**Q1.**

**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.**

Ans-

Business probem: whether there is any significant difference in the diameter of the cutlet between two units.

Ho: there is no difference.

Ha: there is difference.

|  |
| --- |
| fb <- read.csv(file.choose())  > View(fb)  > #normality check  > library(nortest)  > ad.test(fb$Unit.A)  Anderson-Darling normality test  data: fb$Unit.A  A = 0.43309, p-value = 0.2866  > ad.test(fb$Unit.B)  Anderson-Darling normality test  data: fb$Unit.B  A = 0.26123, p-value = 0.6869 |
|  |
| |  | | --- | |  | |

#both unit.A and unit.B are normal. So now we will check the external condition.

Here, external conditions are not same. So, we will check for variance

|  |
| --- |
| #checking variance  > library(car)  > var.test(fb$Unit.A,fb$Unit.B)  F test to compare two variances  data: fb$Unit.A and fb$Unit.B  F = 0.70536, num df = 34, denom df = 34, p-value = 0.3136  alternative hypothesis: true ratio of variances is not equal to 1  95 percent confidence interval:  0.3560436 1.3974120  sample estimates:  ratio of variances  0.7053649 |
| Here, p-value>0.05 so variance is equal. We will go for 2 sample t-test  #2 sample t test  > t.test(fb$Unit.A,fb$Unit.B,alternative = "two.sided",conf.level = 0.95,correct = TRUE)  Welch Two Sample t-test  data: fb$Unit.A and fb$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  Here, p-value>0.05. so, p high null fry. We will go with Ho.  Conclusion:  there is no difference in the diameter of the cutlet between two units |
| |  | | --- | |  | |

**Q2.**

**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.**

Ans-

Business problem: To determine if there is any differencein average TAT among the different laboratories

Ho: no action if normal

Ha: take action if not normal

|  |
| --- |
| lab <- read.csv(file.choose())  > View(lab)  > library(nortest)  > stack\_lab <- stack(lab)  > View(stack\_lab)  > attach(stack\_lab)  The following objects are masked from stack\_sales:  ind, values  The following objects are masked from stack\_data (pos = 5):  ind, values  The following objects are masked from stack\_data (pos = 6):  ind, values  #normality test  > ad.test(values)  Anderson-Darling normality test  data: values  A = 0.7495, p-value = 0.05072  Normality is equal since p-value>0.05  > #checking variance  > library(car)  Loading required package: carData  Attaching package: ‘car’  The following object is masked from ‘package:dplyr’:  recode  > leveneTest(stack\_lab$values~stack\_lab$ind, data = stack\_lab)  Levene's Test for Homogeneity of Variance (center = median)  Df F value Pr(>F)  group 3 2.5996 0.05161 .  476  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Variance is equal since p-value>0.05  > #one way annova test  > Anova\_results <- aov(values~ind,data = stack\_lab)  > summary(Anova\_results)  Df Sum Sq Mean Sq F value Pr(>F)  ind 3 79979 26660 118.7 <2e-16 \*\*\*  Residuals 476 106905 225  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1 |
|  |
| |  | | --- | |  | |

Here, p value < 0.05 so, p low null go. We will go for Ha.

Conclusion,

There is difference in the average so action is required.

**Q3.**

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

Ans-

Business Problem - to find the buyers ratios similar or not

Ho: all proportions are equal

Ha: all proportions are not equal

|  |
| --- |
| sales <- read.csv(file.choose())  > View(sales)  > View(sales)  > stack\_sales <- stack(sales)  Warning message:  In stack.data.frame(sales) : non-vector columns will be ignored  > View(stack\_sales)  > attach(stack\_sales)  The following objects are masked from stack\_data (pos = 3):  ind, values  The following objects are masked from stack\_data (pos = 4):  ind, values  > table(values,ind)  ind  values East West North South  50 1 0 0 0  70 0 0 0 1  131 0 0 1 0  142 0 1 0 0  435 1 0 0 0  750 0 0 0 1  1356 0 0 1 0  1523 0 1 0 0  > chisq.test(table(values,ind))  Pearson's Chi-squared test  data: table(values, ind)  X-squared = 24, df = 21, p-value = 0.2931  Here, p value> 0.05 so p high null fry  We will go with null hypothesis  Conclusion:    No action Is required as all proportions are equal. |
|  |
| |  | | --- | |  | |

**Q4.**

**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**

Ans-

Since, here the Y is discrete and there are more than 2 populations so we will go for chi-square test.

