# Hwk 4-3

Andrew Guo

10/7/2021

3.)

(a) Read in the data using the read.csv function

```
pharmaLSRL = read.csv("Hwk4-3.csv")
```

(b) Fit a linear model using the lm function with "Ingredients Purchased Directly" as the predictor and "Sales Volume" as the response. Use the summary function to output the resulting fit

```
attach(pharmaLSRL)

pharmaLSRL.fit <- lm(Sales.Volume~Ingredients.Purchased.Directly, data = pharmaLSRL)

summary(pharmaLSRL.fit)
```

```
##
## Call:
## lm(formula = Sales.Volume ~ Ingredients.Purchased.Directly, data = pharmaLSRL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.074  -4.403  -1.607   5.719  14.834
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      4.6979     5.9520   0.789    0.453
## Ingredients.Purchased.Directly   1.9705     0.1545  12.750 1.35e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.022 on 8 degrees of freedom
## Multiple R-squared:  0.9531, Adjusted R-squared:  0.9472
## F-statistic: 162.6 on 1 and 8 DF,  p-value: 1.349e-06
```

(c) Confirm your estimates for b0 and b1 from Problem 1

Basd on the data in the summary() command, we note that the estimated b0 and b1 values are 4.6979 and 1.9705, respectively. These values are very close to the b0 and b1 values calculated by hand in Problem 1, which were approximately 4.714 and 1.97, respectively. There is a slight discrepency in both of the values, but this can be attributed to rounding. 1.9705 rounded to the nearest hundreth is 1.97, and if 1.9705 is used in place of 1.97 to calculate b0, we get a value of of 4.6971. This is much closer to the R-estimated value of 4.6979, but this variation can also be attributed to the rounded values for (xbar * ybar), summation(xi*yi), summation(xi^2), and summation(xbar^2). If rounded farther out, such as to the thouandth instead of the hundreth, we could expect the hand calculations and R-calculations to be even closer.

(d) Use the output to report the estimate for variance(epsilon) = theta^2

In the summary() output, we see that the residual standard error is 9.022. Because the residual standard error is the square root of the sample variance, and because the sample variance is also the mean square error, which is also an unbiased estimator for theta^2, then the square of the residual standard error would also be the expected value of epsilon.

Thus, we have v(epsilon) = 9.022^2 = 81.396484

(e) How are the 8 degrees of freedom obtained?

Because there were two unknown variables that were calculated, then the degrees of freedom become n-2. Because the sample size is 10, then the degrees of freedom become:

10 - 8 = 2

# Hwk 4-4

Andrew Guo

10/7/2021

4.)

(a) Read the data using the read.csv function

```
crimeLSRL = read.csv("Hwk 4-4.csv")
```

