Part II

Question 1 (5 marks)

A food store has determined that daily demand for milk cartons has a normal distribution, with a mean of 55 cartons and a standard deviation of 6 cartons.

a. On Saturdays, the demand for milk is known to exceed 60 cartons. On the coming Saturday, what is the probability that it will be at least 70 cartons?

when $x=60$

$$z = \frac{60-55}{6} = 0.833$$

$$P(X>60) = 1-0.7967 = 0.2033$$

when $x=70$

$$z = \frac{70-55}{6} = 2.5$$

$$P(X>70) = 1-0.9938 = 0.0062$$

$$\Rightarrow P(x>70 \mid x>60) = \frac{0.0062}{0.2033} = 0.0305$$

b. The store receives 50 cartons of milk each morning, from which one is put aside for the manager's personal use. What is the probability of having an insufficient number of milk cartons to meet demand?

Find $P(x>49)$

when $x=49$ $\quad z = \frac{49-55}{6} = -1$

$$P(x>49) = P(z>-1) = 1-0.1587$$
$$= 0.8413$$

Question 2 (10 marks)

The Federal Environmental Agency has warned the city of about recurring poor air quality indicators due to the presence of an air pollutant called Carbon Monoxide (CO). CO is a colorless poisonous gas that is emitted directly from automobile tailpipes. Some years ago, the Agency imposed limits of 2.1 g/km on exhaust gas emissions at the tailpipe. The city is planning to launch a massive campaign of gas emission controls. A prior pilot set of controls was achieved on a random set of 30 cars.

2	2.03	1.55	1.38	2.58	1.62	0.63	1.14	1.69	0.88
1.9	2.98	2.39	1.07	2.89	0.51	2.69	2.04	2.32	1.99
1.24	0.08	1.49	1.91	2.92	1.7	2.03	1.36	2.49	1.62

$\sum x_i = 53.12, \sum x_i^2 = 109.359$

a. Estimate with a 95% confidence interval, the true mean CO emission per car. Interpret.

$$\bar{x} = \frac{\Sigma x}{n} = \frac{53.12}{30} = 1.77$$

$$S_x^2 = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 109.359 - \frac{1}{30} \times (53.12)^2$$
$$= 15.3$$

$$S_x = \sqrt{15.3} = 3.912$$

95% confidence interval is

$$\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{S_x}{\sqrt{n}}$$

$z_{\frac{\alpha}{2}} = 1.96$

$$\Rightarrow 1.77 \pm 1.96 \cdot \frac{3.912}{\sqrt{30}}$$

$\Rightarrow$ 95% confidence interval is $[0.974, 4.514]$

b. The city of wants to know with a confidence level of 90% the true mean CO emission per car with a margin of error of maximum 0.025. Assuming that the population standard deviation is unknown, how many additional cars should be tested to provide such an estimate?

Confidence level $= 90\%$

$z_{\frac{\alpha}{2}} = 1.645$

$$n = \frac{(z_{\frac{\alpha}{2}})^2 \cdot S^2}{E^2}$$

$$= \frac{(1.645)^2 \cdot 3.912^2}{(0.025)^2}$$

$$\approx 66260$$

Additional sample should be 66230.

Question 3 (10 marks)

A travel company wishes to determine if the type of vacation purchased in its market area is independent of income level of purchasers. A random survey of purchasers gave the following results:

Vacation Type	Income Level			
	High	Medium	Low	
Domestic	50	120	65	235
Foreign	25	30	10	65
	75	150	75	300

a. At the 0.05 level of significance, can it be concluded that vacation preference and income level are statistically independent? Interpret the result in the context of the problem.

H_0: vacation preference & Income level are statically Independent

H_A: Vacation preference and Income level are not statistically independent.

χ^2 test

$$\chi^2_{stats} = \frac{(50-58.75)^2}{58.75} + \frac{(120-117.5)^2}{117.5} + \frac{(65-58.75)^2}{58.75}$$

Expected value table

$$+ \frac{(25-16.25)^2}{16.25} + \frac{(30-32.5)^2}{32.5} + \frac{(10-16.25)^2}{16.25}$$

	H	M	L
D	58.75	117.5	58.75
F	16.25	32.5	16.25

$$= 9.33$$

$$\chi^2_{\alpha,(r-1)(c-1)} = 5.99147$$
$$5\%,(2-1)(3-1)$$

$$\chi^2_{stats} > \chi^2_{\alpha,(r-1)(c-1)}$$

we reject H_0

We have enough evidence to show that the variable vacation preference is not statistically Independent with variable Income level at 5% significance level

b. At 5% level of significance, is there sufficient evidence to conclude that more than a third of the vacationers are high-income purchasers. Interpret the result in the context of the problem.

