**Tutorial 2 Solutions**

**Statistical Analysis 1, CBA**

1. Following table contains data on water absorbency of cotton and Acetate fibers (in %). Based on this data, can one conclude that the difference in mean water absorbency of cotton fiber and acetate fiber is statistically significant at 5%? Assume the population variances are equal.

|  |  |  |  |
| --- | --- | --- | --- |
| Fiber | Sample Size, n | Sample mean | Sample Standard deviation |
| Cotton | 28 | 19.93 | 1.51 |
| Acetate | 25 | 12.07 | 1.25 |

**Solution:**

Let denote the mean absorbency (in %) for cotton and denote the mean absorbency (in %) for Acetate.

Our Null and Alternative hypothesis are as follows:

We will use a two sided test. Since the population variances are assumed equal, we use pooled standard deviation:

On calculation we get .

The degrees of freedom are

Our test statistic is . The strength of the evidence, P-value = \*pt(-20.51,51) < 0.05. We reject the null hypothesis and conclude that there is statistically significant difference in the absorbency of the two fibres.

1. Twelve cars were equipped with radial tires and driven over a test course. Then the same 12 cars (with the same drivers) were equipped with regular belted tires and driven over the same course. After each run, the cars’ gas economy (in km/l) was measured. Is there evidence that radial tires produce better fuel economy? (Use α= 0.05.)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gas Economy|Cars | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Y1 Radial | 4.2 | 4.7 | 6.6 | 7.0 | 6.7 | 4.5 | 5.7 | 6.0 | 7.4 | 4.9 | 6.1 | 5.2 |
| Y2 Belted | 4.1 | 4.9 | 6.2 | 6.9 | 6.8 | 4.4 | 5.7 | 5.8 | 6.9 | 4.7 | 6.0 | 4.9 |

**Solution:**

Let denote the mean fuel economy of radial tires and denote the mean fuel economy of regular belted tired. Then, our hypotheses are:

Let *D* denote the difference between the gas economies for a randomly chosen car. The table below includes the differences values.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gas Economy|Cars | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Y1 Radial | 4.2 | 4.7 | 6.6 | 7.0 | 6.7 | 4.5 | 5.7 | 6.0 | 7.4 | 4.9 | 6.1 | 5.2 |
| Y2 Belted | 4.1 | 4.9 | 6.2 | 6.9 | 6.8 | 4.4 | 5.7 | 5.8 | 6.9 | 4.7 | 6.0 | 4.9 |
| Differences *di* | 0.1 | -0.2 | 0.4 | 0.1 | -0.1 | 0.1 | 0 | 0.2 | 0.5 | 0.2 | 0.1 | 0.3 |

**T**est statistic **=**

We will use this to calculate the p-value =

Hence the evidence is strong enough to conclude that radial tires produce better fuel economy.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Machine 1** | **Machine 2** | **Machine 3** |
| **Observations** | 25 | 31 | 24 |
| 30 | 39 | 30 |
| 36 | 38 | 28 |
| 38 | 42 | 25 |
| 31 | 35 | 28 |

1. A manufacturing company has purchased three new machines of different makes and wishes to determine whether one of them is faster than the others in producing a certain output. Five hourly production figures are observed at random from each machine and the results are given in the table below. Determine whether the machines are significantly different in their average speeds.

Use 5% level of significance.

**Solution:**  Let be the mean speed of the machine i.

and

Now, let us solve this in R using the *aov* command. The output is given below:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| summary(model2) | | | | | | |
|  | Df | Sum of Squares | Mean Sum of Squares | F value | Pr(>F) |  |
| factor(Machine) | 2 | 250 | 125 | 7.5 | 0.00771 | \*\* |
| Residuals | 12 | 200 | 16.67 |  |  |  |
| --- |  |  |  |  |  |  |
| Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 | | | | | | |

From the table, p-value is 0.00771 < 0.01. Therefore we can reject the null hypothesis at 5% level of significance and conclude that the output speeds of the three machines differ significantly. To determine which of the machines is faster than the other, we perform the **Tukey’s test for pairwise comparison**. Below is the output table in R.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| TukeyHSD(model2) | | | | |
| Tukey multiple comparisons of means | | | | |
| 95% family-wise confidence level | | | | |
|  |  |  |  |  |
| Fit: aov(formula = Output ~ factor(Machine)) | | | | |
|  |  |  |  |  |
|  | diff | lwr | upr | p adj |
| Machine 2 - Machine 1 | 5 | -1.88839 | 11.88839 | 0.17095 |
| Machine 3 - Machine 1 | -5 | -11.8884 | 1.888394 | 0.17095 |
| Machine 3 - Machine 2 | -10 | -16.8884 | -3.11161 | 0.005803 |

We see that **only Machine 3 – Machine 2** has significant difference and the difference value is negative. Therefore we conclude at 5% level of significance; that Machine 2 gives an output significantly faster than Machine 3.

