HW8

Darnell Chen

2024-11-11

8.1.1

a) A z-score of 2.14 occurs when the lower tail is 0.9838 and the upper tail is 0.0162. Thus our confidence level is simply $1 - 2(0.0162) = 0.9676$.

b) This will be answered with the assumption that the book had a typo that resulted in a third term in the lower bound:

A z-score of 2.49 has a lower tail probability of 0.9936 and upper tail of 0.0064. Thus, our confidence level is simply $1 - 2(0.0064) = 0.9872$.

c) A z-score of 1.85 has a lower tail probability of 0.9678 and upper tail of 0.0322. Thus our confidence level is simply $1 - 2(0.0322) = 0.9356$.

d) The z-score of 2.00 has a lower tail probability of 0.9772 and upper tail probability of 0.0228. However, since we only need to exclude the upper tail probability from out interval, our confidence interval is just 0.9772

e) The z-score of 1.96 has a lower tail probability of 0.9750 and upper tail probability of 0.0250. However, since only need to exclude the lower tail probability, our confidence interval only consist of the range below the mean, which is just 0.9750

8.1.5

a) A 99 CI from the sample data would be longer since it would cover a larger range of numbers (i.e. outliers).

b) No, the answer is incorrect. Our true population mean μ can either be in the interval, or it won't be. Thus, the probability of μ being in our CI is either 0% or 100% since it's a static constant.

c) This answer is partially correct since a 95% CI means that the sample mean would typically fall into the CI 95% of the time. This means if we were to take 1000 means and 1000 CIs of 95%, then approximately 950 of those means would fall into that interval, since each once has a 95% chance of being in the CI. Of course, this all relies on it being a "perfect world" since taking 1000 means and 1000 CIs might not guarantee exactly 950 are within the CI, but rather approximately 950 should.

8.1.7

$\bar{x} = 74.036, \ \sigma = 0.001, \ n = 15$

a)

A two-sided confidence interval of 99% means $\alpha = 0.01$, thus we're looking for:
$\bar{x} - z_{0.005}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{0.005}\frac{\sigma}{\sqrt{n}}$

where:
$z_{0.005} \approx 2.58$

And plugging in our known values, we get:

$$\begin{aligned}\bar{x} \pm z_{0.005}\frac{\sigma}{\sqrt{n}} &= 74.036 \pm 2.58\frac{0.001}{\sqrt{15}} \\ &= 74.036 \pm 0.000666153135548\end{aligned}$$

Thus our two-sided interval with 99% CI is $[74.0353338469, 74.0366661531]$

b) $z_{0.01} \approx 2.33$, where our subscript interval is 0.01 because it's a single-sided bound.

And lower confidence bound is only looking for where $\bar{x} - z_{0.01}\frac{\sigma}{\sqrt{n}} \leq \mu$, thus we get the following:

$$\begin{aligned}\bar{x} - z_{0.01}\frac{\sigma}{\sqrt{n}} &= 74.036 - 2.33\frac{0.001}{\sqrt{15}} \\ &= 74.036 - 0.000601603413111 \\ &= 74.0353983966\end{aligned}$$

So our lower confidence bound is $[74.0353983966, \infty)$.

