**MG315**

Chapter 12:

1. What two reasons do we use for the F-Statistic for? (10/200 points)

For testing the hypothesis of equality of two population variances.

For testing the equality of more than two means

1. What is the **critical** F value for a hypothesis test with a sample from 5 populations with 20 observations total. Use a .05 significance level. (10/200 points)

We have to see (N-k) and (k-1) degrees of freedom. This makes (20-5) and (5-1). From the table it is 3.87.

1. Given a calculated F value of 3.5, would you reject or fail to reject the null hypothesis with the critical F value from question 2? (5/200 points)

We will reject the null hypothesis as the calculated value is greater than the critical value.

1. Arbitron Media Research Inc. conducted a study of the iPod listening habits of men and women. One facet of the study involved the **mean** listening time. It was discovered that the mean listening time for men was 25 minutes per day. The standard deviation of the sample of the 13 men studied was 8 minutes per day. The mean listening time for the 12 women studied was also 25 minutes, but the standard deviation of the sample was 12 minutes. At the 0.05 significance level.

Given that the null hypothesis is the two variances are equal, do we reject the null hypothesis or fail to reject? (15/200)

We take v1 as the men and v2 as the women

Then df numerator are 13-1 = 12 and df denominator are 25-2 = 23.

The F calculated is s1/s2 that is 8/12= 0.67

The F critical value is 2.04. Thus we accept the null hypothesis of equality of variances.

Chapter 14:

1. Suppose we wanted to see the cost of gas a person pays per month. We use miles per gallon of their vehicle as the first independent variable, the average distance driven as the second independent variable, and cost of tires as the third independent variable.

Use the following regression output to build a multiple regression equation: (10/200 points)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficient | Standard Error | T | P-value |
| Constant | 84.998 | 1.863 | 45.61 | 0.000 |
| X1 | 2.391 | 1.2 | 1.99 | 0.051 |
| X2 | -0.4086 | 0.1717 | -2.38 | 0.02 |
| X3 | 1.153 | .789 | 1.54 | 0.1587 |

Cost of gasoline = 2.391\* miles per gallon + -0.4086\* average distance travelled + 1.153\* cost of tires

1. Which independent variable(s) are statistically significant? (5/200 points)

At 0.05 level of significance only X2 namely the average distance travelled is statistically significant.

1. Interpret the data in terms of the variables given the output. (10/200)

A unit change in miles per gallon will bring a 2.391 units change in the cost of gasoline. A unit change in average distance travelled will bring an opposite direction change in the cost of gasoline. A change in the cost of tires will bring a 1.153 units change in the cost of gasoline.

1. The director of marketing at Reeves Wholesale Products is studying monthly sales. Three independent variables were selected as estimators of sales: regional population, per capita income, and regional unemployment rate.

Yhat = 64,100 + 0.394 *X*1 + 9.6*X*2 – 11,600*X*3

Note: Here, the variables *X*1, *X*2, and *X*3 refer to regional population, per capita income, and regional unemployment rate respectively.

What does the 64,00 mean in terms of the equation (aka what is it called?) (5/200 points)

 This is the intercept of this equation showing the value of sales when the values of all independent variables is zero.

1. What are the estimated monthly sales for a particular region with a population of 796,000, per capita income of $6,940 and an unemployment rate of 6.0%? (15/200 points)

= 796000\*0.394+6940\*9.6-11600\*0.06

= 313624+ 380248-696

= $ 693176

|  |
| --- |
|  |

Chapter 1:

1. Describe the difference between inferential and descriptive statistics. (10/200 points)

Descriptive statistics provide the numerical information about the data in the form of mean, median or mode. The inferential statistics help to make inferences about the data and to test these inferences.

1. Why don’t we just measure populations? Why do we use samples to infer about populations? (10/200 points)

Using populations will be very expensive and time consuming. Samples are groups taken from the populations that are assumed to have same characteristics as the population.

Chapter 2:

Ten people were asked how many siblings they have. Below is the data:

2, 4, 1, 2, 1, 3, 5, 0, 1, 3, 0

1. Create a frequency distribution table. (20/200 points)

The data is discrete in nature and the range Is less than 10 so we will use the numbers to show the groups.

|  |  |
| --- | --- |
| Group | Frequency |
| 0 | 2 |
| 1 | 3 |
| 2 | 2 |
| 3 | 2 |
| 4 | 1 |
| 5 | 1 |

1. Add on a cumulative frequency column and compute the cumulative frequencies. (15/200 points)

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Frequency | Cumulative Frequency | Relative Frequency |
| 0 | 2 | 2 | 2/11 |
| 1 | 3 | 5 | 3/11 |
| 2 | 2 | 7 | 2/11 |
| 3 | 2 | 9 | 2/11 |
| 4 | 1 | 10 | 1/11 |
| 5 | 1 | 11 | 1/11 |

1. Add on a relative frequency column and compute the relative frequencies. (15/200 points)

Chapter 3:

1. A sample of retailers reported that they had the following number of copies of statistics textbooks in inventory:

57, 81, 69, 84, 85, 79, 71, 74, 55.

1. What is the mean number of textbooks in inventory? (10/200 points)

Mean is 72.78 calculated by dividing sum of values on number of values.

1. What is the median number of textbooks in inventory? (10/200 points)

Median is the middle value of arranged data which comes out to be 74.

1. What is the range of the number of textbooks in inventory? (5/200 points)

Range is calculated as the difference of highest and lowest value which is 85-55 = 30

1. The standard deviation is 10.98. What is the variance? (round to two decimals) (5/200 points)

Variance is the square of SD thus, 10.98^2 = 120.56

Chapter 4:

1. Create a Box and Whisker Plot using the following data. Be sure to give the 5 quartiles. (15/200 points)

12, 20, 15, 17, 32, 21, 45, 15



The lower boundary of the above box is the lower quartile of the data. The black line in the box is depicting median. The top boundary of the box shows the third quartile. The lines above and below the tables show the maximum and minimum values in the data.