INDEX

Sr. No.	Practical Name	Date	Signature
1.	Import the data from excel to R and perform basic inbuild command in R		
2.	Write a program in R for finding Mean, variance, standard deviation, covariance.		
3.	Write a program on Regression analysis in R		
4.	Prepare Google Questionnaire		
5	Prepared a Traditional Questionnaire		
6.	Small Sample Test (T-test) in R		
7.	Large Sample Test (Z test) in R		
8.	ANNOVA		
9.	Chi-Square Test of Association		
10.	Primary data Analysis		

Practical Number 1

Import the data from excel to R and perform basic inbuild command in R?

Code:

```
> data<-read.csv(file.choose())
> data
   marks
1     12
2     18
3     25
4     42
5     29
6     35
7     21
8     28
9     26
10    24
11    18
12    10
13    16
14    12
15    19
> marks<-data$marks
> marks
 [1] 12 18 25 42 29 35 21 28 26 24 18 10 16 12 19
> min(marks)
[1] 10
> max(marks)
[1] 42
> length(marks)
[1] 15
> mean(marks)
[1] 22.33333

> sd(marks)
[1] 8.877473

> var(marks)
[1] 78.80952
> summary(marks)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  10.00   17.00   21.00   22.33   27.00   42.00
> sqrt(marks)
 [1] 3.464102 4.242641 5.000000 6.480741 5.385165 5.916080 4.582576 5.291503
 [9] 5.099020 4.898979 4.242641 3.162278 4.000000 3.464102 4.358899
```

```
> round(marks)
 [1] 12 18 25 42 29 35 21 28 26 24 18 10 16 12 19
> round (sqrt(marks),2)
 [1] 3.46 4.24 5.00 6.48 5.39 5.92 4.58 5.29 5.10 4.90 4.24 3.16 4.00 3.46 4.36
> s=sqrt(marks)
> round(s,3)
 [1] 3.464 4.243 5.000 6.481 5.385 5.916 4.583 5.292 5.099 4.899 4.243 3.162
[13] 4.000 3.464 4.359
> round(s,0)
 [1] 3 4 5 6 5 6 5 5 5 5 4 3 4 3 4
> round(s,5)
 [1] 3.46410 4.24264 5.00000 6.48074 5.38516 5.91608 4.58258 5.29150 5.09902
[10] 4.89898 4.24264 3.16228 4.00000 3.46410 4.35890
> s=sum(marks)
> s
[1] 335
> cat("sum of marks=",s)
sum of marks= 335>
```

Practical Number 2

Write a program in R for finding Mean, variance, standard deviation, covariance.

Code:

```
> x=c(15,18,35,68,45,45,11)
> mean=mean(x)
> mean
[1] 33.85714
> varients=var(x)
> varients
[1] 424.1429
> sd(x)
[1] 20.59473
> sd=sd(x)
> sd
[1] 20.59473
> cv=sd/mean*100
> cv
[1] 60.82831
> x=c(11,12,13,14,15,16,17)
> f=c(5,7,12,15,10,3,2)
> y=rep(x,f)
> y
 [1] 11 11 11 11 11 12 12 12 12 12 12 12 13 13 13 13 13 13 13 13 13 13 13 13 14
[26] 14 14 14 14 14 14 14 14 14 14 14 14 14 14 15 15 15 15 15 15 15 15 15 15 16
[51] 16 16 17 17
> mean=mean(y)
> mean
[1] 13.64815
> var=var(y)
> var
[1] 2.232355
> sd=sd(y)
> sd
[1] 1.494107
> cv=sd/mean*100
> cv
[1] 10.94732
> lb=seq(120,145,5)
> ub=seq(125,150,5)
> f=c(5,12,15,20,12,8)
> x=(lb+ub)/2
> data.frame(lb,ub,f,x)
  lb  ub  f     x
```

```
1 120 125  5 122.5
2 125 130 12 127.5
3 130 135 15 132.5
4 135 140 20 137.5
5 140 145 12 142.5
6 145 150  8 147.5
> fx=f*x
> fx
[1]  612.5 1530.0 1987.5 2750.0 1710.0 1180.0
> data.frame(lb,ub,f,x,fx)
   lb  ub  f     x     fx
1 120 125  5 122.5  612.5
2 125 130 12 127.5 1530.0
3 130 135 15 132.5 1987.5
4 135 140 20 137.5 2750.0
5 140 145 12 142.5 1710.0
6 145 150  8 147.5 1180.0
> mean=sum(fx)/sum(f)
> mean
[1] 135.6944
> y=rep(x,f)

> y
 [1] 122.5 122.5 122.5 122.5 122.5 127.5 127.5 127.5 127.5 127.5 127.5 127.5
[13] 127.5 127.5 127.5 127.5 127.5 132.5 132.5 132.5 132.5 132.5 132.5 132.5
[25] 132.5 132.5 132.5 132.5 132.5 132.5 132.5 132.5 137.5 137.5 137.5 137.5
[37] 137.5 137.5 137.5 137.5 137.5 137.5 137.5 137.5 137.5 137.5 137.5 137.5
[49] 137.5 137.5 137.5 137.5 142.5 142.5 142.5 142.5 142.5 142.5 142.5 142.5
[61] 142.5 142.5 142.5 142.5 147.5 147.5 147.5 147.5 147.5 147.5 147.5 147.5

> mean=mean(y)
> mean
[1] 135.6944
> variance=var(y)
> variance
[1] 50.21518
> sd=sd(y)
> sd
[1] 7.086267
> cv=sd/mean*100
> cv
[1] 5.222223
```

Practical No. 3

Write a program on Regression analysis in R?

Q1. Estimate Y when X=110 also estimate X when y=22.

