Testing.R

monca016

Sun Sep 30 16:23:38 2018

```
## ex 8.4.10

## sample mean - H0 mean
250000 - 210250
```

```
## [1] 39750
```

```
## diff divided by square root of sample size 5
37500 / sqrt(5)
```

```
## [1] 16770.51
```

```
## t-value - rejection region is (1.533, infinity)
39750 / 16770.51
```

```
## [1] 2.370232
```

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 3.4.4
```

```
Congress <- read_excel("Piracy.xlsx", col_names = TRUE)
# two-tailed test of H0: mu(Years) = 10
t.test(Congress$years, alternative = "two.sided", mu = 10)
```

```
##
##  One Sample t-test
##
## data:  Congress$years
## t = 4.1597, df = 533, p-value = 3.713e-05
## alternative hypothesis: true mean is not equal to 10
## 95 percent confidence interval:
##  10.9290 12.5916
## sample estimates:
## mean of x
##   11.7603
```

```
# one-tailed test of H0: mu(money_pro) <= $25,000
t.test(Congress$money_pro, alternative = "greater", mu = 25000)
```

```
##
##  One Sample t-test
##
## data:  Congress$money_pro
## t = 0.27262, df = 533, p-value = 0.3926
## alternative hypothesis: true mean is greater than 25000
## 95 percent confidence interval:
##  21792.3     Inf
## sample estimates:
## mean of x
##  25635.94
```

```
# one-tailed test of H0: mu(money_con) >= 25,000
t.test(Congress$money_con, alternative = "less", mu = 25000)
```

```
##
##  One Sample t-test
##
## data:  Congress$money_con
## t = -1.7363, df = 533, p-value = 0.04155
## alternative hypothesis: true mean is less than 25000
## 95 percent confidence interval:
##      -Inf 24830.33
## sample estimates:
## mean of x
##  21672.77
```

```
# two-tailed test for H0: (proportion with stance = "yes") = .10
table(Congress$stance)
```

```
##
## leaning no         no  undecided    unknown        yes
##         44        122         11        294         63
```

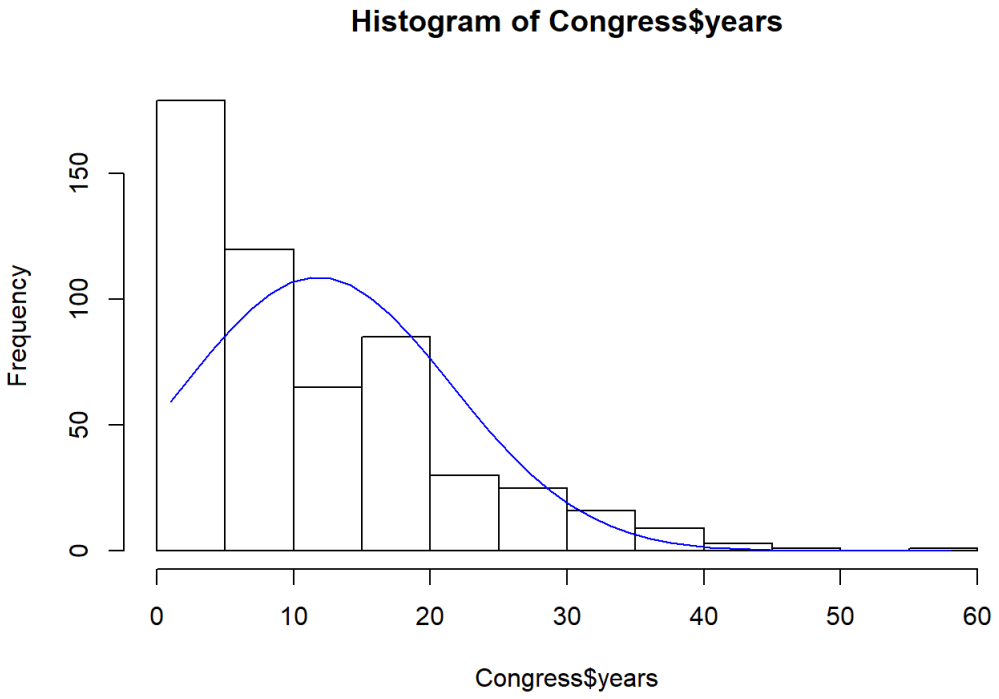
```
p_hat <- 63/534

