

Assignment 11 (S-520)

FNU Anirudh

December 2, 2015

Solution 1:-

```
# 144 Races
observed<- c(29,19,18,25,17,10,15,11)
# Ho:- Horse's starting position does not affect winning
expected<-rep(144/8, 8)
# LR chi-squared test
G2 = 2 * sum(observed * log(observed/expected))
1 - pchisq(G2, df=7)

## [1] 0.02388413

# Pearson's chi-squared
X2 = sum((observed - expected)^2 / expected)
1 - pchisq(X2, df=7)

## [1] 0.02223948
```

We can see that both results yield value less than 0.05 hence we can reject our Null Hypothesis that horse's starting position does not affect it's chance of winning. (I am assuming 95% Confidence Interval) since it's not mentioned in problem.

Solution 2:-

	Height	Leaves
Dominant	Tall(3/4)	Cut(3/4)
Recessive	Dwarf(1/4)	Potato(1/4)

```
height<- c(3/4,1/4)
leaves<- c(3/4,1/4)
dominant<- c(3/4,3/4)
```

```

recessive<- c(1/4,1/4)
# a) Probability of Each Ej
E1<- 3/4*3/4
E1
## [1] 0.5625

E2<- 3/4*1/4
E2
## [1] 0.1875

E3<- 1/4*3/4
E3
## [1] 0.1875

E4<- 1/4*1/4
E4
## [1] 0.0625

# b)
observed<- c(926,288,293,104)
observed
## [1] 926 288 293 104

# n=1611 given in the problem
expected<-c(E1,E2,E3,E4)*1611
expected
## [1] 906.1875 302.0625 302.0625 100.6875

# LR chi-squared test
G2 = 2 * sum(observed * log(observed/expected))
G2
## [1] 1.477587

# Degrees of Freedom= 3
1 - pchisq(G2, df=3)
## [1] 0.6874529

# Pearson's chi-squared
X2 = sum((observed - expected)^2 / expected)
X2
## [1] 1.468722

1 - pchisq(X2, df=3)
## [1] 0.6895079

```

From P-value we cannot reject our Null Hypothesis hence both observed and expected values are more or less similar which shows the correctness of cell probabilities calculated.

Solution 3:-

```
threeft<- c(173,150)
thirtyfive<- c(125,73)
male<- c(173,125)
female<- c(150,73)
observed<- c(173,125,150,73)
N=sum(observed)
e1=(sum(male)*sum(threeft))/N
e2=(sum(male)*sum(thirtyfive))/N
e3=(sum(female)*sum(threeft))/N
e4=(sum(female)*sum(thirtyfive))/N
expected<- c(e1,e2,e3,e4)
expected

## [1] 184.74856 113.25144 138.25144 84.74856

observed

## [1] 173 125 150 73

# Degree of Freedom= (r-2)(c-1)= 1
# LR chi-squared test
G2 = 2 * sum(observed * log(observed/expected))
G2

## [1] 4.623063

1 - pchisq(G2, df=1)

## [1] 0.03154486

# Pearson's chi-squared
X2 = sum((observed - expected)^2 / expected)
X2

## [1] 4.59297

1 - pchisq(X2, df=1)

## [1] 0.03210333
```

From our p-value, we can conclude that sex ratio of Panamanian sandflies varies with height above ground and we can reject our null hypothesis of Independence across categories.

