

# STAT 22000: Homework 11

Zixi Li

## Problem 1 Confidence interval and significance level

- (a)  $H_o$  will be rejected at significance level  $\alpha = 0.05$ , because  $\alpha = 0.05$  corresponds to a 95% confidence level, and  $\mu = 3$  is not in the 95% confidence level.
- (b)  $H_o$  will not be rejected at significance level  $\alpha = 0.01$ , because  $\alpha = 0.01$  corresponds to a 99% confidence level, and  $\mu = 3$  is in the 99% confidence level.

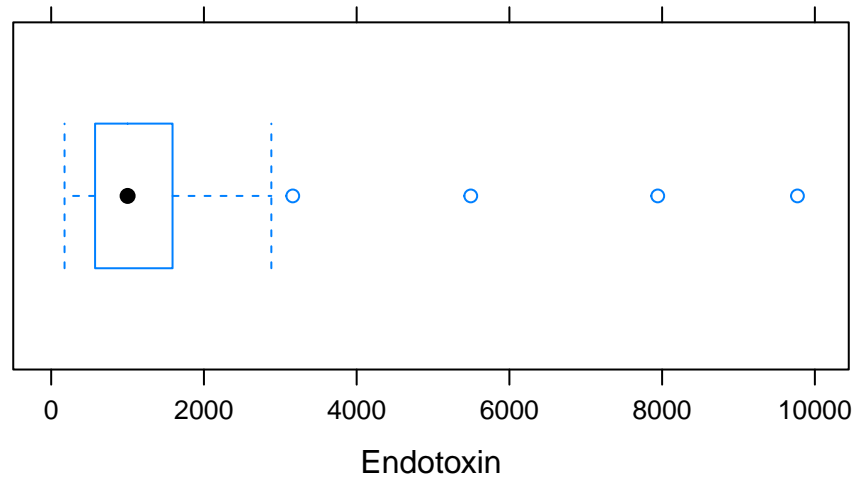
## Problem 2 P-value and confidence interval

- (a)  $\mu = 10$  is not included in the 95% confidence interval, because 95% confidence interval corresponds to 0.05 significance level, and p-value 0.03 will make  $H_0$  be rejected at 0.05 significance level.
- (b)  $\mu = 10$  is included in the 99% confidence interval, because 99% confidence interval corresponds to 0.01 significance level, and p-value 0.03 will make  $H_0$  fail to be rejected at 0.01 significance level.

## Problem 3 Allergies and antigens levels

(a)

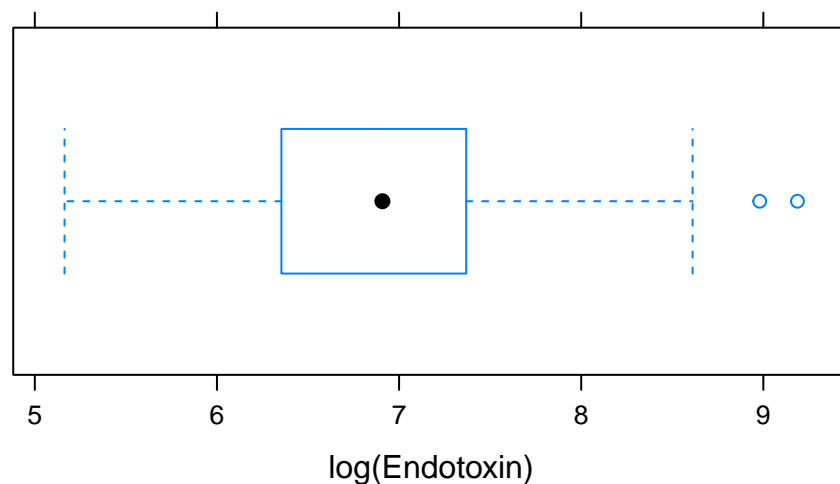
```
Endotoxin = c(708.23, 911.60, 976.81, 1316.63, 262.74, 9772.08,  
              370.76, 229.16, 2570.51, 891.19, 3163.20, 1777.65,  
              1288.57, 436.23, 2631.63, 1173.52, 911.67, 7942.42,  
              740.32, 356.92, 1175.48, 1480.55, 2754.61, 575.62,  
              573.89, 468.26, 1000.71, 364.22, 1025.26, 1022.04,  
              645.41, 363.57, 977.47, 1022.75, 1860.63, 371.13,  
              174.73, 399.68, 1479.77, 2882.96, 601.99, 1697.32,  
              2291.00, 646.49, 1176.27, 1995.43, 955.54, 1480.05,  
              456.71, 1174.70, 5494.22)  
bwplot(Endotoxin, horizontal=T)
```