binom.test(63, 534, p = .10, alternative = "two.sided")
```

```
##
##  Exact binomial test
##
## data:  63 and 534
## number of successes = 63, number of trials = 534, p-value = 0.1702
## alternative hypothesis: true probability of success is not equal to 0.1
## 95 percent confidence interval:
##  0.09185852 0.14841005
## sample estimates:
## probability of success
##              0.1179775
```

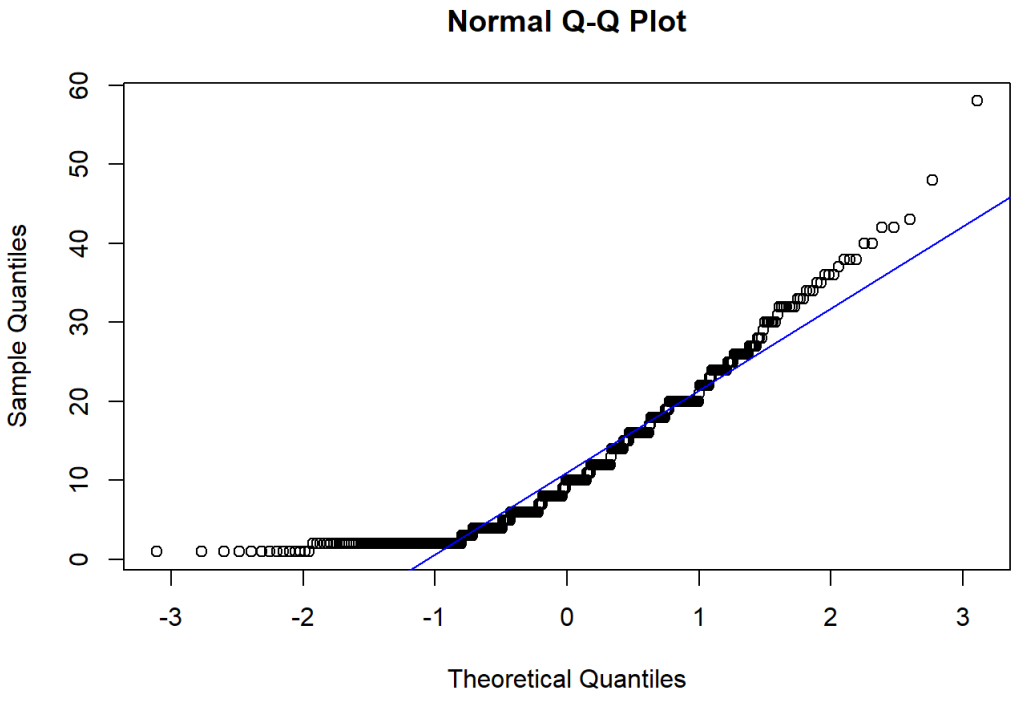
```
prop.test(63, 534, p = .10, alternative = "two.sided")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  63 out of 534, null probability 0.1
## X-squared = 1.7231, df = 1, p-value = 0.1893
## alternative hypothesis: true p is not equal to 0.1
## 95 percent confidence interval:
##  0.09247091 0.14912640
## sample estimates:
##         p
## 0.1179775
```

```
# Checking normality of Years distribution
  # histogram: years, with normal curve
h <- hist(Congress$years)
  # code to add normal curve
x <- Congress$years
xfit <- seq(min(x), max(x), length = 40)
yfit <- dnorm(xfit, mean = mean(x), sd = sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col = "blue")
```




