0022-TUS **Answer form for multiple choice. You can make your own form, put please be clear and answer on <u>one</u> page. Do <u>not</u> submit solutions to the multiple-choice problems.**

Number	Part	A	B	C	D	E
1	a.	■				
1	b.	■				
2	a.		■			
2	b.			■		
3	a.					■
3	b.				■	
4	a.				■	
4	b.			■		
5	a.			■		
5	b.		■			

1.

A) (A) 0

B) (A) 1.16

2.

a) (B) 0.0028

b) (C) 0.10

3.

A) (E) 0.29
B) (D) 0.67

4.

A) (D) (-5,36, 7.76)
B) (C) 3.1%

5.

A) (C) 6.635
B) (B) 1.34

Q.6. 331, 329, 328, 328, 328, 329, 331, 329, 328, 329.

a) Sample mean $\bar{x} = \frac{\sum x_i}{n}$

$$= \frac{331 + 329 + 328 + 328 + 328 + 329 + 331 + 329 + 328 + 329}{10}$$

$$= \frac{3290}{10} = 329$$

Sample var. $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$

$$= \frac{2^2 + 0^2 + 1^2 + 1^2 + 1^2 + 0^2 + 2^2 + 0^2 + 1^2 + 0^2}{10-1}$$

$$= \frac{4 + 1 + 1 + 1 + 4 + 1}{9} = \frac{12}{9} = \frac{4}{3}$$

Sample mean $\bar{x} = 329$
Sample Variance $s^2 = \frac{4}{3}$

b) $\alpha = 10\%$ or 0.10

Test whether mean content < 330 ml,

Q6. b) Assumptions:

- Data is from a normally distributed population. (given)
- Data is continuous (volume is continuous)
- ~~A~~ simple random sample is chosen. (given).

∴ Assumptions are ~~not~~ satisfied.

Hypothesis:

$H_0: \mu = 330$ ml. — Null Hypothesis

$H_a: \mu < 330$ ml. — Alternate Hypothesis.

Null Hypothesis: The mean content is 330 ml.

Alternate " : The mean content is less than 330 ml.

Test variable: Small population, unknown σ, normal population.

We will use ~~t-variable~~ t-statistic for testing.

Degrees of freedom = 10 − 1 = 9.

c) Critical value: $t_{crit} = t_{0.1}$ with df = 9

$= -1.383$

[Value of t for which $P(T<t) = 0.1$]

Decision rule: If $t_{statistic} < t_{crit}$: reject H_0.

If $t_{statistic} \geq t_{crit}$: fail to reject H_0.

d) $t_{statistic} = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{329 - 330}{\sqrt{4/3}/\sqrt{10}} = -2.7386 \approx -2.739$

Since the $t_{statistic} < t_{crit}$, we reject H_0.

Thus, there is sufficient evidence to claim that the mean content is less than 330 ml.

Q.6- e) Total number of possible selections of 4 cola out of 8 possible colas.

$$= {}^{8}C_4 = \frac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2 \times 1} = 70$$

Number of cases in which all 4 are correct = 1.

Upon just guessing, each choice of 4 colas is equally likely.

$\therefore$ P(all 4 correct) $= \frac{1}{70}$.

Q.7. a) 95% CI for BEDROOMS in model 3.

$\hat{\beta} = 14.7904$ $\quad SE(\hat{\beta}) = 22.4612$ [approx.]

To get 95% CI, we need $\pm t_{0.025, n-2}$ $\quad df = n-2 = 50-2 = 48$.

$t_{0.025, 48}$ = t-value for which $P(T > t) = 0.025$.

$= 2.011$

Margin of error $= t_{0.025, 48} * SE(\hat{\beta})$

$= 2.011 * (22.4612)$

$= 45.1695$

$\therefore$ 95% CI $\Rightarrow [14.7904 - 45.1695, \; 14.7904 + 45.1695]$

$= [-30.4655, \; 59.8735]$

Interpretation: Since 0 is part of the CI, we can not say that the coefficient of BEDROOMS in model 3 is different from 0 at 0.05 level of significance.

~~Given that SQFT is included in the model~~
~~$\Rightarrow$ Model 2 has to be analyzed~~

Q.7. b)

Hypothesis: $H_0: \beta_{(LONG)} = 0$ [coefficient is not different from 0]

$H_a: \beta_{(LONG)} \neq 0$. [Coefficient is different from 0]

→ Test variable is the t-statistic.

→ Critical value: It is given that SOFT is included. This means that the result for SOFT is significant.

$T_{model\ 2\ for\ SOFT} = \frac{0.231142}{0.031807} = 7.267$.

$T_{model\ 3\ for\ SOFT} = 6.4633$

Since we are to check the significance of LONG given SOFT is included, we can check either model 2 or 3. But model 3 is more explanatory of the variation because of higher R^2. So we will go with model 3 for this test.

$T_{crit} = 6.4633$

(Since Confidence level or other info is not provided and only inclusion of SOFT is known)

→ Decision Rule: $|T| \geq |T_{crit}|$: reject H_0.

$|T| \leq |T_{crit}|$: fail to reject H_0.

Q.7. c) Test statistic: $t = \frac{-4930.9242}{1485.722}$

$= -3.3189$.

$|t| = 3.3189$.

$|t| < 6.6633$ (Critical value)

∴ we fail to reject the null Hypothesis.

∴ we can not conclude that the coefficient of LOVG is significantly different than zero.

Q.7. d) Area = 1500 sqft. 90% CI for prediction interval

Model 1: Price = −83.26899 + 0.27125 · SQFT.

For SQFT = 1500,

$\widehat{Price} = 323.60601$

$\widehat{Price} \approx 323.606$

t_{crit} = t value for which $Pr(-t < T < t) = 0.9$

$t_{crit} \Rightarrow Pr(T > t)$ for $df = n-2 = 48$.

α is 0.05.

$\Rightarrow t_{crit} = 1.677$.

~~CI for Price predicted =~~

$\bar{x} = 1650$

$s^2 = 552284$

$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$

$SE = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$

$= \sqrt{552284 + \frac{552284}{50} + \frac{(1500-1650)^2}{49}}$

$= 750.8588$

Q.7. d) continued.

Margin of error = $t_{crit} * SE$

$= 1.677 * 750.8588$

$= 1259.190$

90% CI for Price is 323.606 ± 1259.190

$\Rightarrow [-935.584, 1582.796]$

Interpret: We are 90% confident that the predicted price for SQFT = 1500 falls in $[-935.584, 1582.796]$ if model 1 is followed. Since the price also includes negative values, this is not a good prediction.

---

e) Correlation = -0.55 (negatively correlated)

$R_{model\ 1} = 0.27125$

$R_{model\ 2} = 0.23114$

As SQFT increases, LONG decreases, and vice-versa.

R in model 2 is smaller since the negative correlation makes it such that some effects / variation is explained by LONG and so the coefficient for SQFT is smaller.