Group 2

Silver Lee-A-Fong Jakub Lewkowicz Domantas Pukelevičius

2023-02-20

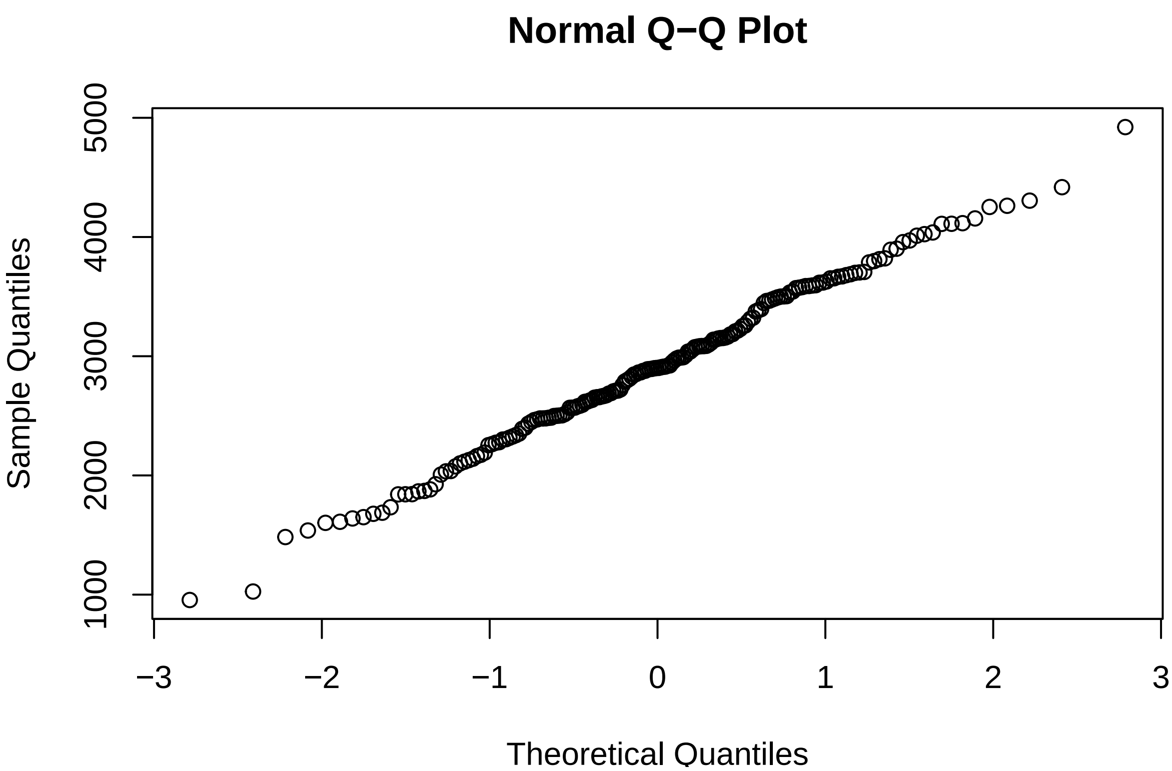
Exercise 1. Birthweights

```
df <- read.csv('Data/birthweight.txt')
shapiro.test(df$birthweight)[2]
```

a)

```
## $p.value
## [1] 0.8995
```

```
qqnorm(df$birthweight)
```




```
n = nrow(df)
mu = mean(df$birthweight)
s = sd(df$birthweight)
z_98p = 2.05 # value from z score table for 98th percentile
m = z_98p*s/sqrt(n) # m = 1.96s/sqrt(n)
bounded_CI = c(mu - m, mu + m); bounded_CI #bounded 96% CI for mu
```

```
## [1] 2809.0082 3017.5769
```

```r
get_m = function(n) {
  s = sd(df$birthweight)
  z_98p = 2.05 # value from z score table for 98th percentile
  m = z_98p*s/sqrt(n)

  return(m)
}

for (sample_size in 1:1000) {
  lower_bound = mu - get_m(sample_size)
  upper_bound = mu + get_m(sample_size)
  CI_length = upper_bound - lower_bound
  if (CI_length <= 100) {
    break
  }
}
sample_size
```

```
## [1] 818
```

```r
# bootstrap 96%-CI:
B = 1000
Tstar = 1:B
for (i in 1:B){
  Xstar = sample(df$birthweight, replace=TRUE)
  Tstar[i] = mean(Xstar)
}
Tstar20 = quantile(Tstar, 0.020)
Tstar980 = quantile(Tstar, 0.980)
sum(Tstar<Tstar20)
```

```
## [1] 20
```

```r
bootstrap_CI = c(2*mu-Tstar980,2*mu-Tstar20)
bootstrap_CI
```

```
##     98%      2%
## 2809.17 3025.00
```

```r

# H0 mean <= 2800
t.test(df$birthweight, mu=2800, alt="g")
```

b)

```
##
##  One Sample t-test
##
## data:  df$birthweight
## t = 2.227, df = 187, p-value = 0.0136
## alternative hypothesis: true mean is greater than 2800
## 95 percent confidence interval:
##  2829.2    Inf
## sample estimates:
## mean of x
```

```
##   2913.29
```

```r
# p value 0.01357 means that H0 has to be rejected in favor of h1
# which means that true mean is greater than 2800

# sign test
binom.test(sum(df$birthweight > 2800), length(df$birthweight), alt='l')[3]
```

```
## $p.value
## [1] 0.975678
```

