.1)(80-65.5) + \\ &(6-7.1)(58-65.5) + (8-7.1)(72-65.5)\end{aligned}}{10}$$

$= (-2.1)(-13.5) + (-0.1)(-0.5) + (-1.1)(-5.5) +$

$+(0.9)(4.5) + (-2.1)(-10.5) + (1.9)(9.5) +$
$(0.1)(2.5) + (2.9)(+14.5) + (-1.1)(-7.5) +$
$(0.9)(6.5)$

$\overline{\qquad\qquad\qquad\qquad}$
9

$$= \frac{28.35 + 0.05 + 6.05 + 4.05 + 22.05 + 18.05 + 0.25 + 42.05 + \cancel{7.5}\ 8.25 + \cancel{7.15}\ 5.85}{9}$$

$$= \frac{135}{9} = 13.5 \qquad 15$$

c) Pearson correlation coefficient between advertising spend & sales revenue.

$$\text{Pearson corr. coeff} = \frac{cov(x,y)}{\sigma x \, \sigma y}$$

$SD(x) = \sigma x = 1.663$ (using excel)

$SD(y) = \sigma y = 9.119$

$$P.C.C = \frac{13.5}{(1.663)(9.119)} = \frac{15}{15.164} = 0.98$$

d) ~~Conversion~~

Advertising spend & Selves Revenue share a positive linear relationship.

(a) Avg daily commute = 45 mins.

H_0 = $\mu = 45$.

(Null Hypothesis)

H_1 = $\mu > 45$

(Alt Hypothesis

Sample Mean = $\bar{X}$ = 48 mins.

SD = 12 mins

Sample size = $n = 100$

Confidence level = 95%.

Tail area (100-95) = 5%.

2.5% each tail.

cumulative area = 95 + 2.5 = 97.5

(b) i) Z score in Z table = 1.96.

Std error SE = $\frac{SD}{\sqrt{n}} = \frac{12}{\sqrt{100}} = \frac{12}{10} = 1.2$

Margin error ME = $Z_{\alpha/2} * SE$

$= 1.96 \times 1.2 = 2.352$

CI = $\bar{x} \pm ME$ = 48 ± 2.352

c) $CI = (48 - 2.352,\ 48 + 2.352)$

$= (45.648,\ 50.352)$

d) <u>Conclusion</u>

Avg comute time is > 45 mins

Alt Hypothesis is true.

So, the claim of avg daily commute

Null Hypothesis is rejected.

Problem - 5

$n = 36$

sample mean $\bar{x} = 150$

$SD = 15$

a) Point of estimate for true avg no. of units proo
per hour = sample mean = 150 units

b) Confidence ~~for~~ interval 99%.

$$CI = \bar{x} \pm Z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}}\right)$$

1% / 2

$\bar{x} = 150$

$Z_{\alpha/2} =$ 99% + 0.005 = 0.995

Zscore from Z table = 2.58

$\sigma = 15$ $n = 36$.

$CI = 150 + 2.58\left(\frac{15}{\sqrt{36}}\right) = 150 \pm \underbrace{2.58\left(\frac{15}{6}\right.}_{6.4}$

(143.55, 156.45)

CI = 99%. Margin error = 3 units.

~~Estimate = ± 3~~

SD = 15

n = ?

$$M.E = Z_{\alpha/2} * \frac{SD}{\sqrt{n}}$$

99%

Z score = 2.58

$$E = Z_{\alpha/2} * \frac{15}{\sqrt{n}}$$

$$3 = 2.58 * \frac{15}{\sqrt{n}}$$

$$\sqrt{n} = \frac{2.58 \times 15}{3}$$

$$n = \left(\frac{2.58 \times \cancel{15}^{5}}{\cancel{3}}\right)^2 = 166.41$$

$$\boxed{n \simeq 166}$$

Problem-6

$\mu_{old} = 3.5$

$\mu_{new} = 3.8$.

$SD = 1.0$ $\qquad n = 50$.

