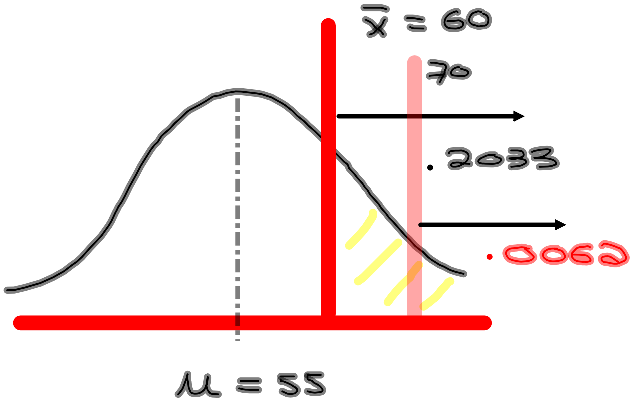
**Question 1 (5 marks)**

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?




$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{60 - 55}{6} \cong \boxed{.83}$$

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

Let A be the event that $x \geq 70$

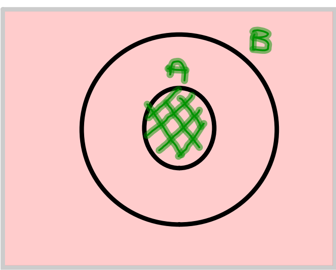
Let B " " " " $x \geq 60$

Joint

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{.0062}{.2033}$$

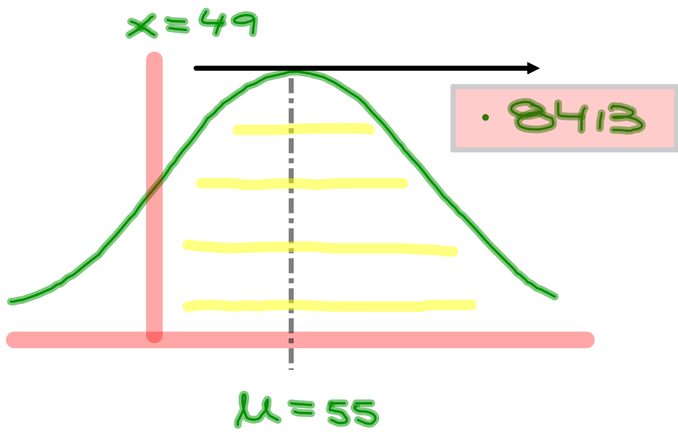
conditional

marginal




$$= \boxed{.0305}$$

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?




$$Z = \frac{49-55}{6} = -1$$

**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 |

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

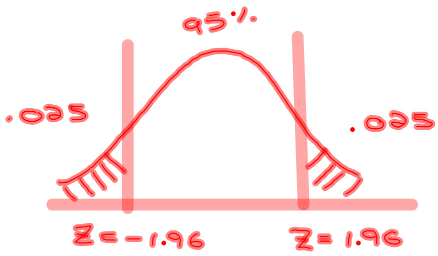
a. Estimate with a 95% confidence interval, the true mean CO emission per car. Interpret.

$$\bar{x} \pm z \cdot \sigma_{\bar{x}}$$

$$\sigma_{\bar{x}} \sim \frac{s}{\sqrt{n}}$$

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{109.359 - \frac{(53.12^2)}{30}}{30-1}} = 0.7264$$

$$\bar{x} = \frac{\sum x}{n} = \frac{53.12}{30} = 1.7707$$



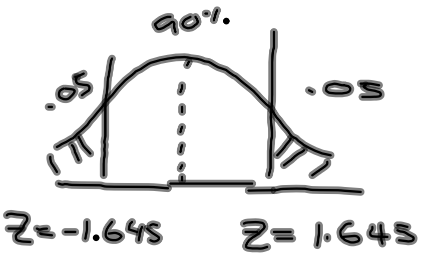

$$1.7707 \pm 1.96 \left[\frac{.7264}{\sqrt{30}}\right]$$

$$[\ 1.510761\ ,\ 2.030639\ ]$$

We are 95% confident that the population mean emission level is contained in the above interval.

