Quiz 2

Alyssa Sharma

2023-09-14

```r
x <- c(1,0,2,0,3,1,0,1,2,0)
y <- c(16,9,17,12,22,13,8,15,19,11)

timetransvsbroke <- data.frame(x,y)
head(timetransvsbroke)
```

```
##   x  y
## 1 1 16
## 2 0  9
## 3 2 17
## 4 0 12
## 5 3 22
## 6 1 13
```

```r
tbmod <- lm(y ~ x, data = timetransvsbroke)
summary(tbmod)
```

```
##
## Call:
## lm(formula = y ~ x, data = timetransvsbroke)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -2.2   -1.2    0.3    0.8    1.8
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.2000     0.6633  15.377 3.18e-07 ***
## x             4.0000     0.4690   8.528 2.75e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.483 on 8 degrees of freedom
## Multiple R-squared:  0.9009, Adjusted R-squared:  0.8885
## F-statistic: 72.73 on 1 and 8 DF,  p-value: 2.749e-05
```

Part a:

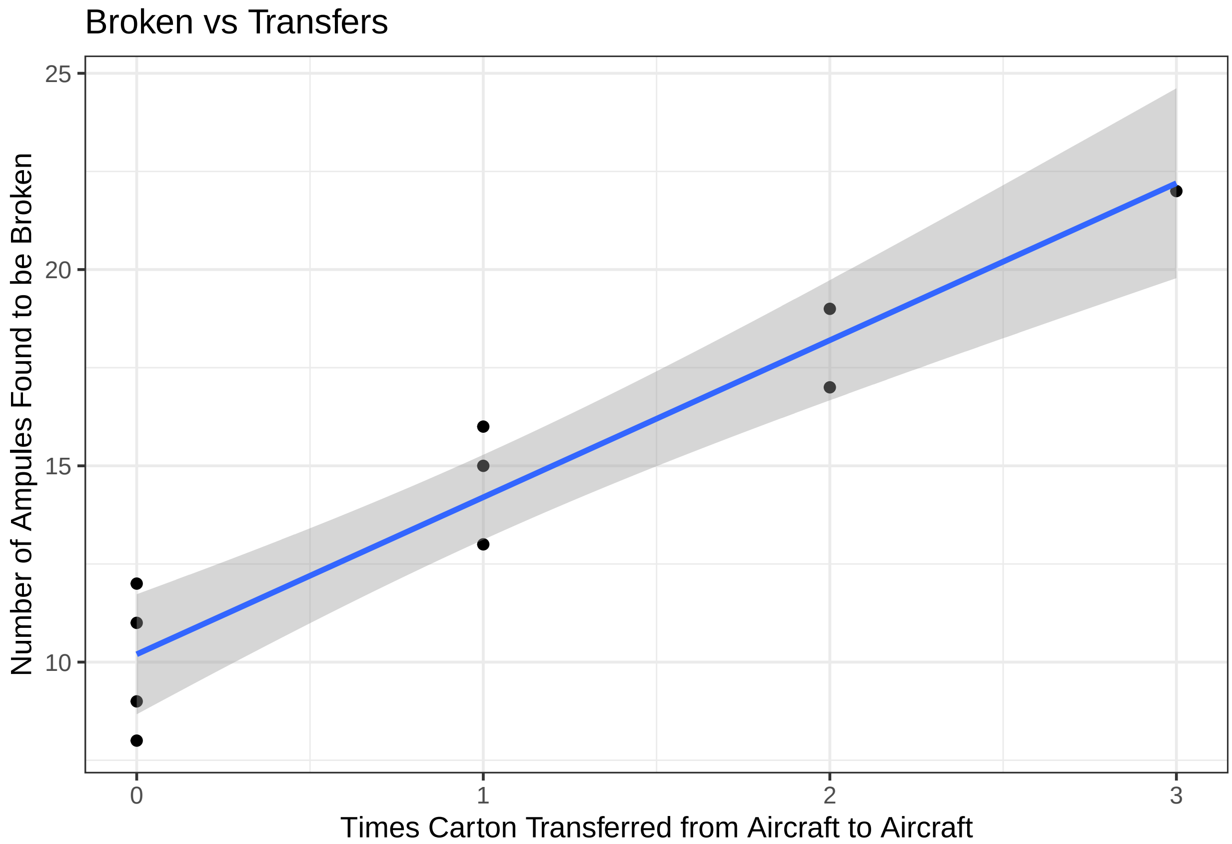
The estimated regression function is $\hat{Y} = 10.2000 + 4.0000(X_i)$

Part b:

```r
library(moderndive)
library(ggplot2)
ggplot(data = timetransvsbroke, aes(x = x, y = y)) +
```

```
  geom_point() +
  geom_smooth(method = lm) +
  theme_bw() +
  labs(x = "Times Carton Transferred from Aircraft to Aircraft", y = "Number of Ampules Found to be Brol
```

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



Part c: The linear regression function appears to be a good fit here.

Part d: A point estimate of the expected numnber of broken ampumles when X = 1 transfers are made is 14.2 broken ampules.

```
point <- get_regression_points(tbmod)
point
```

```
## # A tibble: 10 x 5
##       ID     y     x y_hat residual
##    <int> <dbl> <dbl> <dbl>    <dbl>
##  1     1    16     1  14.2      1.8
##  2     2     9     0  10.2     -1.2
##  3     3    17     2  18.2     -1.2
##  4     4    12     0  10.2      1.8
##  5     5    22     3  22.2     -0.2
##  6     6    13     1  14.2     -1.2
##  7     7     8     0  10.2     -2.2
##  8     8    15     1  14.2      0.8
##  9     9    19     2  18.2      0.8
```

```
## 10     10     11      0   10.2       0.8
```

Part e: The estimated increase in the expected number of ampules broken when there are 2 transfers is going to be $b_i * 2$, which equals 8.

```
2*4
```

```
## [1] 8
```

Part f:

Because $\hat{Y_1} = 14.2$ according to the get_regression_points function, and the coordinate $(\bar{X}, \bar{Y}) = (1, 14.2)$, the fitted regression line goes through the point $(\bar{X}, \bar{Y}) = (1, 14.2)$.

```
xbar <- mean(x)
ybar <- mean(y)
xbar
```

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

```
ybar
```

```
## [1] 14.2
```

Part g: The residual for the first case is 1.8 broken ampules.

Part h: The two values are equal. $e_i = y_i - \hat{y} = \epsilon_i$

Part i:

The sum of square residuals is 2.773339e-32. The MSE is 3.466674e-33

```
sumris <- sum(point$residual)
sumrissq <- sumris^2
sumrissq
```

```
## [1] 2.773339e-32
```

```
mse <- sumrissq/(nrow(timetransvsbroke) - 2)
mse
```

```
## [1] 3.466674e-33
```

Part j:

The estimated MSE is 3.466674e-33.