STAC67A1Q1Q6

Yulun Wu

29/01/2022

Q1

(a)

```
set.seed(1004912785)
Y = rnorm(20, mean=2, sd=5)
Y_bar = mean(Y)
Y_bar
```

```
## [1] 2.820138
```

```
sd_Y_bar = sqrt(sum((Y-Y_bar)^2)/20)
sd_Y_bar
```

```
## [1] 5.242586
```

```
StandardErr = sd_Y_bar/(sqrt(19)) # here is SE of mean
StandardErr
```

```
## [1] 1.202732
```

(b)

```
model = lm(Y ~ 1)
summary(model)
```

```
##
## Call:
## lm(formula = Y ~ 1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.4517  -4.1210   0.0289   4.1078   9.4088
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.820      1.203   2.345   0.0301 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.379 on 19 degrees of freedom
```

These provide the same mean as (a) because $\mathbf{E(Y) = \beta_0}$.

(c)

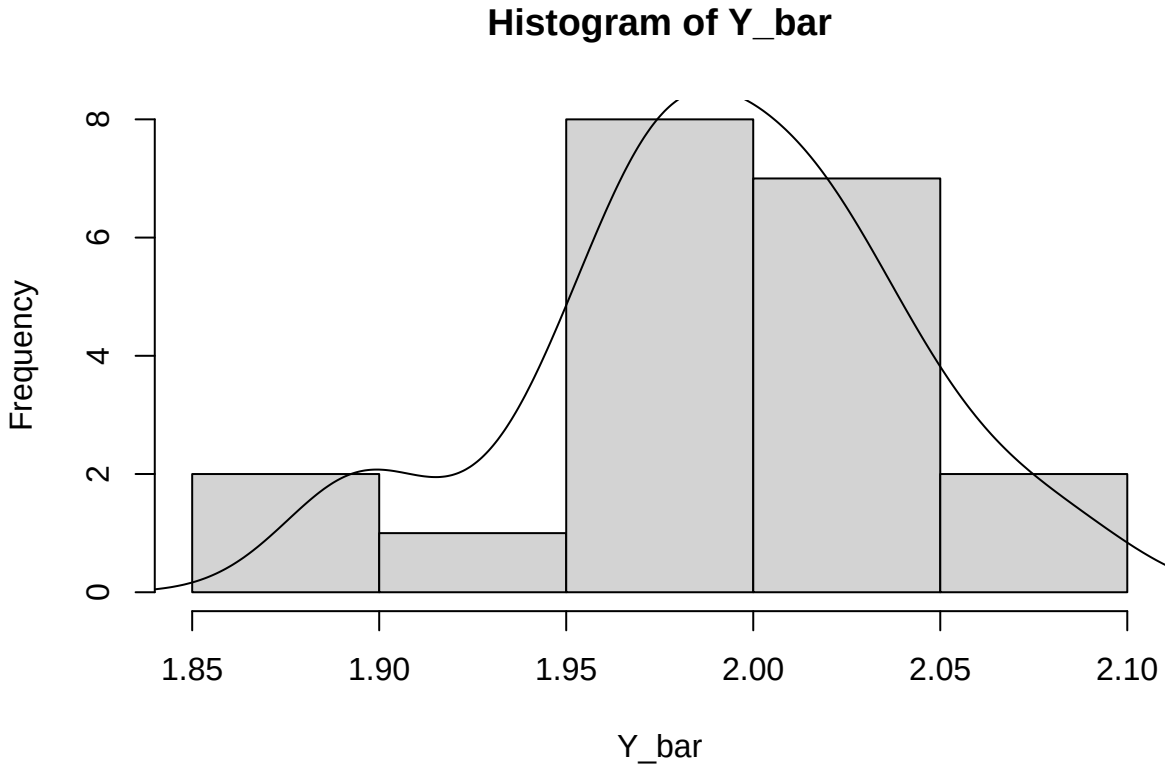
```
Y = replicate(10000,rnorm(20, mean=2, sd=5))
Y_bar = rowMeans(Y)
Y_bar
```

```
##  [1] 1.968585 2.083917 1.980719 2.055946 1.951725 1.967888 1.893764 2.023605
##  [9] 2.009004 1.897203 2.006770 2.023658 2.044372 2.002036 1.941738 1.971380
## [17] 1.995883 1.980652 1.983320 2.026579
```

```
Y_bar_hist = hist(Y_bar)
Y_bar_hist
```

```
## $breaks
## [1] 1.85 1.90 1.95 2.00 2.05 2.10
##
## $counts
## [1] 2 1 8 7 2
##
## $density
## [1] 2 1 8 7 2
##
## $mids
## [1] 1.875 1.925 1.975 2.025 2.075
##
## $xname
## [1] "Y_bar"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

