STAT 201: REVIEW

November 29th, 2004

Problem 1:

Data here displays the mathematics achievement test scores for a random sample for $n = 10$ college freshmen, along with their final calculus grades. The aim is to check if there is a linear relationship between the mathematics achievement test scores and the final calculus grades.

Table 1

Student	Mathematics Achievement Test score (x)	Final Calculus Grade (y)
1	39	65
2	43	78
3	21	52
4	64	82
5	57	92
6	47	89
7	28	73
8	75	98
9	34	56
10	52	75

Those are the output of a linear regression

$R^2 = 0.7552$, $n = 10$, intercept$=30.78$, slope$=0.8655$

1. What is are the explanatory variable (X) and what is the response variable (Y)?
2. Make a scatterplot of X and Y
3. What are the regression coefficients: β_0 and β_1?
4. What is the regression equation summarizing the relationship between Test score and Grade.
5. Calculate the correlation coefficient between *Test Score* and *Grade*.
6. What is the coefficient of determination.
7. Calculate the predicted Final grade of a student who had 54 in the Test Score.
8. Calculate the residual value of the 8th student.

Answer:

1. The explanatory variable X is the Test score, The response variable Y is Final

calculus grade

2. For the scatterplot, make sure to put the response variable on the Y axis and the explanatory variable on the X-axis
3. The intercept $\beta_0 = 30.78$ and the slope $\beta_1 = 0.8655$
4. The regression equation $Grade = 30.78 + 0.8655\ Test\_score$
5. The correlation coefficient $r = \sqrt{R^2} = \sqrt{0.75} = 0.866$
6. The coefficient of determination $R^2 = 75\%$
7. The predicted Final grade $\hat{Y} = 30.78 + 0.8655 \times 54 = 77.517$
8. The residual value for the 8^{th} student is $\hat{e}_8 = Y_8 - \hat{Y}_8$, $\hat{Y}_8 = 30.78 + 0.8655 \times 75 = 95.693$, so $\hat{e}_8 = 98 - 95.69 = 2.31$

Problem 2:

The Lifetime of a particular brand of a $Television,$ is a random variable X which has a normal distribution with mean and standard deviation

μ_X= **14** years and σ_X= **3** years

- Calculate the probability that a $Television$ lasts at least 11.45 years.
- Calculate the probability that a $Television$ lasts at most 15 years.
- Calculate the probability that a $Television$ lasts at least 11 and at most 15 years.

- Probablility that a television lasts at least 11.45 years is

$$P(X > 11.45) =$$
$$P(Z > \frac{15.45 - 14}{3}) =$$
$$P(Z > 0.48) = 0.31$$

- Probability that a television lasts at most 14 years is

$$p(X < 14) =$$
$$P(Z < \frac{15 - 14}{3}) =$$
$$P(Z < 0.33) = 0.62$$

- Probability that a television lasts at least 11 and at most 15 is

$$P(11 < X < 15) =$$
$$P(\frac{11 - 14}{3} < Z < \frac{15 - 14}{3}) =$$
$$P(-1 < Z < 0.33) = 0.62 - 0.15 = 0.47$$

Problem 3:

Consider the following sample:

50	52	79	76	44	49	63	77	53	45	91	49	63

a. Calculate the sample mean.
b. Calculate the median.
c. Calculate the first and third quartile and make a boxplot.

Answer:
The ranked data is:

44	45	49	49	50	52	53	63	63	76	77	79	91

a. Mean$= 60.846$
b.Median$= 53$
c.$Q1 = 49, \quad Q2 = 76.5$

Problem 4:

The owner of a haircut shop thinks that the average waiting time for a customer is μ_x= **25 min** and that the standard deviation is σ_x= **6 min**. To check his claim, we have selected a random sample of **30** customers and calculated the average waiting time $\bar{x}$ = **28 min.** If the waiting time of the whole customer is supposed to be normally distributed:

- 1- Calculate the **97**% confidence interval of the customers average waiting time μ?
- 2- If the standard deviation of the whole customer waiting time is unknow and that the standard deviation of the sample is provided to be equal to $s = 5.5$, find the 97% confidence interval of the customers average waiting time μ.
- 3- In both cases, do you think that the manager claim is justified.

Answer:

1- 96% Confidence interval for the average waiting time for the whole customer is:

$$\bar{X} \pm z^* \frac{\sigma}{\sqrt{n}}, \text{ where } z^* = 2.054$$

The 97% Confidence interval is $[28 - 2.054\frac{6}{\sqrt{30}}; \quad 28 + 2.054\frac{6}{\sqrt{30}}] = [25.75; \ 30.25 \ ]$

2- If $s = 5.5\,\text{min}$, then the 96% Confidence interval is $[28 - t^*_{(29,96\%)}\frac{5.5}{\sqrt{30}};$ $28 + t^*_{(29,96\%)}\frac{5.5}{\sqrt{30}}]$

$[28 - 2.15\frac{5.5}{\sqrt{30}};$ $28 + 2.15\frac{5.5}{\sqrt{30}}]$, $\quad [25.841; \ 30.159]$

3- The manager claim is not justified since μ_x= **25 min**, does not belong the both confidence intervals

Problem 5:

The University of Tennessee uses thousands of fluorescent light bulb each year. The brand of bulb it currently uses has a mean life of $\mu = 805$ hours. A manufacturer claims that its new brand of bulbs, which cost the same as the brand the university currently uses, has a mean life **different** than **805 hours**. The university has decided to purchase the new brand if, when tested, the test evidence supports the manufacturer's claim at the $\alpha = .05$ significance level. Suppose that **30** bulbs were tested with the following results:

$\bar{X} = 800$ **hours,** $s_X = 47$ **hours**.

Will the University of Tennessee purchase the new brand of fluorescent bulbs?

Answer:

1) $H_0 : \mu = 805$ and $H_a : \mu \neq 805$

2) Test statistics

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{800 - 805}{47/\sqrt{30}} = -0.58$$

3)P-value is

$$2 \times P(T_{(df=29)} > |-0.58|) = 2 \times P(T_{(df=29)} > 0.53) \simeq 2 \times 0.25$$

4) Decision, reject the null hypothesis if P-value is smaller than 5%. *Here* we do not reject H_0

Example 6:

The following table describes the number of Teenagers and Women surveyed this year asking them if they a particular skin care product:

Table:

Skin Care use (rows) / Female (columns)

	Teenagers	Women	Total
Yes	8000	1200	
No	1500	1000	
Indecided	600	800	
Total			

1. How many variable do we have here?.
2. How many categories do we have for each variable? Describe them?
3. What do we call the table above?
4. Fill in the empty (calculate total rows, total columns, and the total number of female surveyed).
5. Provide the same table with frequencies instead of counts
6. What is the proportion of teenagers who use the particular Skin care?
7. What is the proportion of Women who are indecided
8. Conditional on the number of undecided female, what is the proportion of women?

Answer:

1. We have two variables: Female variable and the use of skin care variable
2. Two categories for Female and three categories for the use of Skin care
3. The table above is called: contingency table or two-way table or classification table
4. Total (Yes)$=9200$, Total(No)=2500, Total(Indecided)$=1400$, Total(Teenagers)$=10100$, Total (Women)$=3000$, Total$=13100$
5.

	Teenagers	Women	Total
Yes	0.61	0.091	0.702
No	0.11	0.076	0.190
Indecided	0.04	0.061	0.106
Total	0.77	0.229	1

,

6. prop of Teenagers who use the particular skin care is 0.61
7. prop of Women who are indecided is 0.061
8. Given the indecided female, the prop of women is $\frac{800}{1400} = 0.57$