$\mu \leq 3.5$

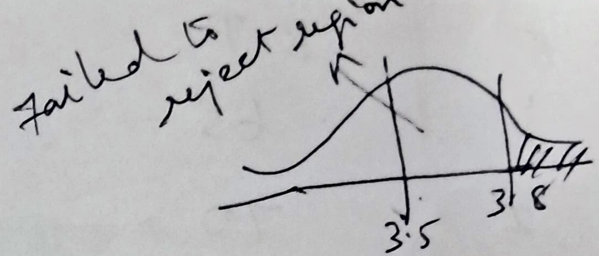
a) $\alpha = 0.05$

H_0 = Redesign did not help $\left(\cancel{\mu_{old} = \mu_{new}}\right)$

H_1 = Redesign increased avg time $\left(\cancel{\mu_{new} > \mu_{old}}\right)$

$\mu > 3.5$.

b) $Z = \dfrac{x-\mu}{\frac{\sigma}{\sqrt{n}}}$



$$= \frac{3.8-3.5}{1/\sqrt{50}} = \frac{0.3}{0.1414} = 2.1216$$

$P = P(Z > 2.12)$

$= 1 - P(Z < 2.12)$ (Z table)

$= 1 - 0.9830$

$= 0.017$

P is less than 0.05

So, null hypothesis is rejected.

Hence Redesign definitely increased the avg time spent in website.

$= 1 - 0.9830$

$= 0.017$

P is less than 0.05

c) So, null hypothesis is rejected.

Hence Redesign definitely increased the avg time spent in website.

Problem - 7

$n = 15$

a) Ho (null Hypothesis) = Training pgm did not improve productivity $\mu = 0$

H_1 (Alt Hypothesis) = Training pgm improved product $\mu > 0$

b) $\alpha = 0.05$

T-test & p-value

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

E	B.T	AF	D (AF − BF)
1	65	70	5
2	70	72	2
3	60	68	8
4	75	73	−2
5	80	~~80~~ 85	5
6	68	75	2
7	72	65	3
8	62	80	3
9	78	72	2
10	70	69	2
11	66	76	3
12	73	67	3
13	64	74	3
14	71	71	3
15	69		2

44

Mean diff $= \frac{44}{15} = 2.93$

SD of diff $= 2.120$ (By excel)

$$t = \frac{\bar{D} - M_D}{\frac{S_D}{\sqrt{n}}} = \frac{2.93 - 0}{\frac{2.120}{\sqrt{15}}} = \frac{2.93\cancel{\times 3.872}}{\frac{2.12}{3.872}}$$

$$= \frac{2.93}{0.547}$$

$$= 5.356$$

p-value

degrees of freedom $= 15-1 = 14$

$\alpha = 0.05$

Critical Value (T_table) $= 1.761$

c) conclusion

T_statistic > critical value

5.356 > 1.761, null hypothesis is rejected. There is statistically significant evidence at 0.05 level to conclude training pgm significantly improved productivity.

Problem -8.

a) Ho (null hypo) = Mean life of Prolife is not longer than std. batteries.

$M_1 \leq M_2$

H1 (alt Hypo) = Mean life of prolife greater than std batteries $M_1 > M_2$

Prolife	STD.
$n_1 = 40$	$n_2 = 50$
$\overline{x_1} = 120$	$\overline{x_2} = 115$
$s_1 = 10$	$s_2 = 12$

$\alpha = 0.02$

Z-test

$$Z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}}$$

$$= \frac{120-115}{\sqrt{\frac{(10)^2}{40} + \frac{(12)^2}{50}}} = \frac{5}{\sqrt{\frac{100}{40} + \frac{144}{50}}}$$

$$= \frac{5}{\sqrt{2.5 + 2.88}} = \frac{5}{2.319}$$

$$Z = 2.156$$

p value

$$p = P(z > 2.156)$$
$$= 1 - P(z < 2.156)$$
$$= 1 - 0.98422$$
$$= 0.01578$$

Conclusion

$2.156 > 0.0157$

$Z >$ p value & $P <$ significan level 0.02

∴ Reject the null hypothesis.

Mean life of 'Prolife' batteries is greater than the standard batteries.

Problem - 7

Sample mean = 24

Sample S.D = 3.93

b) one sided t-test

$\alpha = 0.01$

c) H_0 = Avg waiting time ≤ 20 mins.

H_1 = Avg waiting time > 20 mins.

$n = 30$; $\mu = 20$

$$t\text{-test} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$$= \frac{24 - 20}{\frac{3.93}{\sqrt{30}}} = \frac{4}{\frac{3.93}{5.477}}$$

$$= \frac{4}{0.717} = 5.578.$$

Critical value

dof = 30 - 1 = 29

$\alpha = 0.01$

$p = 2.462$

$t > p$ ∴ Null hypothesis is rejected.

So Avg waiting time for a table on weekend is longer than 20 minutes.

