ST2137 Tutorial 9

Wei Hao Khoong

17 April 2019

Q1

Part a

PDF of t-Distribution

```
fcn <- function(x, n){
  (gamma((n+1)/2)/gamma(n/2))*(1/sqrt(n*pi))*(1/((1+(x^2)/n)^((n+1)/2)))
}
```

CDF When n=5, x=2.5

```
integrate(fcn, lower=-Inf, upper=2.5, n=5)
```

```
## 0.972755 with absolute error < 2.9e-06
```

Part b

CDF with R function pt with n=4

```
integrate(pt, lower=-Inf, upper=2.5, df=4)
```

```
## 2.538434 with absolute error < 4.1e-05
```

Q2

Data

```
t9q2 <- read.table("C:/Users/Wei Hao/Desktop/ST2137/Tutorials/Data/beta30.txt", header=FALSE)
```

Calculation of Beta Distribution Parameters

Mean of Sample

```
mu <- mean(t9q2$V1)
mu
```

```
## [1] 0.3806915
```

Variance of Sample

```
vars <- var(t9q2$V1)
vars
```

```
## [1] 0.03782419
```

Function to Compute Parameters of Beta Distribution

```
estBetaParams <- function(mu, var) {
  alpha <- ((1 - mu) / var - 1 / mu) * mu ^ 2
  beta <- alpha * (1 / mu - 1)
  return(params = list(alpha = alpha, beta = beta))
}
```

```
params <- estBetaParams(mu, vars)
```

```
alpha <- params$alpha
beta <- params$beta
```

Log-likelihood Function

```
LL <- function(theta, slogx, sloglx, n){
  alpha <- theta[1]
  beta <- theta[2]
  loglik <- n*(log(gamma(alpha + beta)) - log(gamma(alpha)) - log(beta)) + (alpha - 1)*slogx + (beta -1)
*sloglx
  return(-loglik)
}
```

```
n <- 30
x <- rbeta(n, shape1=alpha, shape2=beta)
theta.start <- c(1,1)
# out <- optim(theta.start, LL, slogx=sum(log(x)), sloglx=sum(log(1-x)), n=n)
```

Q3

Data Import

```
t9q3 <- read.table("C:/Users/Wei Hao/Desktop/ST2137/Tutorials/Data/rent.txt", header=TRUE)
```

The Linear Model

```
rent <- t9q3$rent
size <- t9q3$size
model1 <- lm(rent~size)
model1
```

```
##
## Call:
## lm(formula = rent ~ size)
##
## Coefficients:
## (Intercept)         size
##     177.121        1.065
```

So we have the fitted model: $\hat{Rent} = 177.121 + 1.065 \cdot size$.

One-Way ANOVA

```
anova(model1)
```

```
## Analysis of Variance Table
##
## Response: rent
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## size       1 2268777 2268777  59.914 7.518e-08 ***
## Residuals 23  870949   37867
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value is $7.5 \times 10^{-8} << 0.05$, we reject $H_0 : \beta_1 = 0$ and conclude that there is evidence of a linear relationship between the size of the apartment and monthly rent.

Summary Statistics

```
summary(model1)
```

```
##
## Call:
## lm(formula = rent ~ size)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -442.26  -58.86  -15.42  104.17  365.13
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 177.1208   161.0043    1.10    0.283
## size          1.0651     0.1376    7.74 7.52e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 194.6 on 23 degrees of freedom
## Multiple R-squared:  0.7226, Adjusted R-squared:  0.7105
## F-statistic: 59.91 on 1 and 23 DF,  p-value: 7.518e-08
```

From the summary statistics, we observe that $R^2 = 0.7226$.

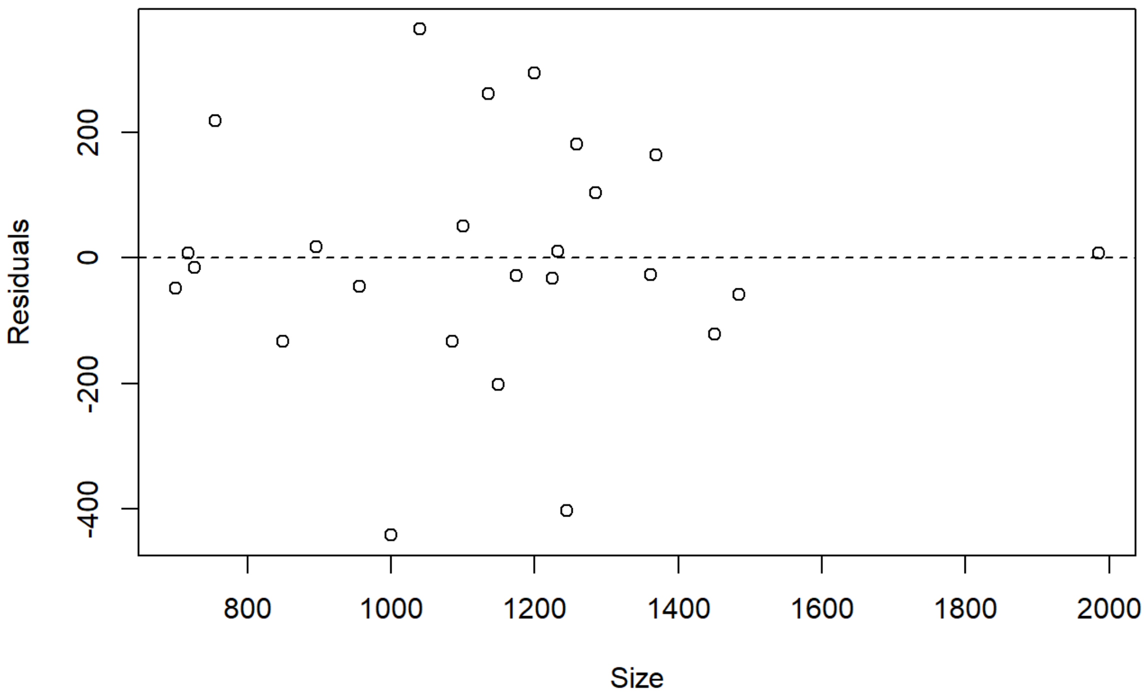
Plot of Residuals Against Fitted Values

Scatter Plot

