Question1

Below you are given the examination scores of 20 students (data set also provided in accompanying MS Excel file).

52	99	92	86	84
63	72	76	95	88
92	58	65	79	80
90	75	74	56	99

a.   Construct a frequency distribution, cumulative frequency distribution, relative frequency distribution, cumulative relative frequency distribution and percent frequency distribution for the data set using a class width of 10. **(5 marks)**

Answer:

The lowest value is 52 and the highest is 99.

The smallest number is 52, and the largest is 99, so the range is 99-52=47.

The class width will be 47/5 = 9.4 approx. 10,

Our classes will be 52-61, 62-71, 72-81, 82-91, and 92-101.

Then, counting the number of entries in each class, we get:

Class	Frequency (f)
52-61	3
62-71	2
72-81	6

82-91	4
92-101	5

Note that the sum of the frequencies is 20, which is the same as number of data entries that we had.

The "midpoint" of each class can be calculated as:

Midpoint = (Lower class limit + Upper class limit)/2

The "relative frequency" of each class is the proportion of the data that falls in that class. It can be calculated for a data set of size n by:

Relative frequency = Class frequency Sample size = f / n.

The "cumulative relative frequency" is the sum of the relative frequencies of that class and all previous classes.

**Percentage frequency = relative frequency X 100 = (f / n)* 100**

Class	Frequency	Mid-Point	Relative frequency	Cumulative relative frequency	Percent Frequency
52-61	3	56.5	0.15	0.15	15
62-71	2	66.5	0.1	0.25	10
72-81	6	76.5	0.3	0.55	30
82-91	4	86.5	0.2	0.75	20
92-101	5	81.5	0.25	1	25
Total	20				

b.   Construct a histogram showing the percent frequency distribution of the examination scores. Comment on the shape of the distribution.

Answer: Using Excel

Bin(Upper class Limit)	Percent Frequency
61	15
71	10
81	30
91	20
101	25
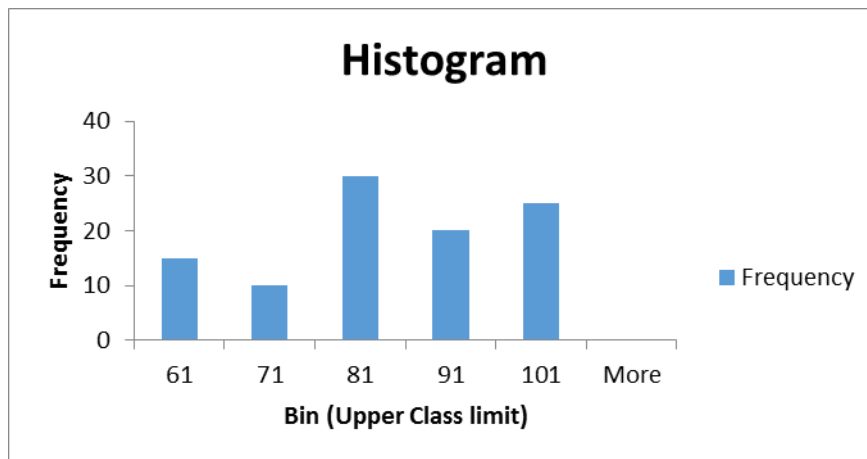
From the histogram, distribution looks negatively skewed.

Question 2

Shown below is a portion of a computer output for a regression analysis relating supply (Y in thousands of units) and unit price (X in thousands of dollars).

ANOVA

	df	SS
Regression	1	354.689
Residual	39	7035.262

	Coefficients	Standard Error
Intercept	54.076	2.358
X	0.029	0.021

a.   What has been the sample size for this problem? (1 mark)

Answer :

In Example

Total degrees of freedom,

$Df_T = df_R + df_E = 1 + 39 = 40$

But $df_T = n-1$

n=40+1=41

Therefore, sample size for the given analysis is 41

b.   Determine whether or not demand and unit price are related. Use α = 0.05. (2 marks)

Answer :

Null hypothesis: H_0: β=0

Alternative hypothesis: H_0: β0≠0

We are given level of significance α= 0.05

Decision Rule: Reject H_0 if P-value <= α= 0.05

Summarize the available information:

Slope b=0.029 ,SE_b=0.021 and n=41

To test the hypothesis, define the test statistic as follows:

$t = b-β/ES_b$

=0.029-0.00/0.021

=1.38

Degrees of freedom = n-2=41-2=39

We calculate p-value for two tailed test using following formula:

p-value = $2*P(t_{39}>=1.38)$

$\qquad$ =TDIST(1.38,39,2) (Using Excel function)

$\qquad$ =0.1754

Conclusion: P-value 0.1754 is greater than level of significance 0.05 so we do not reject null hypothesis, So we can conclude that supply and unit are not related.

c. $\qquad$ Compute the coefficient of determination and fully interpret its meaning. Be very specific. (2 marks)

Answer :

From given ANOVA we have

SSR = 354.689

SSE=7035.262

SST=SSR+SSE

$\qquad$ =354.689+7035.262

$\qquad$ =7389.951

Coefficient of determination $=R^2$

$\qquad$ $= SSR/SST$

$\qquad$ $= 354.689/7035.262$

$\qquad$ =0.04799

$\qquad$ =0.048

It Means that, 4.8 % of the sample variation in supply (y) can be explained by using the independent variable unit price.