Solution 4:-

```
observed<-c(74,18,12,68,16,12,154,54,58,18,10,44)
N=sum(observed)
LP<-sum(74,18,12)/N
NS<-sum(68,16,12)/N
MC<-sum(154,54,58)/N
LD<-sum(18,10,44)/N
type<- c(LP,NS,MC,LD)
positive<- sum(74,68,154,18)/N
partial<- sum(18,16,54,10)/N
none<- sum(12,12,58,44)/N
response<- c(positive,partial,none)
expected<- (type*response)*N

## Warning in type * response: longer object length is not a multiple of
## shorter object length

X2 = sum((observed - round(expected,2))^2/round(expected,2))
# Degrees of Freedom = (r-2)(c-1)= 2*3= 6
1 - pchisq(X2, df=6)

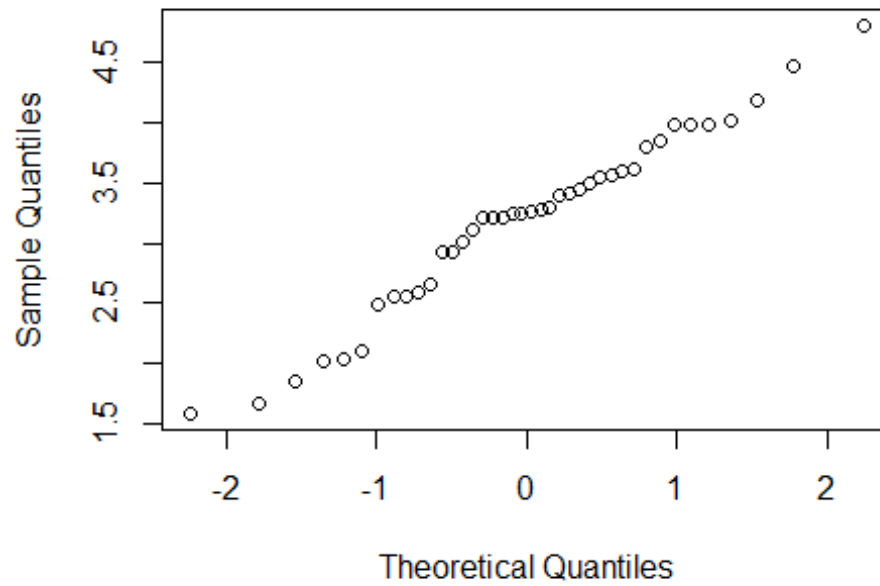
## [1] 0
```

P- value is very low which suggests dependence i.e Patient's response to treatment for Hodgkin's disease varies by histological type.

Solution 5:-

```
source('binorm.R')
x<- c(4.813,3.449,2.558,1.657,3.988,3.250,3.568,3.248,2.921,3.116,3.208
      ,4.025,3.302,2.017,3.394,4.482,2.583,3.271,2.107,3.551,2.034,4.189
      ,3.493,3.268,3.985,2.492,3.006,3.613,3.203,3.415,3.804,1.578,3.848
      ,3.211,2.925,2.549,3.598,3.985,1.843,2.651)
y<- c(5.505,3.576,2.316,1.042,4.339,2.988,2.538,2.991,3.453,3.178,3.235
      ,4.391,3.368,1.551,3.498,1.246,3.931,2.959,4.604,2.563,1.575,4.623
      ,3.639,3.320,4.334,4.060,3.333,2.475,3.055,2.756,4.078,0.930,4.140
      ,3.240,2.835,3.980,2.495,1.949,4.978,3.836)
# a)
qqnorm(x)
```

Normal Q-Q Plot

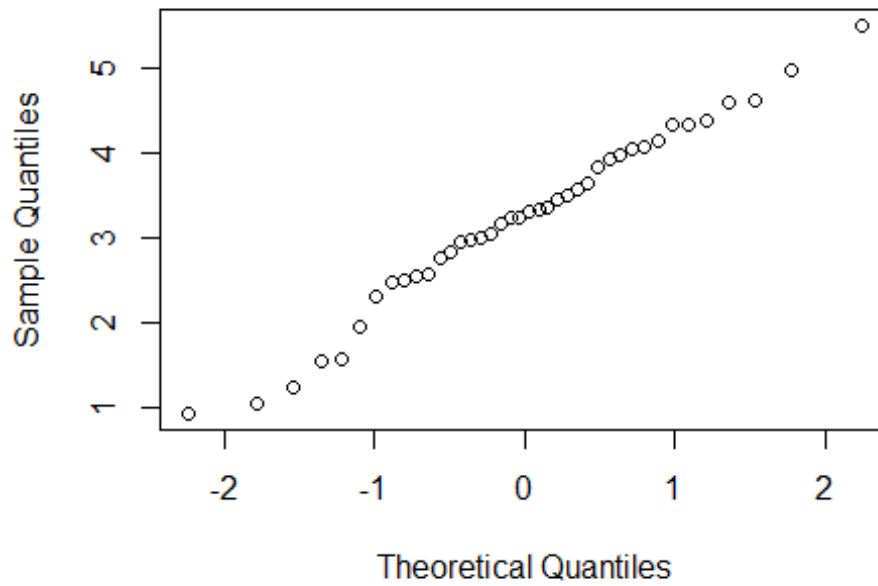


From QQ Plot, We can say that X is drawn from population close to Normal Distribution.