Based on the above boxplot, we can see there are some outliers, and the shape is right-skewed, so it's not appropriate to construct a t-confidence interval for the mean endotoxin level at the homes of children without allergy.

(b)

```
bwplot(log(Endotoxin), horizontal=T)
```



Based on the above boxplot, we can see there are only two outliers, and the shape is not skewed, so it's more appropriate to construct a t confidence interval for the mean of the log endotoxin level at the homes of children without allergy.

(c)

We can find the sample mean  $\bar{x}=6.91658$  and the sample SD  $s = 0.858441$  and the critical value  $t^* = 2.00856$  using the following R codes:

```
mean(log(Endotoxin))
```

```
## [1] 6.91658
```

```
sd(log(Endotoxin))
```

```
## [1] 0.858441
```

```
qt(0.05/2,df=51-1,lower.tail = F)
```

```
## [1] 2.00856
```

Then the 95% t-confidence interval can be calculated as:

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = 6.91658 \pm 2.00856 \times \frac{0.858441}{\sqrt{51}} = (6.67514, 7.15802)$$

## Problem 4 Measurements of body dimensions

(a)

Here I make a histogram for the height of the 260 women and calculate the population mean and SD using the following R codes:

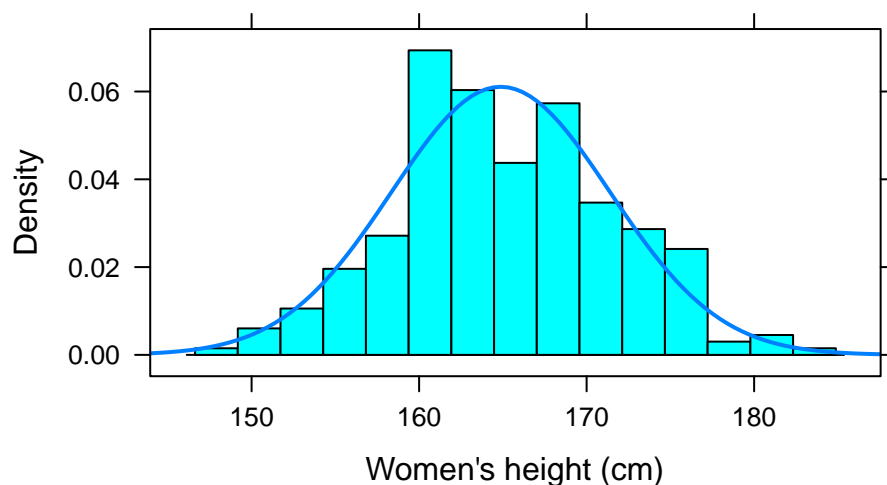
```
download.file("http://www.openintro.org/stat/data/bdims.RData",  
             destfile = "bdims.RData")  
load("bdims.RData")  
fdims = subset(bdims, sex == 0)  
population = fdims$hgt  
mu = mean(population); mu
```

```
## [1] 164.872
```

```
sigma = sd(population); sigma
```

```
## [1] 6.5446
```

```
histogram(population, fit="normal", nint=15, xlab="Women's height (cm)")
```

