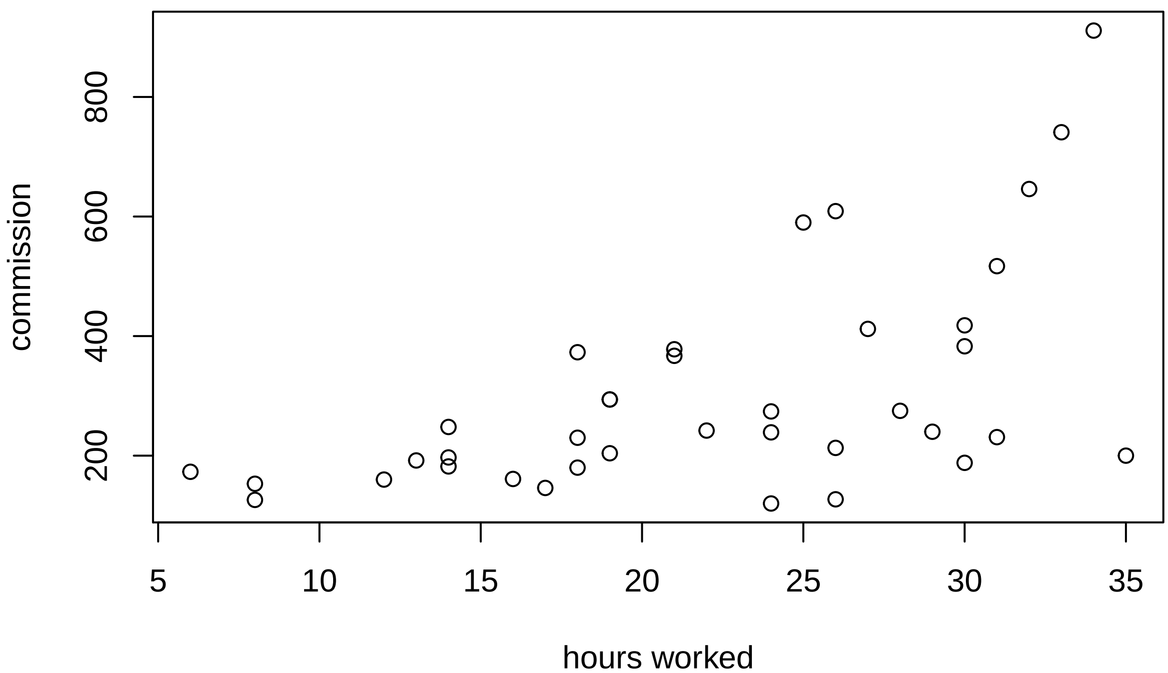
a68

Mushi Wang

15/07/2020

1.(a)

```
com = read.csv("commissions.csv")
plot(com$Hours_Worked, com$Commission, xlab = "hours worked", ylab = "commission")
```



Since there is a roughly linear relationship. Also the plot is bell shaped, the variance of residual is not a constant. Hence we want to use a ratio estimation.

1.(b)

```r
mean(com$Commission)
```

```
## [1] 306.1579
```

```r
mean(com$Hours_Worked)
```

```
## [1] 22.15789
```

```r
length(com$Unit)
```

```
## [1] 38
```

```r
306.1579 / 22.15789 * 21
```

```
## [1] 290.1592
```

$\hat{\mu}_{ratio} = \frac{\bar{y}}{\bar{x}}\mu_x = \frac{306.1579}{22.15789} \cdot 21 = 290.1592$

```r
sqrt_hours = sqrt(com$Hours_Worked)
com_ratio = data.frame(unit = com$Unit, hours_worked = com$Hours_Worked / sqrt_hours,
                       commission = com$Commission / sqrt_hours)
com_ratio_model = lm(com_ratio$commission ~ com_ratio$hours_worked - 1)
summary(com_ratio_model)
```

```
##
## Call:
## lm(formula = com_ratio$commission ~ com_ratio$hours_worked -
##     1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -47.94 -18.23  -0.52  15.70  75.67
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## com_ratio$hours_worked   13.817      1.002   13.79  3.7e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.07 on 37 degrees of freedom
## Multiple R-squared:  0.8372, Adjusted R-squared:  0.8328
## F-statistic: 190.2 on 1 and 37 DF,  p-value: 3.702e-16
```

From summary, $\hat{\sigma}_{ratio} = 29.07$

So, a 95% C.I. for the mean commission is

$$\begin{aligned}
&\hat{\mu}_{ratio} \pm \frac{c\hat{\sigma}_{ratio}}{\sqrt{n}}\sqrt{1-\frac{n}{N}} \quad \text{where } C \sim N(0,1) \\
=&290.1592 \pm 1.96\frac{29.07}{\sqrt{38}}\sqrt{1-\frac{38}{112}} \\
=&[282.6462, 297.6722]
\end{aligned}$$

We are 95% confident that the mean commission lies in that interval.

Since $\tilde{Y}_{tot} = N \cdot \tilde{\mu}_{ratio}$.

$\hat{y}_{tot} = N \cdot \hat{\mu}_{ratio} = 32497.83$, $sd(\hat{y}_{tot}) = \sqrt{N^2 \cdot var(\mu_{ratio})} = N \cdot sd(\mu_{ratio})$

a 95% C.I. for the total commission is

$$
\begin{aligned}
&\hat{\mu}_{ratio} \pm \frac{cN\hat{\sigma}_{ratio}}{\sqrt{n}}\sqrt{1-\frac{n}{N}} \quad \text{where } C \sim N(0,1) \\
=&32497.83 \pm 1.96 \times 112\frac{29.07}{\sqrt{38}}\sqrt{1-\frac{38}{112}} \\
=&[31656.37, 33339.29]
\end{aligned}
$$

We are 95% confident that the total commission lies in that interval.

1.(c) $\frac{19}{20} = 0.95$

$$
\begin{aligned}
\frac{c\hat{\sigma}_{ratio}}{\sqrt{n}}\sqrt{1-\frac{n}{N}} &\leq 1 \quad \text{where } C \sim N(0,1) \\
\sqrt{\frac{1}{n}-\frac{1}{112}} &\leq \frac{1}{1.96 \times 29.07} \\
\frac{1}{n} &\leq \frac{1}{3246.401} + \frac{1}{112} \\
n &\geq 108.2649
\end{aligned}
$$

So we need at least 109 employers

1.(d)

```r
sqrt(var(com$Commission))
```

```
## [1] 185.2197
```

$\hat{\sigma}_y = 185.2197$, $\hat{\mu}_y = \bar{y} = 306.1579$

So, a 95% C.I. is

$$
\begin{aligned}
&\hat{\mu}_y \pm \frac{c\hat{\sigma}_y}{\sqrt{n}}\sqrt{1-\frac{n}{N}} \quad \text{where } C \sim N(0,1) \\
=&306.1579 \pm 1.96\frac{185.2197}{\sqrt{38}}\sqrt{1-\frac{38}{112}} \\
=&[258.2885, 354.0273]
\end{aligned}
$$

This C.I. is much wider than the ratio intervals since $\hat{\sigma}_y$ is much larger. Hence the ratio interval is more accurate.