Homework 1: Solutions

Due: Thursday 1/30/20 by 8:30am

Rubric:

- 1 point for 1. if "I have installed R" was printed/written somewhere on the homework.
- Maximum of 3 points for 2., determined as follows:
  - 0 points for no solutions whatsoever or R output only;
  - 1 point for an honest effort but very few correct answers or R output only plus a figure;
  - 2 points for mostly correct answers but at least one substantial issue;
  - 3 points for nearly/exactly correct.
- Maximum of 2 points each for 3. and 4., determined as follows:
  - 0 points for no solutions whatsoever or R output only;
  - 1 point for an honest effort but at least one substantial issue;
  - 2 points for nearly/exactly correct.

2.

(a) Obtain a 90 percent confidence interval for μ when $n = 424$, $\bar{Y} = 6.09$, and $s = 2.78$.

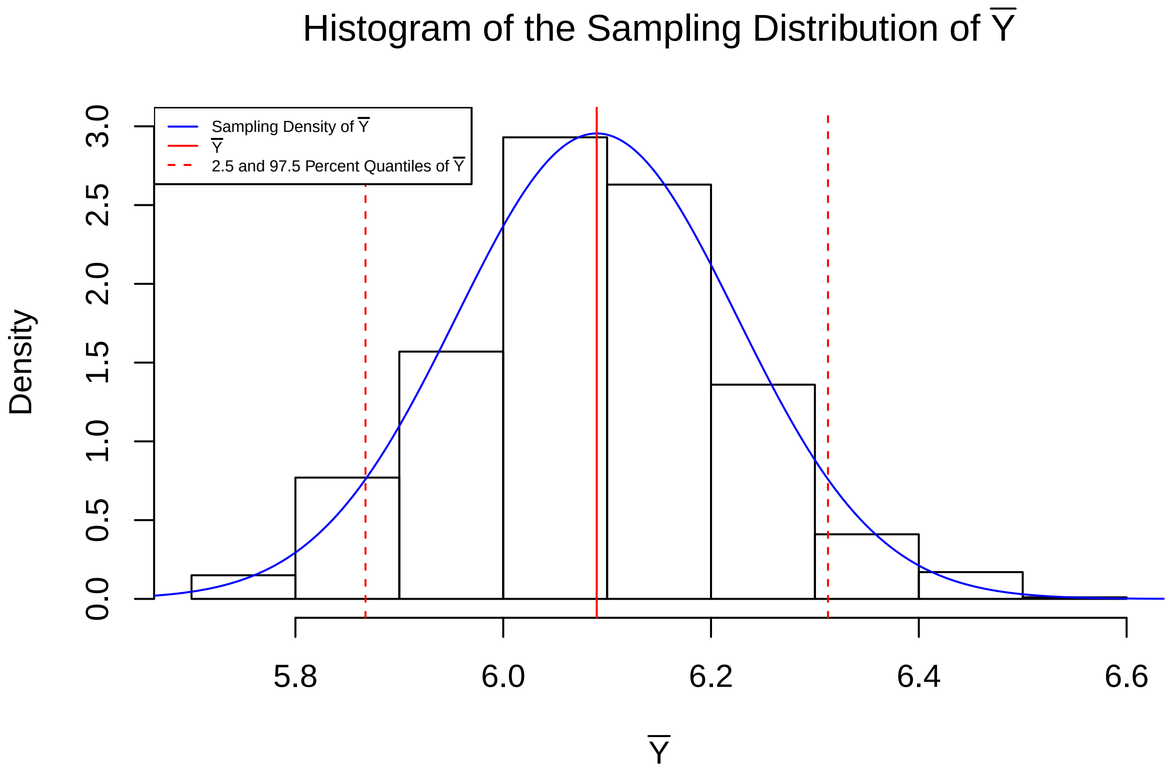
```
n <- 424
y.bar <- 6.09
s <- 2.78

alpha <- 0.1


int <- y.bar + qt(c(alpha/2, 1 - alpha/2),
                  df = n - 1)*s/sqrt(n)
```

We obtain a 95 percent confidence interval of (5.87, 6.31).

