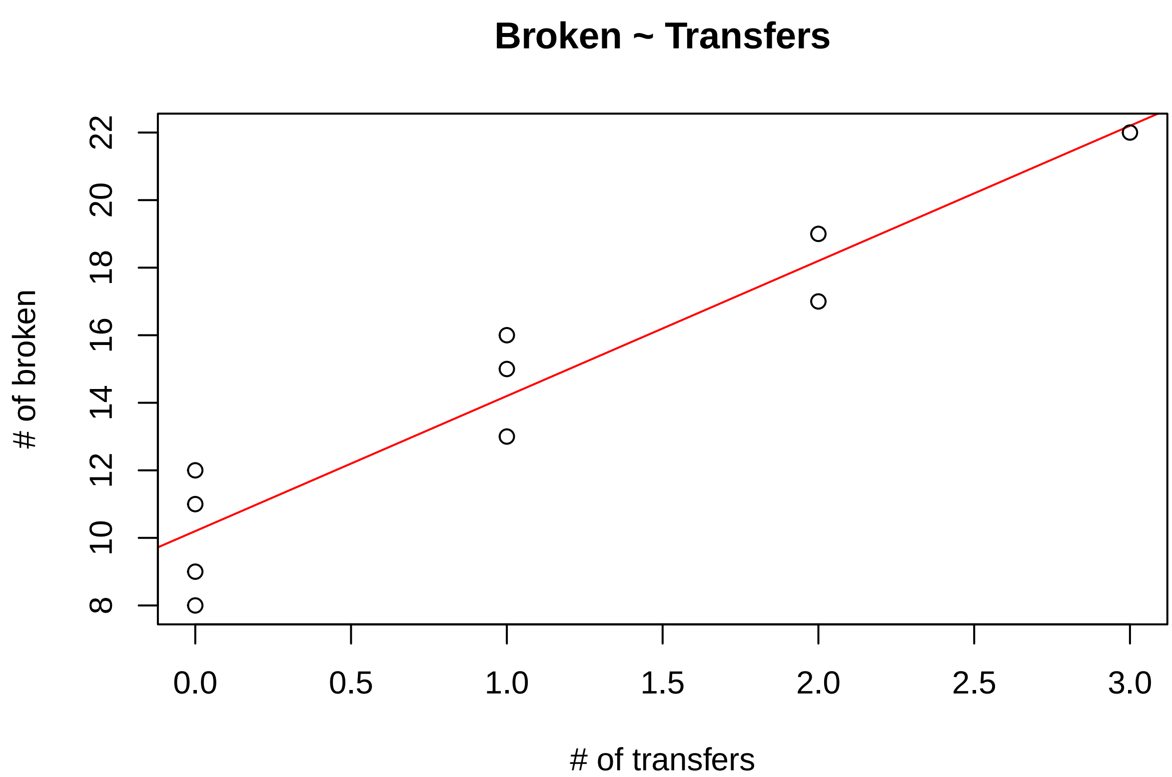
hw2

Dylan M Ang

1/26/2022

1.21

a



$\hat{y} = 10.2 + 4X$

Linear regression seems like a good fit. The epsilon values seem to be random.

b

$y = 10.2 + 4(1)$

$y = 14.2$

c

Based on the β_1 value for slope, the model estimates an increase of 4 broken ampules per each extra transfer.

d

```
mean(x)
```

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

```
mean(y)
```

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

$14.2 = 10.2 + 4(1)$

$14.2 = 14.2 \implies (\bar{X}, \bar{Y})$ is on the line.

1.29

$Y_i = \beta_1 X_i + \beta_0 + \epsilon_i$

In terms of the graph, β_0 is the y-intercept. So, if $\beta_0 = 0$ and $X = 0$ is in the domain of the model, then the point (0,0) would appear on the regression line. The line would go through the origin of the graph.

2.1

a

The 95% confidence interval for the slope estimate is $[0.452886, 1.05721]$

This conclusion is warranted since the limits do not cover 0. Both values are > 0 so it can be safely concluded that X and Y are correlated. The implied level of significance is $1 - 0.95 = 0.05$.

b

This is okay since there shouldn't be any cities with 0 population, therefore $X = 0$ isn't in the domain of the model. This means that β_0 has not practical interpretation.

2.3

The regression equation does suggest that what the student stated was true, but the P-value is given as 0.91, which would not be considered significant for $\alpha = 0.05$ or $\alpha = 0.01$. Therefore the student is incorrect, since there is insufficient evidence to assume the model is correct.

