hw05

Part 1

```r
library(tidyverse)
```

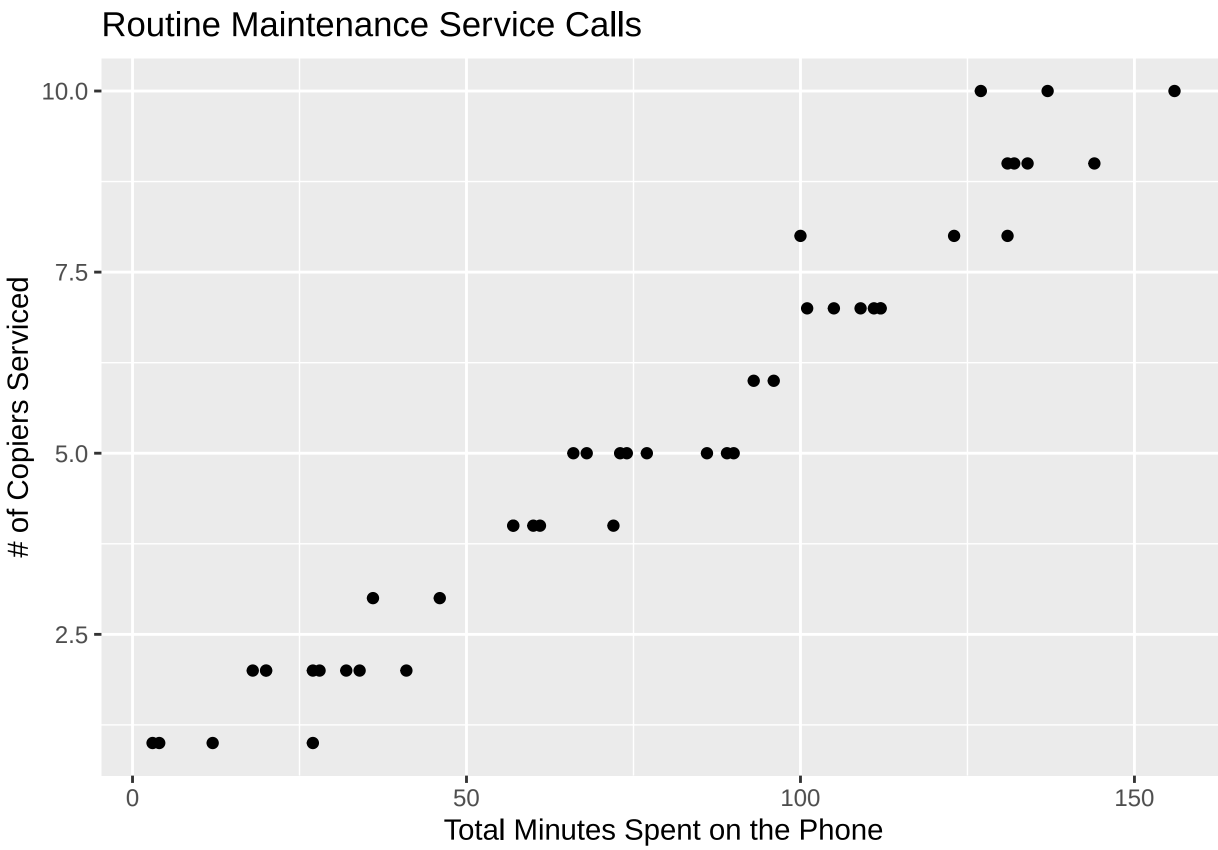
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.1
## v readr   2.1.2     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
copier = read.table("copier.txt", header = TRUE)
```

1A

#create scatter plot

```r
#create scatter plot
ggplot(copier, aes(x=Minutes,y=Serviced))+
geom_point()+
labs(x="Total Minutes Spent on the Phone",
y="# of Copiers Serviced",
title="Routine Maintenance Service Calls")
```