```
lines(density(Y_bar))
```

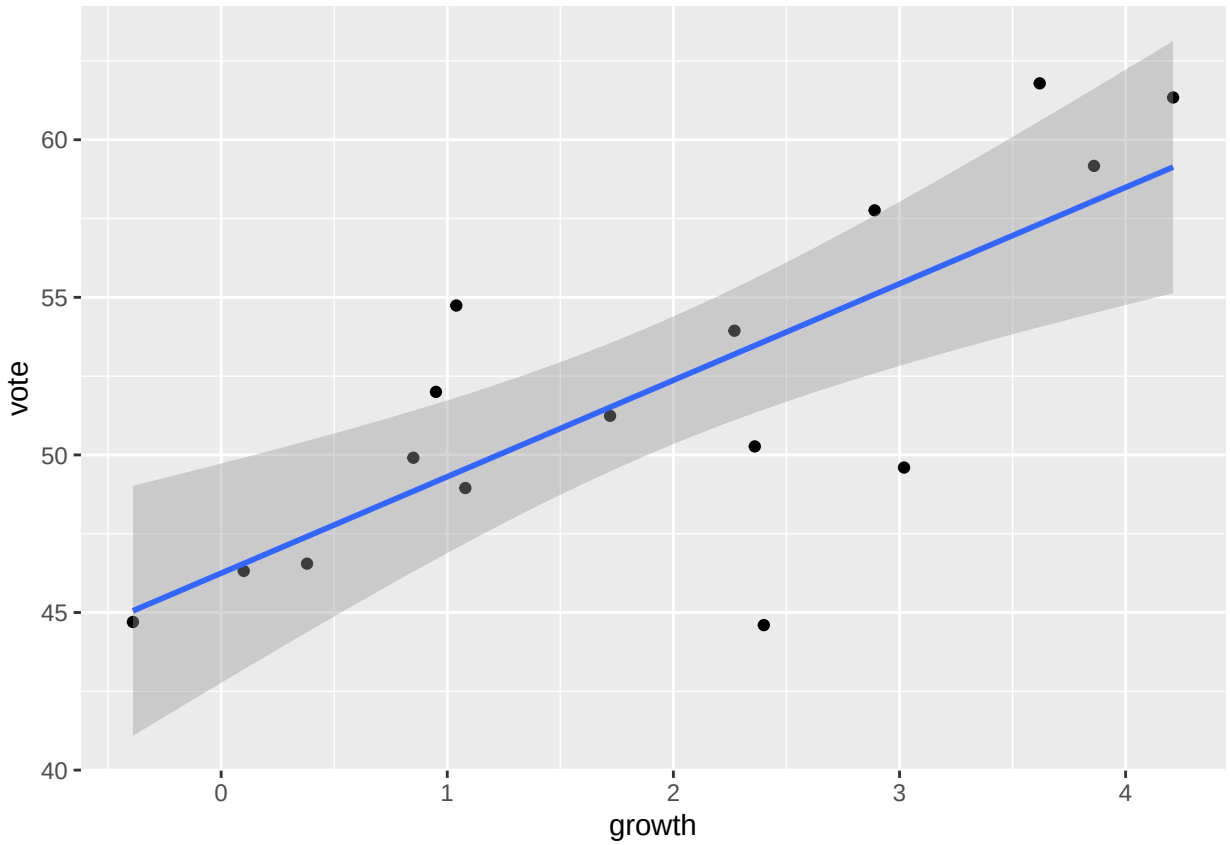


Q6

(a)

```
Vote = read.table("vote.txt", header=T)
library(ggplot2)
ggplot(data=Vote,aes(growth,vote)) + geom_point() + stat_smooth(method="lm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



(b)

```
fit = lm(vote~growth, data=Vote)
summary(fit)
```

```
##
## Call:
## lm(formula = vote ~ growth, data = Vote)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9929 -0.6674  0.2556  2.3225  5.3094
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  46.2476     1.6219  28.514 8.41e-14 ***
## growth        3.0605     0.6963   4.396  0.00061 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.763 on 14 degrees of freedom
## Multiple R-squared:  0.5798, Adjusted R-squared:  0.5498
## F-statistic: 19.32 on 1 and 14 DF,  p-value: 0.00061
```

$\hat{\beta}_0 = \mathbf{46.2476}$

$\hat{\beta}_1 = \mathbf{3.0605}$

(c)

```
46.2476 + 0.1*3.0605
```

```
## [1] 46.55365
```

```
predict(fit,data.frame(growth = 0.1))
```

```
##       1
## 46.5537
```

The incumbent party's vote in 2008 is predicted to be 46.55365 by hand and 46.5537 by R, based on this result, "Obama" will win the election.

(d)

$H_0 : \beta_1 = 0$ vs $H_a : \beta_1 > 0$ $t = \hat{\beta}_1 / SE(\hat{\beta}_1) = 3.0605/0.6963 = 4.4$
$P - value = P(t > 4.40) = 3.0240706 \times 10^{-4}$
So, there is an extreamly strong evidence that there is a signficant positive association between incumbent party's vote share and economical growth.

(e)

```
3.0605 + c(-1,1)*qt(0.975,14)*0.6963
```

```
## [1] 1.567085 4.553915
```

95% CI by hand = (1.567085,4.553915)

95% CI by R:

```
confint(fit)
```

```
##                 2.5 %    97.5 %
## (Intercept) 42.768951 49.726345
## growth       1.567169  4.553887
```

(f)

```
1-pt(1/0.6963,14)
```

```
## [1] 0.08646192
```

```
pt(-1/0.6963,14)
```

```
## [1] 0.08646192
```

The probability that $P(|\hat{\beta}_1 - \beta_1| > 1)$ is 0.08646192.