CODE:

```
> x=c(10,20,30,40,50,60,70,80,90,100)
> x=seq(10,100,10)
> x
 [1]  10  20  30  40  50  60  70  80  90 100
> y=c(4.21,5.91,7.43,9.83,11.5,31.32,14.99,16.74,18.61,20.30)
> length(y)
[1] 10
> r1=lm(y~x)
> r1

Call:
lm(formula = y ~ x)
Coefficients:
(Intercept)            x
     3.5653       0.1912

> summary(r1)

Call:
lm(formula = y ~ x)
Residuals:
   Min     1Q Median     3Q    Max
-2.390 -2.085 -1.750 -1.409 16.280
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.56533    4.15311   0.858   0.4156
x            0.19125    0.06693   2.857   0.0212 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 6.08 on 8 degrees of freedom
Multiple R-squared:  0.5051,    Adjusted R-squared:  0.4432
F-statistic: 8.164 on 1 and 8 DF,  p-value: 0.02124

> co=coef(r1)
> com=matrix(co)
> com
          [,1]
[1,] 3.5653333
[2,] 0.1912485
```

```
> a=com[1,1]
> b=com[2,1]
> a
[1] 3.565333
> b
[1] 0.1912485
> x1=110
> ey=a+b*x1
> cat("Estimate value of y when x=110 is:",ey)
Estimate value of y when x=110 is: 24.60267>
>
>
> r2=lm(x~y)
> r2

Call:
lm(formula = x ~ y)
Coefficients:
(Intercept)            y
     17.805        2.641

> summary(r2)

Call:
lm(formula = x ~ y)
Residuals:
    Min      1Q  Median      3Q     Max
-40.519 -11.916  -0.971  16.641  28.584
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  17.8049    14.8491   1.199   0.2648
y             2.6409     0.9243   2.857   0.0212 *

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 22.59 on 8 degrees of freedom
Multiple R-squared:  0.5051,    Adjusted R-squared:  0.4432
F-statistic: 8.164 on 1 and 8 DF,  p-value: 0.02124

> co=coef(r2)
> com=matrix(co)
> com
          [,1]
[1,] 17.804929
[2,]  2.640945
> a=com[1,1]
```

```
> b=com[2,1]
> a
[1] 17.80493
> b
[1] 2.640945
> y1=22
> ex=a+b*y1
> cat("Estimate value of x when y=22 is:",ex)
Estimate value of x when y=22 is: 75.90572>
```

Practical number 4

Prepare Google Questionnaire

Customer Satisfaction Survey at Restaurant

Please spare few minutes of your valuable time to answer the simple questionnaire.

* Required

1. **Name** *

2. **Age** *

3. **Gender** *
   Mark only one oval.
   - Female
   - Male
   - Prefer not to say

4. **Considering your complete experience at the restaurant, how likely would you be to recommend us to a friend or a colleague?** *
   Mark only one oval.

1	2	3	4	5

5. **The food is served hot and fresh.** *
   Mark only one oval.
   - Strongly Agree
   - Agree
   - Neutral
   - Disagree
   - Strongly disagree

6. **The quality of food is excellent.** *
   Mark only one oval.
   - Strongly Agree
   - Agree
   - Neutral
   - Disagree
   - Strongly Disagree

7. **The menu has a large variety.** *
Mark only one oval.

- Agree
- Strongly agree
- Neutral
- Disagree
- Strongly disagree

8. **Sauces, utensils, napkins, etc., were readily available.** *
Mark only one oval.

- Yes
- No

9. **Employees are patient when taking my order.** *
Mark only one oval.

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

10. **Employees are friendly and courteous.** *
Mark only one oval.

1	2	3	4	5

11. **The value for price paid was excellent.** *
Mark only one oval.

- Agree
- Strongly agree
- Neutral
- Disagree
- Strongly disagree

12. **Please provide any other feedback you may have**

Impact of Social Media of Consumers preference in fashion trends

The enormous infusion of social media, in terms of bloggers, internet celebrities, and e-commerce websites, especially targeting the younger generation, influences the way we dress and what we feel is 'fashionable'. Fashion and lifestyle bloggers today are highly reputable and influential individuals and those extremely committed to their craft have, earned industry recognition.

* Required

1. **Name** *

2. **Age** *

3. **Gender** *

   Mark only one oval.

   - Female
   - Male
   - Prefer not to say

4. **Occupation** *

   Mark only one oval.

   - Business
   - Employee
   - House maker
   - Student
   - Others

5. **Do you use social media?** *

   Mark only one oval.

   - Yes
   - No

6. **Which sites do you use?** *
Check all that apply.

- [ ] Facebook
- [ ] Instagram
- [ ] Twitter
- [ ] Pintrest
- [ ] Blogger

7. **Do you follow fashion brands on social networking sites?** *
Mark only one oval.

- ( ) Yes
- ( ) No

8. **Have you notice fashion apparel advertisement on social media?** *
Mark only one oval.

- ( ) Yes
- ( ) No

9. **How do you observed the latest fashion trends?** *
Check all that apply.

- [ ] Social Networking sites
- [ ] Print media
- [ ] Television Ads
- [ ] Observation

10. **How much do you agree that social helps in acquiring information about fashion trends?** *
Mark only one oval.

- ( ) Strongly agree
- ( ) Agree
- ( ) Neutral
- ( ) Disagree
- ( ) Strongly Disagree

11. **Do you follow fashion related blogs?** *
Mark only one oval.

- ( ) Yes
- ( ) No
- ( ) Sometimes

12. **How often do you buy products that you see on these sites?** *

Mark only one oval.

- Never
- Sometimes
- Rarely
- Always

13. **Do reviews and ratings affect your buying decision?** *

Mark only one oval.

- Yes
- No

14. **Do you copy fashion from social media?** *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Never											Mostly

Student Survey on Autonomous college

An autonomous university typically refers to a university that exercises independent control over its day-to-day operations and curriculum. please spare your valuable time to answer the simple questionnaire sample

* Required

1. **Infrastructure of College** *

Mark only one oval.

- very good
- Good
- Satisfactory
- Very Satisfactory

2. **How well did the teachers prepare for the classes** *

Mark only one oval.

- Thoroughly
- Satisfactorily
- Poorly
- Wont teach at all

3. **The institution provides multiple opportunities to learn and grow**

Mark only one oval.

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

4. **Extent of coverage of course** *

Mark only one oval.

- Very Good
- Good
- Satisfactory
- Very Sastisfactory

5. **Applicability/relevance to real life situations** *

Mark only one oval.

- Very Good
- Good
- Satisfactory
- Very Sastisfactory

6. **Learning value (in terms of knowledge, concepts, manual skills, analytical abilities and broadening perspectives)** *

Mark only one oval.

- Very Good
- Good
- Satisfactory
- Very Sastisfactory

7. **Clarity and relevance of textual reading material** *

Mark only one oval.

- Very Good
- Good
- Satisfactory
- Very Sastisfactory

8. **Relevance of additional source material (Library)** *

Mark only one oval.

- Very Good
- Good
- Satisfactory
- Very Sastisfactory

9. **Extent of effort required by students** *

Mark only one oval.

- Very Good
- Good
- Satisfactory
- Very Sastisfactory

10. **Overall rating** *

Mark only one oval.

- Very Good
- Good
- Satisfactory
- Very Sastisfactory

11. **The syllabus was** *

Mark only one oval.

- challenging
- adequate
- dull
- inadequate

12. **What is your opinion about the library materials for the course?** *
Mark only one oval.

- challenging
- adequate
- dull
- inadequate

13. **Were outsider experts invited to address you?** *
Mark only one oval.

- Yes
- No
- Rarely
- Frequently

14. **Did you visit industries, laboratories, banks and outside Universities?** *
Mark only one oval.

- Yes, frequently
- Some times
- No
- Yes, rarely

Practical number 5

Prepared a Traditional Questionnaire

Aim: Prepare a Traditional Questionnaire. For Job Satisfaction Survey.