(b) Simulate 1,000 draws from the sampling distribution of $\bar{Y}$. Plot a histogram of $\bar{Y}$ on the density scale. Add vertical lines at $\bar{Y}$ and the upper and lower bounds of the interval derived in (a), and overlay the density of the sampling distribution of $\bar{Y}$ on the histogram.

```
sims <- rnorm(1000, mean = y.bar, sd = s/sqrt(n))
hist(sims, freq = FALSE,
     xlab = expression(bar(Y)),
     main = expression(paste("Histogram of the Sampling Distribution of ",
                             bar(Y), sep = "")),
     ylim = c(0, 3))
vals <- seq(5, 7, length.out = 500)
lines(vals, dnorm(vals, mean = y.bar, sd = s/sqrt(n)),
      col = "blue")
abline(v = y.bar, col = "red")
abline(v = int, col = "red", lty = 2)
legend("topleft",
       col = c("blue", "red", "red"),
       lty = c(1, 1, 2),
       legend = c(expression(paste("Sampling Density of ",
```

```
                                  bar(Y), sep = "")),
                 expression(bar(Y)),
                 expression(paste("2.5 and 97.5 Percent Quantiles of ",
                                  bar(Y), sep = ""))),
       cex = 0.5, bg = "white")
```



3. Choose between the alternatives $H_0 : \mu \geq 5$ and $H_a : \mu < 5$, where α is to be controlled at 0.03 and $n = 84$, $\bar{Y} = 5.08$, and $s = 9.51$. Justify your answer with reference to the value of the test statistic, the decision rule, and the p-value.

```
alpha <- 0.03

n <- 84
y.bar <- 5.08
s <- 9.51

t.star <- (y.bar - 5)/(s/sqrt(n))
int <- qt(alpha,
          df = n - 1)
pval <- pt(t.star, df = n - 1)
```

We fail to reject the null hypothesis at level $\alpha = 0.03$, because the test statistic is equal to 0.08, which is inside of the acceptance region (-1.91, ∞), and the p-value, which is 0.53 is much higher than $\alpha = 0.03$.

4. Choose between the alternatives $H_0 : \mu = -10$ and $H_a : \mu \neq -10$ when α is to be controlled at 0.15 and $n = 13$, $\bar{Y} = -6.81$, and $s = 1.55$. Justify your answer with reference to the value of the test statistic, the decision rule, and the p-value.

```
alpha <- 0.15

n <- 13
```

```r
y.bar <- -6.81
s <- 1.55

t.star <- (y.bar - 10)/(s/sqrt(n))
int <- qt(c(alpha/2, 1 - alpha/2),
          df = n - 1)
pval <- 2*pt(abs(t.star), df = n - 1, lower.tail = FALSE)
```

We reject the null hypothesis at level $\alpha = 0.15$, because the test statistic is equal to -39.1, which is outside of the acceptance region (-1.54, 1.54), and the p-value, which is 5.046509×10^{-14} is much lower than $\alpha = 0.15$.

5. (In class, not to be submitted)

(a) Obtain a 95 percent confidence interval for $\mu_1 - \mu_2$ when $n_1 = 49$, $\bar{Y} = 1.1$, $\sum \left(Y_i - \bar{Y}\right)^2 = 96.04$, $n_2 = 34$, $\bar{Z} = 3.37$, and $\sum \left(Z_i - \bar{Z}\right)^2 = 22.78$.

```r
n1 <- 49
y.bar <- 1.1
yiybarsq <- 94.04
n2 <- 34
z.bar <- 3.37
zizbarsq <- 22.78

alpha <- 0.05

ssq <- (yiybarsq + zizbarsq)/(n1 + n2 - 2)

int <- y.bar - z.bar + qt(c(alpha/2, 1 - alpha/2),
                   df = n1 + n2 - 2)*sqrt(ssq/n1 + ssq/n2)
```

We obtain a 95 percent confidence interval of (-2.8, -1.74).

(b) Choose between the alternatives $H_0 : \mu_1 = \mu_2$ and $H_a : \mu_1 \neq \mu_2$ when α is to be controlled at 0.01. Justify your answer with reference to the value of the test statistic, the decision rule, and the p-value.

```r
alpha <- 0.01

t.star <- (y.bar - z.bar)/(sqrt(ssq/n1 + ssq/n2))
int <- qt(c(alpha/2, 1 - alpha/2),
          df = n1 + n2 - 2)
pval <- 2*pt(abs(t.star), df = n1 + n2 - 2, lower.tail = FALSE)
```