```r
copierlm<-lm(Minutes~Serviced, data=copier)
summary(copierlm)
```

```
##
## Call:
## lm(formula = Minutes ~ Serviced, data = copier)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.7723  -3.7371   0.3334   6.3334  15.4039
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.5802     2.8039  -0.207    0.837
## Serviced     15.0352     0.4831  31.123   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.914 on 43 degrees of freedom
## Multiple R-squared:  0.9575, Adjusted R-squared:  0.9565
## F-statistic: 968.7 on 1 and 43 DF,  p-value: < 2.2e-16
```

1B

Correlation between time and copiers serviced:

```
cor(copier$Minutes, copier$Serviced)
```

```
## [1] 0.978517
```

Context: This value indicates a very strong positive linear association between # of copiers serviced and time spent by the person servicing them.

1C

Is the interpretation reliable: The scatter plot above shows a reasonable linear relationship, which would indicate that the interpretation that the correlation is strong is reliable.

1D

```
confint(copierlm, level=0.95)
```

```
##                  2.5 %    97.5 %
## (Intercept) -6.234843  5.074529
## Serviced    14.061010 16.009486
```

1E

```
copierlm<-lm(Minutes~Serviced, data=copier)
summary(copierlm)
```

```
##
## Call:
## lm(formula = Minutes ~ Serviced, data = copier)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.7723  -3.7371   0.3334   6.3334  15.4039
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.5802     2.8039  -0.207    0.837
## Serviced     15.0352     0.4831  31.123   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.914 on 43 degrees of freedom
## Multiple R-squared:  0.9575, Adjusted R-squared:  0.9565
## F-statistic: 968.7 on 1 and 43 DF,  p-value: < 2.2e-16
```