The term Information and Communication Technology (ICT) is a broad and comprehensive expression. It is not restricted to the computers or the internet alone.

Name: _ _ _ _ _ _ _ _ _ _ _ _ _ _ Organization: _ _ _ _ _ _ _ Designation: _ _ _ _ _ _ _ _ _

1] Do you enjoy our company's culture?
Ⓐ Yes Ⓑ No Ⓒ Partially Ⓓ Not at All

2] Do you find your work meaningful?
Ⓐ Yes Ⓑ No Ⓒ Sometimes Ⓓ Not at All

3] Does our company offer adequate opportunities for promotions and career development?
Ⓐ Yes Ⓑ No Ⓒ Sometimes Ⓓ Not at All

4] Do you feel valued for your contributions?
Ⓐ Yes Ⓑ No Ⓒ May be Ⓓ Not at All

5] Does our company give you the tools and technologies you need to do your job well?
Ⓐ Yes Ⓑ No Ⓒ Sometimes Ⓓ Not at All

6] Do your superiors communicate company news effectively and in a timely manner?
Ⓐ Yes Ⓑ No Ⓒ Sometimes Ⓓ Not at All

7] Do you feel as though your job responsibilities are clearly defined?
Ⓐ Yes Ⓑ No Ⓒ May Be Ⓓ Not at All

8] Do you think that work is distributed evenly across your team?
Ⓐ Yes Ⓑ No Ⓒ May Be Ⓓ Not at All

9] Do you feel connected to your co-workers?
Ⓐ Yes Ⓑ No Ⓒ Sometimes Ⓓ Not at All

10] Do you feel like your job utilizes your skills and abilities as much as it could?
Ⓐ Yes Ⓑ No Ⓒ May Be Ⓓ Not at All

11] Does management seem invested in the success of the team?
Ⓐ Yes Ⓑ No Ⓒ Sometimes Ⓓ Not at All

12] Do your managers value your feedback?
Ⓐ Yes Ⓑ No Ⓒ Sometimes Ⓓ Not at All

13] How transparent do you feel the management is?
Little ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩ Very

14] How open to change are we as an organization?
Unopen ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩ Open

15] How happy are you at work?
Unhappy ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩ Happy

Aim: - Effect of ICT on teaching process

1] is ICT helpful for improving the techniques of Teaching Learning process of TVE

- Agree
- Strongly Disagree
- Strongly Agree
- Disagree

2] Is ICT, teaching learning will be easy process learning?

- Agree
- Strongly Disagree
- Strongly Agree
- Disagree

3] ICT teaching learning process is time saving?

- Agree
- Strongly Disagree
- Strongly Agree
- Disagree

4] ICT tool are too complicated to use in teaching learning process?

- Agree
- Strongly Disagree
- Strongly Agree
- Disagree

5] uses of ICT like internet help staff and student have access to relevant research subject?

- Agree
- Strongly Disagree
- Strongly Agree
- Disagree

6] ICT media in the teaching and learning such as the television and computer and projector ensure better understanding of relevant subject topic

- Agree
- Strongly Disagree
- Strongly Agree
- Disagree

7] Is ICT learning process is help to student learn deeply

- Agree

- Strongly Disagree
- Strongly Agree
- Disagree

8] Is ICT enhance the learning of subject content

- Agree
- Strongly Disagree
- Strongly Agree
- Disagree

Aim: - Statistical Analysis on different types of investment plans

1. **What is the intent of your portfolio?**

   a. To generate income for today
   b. To generate income at a later date
   c. To provide for my dependents (I do not anticipate using these funds)
   d. To fund a large purchase in the future.

2. What is your major goal for your portfolio?

a. To ensure my portfolio remains secure
b. To see my portfolio grow and to avoid fluctuating returns
c. To balance growth and security, and to keep pace with inflation
d. To provide growth potential, and to accept some fluctuation in returns
e. To provide the sole objective of potential long-term growth

3. Which of the following ranges includes your age?

a. Under 30
b. 30 to 39
c. 40 to 49
d. 50 to 59
e. 60 to 69
f. 70 to 79
g. Over 79

4. Which of the following ranges best represents your current annual family income (including pensions) before taxes?

a. Under $30,000
b. $30,000 to $60,000
c. $60,001 to $90,000
d. $90,001 to $120,000
e. More than $120,000

5. When do you anticipate using these funds?

a. Immediately
b. One to three years

c. Four to five years
d. Six to 10 years
e. 11 to 15 years
f. 16 to 20 years
g. More than 20 years

6. If you had picked an investment with potential for large gains but also the risk of large losses how would you feel:

- Panicked and very uncomfortable
- Quite uneasy
- A little concerned
- Accepting of the possible highs and lows
- Excited by the potential for gain

7. **Imagine that you have some money to invest and a choice of two investment products, which option would you choose?**

- A product with a low average annual return but almost no risk of loss of the initial investment
- A product with a higher average annual return but some risk of losing part of the initial investment
- A mixture of the two products

8. **I want my investment money to be safe even if it means lower returns.**

- I strongly agree with this statement.
- I tend to agree with this statement.
- In between.
- I tend to disagree with this statement.
- I strongly disagree with this statement.

9. **I am looking for high investment growth. I am willing to accept the possibility of greater losses to achieve this.**

- I strongly agree with this statement.
- I tend to agree with this statement.
- In between.
- I tend to disagree with this statement.
- I strongly disagree with this statement.

Practical number 6

Small Sample Test (T-test) in R

Q.1 The Random Sample of n balls at following IQ. Do this data support the assumption that population mean IQ is 100?

70, 120, 110, 101, 88,83,95,89,107,125

Code in R:

```
> data<-read.csv(file.choose())
> data
    x
1   70
2  120
3  110
4  101
5   88
6   83
7   95
8   89
9  107
10 125
> x=data$x
> x
 [1]  70 120 110 101  88  83  95  89 107 125
> #H0:The polulation mean is 100
> #H1:The population mean is not 100
> t.test(x)

        One Sample t-test

data:  x
t = 18.244, df = 9, p-value = 2.039e-08
alternative hypothesis: true mean is not equal to 0
```

```
95 percent confidence interval:
  86.54903 111.05097
sample estimates:
mean of x
     98.8

> # p value is less than 0.05 so we reject H0
> t.test(x,alternative="greater")

        One Sample t-test

data:  x
t = 18.244, df = 9, p-value = 1.02e-08
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 88.87257      Inf
sample estimates:
mean of x
     98.8

> t.test(x,alternative="less")
        One Sample t-test
data:  x
t = 18.244, df = 9, p-value = 1
alternative hypothesis: true mean is less than 0
95 percent confidence interval:
     -Inf 108.7274
sample estimates:
mean of x
     98.8
```

```
> save.image("C:\\Users\\student\\Desktop\\6242\\practical_2")
```

Q.2 Certain pesticides is packed into bags by a machine a random sample of 10 bags is drawn and their content are found to weight in kg as follows, values are given at 50,49,52,44,45,48,46,45,49,45 test id the average packing can be taken is greater than 50 kg.