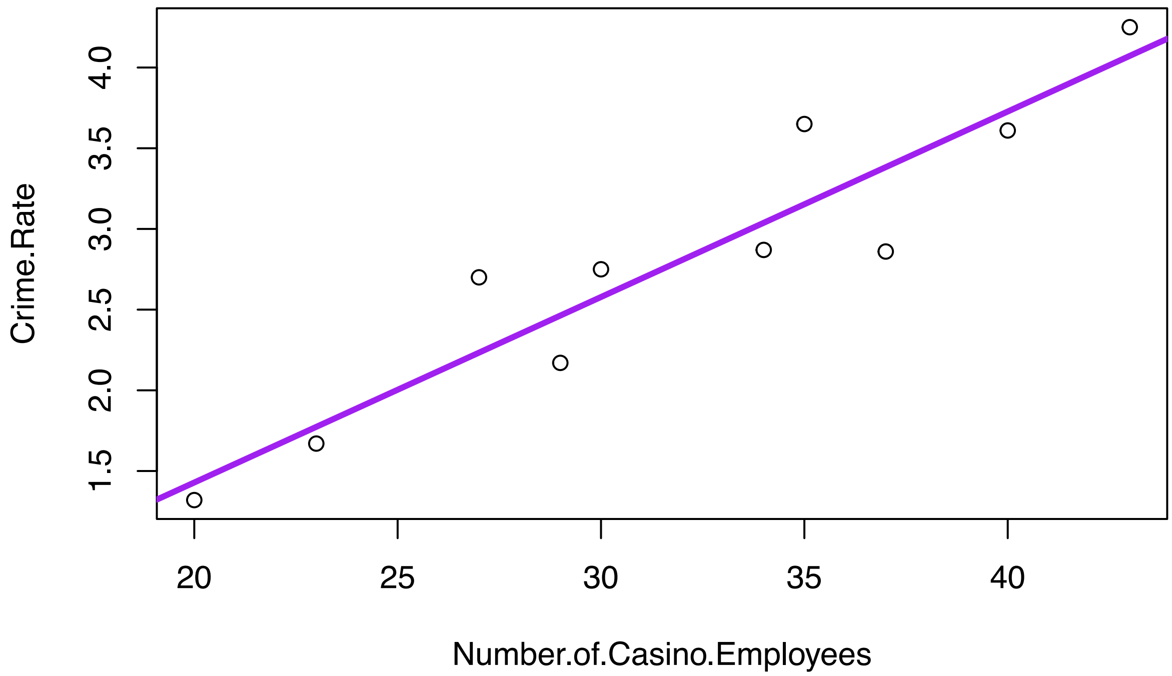
(b) Fit a linear model using the lm function with "Number of Casino Employees" as the predictor and "Crime Rate" as the response. Use the summary function to output the resulting fit.

```
attach(crimeLSRL)
crimeLSRL.fit <- lm(Crime.Rate~Number.of.Casino.Employees, data = crimeLSRL)
summary(crimeLSRL.fit)
```

```
##
## Call:
## lm(formula = Crime.Rate ~ Number.of.Casino.Employees, data = crimeLSRL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5226 -0.1552 -0.1062  0.1763  0.4972
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -0.86977    0.50903  -1.709    0.126
## Number.of.Casino.Employees  0.11493    0.01564   7.350 7.99e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3446 on 8 degrees of freedom
## Multiple R-squared:  0.871,  Adjusted R-squared:  0.8549
## F-statistic: 54.03 on 1 and 8 DF,  p-value: 7.992e-05
```

(c) Plot the data along with the regression line

```
plot(Number.of.Casino.Employees, Crime.Rate)
abline(crimeLSRL.fit, lwd = 3, col = 'purple')
```




(d) Use the output to report the estimate for variance(epsilon) = theta^2

In the summary() output, we see that the residual standard error is 0.3446. Because the residual standard error is the square root of the sample variance, and because the sample variance is also the mean square error, which is also an unbiased estimator for theta^2, then the square of the residual standard error would also be the expected value of epsilon.

Thus, we have v(epsilon) = 0.3446^2 = 0.11943836.

(e) Predict the crime rate when there are 25,000 casino employees

Based on our calculations of b1 and b2, our linear regression equation is approximately:

y = 0.115x - 0.867

In Problem 2, we had calculated that the y-intercept (b0) was -0.872, however, this value was reached by using a rounded b1 value of 0.115. In the summary() function, we notice a closer approximation for b1, which is 0.11493. When this number is used in place of 0.115 in the process of solving for b0, we obtain the y-intercept of approximately -0.869. Therefore, the R-calculated value will be used for the intercept.

We are being asked to estimate the crime rate (y) with 25,000 casino employees (x). So we substitute x with its respective value to calculate y. Note that the number of casino employees is denoted as in thousands, so the value being substituted is not 25,000, but rather, 25. Therefore, we calculate y to be:

y = 0.11493(25) - 0.86977 = 2.00348

This means that the expected crime rate when there are 25,000 casino employees would be 2.00348 (per 1,000 population).