```
# probablity plot
qqnorm(x)
qqline(x, col = "blue")
```



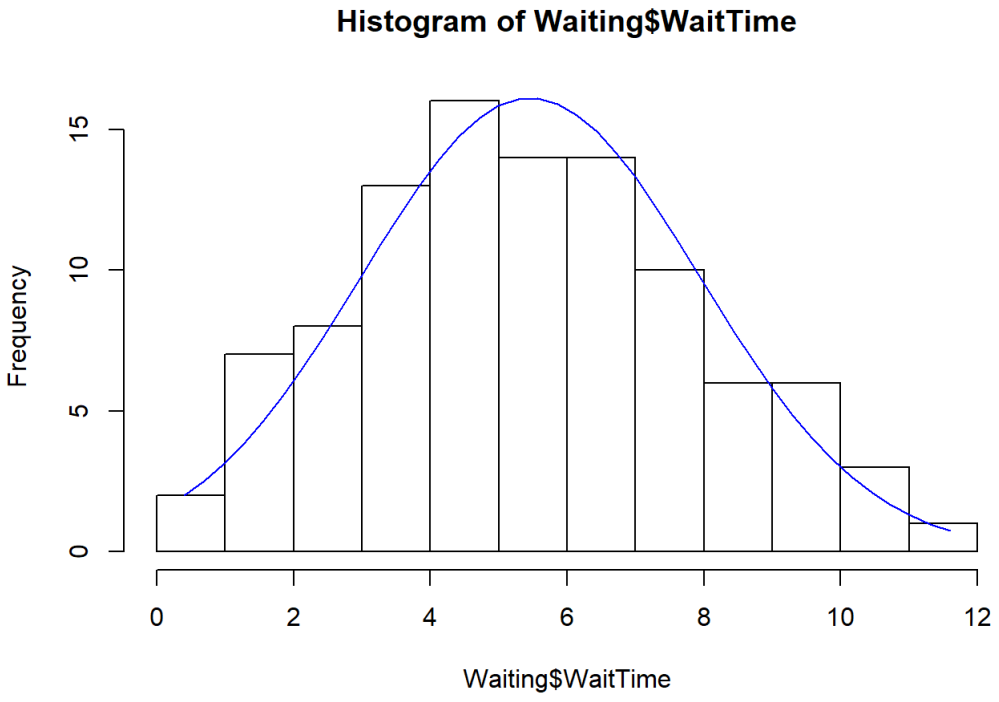

```
# goodness of fit test of H0: normal
# but interpret with caution for large data sets
shapiro.test(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.8852, p-value < 2.2e-16
```

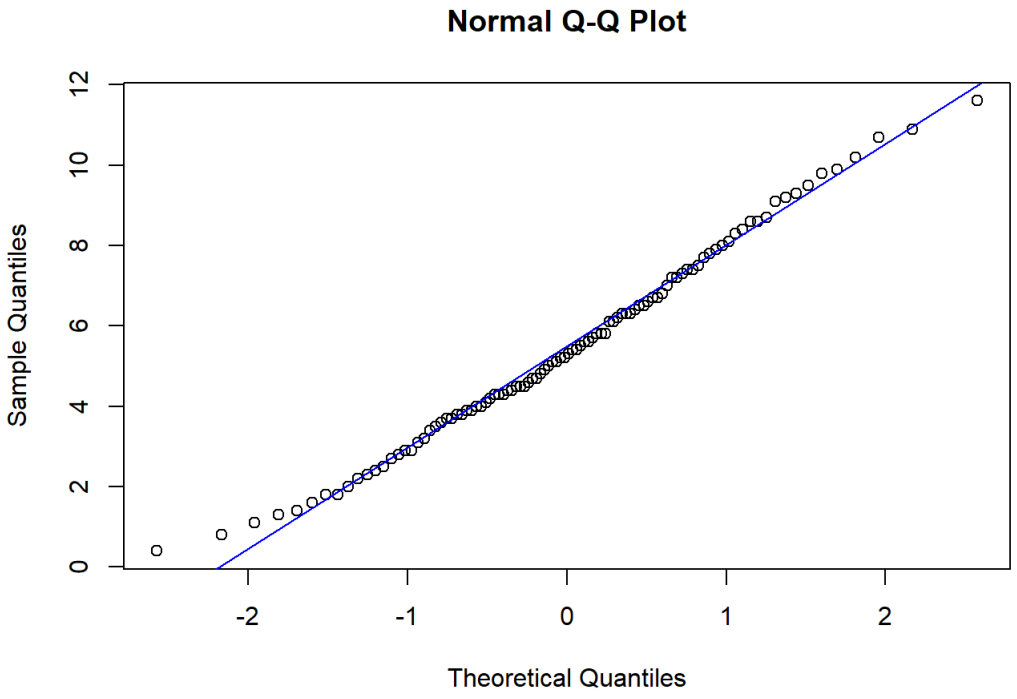
```
Waiting <- read_excel("WaitTime.xlsx", col_names = TRUE)

# one-tailed test of H0: mu(wait time) >= 5 mins
t.test(Waiting$WaitTime, alternative = "greater", mu = 5)
```

```
##
##  One Sample t-test
##
## data:  Waiting$WaitTime
## t = 1.8582, df = 99, p-value = 0.03305
## alternative hypothesis: true mean is greater than 5
## 95 percent confidence interval:
##  5.048978      Inf
## sample estimates:
## mean of x
##      5.46
```

```
# Checking to see if wait times are normally distributed
h <- hist(Waiting$WaitTime)

x <- Waiting$WaitTime
xfit <- seq(min(x), max(x), length = 40)
yfit <- dnorm(xfit, mean = mean(x), sd = sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col = "blue")
```