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?

$$n = \left( \frac{z_{\alpha/2} \cdot \sigma}{\varepsilon} \right)^2$$




$$n = \left[ \frac{(1.645) \cdot (0.7264)}{.025} \right]^2$$

$$= \boxed{2284.565} \rightarrow 2285$$

$\therefore$ 2285 − 30 = $\boxed{\text{2255 additional observations}}$

**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 |
| Foreign | 25 | 30 | 10 |

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.

H0: Income level and type of vacation purchased are independent

Ha: Income level and type of vacation purchased are not independent

$$\chi^2_{OBS} = \sum\sum \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

$$e_{ij} = \frac{\text{Total of Row } i \times \text{Total of column } j}{\text{Sample size}}$$

| | High | Medium | Low | Total |
|---|---|---|---|---|
| Domestic | 50 | 120 | 65 | 235 |
| Foreign | 25 | 30 | 10 | 65 |
| Total | 75 | 150 | 75 | 300 |

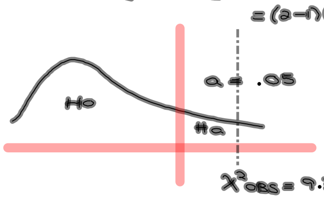
OBS

| | High | Medium | Low | Total |
|---|---|---|---|---|
| Domestic | 58.75 | 117.5 | 58.75 | 235 |
| Foreign | 16.25 | 32.5 | 16.25 | 65 |
| Total | 75 | 150 | 75 | 300 |

Expected

$$\chi^2_{OBS} = \frac{(50-58.75)^2}{58.75} + \frac{(120-117.5)^2}{117.5} + \frac{(65-58.75)^2}{58.75} + \frac{(25-16.25)^2}{16.25} + \frac{(30-32.5)^2}{32.5} + \frac{(10-16.25)^2}{16.25} = 9.328969$$

$\chi^2_c = 5.991$ $\quad d.f = (r-1)(c-1) = (2-1)(3-1) = 2$




$\chi^2_{OBS} = 9.32$

$\because \chi^2_{OBS} > \chi^2_c$ i.e. $9.32 > 5.991$

$\therefore$ we reject Ho at a 5% level of alpha. I.e. Income & vacation type are Not Independent

b) At a 5% level is there sufficient evidence to indicate that the proportion of all vacationers that are from the high income group exceeds one third?

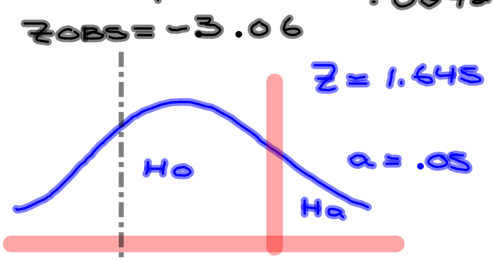
$H_o: \ p \le \frac{1}{3}$

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

$$\bar{p} = \frac{x}{n} = \frac{75}{300} = .25$$

$$\sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(1/3)(2/3)}{300}} = \boxed{0.027217}$$

$$Z_{OBS} = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{.25 - 1/3}{.027217} = \boxed{-3.06186} \approx -3.06$$




$\because$ $Z_{OBS} < Z_c$ i.e. $-3.06 < 1.645$

$\therefore$ we do not reject Ho at a 5% level.

I.e. The proportion of high income travellers does not exceed a third.

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.

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.

**estion 5 (12 marks)**

sed on a survey of 20 firms, a researcher has developed a multiple regression model relating sales (**SALES** in $1,000) to entory investment (**INVEST** in $1,000), advertising expenditures (**AD** in $1,000) and the average bonus paid to ployees (BONUS in $1,000). The table below shows partial results obtained from fitting a multiple regression model ng 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 | | | | | |
| Residual | 3360 | | | | |
| Total | 16800 | | | | |

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.