Business problem: whether to rework or not

Ho: no rework required.

Ha: rework required

|  |
| --- |
| customer <- read.csv(file.choose())  > str(customer)  'data.frame': 300 obs. of 4 variables:  $ Phillippines: Factor w/ 2 levels "Defective","Error Free": 2 2 2 2 2 2 2 2 2 2 ...  $ Indonesia : Factor w/ 2 levels "Defective","Error Free": 2 2 1 2 2 2 1 2 2 2 ...  $ Malta : Factor w/ 2 levels "Defective","Error Free": 1 2 1 2 1 2 2 2 2 2 ...  $ India : Factor w/ 2 levels "Defective","Error Free": 2 1 2 2 2 2 2 2 2 2 ...  > View(customer)  > #converting the factors to numeric  > customer$Malta <- as.numeric(revalue(customer$Malta, c("Error free"="2","Defective"="1")))  The following `from` values were not present in `x`: Error free  > customer$Phillippines <- as.numeric(revalue(customer$Phillippines, c("Error free"="2","Defective"="1")))  The following `from` values were not present in `x`: Error free  > customer$Indonesia <- as.numeric(revalue(customer$Indonesia, c("Error free"="2","Defective"="1")))  The following `from` values were not present in `x`: Error free  > customer$India <- as.numeric(revalue(customer$India, c("Error free"="2","Defective"="1")))  The following `from` values were not present in `x`: Error free  > #stacking the columns  > stack\_data <- stack(customer)  > View(stack\_data)  > View(stack\_data)  > attach(stack\_data)  The following objects are masked from stack\_data (pos = 3):  ind, values  > table(values,ind)  ind  values Phillippines Indonesia Malta India  1 29 33 31 20  2 271 267 269 280  > chisq.test(table(values,ind))  Pearson's Chi-squared test  data: table(values, ind)  X-squared = 3.859, df = 3, p-value = 0.2771 |
|  |
| |  | | --- | |  | |

Here, p> 0.05 i.e p high null fry.

Conclusion:

So, we will go with null hypothesis. i.e no action required.

**Q5.**

**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**

Ans-

Here Y is discrete with 2 populations so we will go for 2 proportion test.

Business problem: male and female differs or not

Ho: male female not differs.

Ha: male female differs.

|  |
| --- |
| fantaloons <- read.csv(file.choose())  > str(fantaloons)  'data.frame': 400 obs. of 2 variables:  $ Weekdays: Factor w/ 2 levels "Female","Male": 2 1 1 2 1 1 1 1 1 1 ...  $ Weekend : Factor w/ 2 levels "Female","Male": 1 2 2 1 1 2 1 2 1 2 ...  > #convert the factors into numeric  > fantaloons$Weekdays <- as.numeric(revalue(fantaloons$Weekdays,c("female"="2","male"="1")))  The following `from` values were not present in `x`: female, male  > View(fantaloons)  > fantaloons$Weekend <- as.numeric(revalue(fantaloons$Weekend,c("female"="2","male"="1")))  The following `from` values were not present in `x`: female, male  > table(fantaloons$Weekdays,fantaloons$Weekend)    1 2  1 167 120  2 66 47  > #2 proportion test  > prop.test(x=c(66,47),n=c(233,167),conf.level = 0.95,correct = FALSE,alternative = "two.sided")  2-sample test for equality of proportions without continuity correction  data: c(66, 47) out of c(233, 167)  X-squared = 0.0015979, df = 1, p-value = 0.9681  alternative hypothesis: two.sided  95 percent confidence interval:  -0.08761305 0.09126240  sample estimates:  prop 1 prop 2  0.2832618 0.2814371 |
|  |
| |  | | --- | | > | |

Here, p value>0.05

That means, p high null fry

Conclusion:

So we will go with null hypothesis and there is no action required.