```
qqnorm(x)
qqline(x, col = "blue")
```




```
# R 4.2 Smoking

# Read in Smoking scores
Smoking <- read_excel("Smoking.xlsx", col_names = TRUE)

# Create a table based on male / female, smoking y or n
table(Smoking$sex, Smoking$smoke)
```

```
##
##           No Yes
##   Female 731 234
##   Male   539 187
```

```
p_hat_fem <- 731 / 965

p_hat_male <- 539 / 726

# testing hypothesis, one-tailed test of H0: p(females who smoke) = 25%
prop.test(731, 965, p = .25, alternative = "two.sided")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  731 out of 965, null probability 0.25
## X-squared = 1322.9, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.25
## 95 percent confidence interval:
##  0.7289496 0.7839952
## sample estimates:
##        p
## 0.757513
```

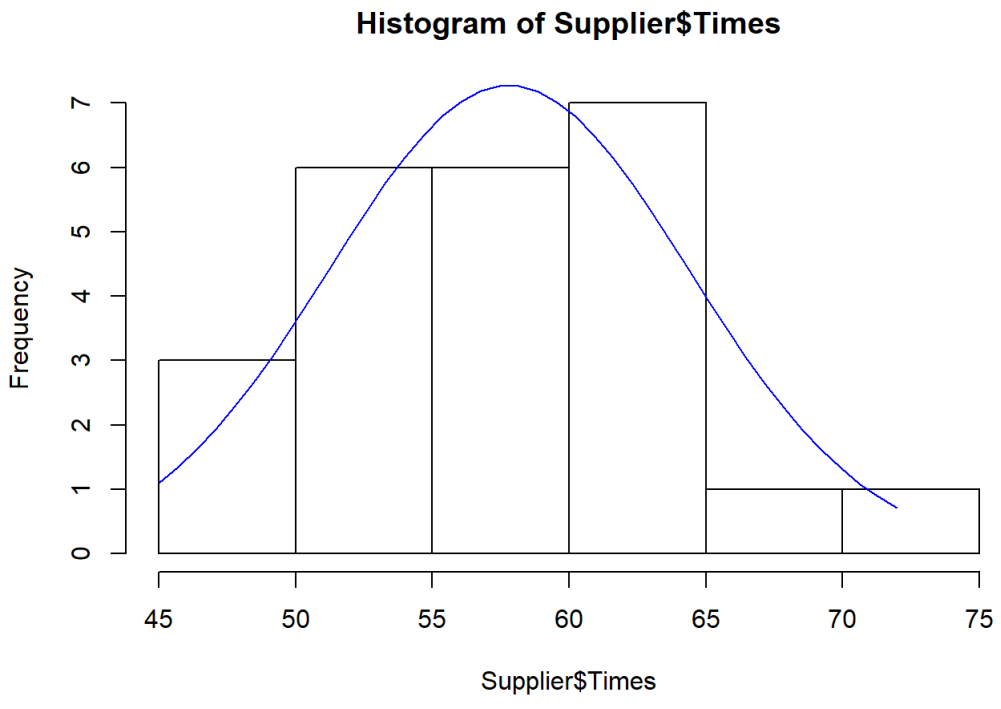
```
# testing hypothesis, one-tailed test of H0: p(males who smoke) = 25%
prop.test(539, 726, p = .25, alternative = "two.sided")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  539 out of 726, null probability 0.25
## X-squared = 936.26, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.25
## 95 percent confidence interval:
##  0.7086843 0.7735572
## sample estimates:
##         p
## 0.7424242
```

```
# R 4.3 Supplier
Supplier <- read_excel("Supplier.xlsx", col_names = TRUE)

# testing hypothesis, one-tailed test of H0: mu(delivery time) = 60 minutes
t.test(Supplier$Times, alternative = "less", mu = 60)
```

```
##
##  One Sample t-test
##
## data:  Supplier$Times
## t = -1.644, df = 23, p-value = 0.05688
## alternative hypothesis: true mean is less than 60
## 95 percent confidence interval:
##      -Inf 60.09379
## sample estimates:
## mean of x
##  57.79167
```

```
h <- hist(Supplier$Times)

x <- Supplier$Times
xfit <- seq(min(x), max(x), length = 40)
yfit <- dnorm(xfit, mean = mean(x), sd = sd(x))
yfit <- yfit*diff(h$mids[1:2])*length(x)
lines(xfit, yfit, col = "blue")
```



```
qqnorm(x)
qqline(x, col = "blue")
```