Coding in R:

```
> data<-read.csv(file.choose())
> data
   x
1  50
2  46
3  52
4  44
5  45
6  48
7  46
8  45
9  49
10 45
> x=data$x
> x
 [1] 50 46 52 44 45 48 46 45 49 45
> #H0: the average package is less than 50 kg
> #H1: the average package is more than 50 kg
> t.test(x)

        One Sample t-test

data:  x
t = 56.627, df = 9, p-value = 8.408e-13
```

```
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 45.12242 48.87758
sample estimates:
mean of x
       47

> t.test(x,alternative="less")

        One Sample t-test

data:  x
t = 56.627, df = 9, p-value = 1
alternative hypothesis: true mean is less than 0
95 percent confidence interval:
     -Inf 48.52147
sample estimates:
mean of x
       47

> t.test(x,alternative="greater")

        One Sample t-test

data:  x
t = 56.627, df = 9, p-value = 4.204e-13
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 45.47853      Inf
sample estimates:
```

mean of x

47

> #since p-value is less thasn 0.05 so we reject H0

> save.image("C:\\Users\\student\\Desktop\\6242\\Q2")

PAIERD T-TEST

Q.3 Two groups of 10 people each were given the digit span subtest from the intelligence scale once group consisted of regular smokers of marijuana while the other group consisted of non-smokers, the scores are given below –

Non-smokers – 18,22,21,17,20,17,23,20,22,21

Smokers - 16,20,14,21,20,18,13,15,17,21

Test the hypothesis that there is no significant effect on scores due to smoking?

Coding in R:

```
> data<-read.csv(file.choose())
> data
smokers non.smokers
1       16          18
2       20          22
3       14          21
4       21          17
5       20          20
6       18          17
7       13          23
8       15          20
9       17          22
10      21          21
> x=data$smokers
> x
 [1] 16 20 14 21 20 18 13 15 17 21
```

```
> y=data$non.smokers
> y
 [1] 18 22 21 17 20 17 23 20 22 21
> #H0: There is no signifiance effect on smokers
> #H1: There is signifiance effect on smokers
> t.test(x,y,var=equal=T)
Error: unexpected '=' in "t.test(x,y,var=equal="
> t.test(x,y,var.equal=T)
        Two Sample t-test
data:  x and y
t = -2.2573, df = 18, p-value = 0.03665
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.0198634 -0.1801366
sample estimates:
mean of x mean of y
     17.5      20.1
> #since p-value is less than 0.05 bwe rejet H0
> #hence we conclude that there is a significant difference between smokers and non-smokers
```

UNPAIRED T-TEST

Q.4 the weight reducing diet program is conducted and observations are noted for 10 participants test whether the program was effective?

Values before – 120,125,115,130,123,119,122,127,128,118

Values after – 111,114,107,120,115,112,112,120,119,112

Coding in R:

```
> #H0: there is no significant difference in weight due to diet program
> #H1: the weight is reduced due to diet program
> data<-read.csv(file.choose())
> data
   before after
1     120   111
2     125   114
3     115   107
4     130   120
5     123   115
6     119   112
7     127   112
8     127   120
9     128   119
10    118   112
> before<-data$before
> before
 [1] 120 125 115 130 123 119 127 127 128 118
> after<-data$after
> after
 [1] 111 114 107 120 115 112 112 120 119 112
> t.test(before,after,paired=T,alternative="greater")

        Paired t-test

data:  before and after
t = 11.023, df = 9, p-value = 7.911e-07
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
```

```
 7.50327     Inf
sample estimates:
mean of the differences
                     9
> #since p-value is less that 0.05 we reject H0
> #hence we conclude that deit program is effective in reducing the weight.
>
```

Practical number 7

Large Sample Test (Z test) in R

Q1) Test the hypothesis H0 , μ=10 against H1 μ≠10, a random sample of size 400 is drone and gives me 10.2 and standard deviation =2.25 used LOS=5%

Code:

```
> n=400
> xbar=10.2
> s=2.25
> m=10
> zcal=(xbar-m)/(s/sqrt(n))
> cat('calculated value of z is :-'zcal)
Error: unexpected symbol in "cat('calculated value of z is :-'zcal"
> cat('calculated value of z is :-',zcal)
calculated value of z is :- 1.777778>
> pvalue=2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.07544036
> #since pvalue is greater then LOS we accept H0 at 5% LOS
```

Q2) Two random samples of sizes 1000 and 200 are drone form 2 population with same standard deviation 2.5 gives mean 67.5 and 68 respectively test the hypothesis at 55 Los.

Code:

```
> n1=1000
> n2=2000
> mx=67.5
> my=68
> sx=2.5
> sy=2.5
> zcal=(mx-my)/(sqrt(sx)/n1+sqrt(sy)/n2)
> cat('calculated value of z is',zcal)
```

```
calculated value of z is -210.8185>
> zcal=(mx-my)/(sqrt(sx)^2/n1+sqrt(sy)^2/n2)
> cat('calculated value of z is',zcal)
calculated value of z is -133.3333>
> zcal=(mx-my)/sqrt(sx^2/n1)+sqrt(sy^2/n2)
> zcal
[1] -6.268654
> zcal=(mx-my)/sqrt(sx^2/n1+sy^2/n2)
> zcal
[1] -5.163978
> pvalue=2*(1-pnorm(abs(zcal)))
> pvalue
[1] 2.417564e-07
> #pvalue is less then LOS so we reject H0
```

Q3) Experienced as shown as 20% of manufactural products is of the top quality in one days production of 400 articles on 50 are of top quality test the hypothesis that Experienced 20% wrong.

Code:

```
> #H0 p=20% H1!= 20% ,n=400, sp(sample propotion)=50/400
> n=400
> sp=50/400
> P=20/100
> zcal=(sp-P)/sqrt(P*(1-P)/n)
> cat('calculated value is',zcal)
calculated value is -3.75>
> pvalue=2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.0001768346
> #since pvalue is less then LOS we Reject H0
```

Q4) from Each of 2 consignment of apples is sampled of size 200 is drone and number of apple are counted test whether the proportional apple in the 2 consignment significantly different

Nos. of apple of consignment

A=44

B=30

Code:

```
> #H0 P1=P2 H1 P1!=P2
> #n1=200,n2=200,p1=44/200,p2=30/200
> n1=200
> n2=200
> p1=44/200
> p2=30/200
> P=((n1*p1)+(n2*p2))/(n1+n2)
> P
[1] 0.185
> q=1-p
Error: object 'p' not found
> Q=1-P
> Q
[1] 0.815
> se=sqrt(P*Q/(1/n1)+(1/n2))
> se
[1] 5.491812
> se=sqrt(P*Q*(1/n1)+(1/n2))
> se
[1] 0.0758543
> se=sqrt(P*Q*(1/n1+1/n2))
> se
[1] 0.03882976
```

```
> zcal=(p1-p2)/se
> cat('zcal value is',zcal)
zcal value is 1.802741>
> pvalue=2*(1-pnorm(abs(zcal)))
> pvalue
[1] 0.07142888
> #we accept the H0 as pvalue is greater then LOS
```

Practical Number: -8

ANNOVA

Analysis of Variance (ANOVA) helps you test differences between two or more group means. ANOVA test is centered around the different sources of variation (variation between and within group) in a typical variable. A primarily ANOVA test provides evidence of the existence of the mean equality between the group. This statistical method is an extension of the t-test. It is used in a situation where the factor variable has more than one group.