b)

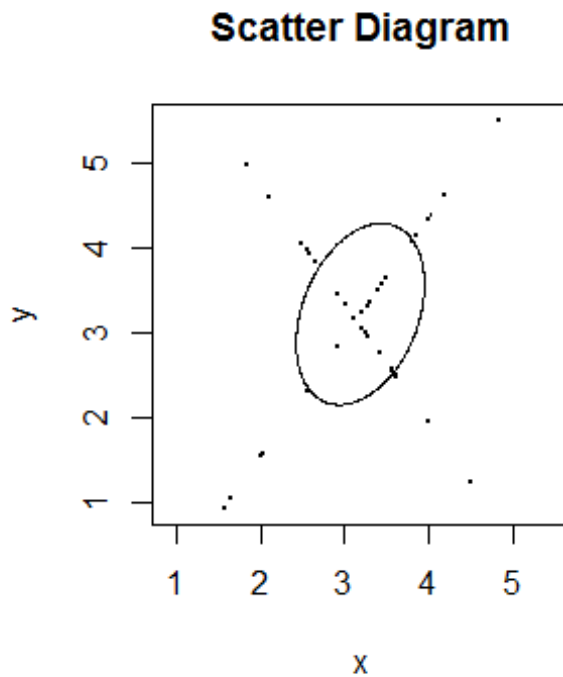
`qqnorm(y)`

Normal Q-Q Plot



From QQ Plot, We can say that Y is drawn from population close to Normal Distribution.

c) Scatter Plot of (x,y)
`binorm.scatter(cbind(x, y))`



```
cor(x, y)
```

```
## [1] 0.3244167
```

Since density of data doesn't fit inside ellipse we can say that they are not from bivariate distribution.

d)

X and Y may be drawn from two different normally distributed population assuming correlation is just by chance.

Solution 6:-

```
count<- c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14)
freq<- c(57,203,383,525,532,408,273,139,45,27,10,4,0,1,1)
# a) Average Observed Count
x_bar=sum(count*freq)/sum(freq)
x_bar

## [1] 3.871549

# b)
count1<- c(0,1,2,3,4,5,6,7,8,9)
freq1<- c(57,203,383,525,532,408,273,139,45,27)
```

```

# Expected Value for 0-9 values
expected1= dpois(count1,x_bar)*2608
expected1

## [1] 54.31442 210.28096 407.05653 525.31311 508.44388 393.69308 254.03368
## [8] 140.50055 67.99435 29.24927

# Expected value for count greater than 10
expected2=(1-ppois(9,x_bar))*2608
expected2

## [1] 17.12015

# Combining all the values
expected=c(expected1,expected2)
expected

## [1] 54.31442 210.28096 407.05653 525.31311 508.44388 393.69308 254.03368
## [8] 140.50055 67.99435 29.24927 17.12015

# Observed Values
observed<- c(57,203,383,525,532,408,273,139,45,27,16)
# LR chi-squared test
G2 = 2 * sum(observed * log(observed/expected))
p=1-pchisq(G2,9)
# Pearson's chi-squared
X2 = sum((observed - expected)^2 / expected)
1 - pchisq(X2, df=9)

## [1] 0.1684555

```

Unrestricted dimension is 10 where as restricted dimension is 1 hence degree of freedom $df=9$. If count of alpha particle scintillations follow a Poisson distribution then $p= 0.12534$ for likelihood ratio test statistics. Hence we cannot dismiss Null Hypothesis that data was drawn from Poisson istribution. Also using Pearson's chisquared test we get larger p value which fails to reject the null hypothesis.