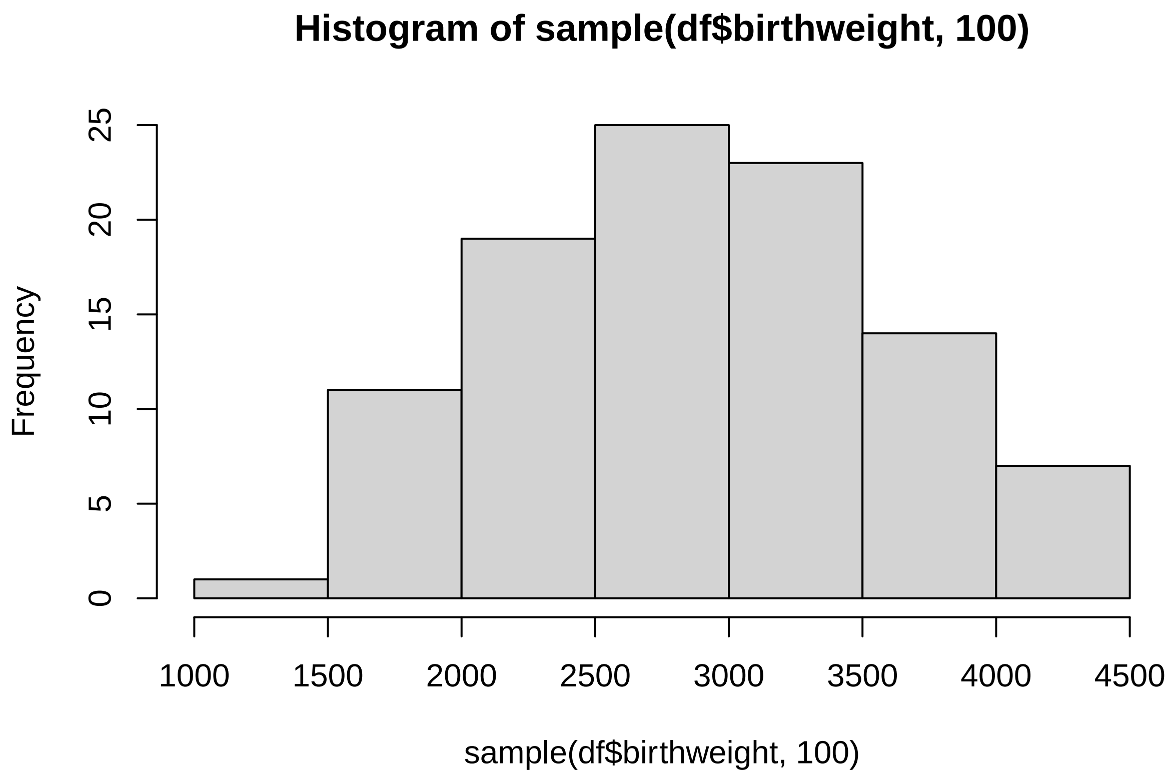
```r

# power of t-test and sign test
B = 1000
psign = numeric(B)
pttest = numeric(B)
n = 50
for(i in 1:B) {
  x = sample(df$birthweight, n)
  psign[i] = binom.test(sum(x>2800), n, alt='g')[[3]]
  pttest[i] = t.test(x, mu=2800, alt='g')[[3]]
}
power_sign = sum(psign<0.05)/B
power_ttest = sum(pttest<0.05)/B
c(power_sign, power_ttest)
```

c)

```
## [1] 0.163 0.283
```

```r
# t-test power (probability of rejecting H0) is bigger, because t-test works better for normal data
hist(sample(df$birthweight, 100))
```



####

d)

```r
n = 100
p_lower = 0.25
sample_probabilities = numeric(n)
for(i in 1:n){
  x = sample(df$birthweight, n)
  sample_probabilities[i] = sum(x < 2600)/n
}
s = sd(sample_probabilities)
p_estimate = mean(sample_probabilities)
# Using asymptotic normality, the expert computed the left end  =0.25
# of the confidence interval for p
# so we know that p_l = p_estimate - m = 0.25
# we also know that m = z_alpha * s/sqrt(n) and m = p_estimate - 0.25
m = p_estimate - p_lower
p_upper = p_estimate + m
c(p_lower, p_estimate, p_upper)
```

```
## [1] 0.250 0.327 0.404
```

```r
p_val = numeric(100)
for(i in 1:100) {
  males_u2600 = 34
  females_u2600 = 28
  males_a2600 = 61
  females_a2600 = 65
  under_2600 = df$birthweight[df$birthweight < 2600]
  above_2600 = df$birthweight[df$birthweight > 2600]
  under_2600
```

```r
  samples_males_u2600_i = sample(1:length(under_2600), males_u2600)
  samples_males_u2600 = under_2600[samples_males_u2600_i]
  samples_females_u2600 = under_2600[-samples_males_u2600_i]
  samples_males_a2600_i = sample(1:length(above_2600), males_a2600)
  samples_males_a2600 = above_2600[samples_males_a2600_i]
  samples_females_a2600 = above_2600[-samples_males_a2600_i]

  samples_males = c(samples_males_a2600, samples_males_u2600)
  samples_females = c(samples_females_a2600, samples_females_u2600)
  p_val[i] = t.test(samples_males, samples_females)[[3]]
}
mean(p_val)
```

e)

```
## [1] 0.485217
```