```
#new.data<-data.frame(y=5)
#predict(copierlm,newdata=new.data,interval='confidence')
```

Could not get code to function. The lower/upper/best fit intervals will likely be close in range.

1F

Value of the residuals for the first observation:

Min 1Q Median 3Q Max -22.7723 -3.7371 0.3334 6.3334 15.4039

Since the median residual is not too high or too low, this seems to make sense contextually since we have a well-fit line in our plots and caccording to our other statistical measures. Our predicted value seems to be fairly accurate accoridng to the residuals.

1G

The average of the residuals is equal to zero. The sum of the residuals is also equal to zero. This make sense because the average of the residuals is equal to the sum of the residuals divided by the number of "n". Since the sum is zero, 0 divided by n (n/0) is also equal to 0.

Part 2

See pictures attached.

Part 3

See pictures attached.

9/26/22

Module 04 Homework

Question #2

$SE_0 = 0.6633$

$\bar{x} = 1, \sum_{i=1}^{10} (x_i - \bar{x})^2 = 10$ $SE_0 = 0.4690$

2a) $\boxed{H_0 : \beta_1 = 0}$ indicates no linear relationship

$\boxed{H_a : B_1 \neq 0}$ indicates a linear

$t = \frac{b_1 - B_1}{SE_1} = \frac{4 - 0}{0.4690} = 8.5287$

$\rightarrow \boxed{t = 8.529}$ w/ $df = n-2$ $\rightarrow n - 2 = 10 - 2 = \boxed{8}$

$\rightarrow$ p-value = 0.00002749 ; p-value < α

$\rightarrow$ So, we reject the null hypothesis ; determine there is a linear relationship between variables.

2b) 95% CI:

$\boxed{t_{\alpha/2, 8} = 2.306}$

$= b_1 \pm t_{\alpha/2} (SE_1)$

$= 4 \pm (2.306)(0.4609)$

$= 4 \pm 1.06$

$\rightarrow \boxed{CI = 2.94 < B_1 < 5.06}$

2c) $H_0: \beta_0 = 9$, indicating the broken # of ampules = 9

$H_a: \beta_0 \neq 9$, indicating the broken # of ampules ≠ 9

$$t = \frac{b_0 - \beta_0}{SE_0} = \frac{10.2 - 9}{0.6633} = \frac{1.2}{0.6633} = \boxed{1.8091}$$

For t = 1.809 ; df = 8 : p-value = 0.10822

At 95% confidence level;

p-value > α

So, we accept the null hypothesis.

Rejecting the alternate hypothesis indicates there is no reason for the consultant to believe the mean # of broken ampules when no transfers are made is different from 9. We believe evidence points to 9 being the mean.

2d) 95% CI:

w/ $t_{\alpha/2, 8} = 2.306$

w/ # of transfers = 2 ; $\hat{y} = t_{\alpha/2} (SE_{\hat{y}_p})$

$\hat{y}_p = b_0 + b_1$ (# of transfers)

$\hat{y}_p = 10.2 + 4(2) = 10.2 + 8 = \boxed{18.2}$

$$SE_{\hat{y}_p} = \sqrt{MSE\left(\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}\right)}$$

w/ # transfers = 2 :

$\Sigma(x_i - \bar{x})^2 = 10$, $n = 10$, $\bar{x} = 1$

ANOVA Table → $\boxed{MSE = 2.2}$

→ $SE_{\hat{y}_p} = \sqrt{(2.2)\left(\frac{1}{10} + \frac{(2-1)^2}{10}\right)} = \sqrt{\frac{4.4}{10}} = \boxed{0.6633}$

→ 95% CI = $18.2 \pm (2.306)(0.6633)$

$= 18.2 \pm 1.53$

→ $\boxed{CI = 16.67 < \hat{y}_p < 19.73}$

95% PI: $\hat{y} \pm t_{\alpha/2} \cdot (SE_i)$

$$SE_i = \sqrt{MSE\left(1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\Sigma x_i - \bar{x})^2}\right)} = \sqrt{2.2\left(1 + \frac{1}{10} + \frac{1}{10}\right)}$$

$= \sqrt{2.2(1.2)} = \boxed{1.625}$ → $PI = 18.2 \pm (2.306)(1.625)$

$= 18.2 \pm 3.75$

→ $\boxed{PI = 14.45 < y_p < 21.95}$

2e) When # of transfers = 1:

Decreasing the # of transfers to 1 causes the value of $\hat{y}$ to decrease.

It will also cause the CI & PI to "decrease" or narrow.

The estimated y-value will dercase as transfers decrease.

2f) F- statistic

$$F = \frac{MS_{Trans}}{MS_{Resid}} = \frac{160.0}{2.2} = \boxed{72.73}$$

2g) R^2:

$$R^2 = 1 - \frac{SS_{Res}}{SS_{Tot}} = 1 - \left(\frac{17.6}{177.6}\right) = 1 - .099 = \boxed{0.901}$$

→ 90.1% of the variation in the "broken" responses can be explained by our linear regression model.

This # is fairly high and indicates that our model's results are usable / trustworthy.

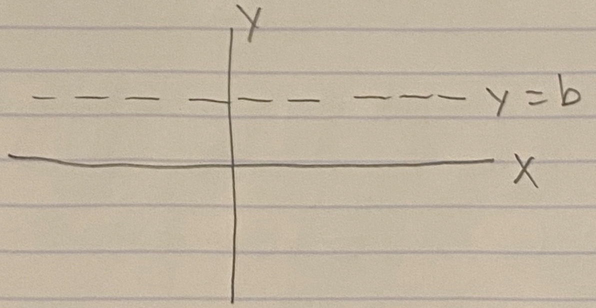
Question 3

3a) $\boxed{y = mx + b}$ for a straight line
w/ m = slope & b = intercept on y-axis

To get a straight line, we set $m = 0$ so that $y = b$, the intercept.

With $y = b$, we get a straight line that is parallel to the x-axis.

Ex:



3b) For $m = 0$:

When $m = 0$, this means $y = c$. Y becomes a constant. As x increases or decreases, y remains constant. This means that x & y are not linearly related if the slope is zero, as they do not impact or change each other.

For $m \neq 0$:

If $m \neq 0$, then $y = mx + b$. A change in m will cause a change in x. As x increases or decreases with m, then y will also increase or decrease. This shows that x & y are linearly related when the slope is not equal to zero.