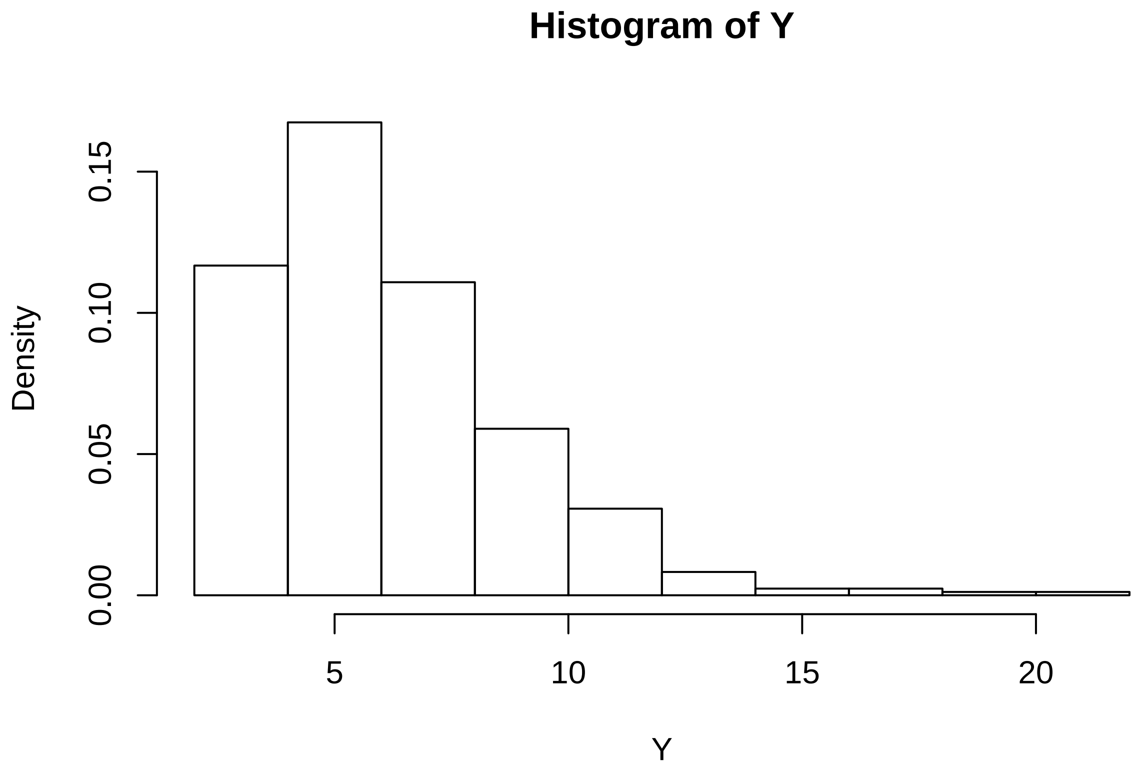
We reject the null hypothesis at level $\alpha = 0.01$, because the test statistic is equal to -8.47, which is outside of the acceptance region (-2.64, 2.64), and the p-value, which is 8.934652×10^{-13} is much lower than $\alpha = 0.01$.

6. (In class, not to be submitted)

(a) Obtain a 85 percent confidence interval for σ^2 when $n = 424$, $\bar{Y} = 6.09$, and $s = 2.78$.

```r
n <- 424
y.bar <- 6.09
s <- 2.78

alpha <- 0.15

int <- (n - 1)*s^2/qchisq(c(1 - alpha/2, alpha/2), n - 1)
```

We obtain a 85 percent confidence interval of (7.02, 8.56).



(b) Suppose a peer who had access to the raw data showed you the above histogram of the values $Y_1, \ldots, Y_n$. In at most one sentence, explain how this would affect your conclusions in (a).

I would be less confident in my conclusions in (a), because confidence intervals for variances are sensitive to violations of normality of the data.

7. (In class, not to be submitted)

(a) Obtain a 99 percent confidence interval for σ_1^2/σ_2^2 using the data from 5.

```
n1 <- 49
y.bar <- 1.1
yiybarsq <- 94.04
n2 <- 34
z.bar <- 3.37
zizbarsq <- 22.78

alpha <- 0.01

ssq1 <- (yiybarsq)/(n1 - 1)
ssq2 <- (zizbarsq)/(n2 - 1)

int <- ssq1/(ssq2)/qf(c(1 - alpha/2, alpha/2), n1 - 1, n2 - 1)
```

We obtain a 99 percent confidence interval of (1.19, 6.4).

(b) Choose between the alternatives $H_0 : \sigma_1^2 = \sigma_2^2$ and $H_a : \sigma_1^2 \neq \sigma_2^2$ when α is to be controlled at 0.2. Justify your answer with reference to the value of the test statistic and the decision rule.

```
alpha <- 0.2
t.star <- ssq1/(ssq2)
int <- qf(c(alpha/2, 1 - alpha/2), n1 - 1, n2 - 1)
```

We reject the null hypothesis at level $\alpha = 0.2$, because the test statistic is equal to 2.84, which is outside of the acceptance region (0.67, 1.53).