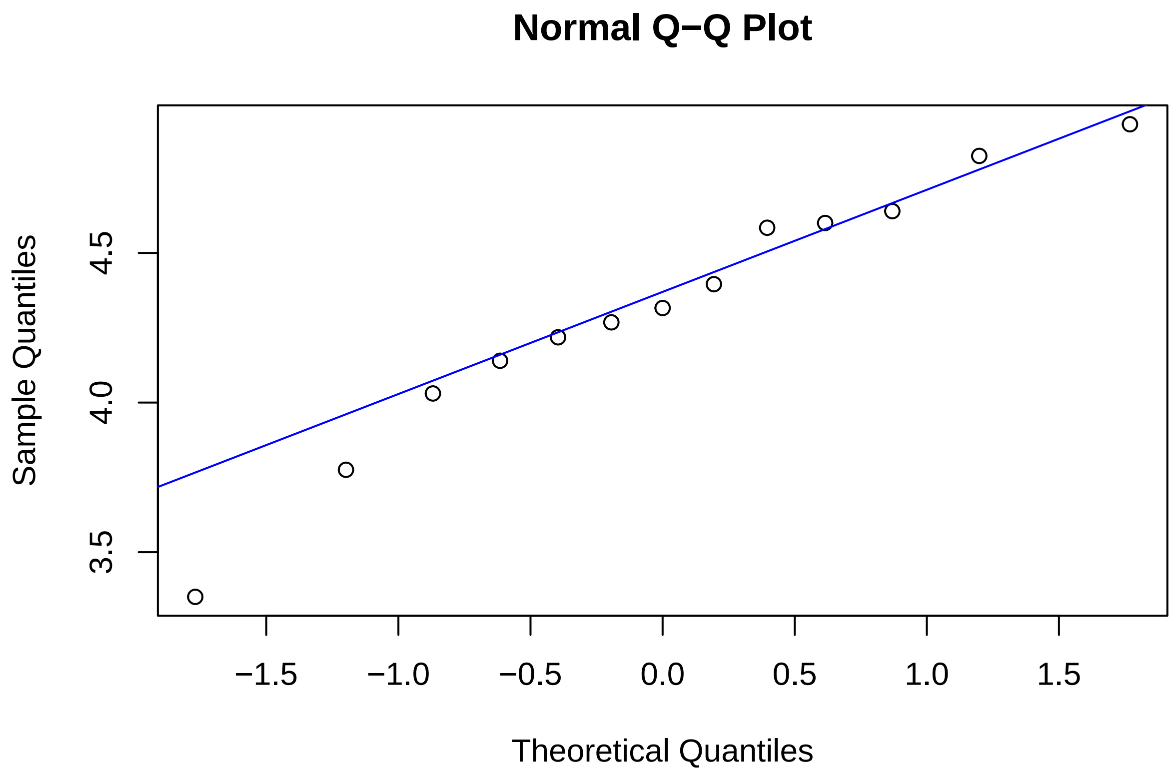
Compared to part (a), our lower bound here is larger since it's compensating for a 1% probability of missing on the lower bound rather than a 0.5% chance.

8.2.10

a) We can test to see if the data is normally distributed by using a QQ-Plot or the Anderson Darling test (both shown below). In our QQ-Plot, most of the data follow our normal line, thus proving that the speed up of CNN is normal. Also, our p-value from the Anderson-Darling test is significant enough for the data to be considered normal.

```r
speedups <- c(3.775302, 3.350679, 4.217981, 4.030324, 4.639692,
              4.139665, 4.395575, 4.824257, 4.268119, 4.584193,
              4.930027, 4.315973, 4.600101)

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




```r
library(nortest)
ad.test(speedups)
```

```
##
##  Anderson-Darling normality test
##
## data:  speedups
## A = 0.23339, p-value = 0.7448
```

b) The mean and standard deviation of our data is:

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=0}^{n} x_i}{n} \\ &= \frac{56.071888}{13} \\ &\approx 4.313222\end{aligned}$$

$$\begin{aligned}\sigma &= \sqrt{\frac{\sum_{i=0}^{n}(x_i - \bar{x})^2}{n-1}} \\ &= \sqrt{\frac{2.2478}{12}} \\ &\approx 0.4328017\end{aligned}$$

A two-sided confidence interval of 95% means we're looking at $z_{0.025} = 1.96$. So we get:

$$
\begin{aligned}
\bar{x} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} &= 4.313222 \pm 1.96\frac{0.4328017}{\sqrt{13}} \\
&= 4.313222 \pm 0.235273684158
\end{aligned}
$$

Thus, our 95% CI is $[4.07794831584, 4.54849568416]$.

c) A single-sided lower bound confidence interval of 95% means we're looking at $z_{0.05} = 1.64$, so we get:

$$
\begin{aligned}
\bar{x} - z_{0.05}\frac{\sigma}{\sqrt{n}} &= 4.313222 - 1.64\frac{0.4328017}{\sqrt{13}} \\
&= 4.313222 - 0.196861654092 \\
&\approx 4.11636034591
\end{aligned}
$$

Thus, our interval for the lower bound confidence interval is $[4.11636034591, \infty)$