The value of R^2 is close to 0, it indicates that the linear regression model is not accurate.

d. $\qquad$ Compute the coefficient of correlation and explain the relationship between supply and unit price. (2 marks)

Answer:

Coefficient of correlation: r=$\sqrt{R^2}$ = $\sqrt{0.048}$ =0.219

There is a weak correlation between supply and unit price, because correlation coefficient is close to 0.

e.   Predict the supply (in units) when the unit price is $50,000.

Answer :

The regression equation is,

   The Predicted supply (in units) when the unit price is $50,000 is

y=54.076+0.029*50

  =54.076+1.45

  =55.526

Question 3 :

Allied Corporation wants to increase the productivity of its line workers. Four different programs have been suggested to help increase productivity. Twenty employees, making up a sample, have been randomly assigned to one of the four programs and their output for a day's work has been recorded. You are given the results below.

Program A	Program B	Program C	Program D
150	150	185	175
130	120	220	150
120	135	190	120
180	160	180	130
145	110	175	175

(a)   Construct an ANOVA table.

Answer :

H0: All the four groups means are equal

H1: At least one group mean is not same

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	8750	3	2916.666667	6.140350877	0.005569864	3.238871517
Within Groups	7600	16	475			
Total	16350	19				

(b)   As the statistical consultant to Allied, what would you advise them? Use a .05 level of significance.

Answer :

Since P-value 0.0056 < 0.05 so we reject the null hypothesis, So we conclude that there are statistical differences between programs.

Question 4:

A company has recorded data on the weekly sales for its product (*y*), the unit price of the competitor's product (*x1*), and advertising expenditures (*x2*). The data resulting from a random sample of 7 weeks follows. Use Excel's Regression Tool to answer the following questions (data set also provided in accompanying MS Excel file).

Week	Price	Advertising	Sales
1	.33	5	20
2	.25	2	14
3	.44	7	22

4	.40	9	21
5	.35	4	16
6	.39	8	19
7	.29	9	15

a.   What is the estimated regression equation? Show the regression output. (2 marks)

Answer:

ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	45.35284	22.67642	6.716801	0.052643614			
Residual	4	13.5043	3.376075					
Total	6	58.85714						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.597615	4.052244	0.887808	0.424805	-7.65321747	14.84845	-7.65322	14.84845
Price	41.32002	13.33736	3.098065	0.036289	4.289567248	78.35048	4.289567	78.35048
Advertising	0.013242	0.327592	0.040422	0.969694	-0.89629843	0.922782	-0.8963	0.922782

Regression line is : Sales Y = 3.597 + 41.3*(Price) + 0.013*(Advertising)

b.   Determine whether the model is significant overall. Use α = 0.10. (2 marks)

Answer:

In above ANOVA output P-value 0.052 < 0.1 so we reject the null hypothesis we can conclude that model is significant overall.

c. Determine if competitor's price and advertising is individually significantly related to sales. Use $\alpha = 0.10$. (2 marks)

Answer:

The p value for price is 0.036. At 0.10 sig level, the variable is **significant since value < 0.10**

The p value for advertising is 0.969. At 0.10 sig level, the variable is **insignificant since value > 0.10**

d. Based on your answer to part (c), drop any insignificant independent variable(s) and re-estimate the model. What is the new estimated regression equation? (2 marks)

Answer:

In part c we get by p-value that price is significantly effecting the sales whereas advertising is not related with the sales.

So in data you should drop advertising covariate and run regression using price on sales by that we will get a regression line here.

Sales Y = 3.584 + 41.6*(Price)

ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	45.34733	45.34733	16.78311	0.009384894			
Residual	5	13.50981	2.701963					
Total	6	58.85714						

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	3.581788	3.608215	0.992676	0.366447	-5.6934245	12.857	-5.69342	12.857
Price	41.60305	10.15521	4.096719	0.009385	15.49824676	67.70786	15.49825	67.70786

e.   Interpret the slope coefficient(s) of the model from part

Answer:

The value of price changes one unit then the effect on sales is 41.6 times.