- **One-way ANOVA**:- There are many situations where you need to compare the mean between multiple groups. For instance, the marketing department wants to know if three teams have the same sales performance. Team: 3 level factors: A, B, and C

  Sale: A measure of performance

The ANOVA test can tell if the three groups have similar performances.

To clarify if the data comes from the same population, you can perform a **one-way analysis of variance** (one-way ANOVA hereafter). This test, like any other statistical tests, gives evidence whether the H0 hypothesis can be accepted or rejected.

Pairwise comparison: - The one-way ANOVA test does not inform which group has a different mean. Instead, you can perform a Tukey test with the function **TukeyHSD().**

Two-way ANOVA: - A two-way ANOVA test adds another group variable to the formula. It is identical to the one-way ANOVA test, though the formula changes slightly:

Q. 2 Raju restaurant near the railway station at Falna has been having average sales of 500 tea cups per day because of the development of the bus stand nearby it expects to increase its sale. During the increase its sale in the first 12 days after the start of the bus stand the daily sales were as under

Values: 550,570,490,615,505,580,570,460,600,580,530,526

On the basis of this sample information can one conclude that Raju's restaurant sales have increased at 5% los.

Code:

```
> x=c(550,570,490,615,505,580,570,460,600,580,530,526)

> #H0:M=500

> #H1:M>500

> t.test(x,mu=500,alternative="greater")
```

```
	One Sample t-test

data:  x
t = 3.5614, df = 11, p-value = 0.002231
alternative hypothesis: true mean is greater than 500
95 percent confidence interval:
 523.7954      Inf
sample estimates:
mean of x
      548
> #since p value is lesss than 0.05 we reject H0.
> #Hence we conclude that Raju restarunt's sale is increased
```

Q 3 A Drug Company tested three formulations of pain relief medicine for migraine headache sufferers. For the experiment 27 patients were selected and 9 were randomly assigned to one of the three drug formulations. They were instructed to take the drug during their next migraine headache and to report their pain on a scale of 1-10. Apply One way ANOVA.

Drug A: 4, 5, 4, 3, 2, 4, 3, 4, 4

Drug B: 6, 8, 4, 5, 4, 6, 5, 8, 6

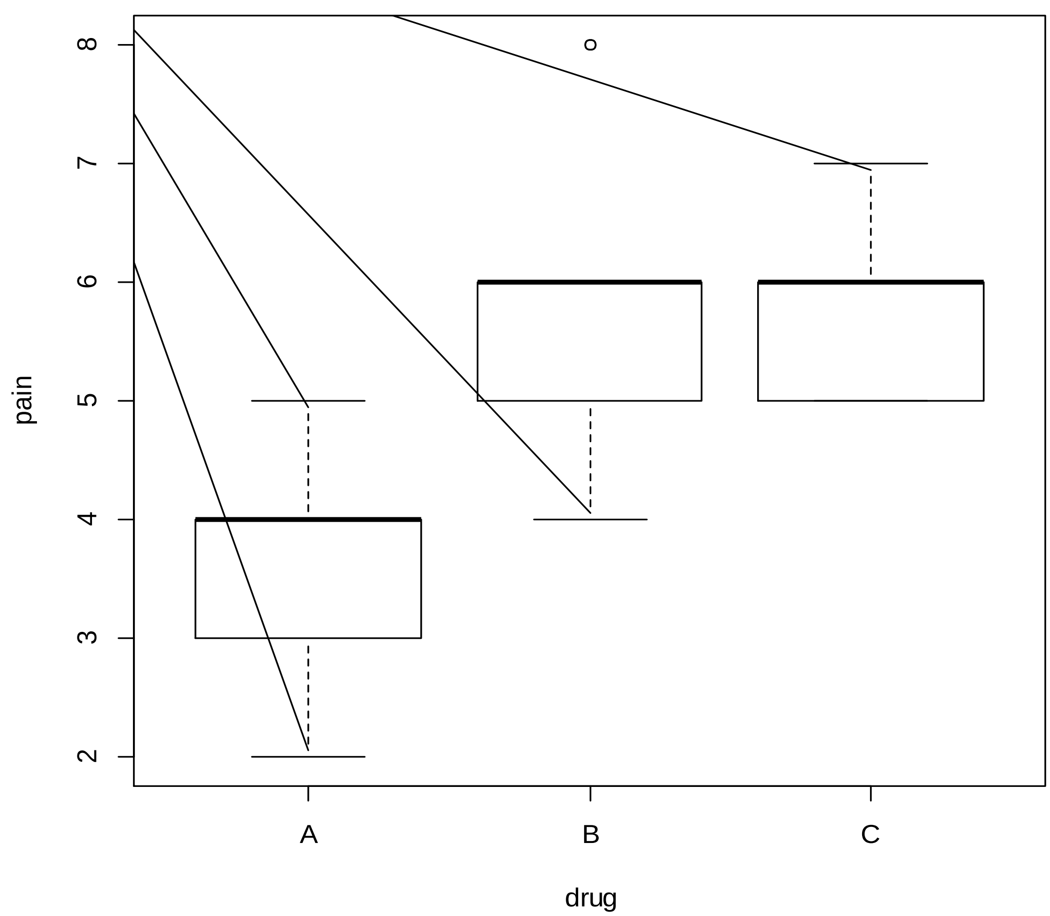
Drug C: 6, 7, 6, 6, 7, 5, 6, 5, 5

Code:

```
> pain=c(4, 5, 4, 3, 2, 4, 3, 4, 4, 6, 8, 4, 5, 4, 6, 5, 8, 6, 6, 7, 6, 6, 7, 5, 6, 5, 5)
> length(pain)
[1] 27
> drug=c(rep("A",9),rep("B",9),rep("C",9))
> migraine=data.frame(pain,drug)
> migraine
   pain drug
1     4    A
```

```
2  5  A
3  4  A
4  3  A
5  2  A
6  4  A
7  3  A
8  4  A
9  4  A
10  6  B
11  8  B
12  4  B
13  5  B
14  4  B
15  6  B
16  5  B
17  8  B
18  6  B
19  6  C
20  7  C
21  6  C
22  6  C
23  7  C
24  5  C
25  6  C
26  5  C
27  5  C
> plot(pain~drug,data=migraine)
> #From the box plot it appears that the mean pain for drug A is lower than that of drug B & C.
> result=aov(pain~drug, data=migraine)
> summary(result)
```

```
          Df Sum Sq Mean Sq F value   Pr(>F)
drug        2  28.22  14.111   11.91 0.000256 ***
Residuals   24  28.44   1.185
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> #since p value is less than 0.05, we reject H0.
> #hence we conclude that there is a significant difference between the means.
>#these result show that B-A and C-A differences are significant while C-B differences is not significant.
```