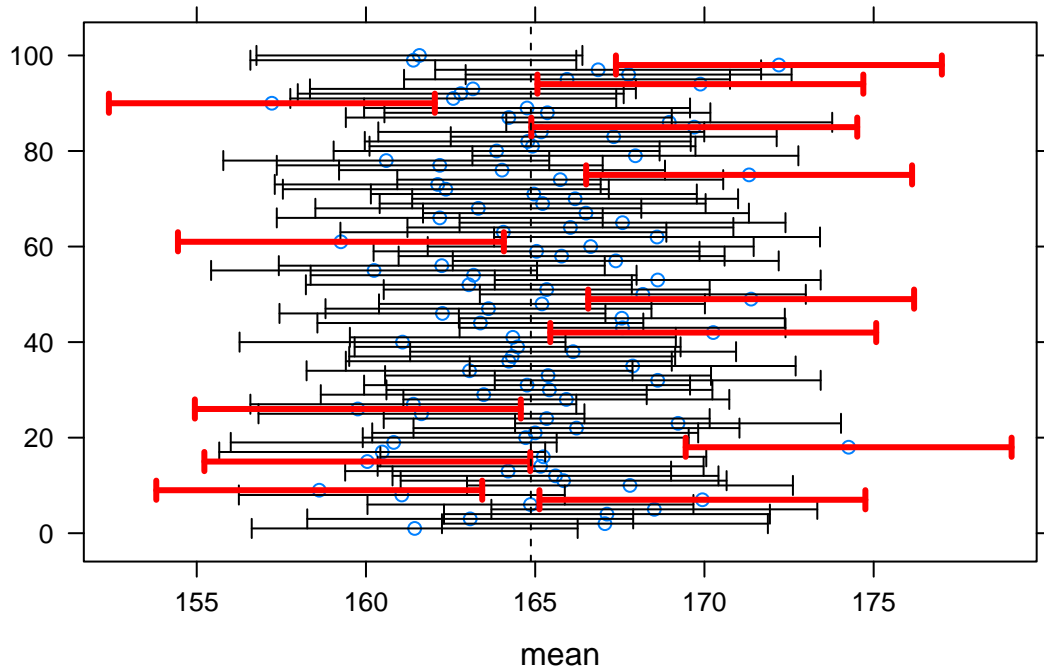


We can see that the population mean  $\mu = 164.872$ , the SD  $\sigma = 6.5446$ . From the shape of histogram we know there's no obvious outlier nor clear skewness.

## (b) (1)

Here I plot the 90% z-CI with known population SD  $\sigma$ :  $\bar{x} \pm 1.645\sigma/\sqrt{n}$  with following codes:

```
samp = do(100)*favstats(sample(population, size=5))
sigma=sd(population)
samp=transform(samp, lower=mean-1.645*sigma/sqrt(5))
samp=transform(samp, upper=mean+1.645*sigma/sqrt(5))
plot_ci(samp, mu=mean(population))
```

