**Hypothesis Testing Assignment**

***Problem 1: A F&B manager wants to determine whether there is any significant difference in the diameter of the cutlet between two units. A randomly selected sample of cutlets was collected from both units and measured? Analyze the data and draw inferences at 5% significance level. Please state the assumptions and tests that you carried out to check validity of the assumptions.***

1. Business problem: To check difference in the diameter of the cutlet between 2 units.
2. Data Collection: 35 samples of diameter for both units

Y = Difference in diameter = Continuous

X = 2 samples (Diameter 1 (Unit A) and Diameter 2 (Unit B) ) = Discrete

1. Follow Flow Chart for Test
2. Normality Test:

Ho: No action, if Unit A & Unit B are normal

Ha: Take action, if Unit A ***or*** Unit B are not normal

*Unit A*

*data: Cutlets$`Unit A`*

*A = 0.43309, p-value = 0.2866*

*Unit B*

*data: Cutlets$`Unit B`*

*A = 0.26123, p-value = 0.6869*

Both data are assume to be normal as p-value for both is high, p-high -> Null fly

As p-high -> Null fly, hence we Fail to reject Null Hypothesis. Data is normal

\*\*Check for the flow chart

1. External conditions: Considering external are different as samples from both units are selected randomly, we will go with Variance.
2. \*\*Check for the flow chart – Check Variance

Hypothesis statement for Variances

Ho -> Variances are equal (Var of Unit A = Var of Unit B)

Ha -> Variances are not equal (Var of Unit A != Var of Unit B)

*F test to compare two variances*

*data: Unit A and Unit B*

*F = 0.70536, num df = 34, denom df = 34, p-value = 0.3136*

*From Variance test in R, p value is 0.3136*

p-value > 0.05 i.e. 0.3136 > 0.05, P High Ho fly. We fail to reject Null hypothesis. So we will accept it as Variances of A is equal to Variances of B.

\*\*Check for the flow chart

1. 2 sample T test for equal variances:

Ho -> Avg of diameters of Unit A is equal to Avg of diameters of unit B

Ha -> Avg of diameters of Unit A is not equal to Avg of diameters of unit B

*Welch Two Sample t-test*

*data: Unit A and Unit B*

*t = 0.72287, df = 66.029, p-value = 0.4723*

*alternative hypothesis: true difference in means is not equal to 0*

*95 percent confidence interval:*

*-0.09654633 0.20613490*

*sample estimates:*

*mean of x mean of y*

*7.019091 6.964297*

P-value>0.05 (i.e. 0.4723) and hence P High and Ho Fly. We fail to reject null hypothesis i.e. **Avg of diameters of Unit A is equal to Avg of diameters of unit B**

***Conclusion:*** there is no significant difference in the cutlet diameters of Unit A and Unit B

***Problem 2: A hospital wants to determine whether there is any difference in the average Turn Around Time (TAT) of reports of the laboratories on their preferred list. They collected a random sample and recorded TAT for reports of 4 laboratories. TAT is defined as sample collected to report dispatch.***

***Analyze the data and determine whether there is any difference in average TAT among the different laboratories at 5% significance level.***

1. Business problem: To check if there is any difference in average Turn around time for preparing reports among laboratories
2. Data Collection: Turn Around Time of Lab 1, Lab 2, Lab 3 and Lab 4

Y => Continuous and X => 4 Samples Discrete

1. Compare 4 population with each other
   1. Normality Test for all 4 Laboratories

**Hypothesis statement for Normality test**

Ho: No action, If data for Lab 1, Lab 2, Lab 3 and Lab 4 are normal

Ha: Take action, if data for Lab 1, Lab 2, Lab 3 and Lab 4 are not normal Hypothesis for Lab 1

Ho= Data is normally distributed

Ha=Data is not normally distributed

P-value is >0.05 (i.e. 0.5322>0.05) P High null Fly. Hence we fail to reject Null (H0) hypothesis.

Hypothesis for Lab 2

Ho= Data is normally distributed

Ha=Data is not normally distributed

P-value is >0.05 (i.e. 0.7331>0.05) P High null Fly. Hence we fail to reject Null (H0) hypothesis.

Hypothesis for Lab 3

Ho= Data is normally distributed

Ha=Data is not normally distributed

P-value is >0.05 (i.e. 0.5768>0.05) P High null Fly. Hence we fail to reject Null (H0) hypothesis.

Hypothesis for Lab 4

Ho= Data is normally distributed

Ha=Data is not normally distributed

P-value is >0.05 (i.e. 0.4194>0.05) P High null Fly. Hence we fail to reject Null (H0) hypothesis.

We fail to reject null hypothesis of Laboratories normality test.