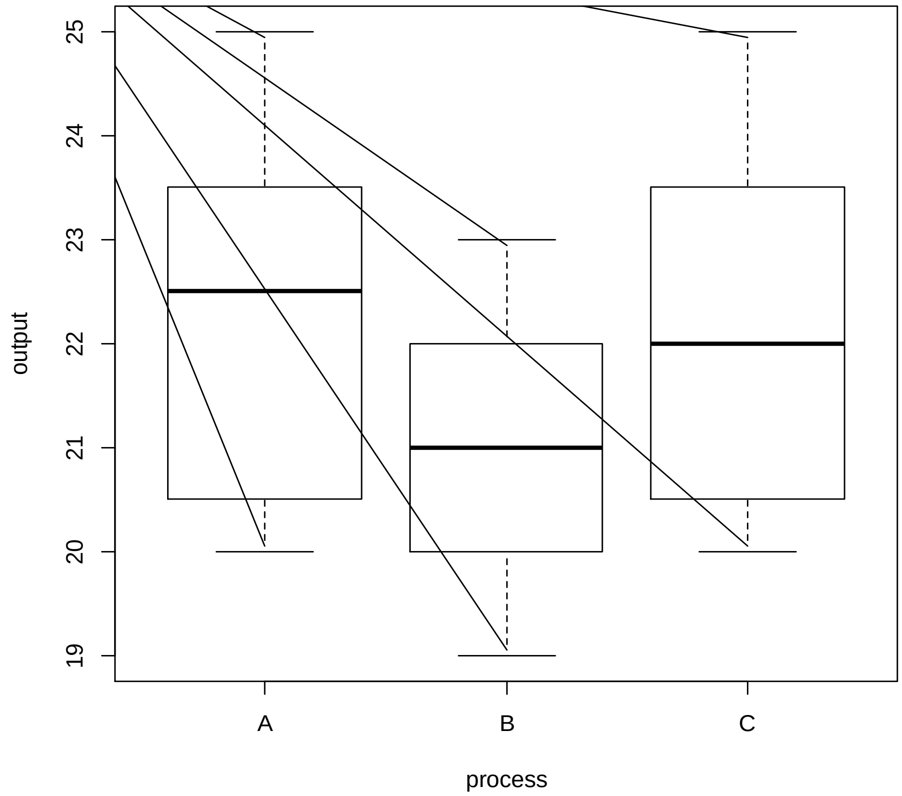


Q 4. Three processes A, B, C are tested to see whether their outputs are equivalent. The following observations of outputs are made. Carry out the analysis of variance and state your conclusion.

Code:

```
> output=c(20,22,23,21,20,24,25,23,19,21,20,22,23,21,20,20,25,24,22,23)
> length(output)
[1] 20
> process=c(rep("A",8),rep("B",5),rep("C",7))
> result=data.frame(output,process)
> result
   output process
1      20       A
2      22       A
3      23       A
4      21       A
5      20       A
6      24       A
7      25       A
8      23       A
9      19       B
10     21       B
11     20       B
12     22       B
13     23       B
14     21       C
15     20       C
16     20       C
17     25       C
18     24       C
19     22       C
20     23       C
> plot(output~process,data=result)
> #mean output of process B is less the that of A and C
```

```
> ans=aov(output~process, data=result)
> summary(ans)
            Df Sum Sq Mean Sq F value Pr(>F)
process      2   5.44   2.721   0.821  0.457
Residuals   17  56.36   3.315
> #since p value is greater than 0.05, we accept H0.
> #hence we conclude that there is no significant difference between mean output of the process.
> TukeyHSD(ans,conf.level=0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = output ~ process, data = result)

$process
          diff       lwr      upr     p adj
B-A -1.2500000 -3.912806 1.412806 0.4667743
C-A -0.1071429 -2.524548 2.310262 0.9929011
C-B  1.1428571 -1.592125 3.877839 0.5435759
```



Q 5. A tea company appoints four salesman A, B, C and D and observes their sale in three seasons- summer, winter and monsoon. The figures (in lakhs) of sales are given in the following table.

Salesman

Season	A	B	C	D
Summer	36	36	21	35
Winter	28	29	31	32
Monsoon	26	28	29	29

- **(i)** **Do the salesman significantly differ in performance?**
- **(ii)** **Is there significant difference between seasons?**

Code:

```
> sales=c(36,36,21,35,28,29,31,32,26,28,29,29)

> f1=c(rep(1:3,rep(4,3)))

> f2=rep(c("A","B","C","D"),3)
```

```
> Season=factor(f1)
> Salesman=factor(f2)
> ans=aov(sales~Season+Salesman)
> summary(ans)
            Df Sum Sq Mean Sq F value Pr(>F)
Season       2     32   16.00   0.706  0.531
Salesman     3     42   14.00   0.618  0.629
Residuals    6    136   22.67
```

> #since p values is greater than 0.05, we accept H0(1), so we conclude that there is no significant difference between the season.

> #since p values is greater than 0.05, we accept H0(2), so we conclude that there is no significant difference between the performance of salesmen.

Q 5. Setup an ANOVA table for the following Two way design

Variety of Fertilizers	**Variety of Seeds**		
	A	**B**	**C**
W	**6**	**5**	**5**
X	**7**	**5**	**4**
Y	**3**	**3**	**3**
Z	**8**	**7**	**4**

Code:

```
> output=c(6,5,5,7,5,4,3,3,3,8,7,4)
> f1=c(rep(1:4,rep(3,4)))
> f2=rep(c("A","B","C"),4)
> fertilizer=factor(f1)
> seed=factor(f2)
> ans=aov(output~fertilizer+seed)
> summary(ans)
            Df Sum Sq Mean Sq F value Pr(>F)
fertilizer   3     18       6       6 0.0308 *
seed         2      8       4       4 0.0787 .
```

```
Residuals    6      6       1
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #hence we conclude that there is a significant difference between the varities of fertilizers
> #hence we conclude that there is no significant difference between the varieties of seed.
> TukeyHSD(ans, conf.level=0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = output ~ fertilizer + seed)

$fertilizer
             diff        lwr       upr     p adj
2-1  8.881784e-16 -2.8264755 2.8264755 1.0000000
3-1 -2.333333e+00 -5.1598088 0.4931422 0.1020317
4-1  1.000000e+00 -1.8264755 3.8264755 0.6353224
3-2 -2.333333e+00 -5.1598088 0.4931422 0.1020317
4-2  1.000000e+00 -1.8264755 3.8264755 0.6353224
4-3  3.333333e+00  0.5068578 6.1598088 0.0249660

$seed
    diff       lwr       upr     p adj
B-A   -1 -3.169598 1.1695977 0.3922561
C-A   -2 -4.169598 0.1695977 0.0673680
C-B   -1 -3.169598 1.1695977 0.3922561
```

Practical number 9

Chi-Square Test in R

Introduction of Chi-Square Test:

Chi-Square test in R is a statistical method which used to determine if two categorical variables have a significant correlation between them. The two variables are selected from the same population. Furthermore, these variables are then categorized as *Male/Female*, *Red/Green*, *Yes/No* etc.

