**Session 3**

**Visualizations in R**

1. The Orlando Sentinel annually lists the top employers in central Florida. Following table is a frequency distribution of five companies with largest number of employees in this area.

|  |  |
| --- | --- |
| Company | Number of employees |
| Disney world | 51600 |
| Florida hospital | 19283 |
| Publix supermarket Inc. | 15325 |
| Wal-mart stores Inc. | 14995 |
| Universal Orlando | 12000 |

Obtain a bar plot with suitable titles

1. Following table gives a partial list of the number of endangered wild life species both inside and outside the US as of April 2004

|  |  |  |
| --- | --- | --- |
| Items | endangered wild life species | |
| US | Foreign countries |
| Mammals | 69 | 251 |
| Birds | 77 | 175 |
| Reptiles | 14 | 64 |
| Amphibians | 12 | 8 |
| Fishes | 71 | 11 |

1. Construct a bar chart of the number of endangered wild life species in the US.
2. Construct a bar chart of the number of endangered wild life species outside the US.
3. Construct a bar chart to compare the number of endangered wild life species in the US to the number of endangered wild life species outside the US.
4. Suppose that an estimate of US federal spending showed that 46% was for entitlement, 18% was for defense, 15% was grants to states and localities, 14% was for interest on debt, 6% was for other federal operations and 1% was for deposit insurance. Construct a pie chart to show this information.
5. Make a histogram with A repeated B times.

A = (3,5,7,9,11,13,15,17)

B = (4,8,15,30,35,28,14,4)

**Session 5**

**Statistical Measures**

1. In an assembling plant 200 people work, 115 of them male and 85 female. When examining the productivity of the plant, the following data were revealed about the performance of the males in a given month. Compute the measures of central tendency.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Number of assembled pieces | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 |
| Number of workers | 2 | 8 | 10 | 14 | 16 | 25 | 14 | 12 | 11 | 3 |

1. Calculate the measures of central tendency for the following data.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Wages (in Rs) | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 |
| No. of employees | 8 | 13 | 16 | 21 | 17 | 12 | 8 | 5 |

**Session 8**

**Testing of Hypothesis**

1. It has been reported that the average credit card debt for college seniors is $3262. The student senate at a large university feels that their seniors have a debt much more than this, so it conducts a study of 50 randomly selected seniors and finds that the average debt is $2995, and the population standard deviation is $1100. Test at the 1% level of significance.
2. We wish to check that whether normal body temperature is more than 98.6 degrees. In a random sample of n =18 individuals, the sample mean was found to be 98.217 and the standard deviation was 0 .684. Assume the population is normally distributed. Use α= 0.05.
3. An educator estimates that the dropout rate for seniors at high schools in Colorado is 12%. Last year in a random sample of 300 Colorado seniors, withdrew from school. At α = 0.05, is there enough evidence to reject the educator’s claim?
4. In 1980, of 750 men 20-34 years old, 130 were found to be overweight. Whereas in 1990, of 700 men, 20-34 years old, 160 were found to be overweight. At 5% significance level, do the data provide sufficient evidence to conclude that for men 20-34 years old, a higher percentage were overweight in 1990 than 10 years earlier?
5. In a packing plant, a machine packs cartons with jars. It is supposed that a new machine will pack faster on the average than the machine currently used. To test that hypothesis, the times it takes each machine to pack ten cartons are recorded. The results (in seconds), are shown in following table

|  |  |
| --- | --- |
| New Machine | 42.1, 41.3, 42.4, 43.2, 41.8, 41.0, 41.8, 42.8, 42.3, 42.7 |
| Old Machine | 42.7, 43.8, 42.5, 43.1, 44.0, 43.3, 43.6, 43.5, 41.7, 44.1 |

1. Trace metals in drinking water affect the flavor and unusually high concentration can pose a health hazard. Ten pairs of data were taken measuring zinc concentration in bottom water and surface water. Does the data suggest that the true average concentration in bottom water exceeds that of surface water?

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Location | | | | | | | | | | |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Zinc concentration in bottom water | 0.43 | 0.266 | 0.567 | 0.531 | 0.707 | 0.716 | 0.651 | 0.589 | 0.469 | 0.723 |
|
|
| Zinc concentration in surface water | 0.415 | 0.238 | 0.39 | 0.41 | 0.605 | 0.609 | 0.632 | 0.523 | 0.411 | 0.612 |

**Session 9**

**Analysis of Variance**

1. A Project manager is interested to test the difference between the process completion time under three different methods. The data is given below. Perform one way Analysis of variance

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Method A | 9.3 | 9.4 | 9.6 | 10 | 12.4 | 13 | 10.4 | 11.1 | 12.2 | 13.5 |
| Method B | 12.2 | 11.4 | 13.2 | 14.4 | 15.4 | 13.4 | 14.2 | 10.5 | 10.8 | 12.4 |
| Method C | 10.2 | 8.7 | 9.7 | 12.1 | 11.4 | 12.4 | 11.8 | 13.4 | 14.5 | 15.4 |

1. A psychologist was interested in whether different TV shows lead to a more positive outlook on life. People were split into 4 groups and then taken to a room to view a program. The four groups saw: The Muppet Show, Futurama, The News, No program. After the program a blood sample was taken and serotonin levels measured (Remember more serotonin means happier!)

|  |  |
| --- | --- |
| The Muppet show | 11, 7, 8, 14, 11, 10, 5 |
| Futurama | 4, 8, 6, 11, 9, 8 |
| The News | 4, 3, 2, 2, 3, 6 |
| No program | 7, 7, 5, 4, 3, 4, 4, 4 |

Carry out a one-way ANOVA to test the hypothesis that some TV shows make people happier than others.

1. In a medical school a new method of teaching in which professional actors played the roles of patients was introduced. The test scores of male and female students who were taught by either the conventional method or by a new form of training using role-play are shown in the table. Find FA, FB.Is there any difference in the mean test score under (i) Gender (ii) Teaching Method, using a two-way ANOVA at α = 0.05. [A : Gender and B : Teaching Method ]

|  |  |  |
| --- | --- | --- |
| Gender | Teaching Method | |
| Conventional | Role play |
| Male | 64 | 38 |
| 75 | 53 |
| 60 | 33 |
| 69 | 46 |
| 42 | 38 |
| Female | 42 | 25 |
| 55 | 23 |
| 50 | 32 |
| 56 | 28 |
| 51 | 43 |

1. Five different methods of analysis are used to determine in parts per million the amount of a certain constituent in a sample. Three analysts use each method two times and the results are given below.

Is there any significance effect of different methods used to determine in parts per million.

Is there any difference in the determination of parts per million depending on the different analyst?

|  |  |  |  |
| --- | --- | --- | --- |
| Method | Analyst | | |
| A | B | C |
| I | 7.1 | 6.4 | 6.6 |
| 6.9 | 6.6 | 6.6 |
| II | 6.8 | 6.8 | 6.2 |
| 7 | 6.6 | 6.2 |
| III | 6.7 | 6.6 | 6.5 |
| 6.9 | 6.4 | 6.3 |
| IV | 7.2 | 6.7 | 6.2 |
| 7 | 6.7 | 6.4 |
| V | 6.9 | 6.5 | 6.5 |
| 6.9 | 6.7 | 6.3 |