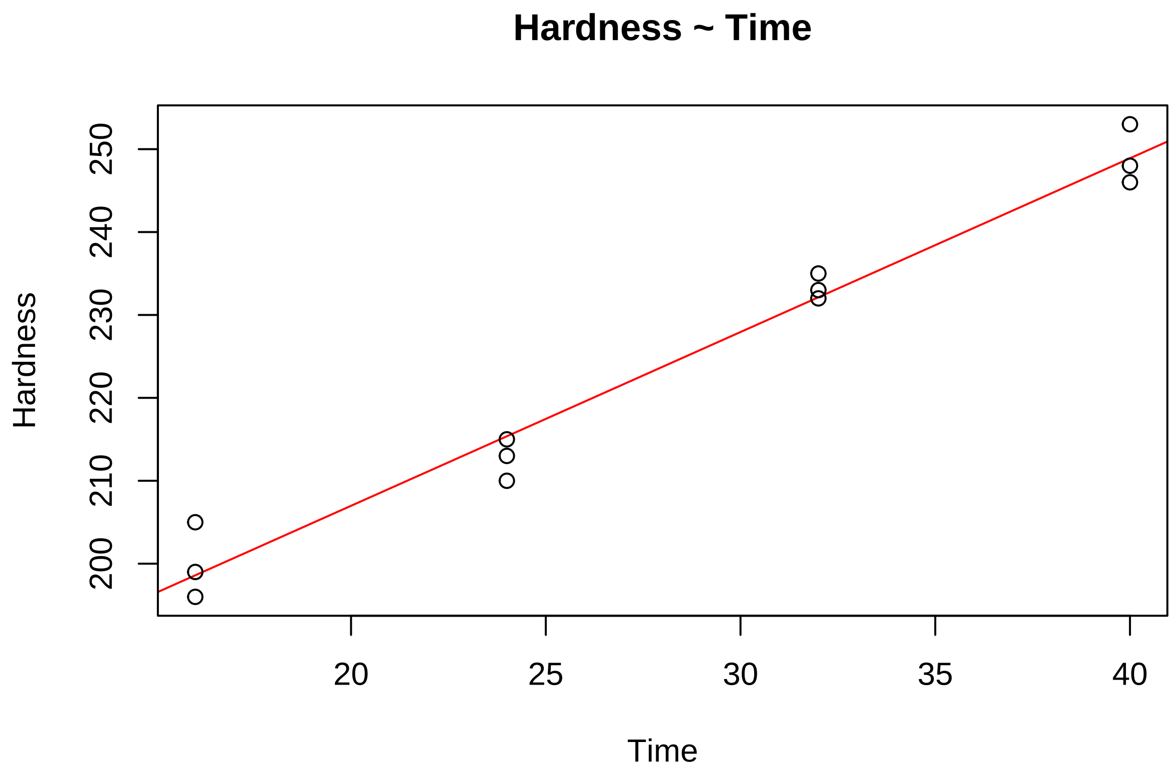
2.7

a

$\hat{\beta}_1 = \frac{\Sigma_{i=1}^{n}(x_i - \bar{x})(Y_i - \hat{Y})}{\Sigma_{i=1}^{n}(x_i - \bar{x})^2}$

1.22 Does not include complete data, but 1.14 says to randomly generate some, so I created my own values that fit roughly around the linear model.

```
x <- c(16, 16, 16, 24, 24, 24, 32, 32, 32, 40, 40, 40)
y <- c(199, 205, 196, 215, 213, 210, 233, 235, 232, 248, 253,
    246)
hardness <- lm(y ~ x)
plot(x, y, main = "Hardness ~ Time", xlab = "Time", ylab = "Hardness",
    abline(hardness, col = "red"))
```



Based on the sample data, the estimated change in mean hardness for each hour $\hat{\beta}_1 = 2.095$. In addition, $SE(\hat{\beta}_1) = 0.1106$. This means that the plastic is expected to get 2.095 Brinell units harder for each hour it is allowed to rest. However, we expect our sample mean to have deviated from the real mean about 0.1106 units.

b

$H_1 : \beta_1 \neq 2$, $H_{10} : \beta_1 = 2$.

$t_1 = \frac{\hat{\beta}_1 - 2}{SE(\hat{\beta}_1)}$

$t_1 = \frac{2.095 - 2}{0.09039}$

$t_1 = 1.051$

The critical value at $\alpha = 0.01$ and $df = 11$ is 2.718079. Our t-value is lower than that, therefore there is insufficient evidence to reject the null.

Therefore, we can say with 99% confidence that the mean hardness increases by 2 Brinell units per hour.

c

power $= P(\text{reject } H_0 | H_1 \text{ is true})$

```
alpha = 0.01
delta = 100
n = 12
cv = qt(1 - alpha/2, n - 2)
power = pt(-cv, n - 2, ncp = delta) + pt(cv, n - 2, lower.tail = FALSE,
```

```
    ncp = delta)
power
```

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

2.15

a

$\hat{y} = 10.2 + 4x$

```
alpha = 1 - 0.99
p = 1 - alpha/2
n = 10
df = n - 2
t = qt(p, df)
se = 0.469  # from summary(transfers)
ME = t * se
ME
```

```
## [1] 1.573677
```

We can say with 99% confidence that the true mean breakage is 18.2 ± 1.573677, so when there are 2 transfers, we can say with 99% confidence that the number of ampules broken is between $[16.62632, 19.77368]$.

For $X = 4$, $Y_4 = 26.2 \pm 1.573677$. Therefore when there are 4 transfers, we can say with 99% confidence that there will be between $[24.62632, 27.77368]$ ampules broken.

b

For $X = 2$, $Y = 18.2 \pm 1.573677(1.625)$. Therefore when there are two transfers, we predict that the number of broken ampules will be between $[15.643, 20.757]$.

c

For $X = 3$, $Y = 18.2 \pm 1.573677(1.083)$. Therefore when there are two transfers, we predict that the number of broken ampules will be between $[16.496, 19.904]$.

So, multiplying by 3 we can predict the total number of ampules will be between $[49.487, 59.712]$.