Syntax of a chi-square test:
chisq.test(data)

We have 105 patients under study and 50 of them were treated with the drug. Moreover, the remaining 55 patients were kept under control samples. Thus, the health condition of all patients was checked after a week.

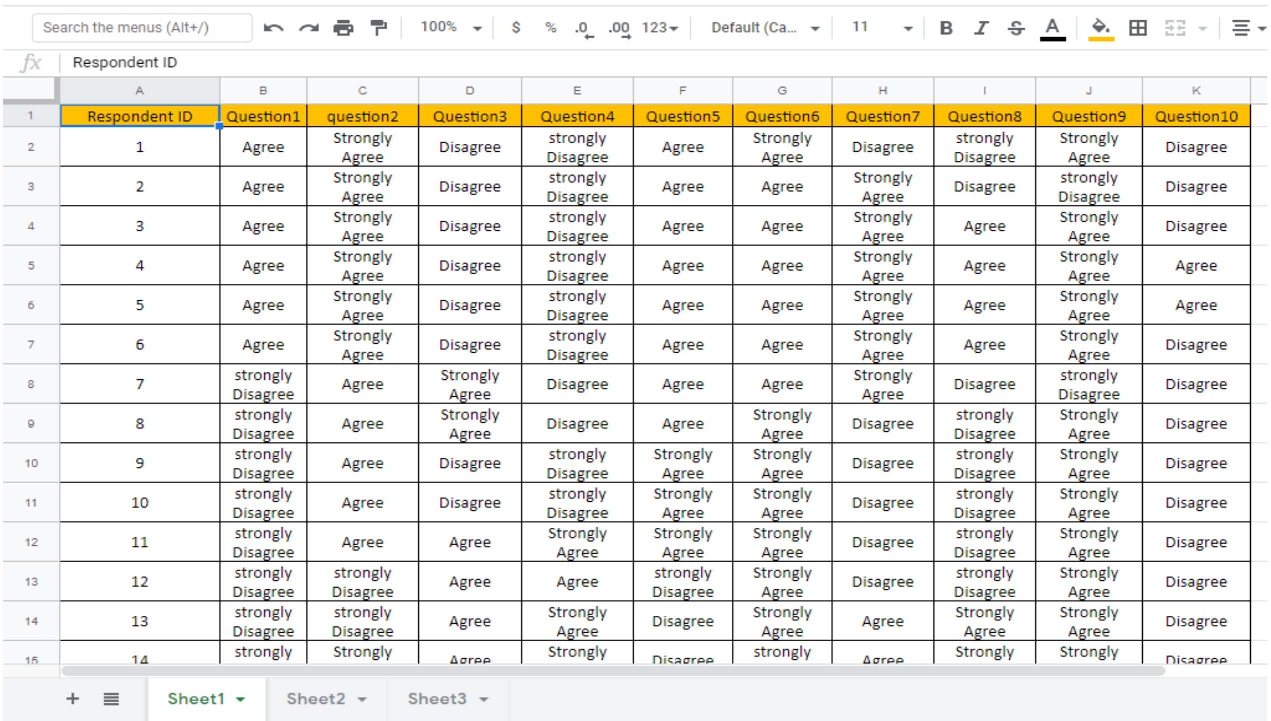
Code:

>data_frame <- read.**csv**("https://goo.gl/j6lRXD") #Reading CSV
>table(data_frame$treatment,data_frame$improvement)
> chisq.**test**(data_frame$treatment, data_frame$improvement, correct=FALSE)

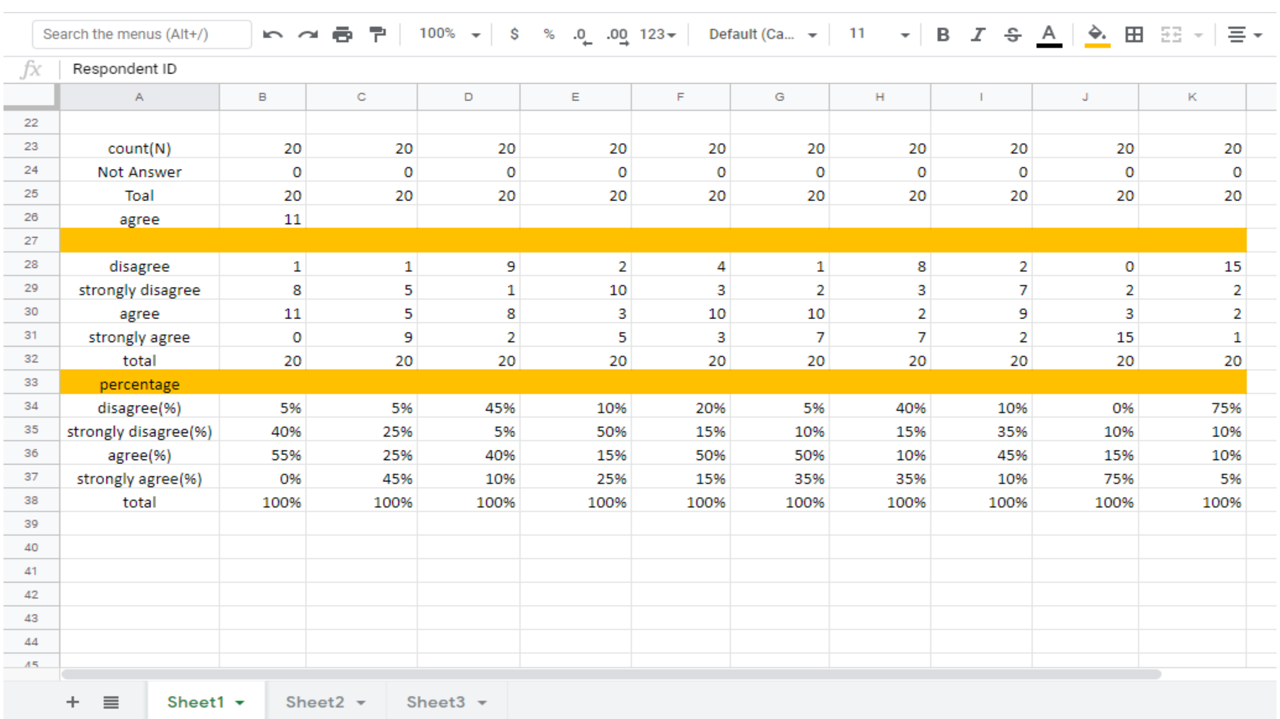
```
> #Author DataFlair
> data_frame <- read.csv("https://goo.gl/j6lRXD")  #Reading CSV
> table(data_frame$treatment, data_frame$improvement)

               improved not-improved
  not-treated        26           29
  treated            35           15
> chisq.test(data_frame$treatment, data_frame$improvement, correct=
FALSE)

        Pearson's Chi-squared test

data:  data_frame$treatment and data_frame$improvement
X-squared = 5.5569, df = 1, p-value = 0.01841
```