```
par(mfrow=c(2,2))
plot(rent~size, pch=16)
abline(model1,lty=2)
title( "Scatter plot and Regression Line")
```

Residual Plot

```
rs <- model1$resid
fv <- model1$fitted
plot(rs~size, xlab="Size", ylab="Residuals")+
abline(h=0,lty=2)
```

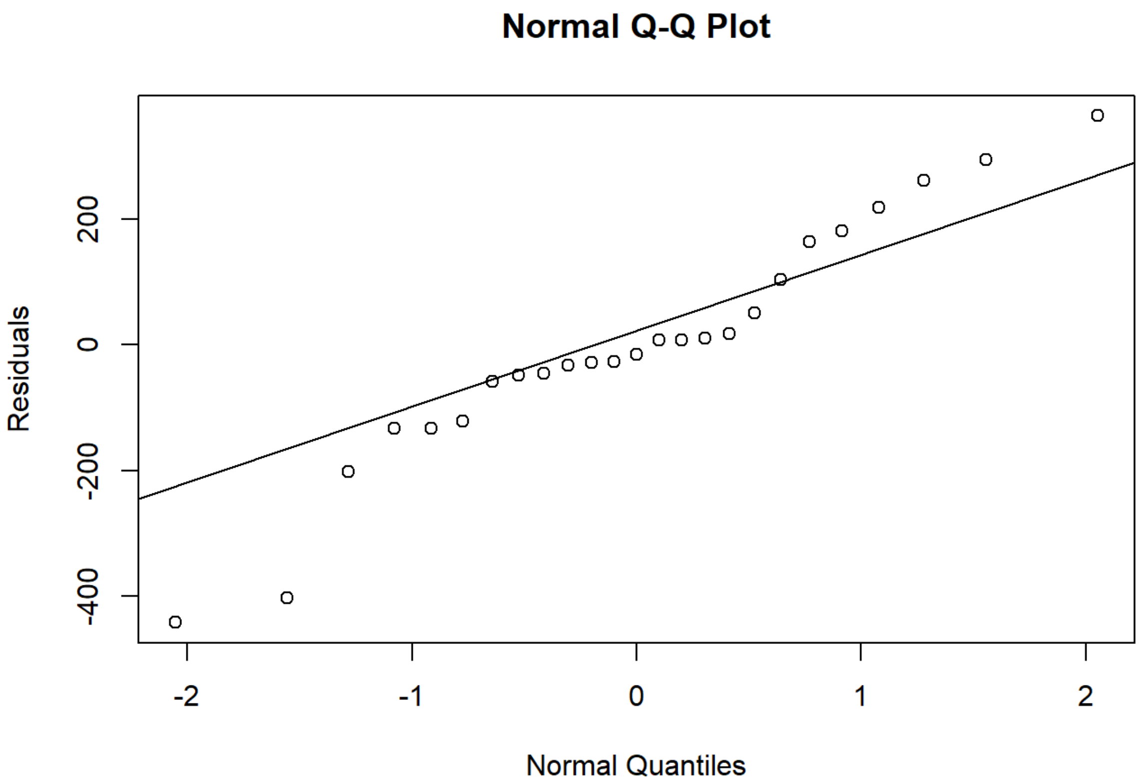



```
## integer(0)
```

From the residual plot, there is no obvious pattern observed.

Normal QQ-Plot

```
qqnorm(rs,ylab="Residuals",xlab="Normal Quantiles")
qqline(rs)
```



```
par(mfrow=c(1,1))
```

From the normal QQ-plot, we observe that the residuals are close to the fitted line, showing signs of normality. So we try to validate this with the KS test.

KS Test

```
ks.test(rs, "pnorm", mean(rs), sd(rs))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  rs
## D = 0.1413, p-value = 0.6493
## alternative hypothesis: two-sided
```

Since the p-value from the KS test is 0.6493, we do not reject the null and conclude that there is no evidence against normality assumption.

Predicting The Monthly Rental Cost With 1000 Square Feet (Size)

```
model1$coefficients[1] + 1000*model1$coefficients
```

```
## (Intercept)        size
##  177297.941    1242.265
```

So the predicted cost is $\$1242.265$.

Recommendation of Apartment

Your friends Jim and Jennifer are considering signing a lease for an apartment in this residential neighborhood. They are trying to decide between two apartments, one with 1000 square feet, for a monthly rent of $\$1275$, and the other with 1200 square feet, for a monthly rent of $\$1425$. What would you recommend to them? Why?

```
# predicted value for 1200 size
model1$coefficients[1] + 1200*model1$coefficients
```

```
## (Intercept)        size
##  212722.105    1455.294
```

Since the predicted value for an apartment with 1000 square feet is $\$1242.265 < \$1275$ (current cost) and for an apartment with 1200 square feet is $\$1455.294 > \$1425$ (current cost), the 1200 square feet apartment is a better option as its cost is lower than its estimated cost.