Week 5 Practice

Jacques Uber

5/2/2021

Problem

Chapter 8 Analyzing the Databases # 1

Using the manufacturing database as a sample, construct a 95% confidence interval for the population mean number of production workers. What is the point estimate? How much is the error of the estimate? Comment on the results.

```
BLACK_DATABASES = "./data/Databases_in_Excel.xlsx"
data = read_excel(BLACK_DATABASES, sheet = "Manufacturer")
```

We don't know the population mean or the population standard deviation, so its likely that a t-distribution is our best option. What does our sample population look like?

```
library(ggplot2)
ggplot(data, aes(x=`No. Prod. Wkrs.`)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```
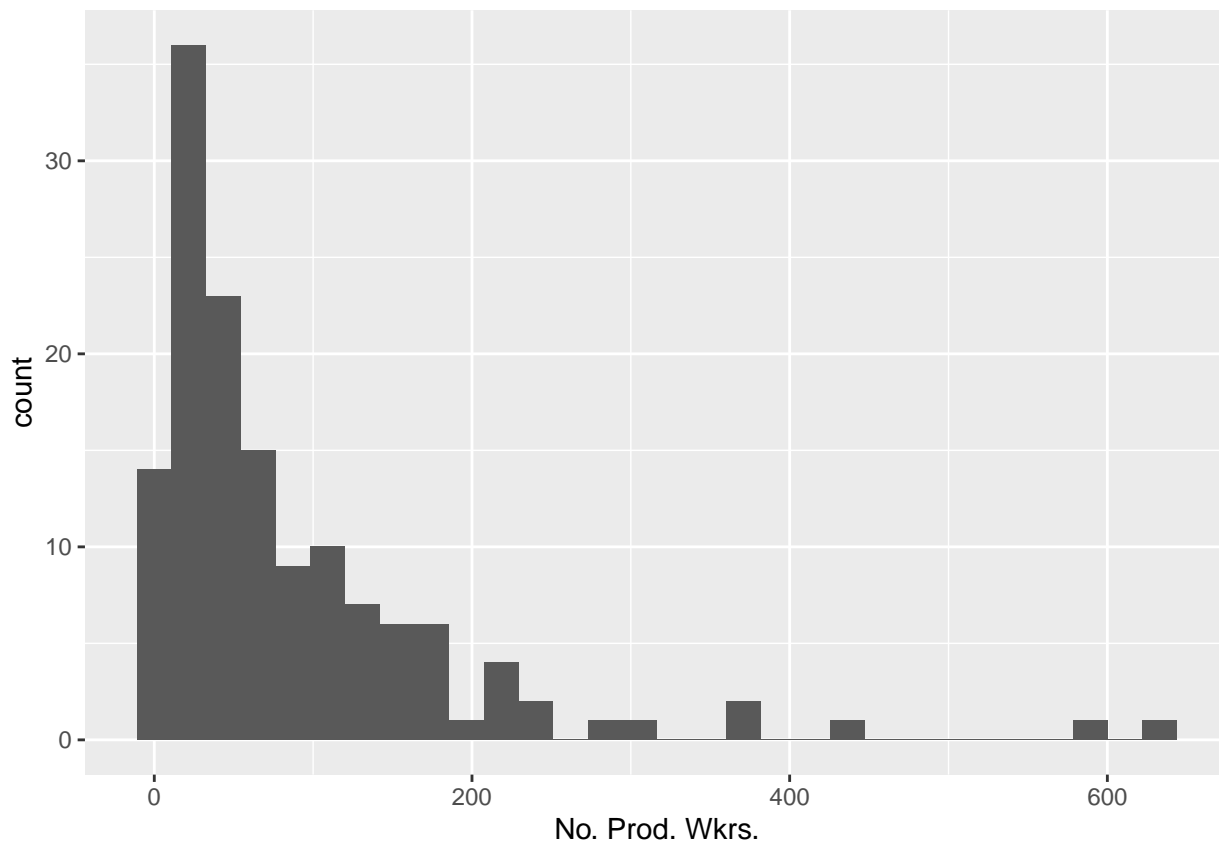
Unfortunately, this doesn't look normal – it is skewed heavily to the right. According to Black, though, the t-distribution is a relatively robust statistic so the fact that our population distribution is not normal should be okay.

Lets now compute our 95% confidence interval using the formula:

$$\overline{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

```
sample_data = data %>% sample_frac(1)
n = nrow(sample_data)
df = n - 1
alpha = 0.95
xbar = mean(sample_data$`No. Prod. Wkrs.`)
s = sd(sample_data$`No. Prod. Wkrs.`)
t = dt(alpha/2, df)
error = t * (s/sqrt(n))
low = xbar - error
high = xbar + error
```

```
sprintf("%s <= population average <= %s with %s precent confidence", round(low,2), round(high, 2), roun
```

```
## [1] "83.89 <= population average <= 90.03 with 95 precent confidence"
```
```
sprintf("The point estimate is %s (which is also the population mean", round(xbar, 2))
```

```
## [1] "The point estimate is 86.96 (which is also the population mean"
```
```
sprintf("The error is %s", round(error, 2))
```

```
## [1] "The error is 3.07"
```

I'm not confident I approached this problem correctly.