$$H_0: \ p \le \frac{1}{3}$$

$$H_A: \ p > \frac{1}{3}$$

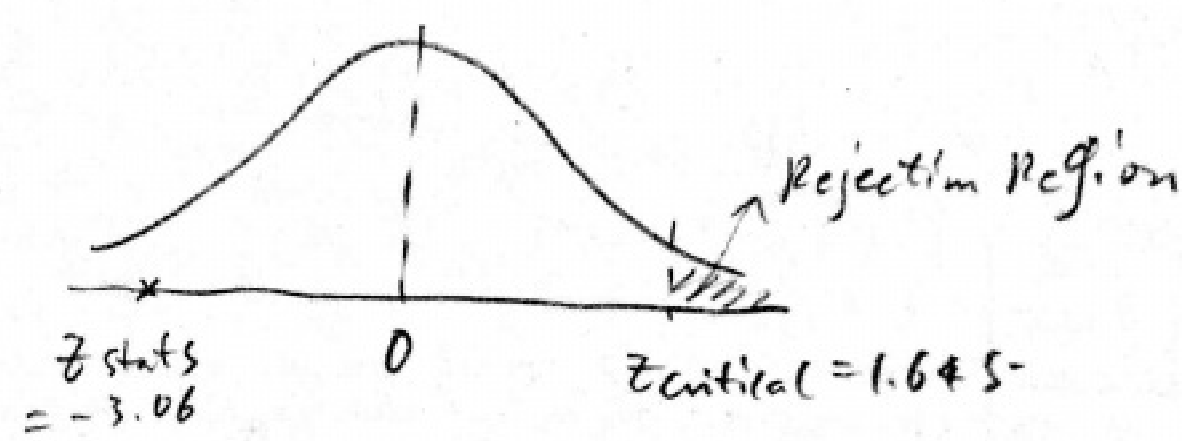
$$\bar{p} = \frac{75}{300} = 0.25$$

z-test right tail

$$z_{stats} = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.25 - \frac{1}{3}}{\sqrt{\frac{\frac{1}{3} \times \frac{2}{3}}{300}}}$$

$$= \frac{-0.08333}{0.02722} = -3.06$$

$$z_{\substack{critical \\ 5\%}} = 1.645$$



Since $z_{stats} < z_{critical}$

we do not reject H_0

$\Rightarrow$ we have enough evidence to show that less than or equal a third of vacationers are high-Income purchasers at 5% significance level.

Question 4 (15 marks)

ZUBRAK Inc. is a computer firm specializing in Web designs. Since ZUBRAK uses its own special web page design software, all newly hired employees have to go through a training course regardless of their work experience in computing. The vice-president of personnel is skeptical about paying a higher salary to hire more experienced people since they have to be retrained. For each of 15 randomly selected new employees, she obtained data on the employee's score in web page design skill Y (after the training course) and also the number of months of computer related work experience X at the time of hiring. Part of the data along with sums, sums of squares and some relevant statistic are given below.

Employee	1	2	3	.	.	.	13	14	15
X_i	19	10	12	.	.	.	18	15	14
Y_i	81	66	70	.	.	.	88	78	73

$\bar{X} = 15.733$, $\bar{Y} = 76.4667$, $\sum(X_i-\bar{X})^2 = 164.934$, $\sum(Y-\bar{Y})^2 = 9376.22$, SSE = 188.03, $S_{b_1} = 0.2961$.

Suppose that a simple linear regression model is appropriate for analyzing the above data and the least squares fit is obtained as $\hat{Y}_i = 49.579 + 1.709X_i$,

a. At 5% significance level, is there sufficient evidence that design skill is positively related to computer related work experience? Interpret the result in the context of the problem.

$H_0: \beta_1 \le 0$

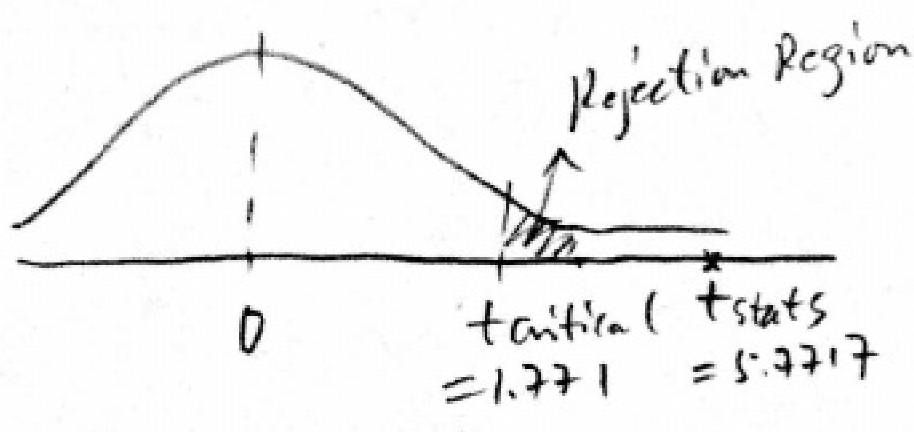
$H_A: \beta_1 > 0$

t-test right tail

$$t_{stats} = \frac{b_1}{S_{b_1}} = \frac{1.709}{0.2961} = 5.7717$$

$$t_{critical} = 1.771$$

$5\%,\ 15-1-1$



Since $t_{stats} > t_{critical}$

we reject H_0

We have enough evidence to conclude that design skill is positively related to computer related work experience at 5% significance level.

b. What proportion of the variation of design skill scores is accounted for by computer related work experience? Interpret the result in the context of the problem.