Problem - 10

a) H_0 - There is no relationship between age group & subscription

H_1 - There is a relationship between age group & subscription.

b) Observed Data

Age grp	Basic	Premium	Elite	Total
18-29	45	60	15	120
30-49	30	50	20	100
50+	25	40	15	80
total	100	150	50	300

Expected Frequency table.

$$E_{row \times col} = \frac{\text{Row Total} \times \text{col. Total}}{\text{Grand Total}}.$$

Age grp	Basic	Premium	Elite	Total
18-29	40	60	20	120
30-49	33.33	50	16.67	100
50+	26.67	40	13.33	80
total	100	150	50	300

c) chi-square test statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Age grp	Basic	Premium	Elite	~~Total~~
18-29	$25/40 = 0.625$	0	$25/20 = 1.25$	
30-49	$(3.33)^2/33.33 = 0.332$	0	$\frac{(3.33)^2}{16.67} = 0.665$	
50+	$(1.67)^2/26.67 = 0.104$	0	$\frac{(1.67)^2}{13.33} = 0.209$	

$$\chi^2 = 0.625 + 1.25 + 0.332 + 0.665 + 0.104 + 0.209$$

$$\chi^2 = 3.185$$

Degrees of freedom

dof = (no. of Rows − 1) × (no. of cols − 1)
$= (3-1) \times (3-1)$
$= 2 \times 2 = 4$

d) Critical chi-square value.

$\alpha = 0.05$ dof $= 4$

Critical value $= 9.488$

e) conclusion

χ^2 (3.185) < Critical value (9.488)

∴ We fail to reject the null hypothesis.

Means, there is no significant evidence to suggest a relationship between age grp & subscription. They are independent.

Bonus Problem

Design A

(Size) $n_A = 30$

(Mean) $\bar{x}_A = 2.85\%$

(SD) $S_A = 0.45\%$

Design B.

$n_B = 30$

$\bar{x}_B = 3.05\%$

$\bar{S}_B = 0.60\%$

a) H_0 -

Null Hypothesis → No difference in avg daily conversions between Design A & B.

b) H_1 → there is a significant diff in avg daily conversions b/w Design A & B.

b) Assuming equal population variance

c) t-test for 2 independent means.

Sp (pooled std deviation)

$$= \sqrt{\frac{(n_A-1)S_A^2 + (n_B-1)S_B^2}{n_A+n_B-2}}$$

$$= \sqrt{\frac{29\times(0.45)^2 + 29\times(0.60)^2}{30+30-2}}$$

$$= \sqrt{\frac{5.8725 + 10.44}{58}} = \sqrt{\frac{16.3125}{58}}$$

$$= \sqrt{0.28125} = 0.5303$$

t - test

$$t = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{S_p \sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

$$\begin{array}{r} \cancel{3}3.05 \\ 2.85 \\ \hline 0.20 \end{array}$$

$$= \frac{(2.85 - 3.05) - 0}{0.5303 \sqrt{\left(\frac{1}{30} + \frac{1}{30}\right)}} = \frac{+\cancel{0.}20}{0.\cancel{1369}}$$

$$t = \frac{0.20}{0.1369} = 1.4609$$

d) Degrees of freedom

$$n_A + n_B - 2 = 30 + 30 - 2$$
$$= 58$$

e) $\alpha = 0.05$ Two-sided $\alpha/2 = 0.025$

Critical value $\pm$ 2.000

f) **conclusion**

$t <$ critical value ∴ It (t) lies within the $\pm$ critical value (2.0)

$(1.4609) < (2.000)$

the null hypothesis.

∴ We fail to reject the null hypothesis.

There is no statistically significant diff in avg daily conversions between Design A & B.

e) 95% confidence ~~interval b/w~~ mean daily conversion rates $(\mu_B - \mu_A)$.

$$C.I = (\overline{x_B} - \overline{x_A}) \pm M.E$$

$$M.E = t_{critical} * S.E$$

$$= 2 * \left(s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \right)$$

$$= 2 * 0.1369$$

$$= 0.2738$$

$$CI = (3.05 - 2.85) \pm 0.2738$$

$$= 0.20 \pm 0.2738$$

$$\begin{array}{r} 0.2738 \\ 0.20 \\ \hline 0.4738 \end{array}$$

Confidence Interval $= (0.0738, 0.4738)$

$$\begin{array}{r} 0.2738 \\ 0.2 \\ \hline 0.0738 \end{array}$$