```r
mean(speedups)
```

```
## [1] 4.313222
```

```r
sd(speedups)
```

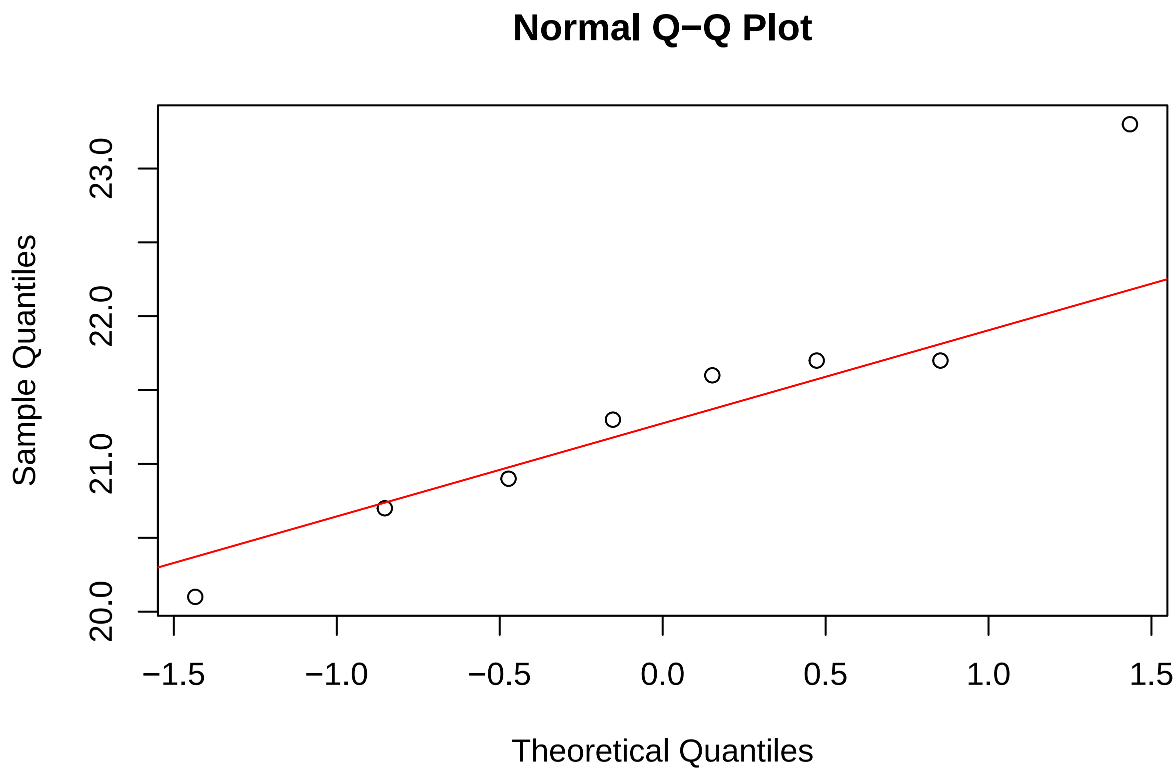
```
## [1] 0.4328017
```

8.3.3

Using the QQ Plot below, we can see that the data is normally distributed as it follows our normal line.

```r
temps <- c(23.3, 21.7, 21.6, 21.7, 21.3, 20.7, 20.9, 20.1)

qqnorm(temps)
qqline(temps, col="red")
```



To calculate mean and standard deviation, we simply find:

$$\begin{aligned}
\bar{x} &= \frac{\sum_{i=0}^{n} x_i}{n} \\
&= \frac{171.3}{8} \\
&= 21.4125
\end{aligned}$$

$$\begin{aligned}
\sigma &= \sqrt{\frac{\sum_{i=0}^{n}(x_i - \bar{x})^2}{n-1}} \\
&= \sqrt{\frac{6.26875}{7}} \\
&\approx 0.9463275
\end{aligned}$$

A 95% CI (assuming it's two-sided confidence interval) will have $z_{0.025} = 1.96$. Thus to find our interval, we simply find:

$$\begin{aligned}
\bar{x} \pm z_{0.025} \frac{\sigma}{\sqrt{n}} &= 21.4125 \pm 1.96 \frac{0.9463275}{\sqrt{8}} \\
&= 21.4125 \pm 0.655771500624
\end{aligned}$$

Thus our 95% CI is [20.75674, 22.06826].

```r
mean(temps)
```

```
## [1] 21.4125
```

```r
sd(temps)
```

```
## [1] 0.9463275
```

```r
mean(temps) - qnorm(.975) * sd(temps) / sqrt(length(temps))
```

```
## [1] 20.75674
```

```r
mean(temps) + qnorm(.975) * sd(temps) / sqrt(length(temps))
```

```
## [1] 22.06826
```

8.4.1

$\hat{p} = \frac{412}{768} \approx 0.536458333333$, since 412 of the 768 graduates voted for Bush.

a) For two-sided 95 CI, we have $z_{0.025} = 1.96$. Thus we get:

$$\begin{aligned}\bar{x} \pm z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.53645833 \pm 1.96\sqrt{\frac{0.53645833(1-0.53645833)}{768}} \\ &\approx 0.53645833 \pm 0.0352685697897\end{aligned}$$

Thus our two-sided 95% CI is [0.50118976021, 0.57172689979].

For our lower bound 95% CI, we simply change the $z-score$ to $z_{0.05} \approx 1.65$:

$$\begin{aligned}\bar{x} - z_{0.05}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.53645833 - 1.65\sqrt{\frac{0.53645833(1-0.53645833)}{768}} \\ &\approx 0.53645833 - 0.0296903776291 \\ &= 0.506767952371\end{aligned}$$

Thus our 95% lower bound CI is: $[0.506767952371, 1]$.