$SST = \Sigma(Y_i - \bar{Y})^2 = 9376.22$

$SSE = 188.03$

$$r^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{188.03}{9376.22} = 0.98$$

98% of variation in design skill scores can be explained by computer related work experience.

c. Find a 95% prediction interval for design skill score of an employee with 15 months of computer related work experience. Interpret the result in the context of the problem.

$$\hat{y} = 49.579 + 1.709\,x$$

$$x_0 = 15$$

$$\Rightarrow \hat{y} = 49.579 + 1.709 \times 15 = 75.214$$

prediction interval

$$\hat{y} \pm t_{\frac{\alpha}{2}, n-k-1} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_{xx}}}$$

$$s = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{188.03}{15-1-1}} = 3.803$$

Check table $\alpha = 5\%$

$$t_{\frac{\alpha}{2}, 13} = 2.160$$

$$\Rightarrow 75.214 \pm 2.16 \times 3.803 \times \sqrt{1 + \frac{1}{15} + \frac{(15 - 15.733)^2}{164.934}}$$

$\Rightarrow$ prediction interval is

$$[66.72, 83.71]$$

Question 5 (12 marks)

Based on a survey of 20 firms, a researcher has developed a multiple regression model relating sales (**SALES** in \$1,000) to inventory investment (**INVEST** in \$1,000), advertising expenditures (**AD** in \$1,000) and the average bonus paid to employees (BONUS in \$1,000). The table below shows partial results obtained from fitting a multiple regression model using Excel.

Regression output

	Coefficients	*Std. error*	*t*	*p-value*
Intercept	25.50	18.802		
INVEST	10.05	4.251		
AD	8.05	3.502		
BONUS	0.125	0.041		

ANOVA table

Source	*SS*	*df*	*MS*	*F*	*p-value*
Regression	13440	3	4480	21.333	
Residual	3360	16	210		
Total	16800	19			

a. Is there sufficient evidence at the 5% level of significance to conclude that the model is useful in predicting sales? Interpret the result in the context of the problem.

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$

HA: At least one equality is not held in Ho.

F-test

$F_{stats} = 21.333$

$F_{critical\ 5\%, 3, 16} = 3.24$

Since $F_{stats} > F_{critical}$

We reject Ho.

We have enough evidance to conclude that the model is useful in predicting sales at 5% significance level.

b. Is there sufficient evidence at the 5% level of significance to conclude that sales are related to expenditure on advertisement, given that inventory investment and bonus paid to employees remain unchanged? What is the p-value of the test?

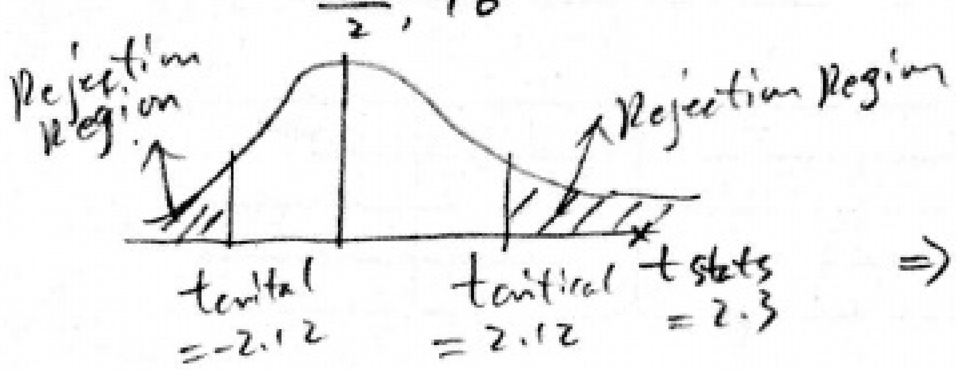
$H_0: \beta_2 = 0$

$H_A: \beta_2 \neq 0$

t-test two tail

$$t_{stats} = \frac{8.05}{3.502} = 2.3$$

$$t_{critical\ \frac{5\%}{2},16} = 2.12$$



$|t_{stats}| > t_{critical}$

We reject H_0

⇒ we have enough evidence to conclude that sales are related to expenditure on advertisement at 5% significance level.

c. Estimate the coefficient of determination and explain its meaning in the context of the problem. Interpret the result in context of the problem.

$$r^2 = \frac{SSR}{SST} = \frac{13440}{16800} = 0.8$$

80% variation of sales can be explained by inventory investment, advertising expenditure, and average bonus paid to employees.

d. Estimate sales based on a planned investment in inventory of \$15,000, an advertising budget of \$10,000 and an average bonus to employees of \$2,000 for the coming year.

$$\hat{y} = 25.5 + 10.05 \times INVEST + 8.05\,AD + 0.125\,BONUS$$

$$\hat{y} = 25.5 + 10.05 \times 15 + 8.05 \times 10 + 0.125 \times 2$$
$$= 257$$