Practical number 10

Primary Data Analysis:



Respondent ID	Question1	question2	Question3	Question4	Question5	Question6	Question7	Question8	Question9	Question10
1	Agree	Strongly Agree	Disagree	strongly Disagree	Agree	Strongly Agree	Disagree	strongly Disagree	Strongly Agree	Disagree
2	Agree	Strongly Agree	Disagree	strongly Disagree	Agree	Agree	Strongly Agree	Disagree	strongly Disagree	Disagree
3	Agree	Strongly Agree	Disagree	strongly Disagree	Agree	Agree	Strongly Agree	Agree	Strongly Agree	Disagree
4	Agree	Strongly Agree	Disagree	strongly Disagree	Agree	Agree	Strongly Agree	Agree	Strongly Agree	Agree
5	Agree	Strongly Agree	Disagree	strongly Disagree	Agree	Agree	Strongly Agree	Agree	Strongly Agree	Agree
6	Agree	Strongly Agree	Disagree	strongly Disagree	Agree	Agree	Strongly Agree	Agree	Strongly Agree	Disagree
7	strongly Disagree	Agree	Strongly Agree	Disagree	Agree	Agree	Strongly Agree	Disagree	strongly Disagree	Disagree
8	strongly Disagree	Agree	Strongly Agree	Disagree	Agree	Strongly Agree	Disagree	strongly Disagree	Strongly Agree	Disagree
9	strongly Disagree	Agree	Disagree	strongly Disagree	Strongly Agree	Strongly Agree	Disagree	strongly Disagree	Strongly Agree	Disagree
10	strongly Disagree	Agree	Disagree	strongly Disagree	Strongly Agree	Strongly Agree	Disagree	strongly Disagree	Strongly Agree	Disagree
11	strongly Disagree	Agree	Agree	Strongly Agree	Strongly Agree	Strongly Agree	Disagree	strongly Disagree	Strongly Agree	Disagree
12	strongly Disagree	strongly Disagree	Agree	Agree	strongly Disagree	Strongly Agree	Disagree	strongly Disagree	Strongly Agree	Disagree
13	strongly Disagree	strongly Disagree	Agree	Strongly Agree	Disagree	Strongly Agree	Agree	Strongly Agree	Strongly Agree	Disagree
14	strongly	Strongly	Agree	Strongly	Disagree	strongly	Agree	Strongly	Strongly	Disagree



Respondent ID										
count(N)	20	20	20	20	20	20	20	20	20	20
Not Answer	0	0	0	0	0	0	0	0	0	0
Toal	20	20	20	20	20	20	20	20	20	20
agree	11									
disagree	1	1	9	2	4	1	8	2	0	15
strongly disagree	8	5	1	10	3	2	3	7	2	2
agree	11	5	8	3	10	10	2	9	3	2
strongly agree	0	9	2	5	3	7	7	2	15	1
total	20	20	20	20	20	20	20	20	20	20
percentage										
disagree(%)	5%	5%	45%	10%	20%	5%	40%	10%	0%	75%
strongly disagree(%)	40%	25%	5%	50%	15%	10%	15%	35%	10%	10%
agree(%)	55%	25%	40%	15%	50%	50%	10%	45%	15%	10%
strongly agree(%)	0%	45%	10%	25%	15%	35%	35%	10%	75%	5%
total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Formula for operation

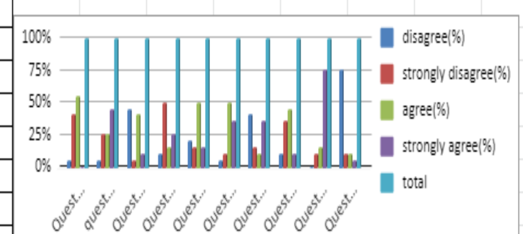
count(N) = COUNTA(B2:B21)
Not Answer = COUNT(D2:D21)
Toal = SUM(C23:C24)

Disagree = COUNTIF(B$2:B$21,$A28)
strongly disagree = COUNTIF(E$2:E$21,$A29)
Agree =COUNTIF(D$2:D$21,$A31)
strongly agree = COUNTIF(D$2:D$21,$A31)
Total = SUM(B28:B31)

Percentage

disagree(%) =B28/B32
strongly disagree(%) = B29/B32
agree(%) = B30/B32
strongly agree(%) = =B31/B32
Total = SUM(B34:B37)

question	disagree(%)	strongly disagree(%)	agree(%)	strongly agree(%)	total
Question1	5%	40%	55%	0%	100%
question2	5%	25%	25%	45%	100%
Question3	45%	5%	40%	10%	100%
Question4	10%	50%	15%	25%	100%
Question5	20%	15%	50%	15%	100%
Question6	5%	10%	50%	35%	100%
Question7	40%	15%	10%	35%	100%
Question8	10%	35%	45%	10%	100%
Question9	0%	10%	15%	75%	100%
Question10	75%	10%	10%	5%	100%



Select the table and click on – click on **Insert Tab** – Select pivot chart from **Charts** section

> d=read.csv(file.choose())

> d

	Resp_id	question1	question2	question3	question4	question5
1	Strongly Agree(%)	5	25	0	0	30
2	Strongly Disagree(%)	15	25	45	45	25
3	Disagree(%)	45	20	20	5	15
4	Agree(%)	35	30	35	50	30

	question6	question7	question8	question9	quesiton10
1	5	5	40	40	0
2	30	25	30	30	0
3	15	35	0	10	50
4	50	35	30	20	50

> q1=d$question1

> q1=d$question1

> q1

[1] 5 15 45 35

pie(q1,col=rainbow(length(q1)),label=q1,main="am satisfied")

> legend("topright",c("Strongly Agree","Strongly Disagree","Agree","Disagree"),cex=0.8,fill=rainbow(length(q1)))