1. The following table shows the lives (in hours) of four batches of electric lamps. Perform an Analysis of Variance on this data to test for whether the significance test can reject their homogeneity. Test at 5% level of significance.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Batch #** | **Life of bulbs in hours** | | | | | | | |  |
| **1** | 1600 | 1610 | 1650 | 1680 | 1700 | 1720 | 1800 | 1860 | **1702.5** |
| **2** | 1580 | 1640 | 1640 | 1700 | 1750 | 1820 | 1830 | 1830 | **1723.75** |
| **3** | 1460 | 1550 | 1600 | 1620 | 1640 | 1660 | 1740 | 1820 | **1636.25** |
| **4** | 1510 | 1520 | 1530 | 1570 | 1600 | 1680 | 1700 | 1720 | **1603.75** |
|  |  |  |  |  |  |  |  |  | **1666.5625** |

**Solution:** Let be the mean life of bulbs from batch i.

and



Totij

ti

eij

xij

Sum of Squares Total (SST) =

Sum of Squares Treatment (SSTR) =

Sum of Squares Error (SSE) =

In the problem, r=4, n=32.

Mean Square Treatment = MSTR = SSTR / (r-1) = 25136.46

Mean Square Error = MSE = SSE / (n-r) = 9318.304

F-ration = MSTR/MSE = 2.6975. Strength of this evidence, p-value = Pr(F(r-1,n-r) ≥ f) = *1-pf(2.6975,3,28)* = 0.0649

Let us our down this information in form of an ANOVA table.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Degrees of Freedom** | **Sum of Squares** | **Mean Sum of Squares** | **F value** | **Pr(>F)** |
| Treatment (Batches) | 3 | 75409.38 | 25136.46 | 2.6975 | 0.0649 |
| Residuals (Error) | 28 | 260913 | 9318.304 |  |  |

We cannot reject the null hypothesis at 5% level of significance. Hence the evidence is not strong enough to conclude that the life (in hours) of bulbs from the given 4 batches is significantly different.

**Note: the same numbers can be verified using R.**

1. A magazine reported the results of a telephone poll of 800 adult Indians; 600 of them non-smokers. They were asked the following question: ‘Should the tax on Cigarettes be raised by 1.25%, to pay for the healthcare reform?’ There are 200 non-smokers and 50 smokers answered yes to this question. Which of the following is APPROXIMATE strength of the evidence to conclude that smokers and non-smokers are different in their opinion about the tax increase?
   1. **0.03**
   2. 0.06
   3. 0.97
   4. 0.94

Solution: (A): H0: π1= π2  and HA: π1≠ π2

n1=600 and n2=200. Let be the proportion of non-smokers who said yes to the question and be the proportion of smokers who said yes to the question. Then, p1=200/600 =0.33 and p2 = 50/200 = 0.25.Pooled proportion, = (0.33\*600 + 0.25\*200)/800 = 0.31.

Test statistic, p-value =

1. A manufacturer uses two different trucking companies to ship its merchandise. The manufacturer suspects that one company is charging more than the other and wants to test it. A random sample of the amounts charged for one truckload shipment from Chicago to Detroit on various days is collected for each trucking company. The data (in dollars) is given below.

Company 1: 2570, 2480, 2870, 2975, 2660, 2380, 2590, 2550, 2485, 2585, 2710

Company 2: 2055, 2940, 2850, 2475, 1940, 2100, 2655, 1950, 2115

Assume that the variation in the prices charged by the two companies is not the same.  Then, the strength of the evidence to test if the two companies charge the same price on an average is

1. **0.069**
2. 0.139
3. 0.049
4. 0.097

Solution: (A) Let be the average amount charged by company i.

From the evidence, 2623.18, . Test statistic, . Since the population variances are different, the degrees of freedom is calculated using the formula,



which is equal to 10.55. Taking df = 11, the p-value for a two-sided test is close to 0.0722. Hence, the closest option is A.