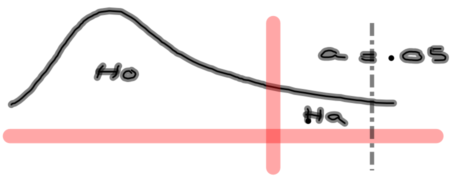
| | Coefficient | Std Error | T-Stat | p-value | |
|---|---|---|---|---|---|
| Intercept | 25.5 | 18.802 | 1.356239 | | |
| Invest | 10.05 | 4.251 | 2.36415 | | |
| Ad | 8.05 | 3.502 | 2.298686 | | |
| Bonus | 0.125 | 0.041 | 3.04878 | | |
| | | | | | |
| ANOVA Table | | | | | |
| Source | SS | df | MS | F | P |
| Regression | 13440 | 3 | 4480 | 21.33333 | |
| Residual | 3360 | 16 | 210 | | |
| Total | 16800 | 19 | | | |

$H_0: B_1 = B_2 = B_3 = 0$

$H_a:$ At least one $B_i \neq 0$

$$F_{OBS} = \frac{MSR}{MSE} = 21.33$$

$F_c = 3.24$




$df_N = 3$

$df_N = 16$

$F_{OBS} = 21.33$

∵ $F_{OBS} > F_c$ i.e. $21.33 > 3.24$

∴ we reject Ho at a 5% level.

I.e. the overall model is significant a linear relationship exists between Sales and at least one independant variable.

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?

$H_o: B_2 = 0$

$H_a: \beta_2 \neq 0$

$$t_{OBS} = \frac{b_2}{s(b_2)} = 2.299$$

$.05 > p\text{-value} > .02$

$\because$ p-value < alpha $\therefore$ we reject Ho.

I.e. A linear relationship exists at a 5% level b/w sales & adv exp.

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

$$R^2 = \frac{SSR}{SST} = \frac{13440}{16800} = 0.8 \text{ or } 80\%$$

The coefficient of determination tells us that 80% of the total variation in sales can be explained by variations in the independent variables that are part of the model.

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.

15 10

$\hat{y} = 25.5 + 10.05x_1 + 8.05x_2 +$

$0.125x_3 =$ 257

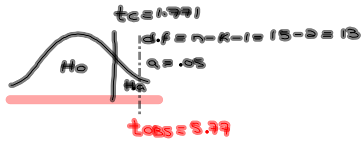
2

$\therefore$ 257 000 is the predicted sales

a) $H_0: B_1 \leq 0$

$H_a: B_1 > 0$

$$t_{OBS} = \frac{b_1}{S(b_1)} = \frac{1.709}{.2961} = \boxed{5.77}$$




$\because t_{OBS} > t_c \quad 5.77 > 1.771$

$\therefore$ we reject Ho at a 5% level

I.e a positive linear relationship exists

B)

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{188.03}{9376.22}$$

$$= \boxed{.9799}$$

The coefficient of determination tells us that 97.99% of the total change in design skill scores can be explained by changes in computer related work experience.

c)

$$\hat{y} = 49.579 + 1.709(15) = \boxed{75.214}$$

$$s = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{188.03}{15-2}} = \boxed{3.803}$$

$$t_{.025,13} \rightarrow \boxed{2.160}$$

$$\text{distance value} = \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_{xx}}$$

$$= \frac{1}{15} + \frac{(15 - 15.933)^2}{164.934}$$

$$= \boxed{0.069924267}$$

$$\hat{y} \pm t \cdot s \cdot \sqrt{1 + \text{dist value}}$$

$$75.214 \pm (2.16)(3.803)\sqrt{1 + .069924}$$

$$[\ 66.71718\ ,\ 83.71082\ ]$$

We are 95% confident that the design skill score for a particular individual with 15 months of computer related work experience is contained in the above interval.

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?

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

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.