Final_Take_Home

Sumeet Mishra

4/28/2019

1.

Null Hypothesis, H_0 : Number of misprints follow Poisson distribution. Alternate Hypothesis, H_1 : Number of misprints does not follow Poisson distribution.

```
observed=c(9,23,40,30,31,26,19,10,5,4,3)
#plot(0:10, observed, type="h")
```

```
num_obs=sum(observed)
total_misprints=sum((0:10)*observed)
ave=total_misprints/num_obs
expected=num_obs*dpois(0:20,ave)
data.frame(total_misprints=0:20,expected=round(expected,1))
```

```
##    total_misprints expected
## 1                0      5.0
## 2                1     18.5
## 3                2     34.1
## 4                3     41.9
## 5                4     38.6
## 6                5     28.4
## 7                6     17.5
## 8                7      9.2
## 9                8      4.2
## 10               9      1.7
## 11              10      0.6
## 12              11      0.2
## 13              12      0.1
## 14              13      0.0
## 15              14      0.0
## 16              15      0.0
## 17              16      0.0
## 18              17      0.0
## 19              18      0.0
## 20              19      0.0
## 21              20      0.0
```

Since the expected count for 8 misprints is less than 5. Hence, we combine expected count for 8 and above.

```
#observed=c(9,23,40,30,31,26,19,10,5,7)
#expected=rep(NA,10)
#expected[1:9] = num_obs * dpois(0:8, ave)
#expected[10] = num_obs * (1 - ppois(8, ave))
#data.frame(total_misprints=0:9,expected=round(expected,1))

observed=c(9,23,40,30,31,26,19,10,12)
expected=rep(NA,9)
expected[1:8] = num_obs * dpois(0:7, ave)
expected[9] = num_obs * (1 - ppois(7, ave))
#data.frame(total_misprints=0:8,expected=round(expected,1))
```

```r
G2 = 2 * sum(observed * log(observed/expected))
print(paste("Test Statistics of the likelihood test is",round(G2, 3)))
```

```
## [1] "Test Statistics of the likelihood test is 13.347"
```

```r
lh_p=1 - pchisq(G2, df=7)
print(paste("pvalue of the likelihood test is",round(lh_p, 3)))
```

```
## [1] "pvalue of the likelihood test is 0.064"
```

Since pvalue from Likelihood test is greater than α= 0.05, we fail to reject null hypothesis because we have insufficient evidence to conclude that the number of misprints does not follow Poisson distribution.

```r
X2 = sum((observed - expected)^2/expected)
p_p=1 - pchisq(X2, df=7)

print(paste("Test Statistics of the Pearson test is",round(X2, 3)))
```

```
## [1] "Test Statistics of the Pearson test is 14.292"
```

```r
print(paste("pvalue of the Pearson test is",round(p_p, 3)))
```

```
## [1] "pvalue of the Pearson test is 0.046"
```

Since pvalue from Pearson test is less than α= 0.05, we reject null hypothesis because we have sufficient evidence to conclude that the number of misprints does not follow Poisson distribution.

2.

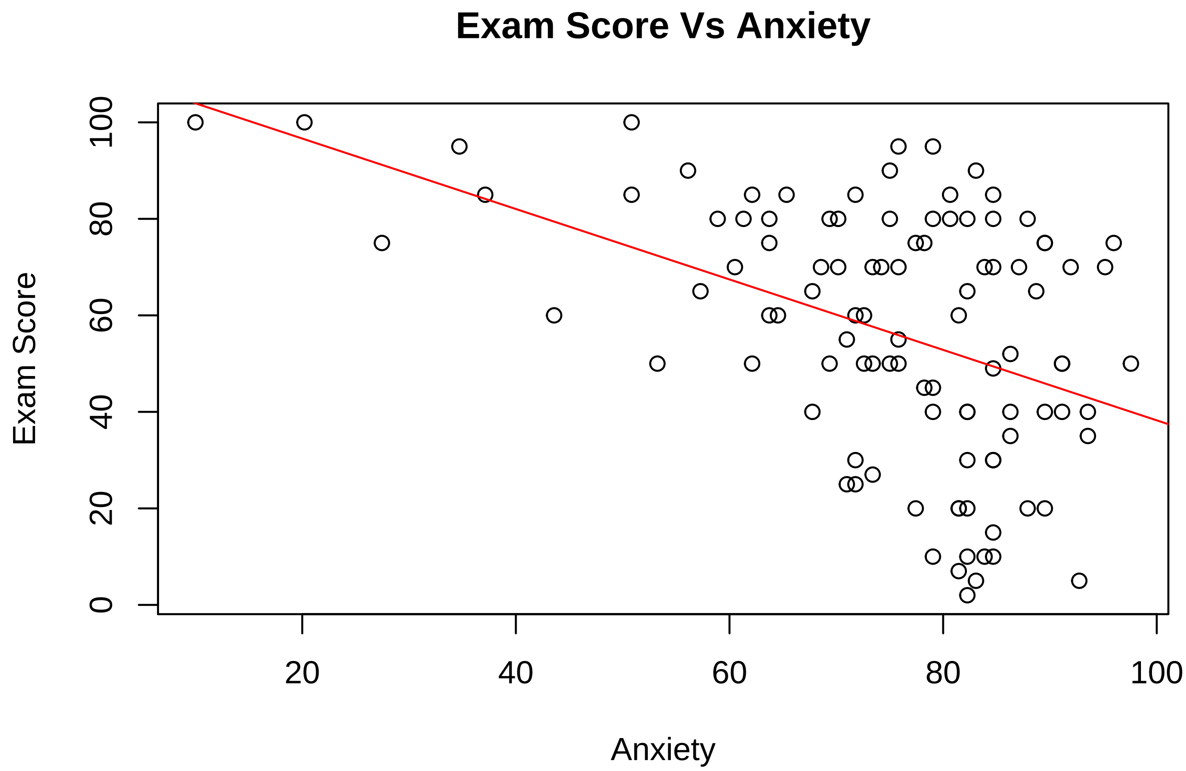
```r
examanxiety=read.table("examanxiety.txt",header=T)
#View(examanxiety)
```

a)

```r
x=examanxiety$Anxiety
y=examanxiety$Exam
r = cor(x, y)
b = r * sd(y) / sd(x)
a = mean(y) - b * mean(x)
#print(c(a,b))
yhat = a + b*((x))
print(paste("The regression line is given by, yhat =",round(a,3),"+", "(",round(b,3),")","* x"," which i
```

```
## [1] "The regression line is given by, yhat = 111.244 + ( -0.73 ) * x  which is in the form y_hat= in
```

```r
plot(x,y,main='Exam Score Vs Anxiety',xlab='Anxiety',ylab='Exam Score')
abline(a,b,col='red')
```



b)

Interpretation of Intercept: When anxiety score is zero, then the predicted or mean exam score is 111.24.

Interpretation of Slope: When anxiety score increases by one unit, the predicted or mean exam score decreases by the factor of slope i.e, 0.73.

Interpretation of yhat: yhat is the regression line which represents the predicted exam score for a given anxiety score.