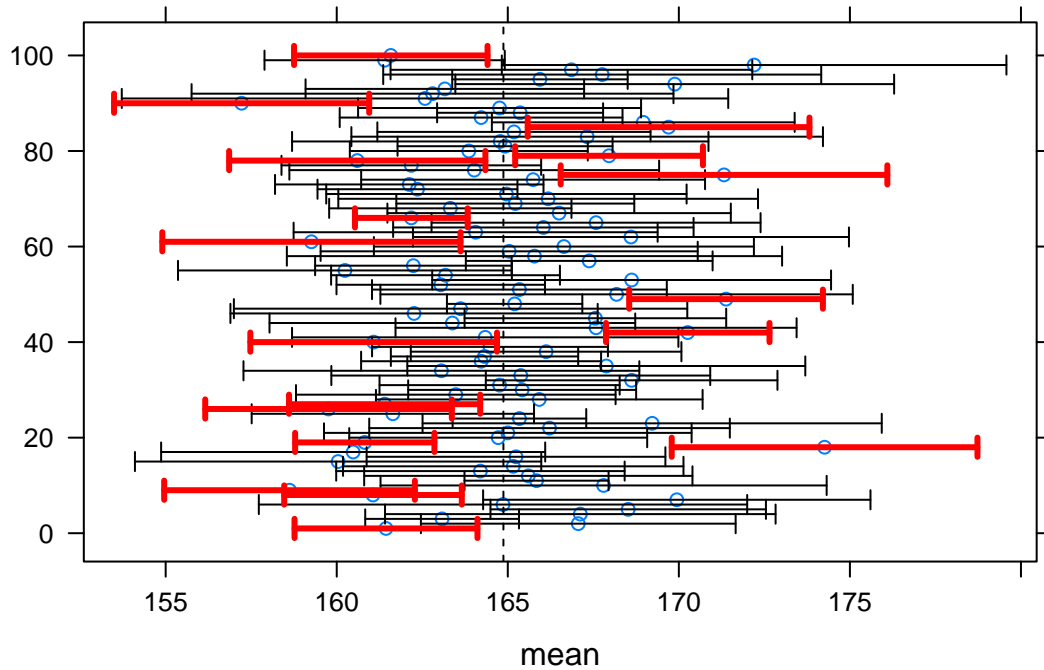


For the 100 CIs constructed, we notice 11 of them have missed the true population mean, the proportion of intervals that include the true population mean is 89%, which is close to nominal level 90%.

## (b) (2)

Here I plot the 90% z-CI with unknown population SD:  $\bar{x} \pm 1.645s/\sqrt{n}$  with following codes:

```
samp=transform(samp, lower=mean-1.645*sd/sqrt(5))
samp=transform(samp, upper=mean+1.645*sd/sqrt(5))
plot_ci(samp, mu=mean(population))
```



For the 100 CIs constructed, we notice 18 of them have missed the true population mean, the proportion of intervals that include the true population mean is 82%, which is lower than nominal level 90%.

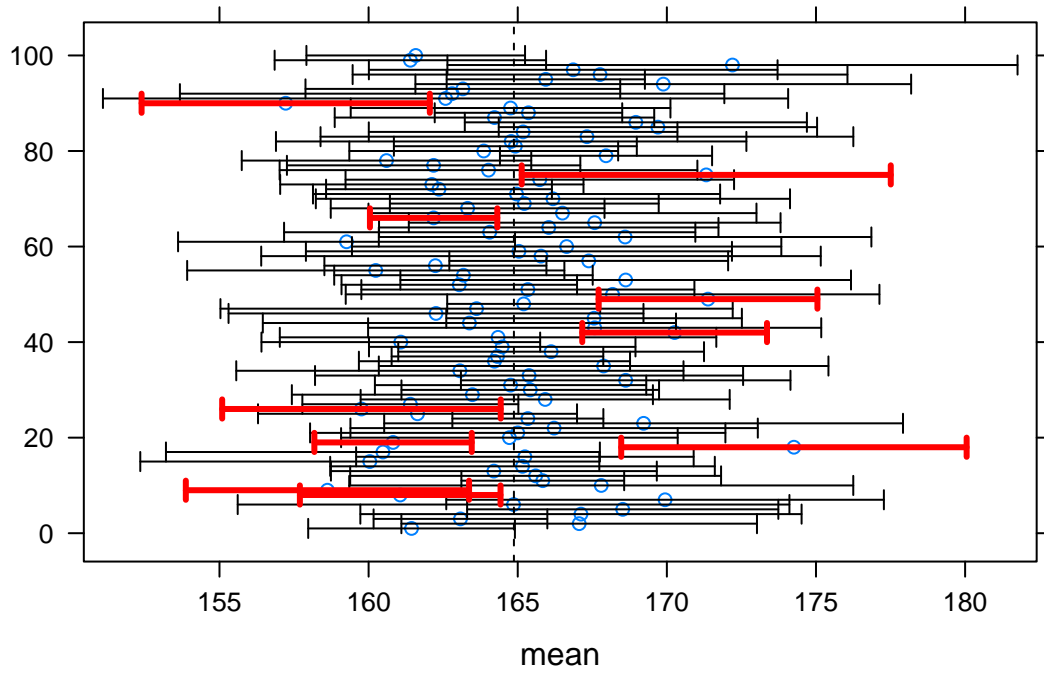
## (b) (3)

Here I plot the 90% t-intervals:  $\bar{x} \pm t^* s / \sqrt{n}$  with following codes, where  $t^* = 2.13185$

```
qt(0.1/2, df=4, lower.tail = F)
```

```
## [1] 2.13185
```

```
samp=transform(samp, lower=mean-2.13185*sd/sqrt(5))
samp=transform(samp, upper=mean+2.13185*sd/sqrt(5))
plot_ci(samp, mu=mean(population))
```



For the 100 CIs constructed, we notice 9 of them have missed the true population mean, the proportion of intervals that include the true population mean is 91%, which is close to nominal level 90%.