* 1. \*\*Check for the flow chart – Check Variance

Hypothesis statement for Variances

Ho -> Variances are equal (Var of Lab 1 = Var of Lab 2 = Var of Lab 3 = Var of Lab 4)

Ho -> Variances are not equal (Var of Lab 1 != Var of Lab 2 != Var of Lab 3 != Var of Lab 4)

***Stacked the data to conduct one test to check the variances of all Laboratories data in one go***

*Levene's Test for Homogeneity of Variance (center = median)*

*Df F value Pr(>F)*

*group 3 2.5996 0.05161 .*

*476*

p-value > 0.05 i.e. 0.05161 > 0.05, P High Ho fly. We fail to reject Null hypothesis. So we will accept it as Variances of Laboratories are equal

* 1. As all Variances are equal, from Flow chart we will go for One Way ANOVA test

Ho= Avg TAT for all the Lab samples is same

Ha= Avg TAT for all the Lab samples is not same

From Rcode:

***summary.aov(Anova\_Results)***

***Df Sum Sq Mean Sq F value Pr(>F)***

***ind 3 79979 26660 118.7 <2e-16 \*\*\****

***Residuals 476 106905 225***

As p <0.05, hence we will reject the null hypothesis, i.e. **there is significant difference in the average TAT for all the labs.**

***Problem 3: Sales of products in four different regions is tabulated for males and females. Find if male-female buyer rations are similar across regions***.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **East** | **West** | **North** | **South** |
| Males | 50 | 142 | 131 | 70 |
| Females | 550 | 351 | 480 | 350 |

1. Business problem: To check if Male and Female buyer ratio is same across region
2. Data Collection:

Y => Discrete, as we are trying to find the proportion across regions for male and female and X => 4 Samples Discrete

1. **Hypothesis statements**

Ho= Proportions of Male and Female are same

Ha= Proportions of Male and Female are not same

**As we have 4 samples as input we will go for Chi Square test**

**Per Chi square Test:**

**Pearson's Chi-squared test**

**data: BuyerRatio**

**X-squared = 11.91, df = 4, p-value = 0.01804**

As P value i.e 0.02<0.05, hence we fail to accept Null hypothesis

***Conclusion: Hence proportion of male and female across regions is not same.***

***Problem 4: TeleCall uses 4 centers around the globe to process customer order forms. They audit a certain % of the customer order forms. Any error in order form renders it defective and has to be reworked before processing. The manager wants to check whether the defective % varies by centre. Please analyze the data at 5% significance level and help the manager draw appropriate inferences***

1. Business Problem: Identify which country required training centres
2. Data Collection:

Y => Discrete, as we are trying to find the proportion across countries and X => 4 Samples Discrete

1. **Hypothesis statements**

Ho: No training requires, if proportions of defective is equal in all countries

Ha: Training requires, if proportions of defective is different in at least one country

**As we have 4 samples as input we will go for Chi Square test**

Rcode output:

Pearson's Chi-squared test

data: table(values, ind)

X-squared = 3.859, df = 3, p-value = 0.2771

As P value i.e 0.2771>0.05, hence we fail to reject Null hypothesis

***Conclusion: Based on above we assume that Defective % across countries is same , hence no need for training center.***

***Problem 5: Fantaloons Sales managers commented that % of males versus females walking in to the store differ based on day of the week. Analyze the data and determine whether there is evidence at 5 % significance level to support this hypothesis.***

1. Business problem: To check if % of males versus females walking in to the store differ based on day of the week.
2. Data Collection:

Y => Discrete and X => 2 Samples Discrete

1. Hypothesis statement

Case 1:

Ho= % of Male and Female are same

Ha= % of Male and Female are not same

2-sample test for equality of proportions with continuity correction

R Code:

data: c(113, 167) out of c(400, 400)

X-squared = 15.434, df = 1, p-value = 8.543e-05

alternative hypothesis: two.sided

5 percent confidence interval:

-0.1395936 -0.1304064

sample estimates:

prop 1 prop 2

0.2825 0.4175

As P value < 0.05 🡺 P low Null go, we will fail to accept Null hypothesis. Hence proportions of Male and Female are not same

Case 2: find out whose proportion is higher. We create another hypothesis

Ho= Proportions of Male is less than or equal to Female

Ha= Proportions of Male is greater than Female

**RCode output**

2-sample test for equality of proportions with continuity correction

data: c(113, 167) out of c(400, 400)

X-squared = 15.434, df = 1, p-value = 1

alternative hypothesis: greater

5 percent confidence interval:

-0.08258261 1.00000000

sample estimates:

prop 1 prop 2

0.2825 0.4175

As P value > 0.05 🡺 P low Null go, we will fail to reject Null hypothesis. Hence proportions of Male is higher than Female.