HW 3 Q1 Solutions

Question (a)

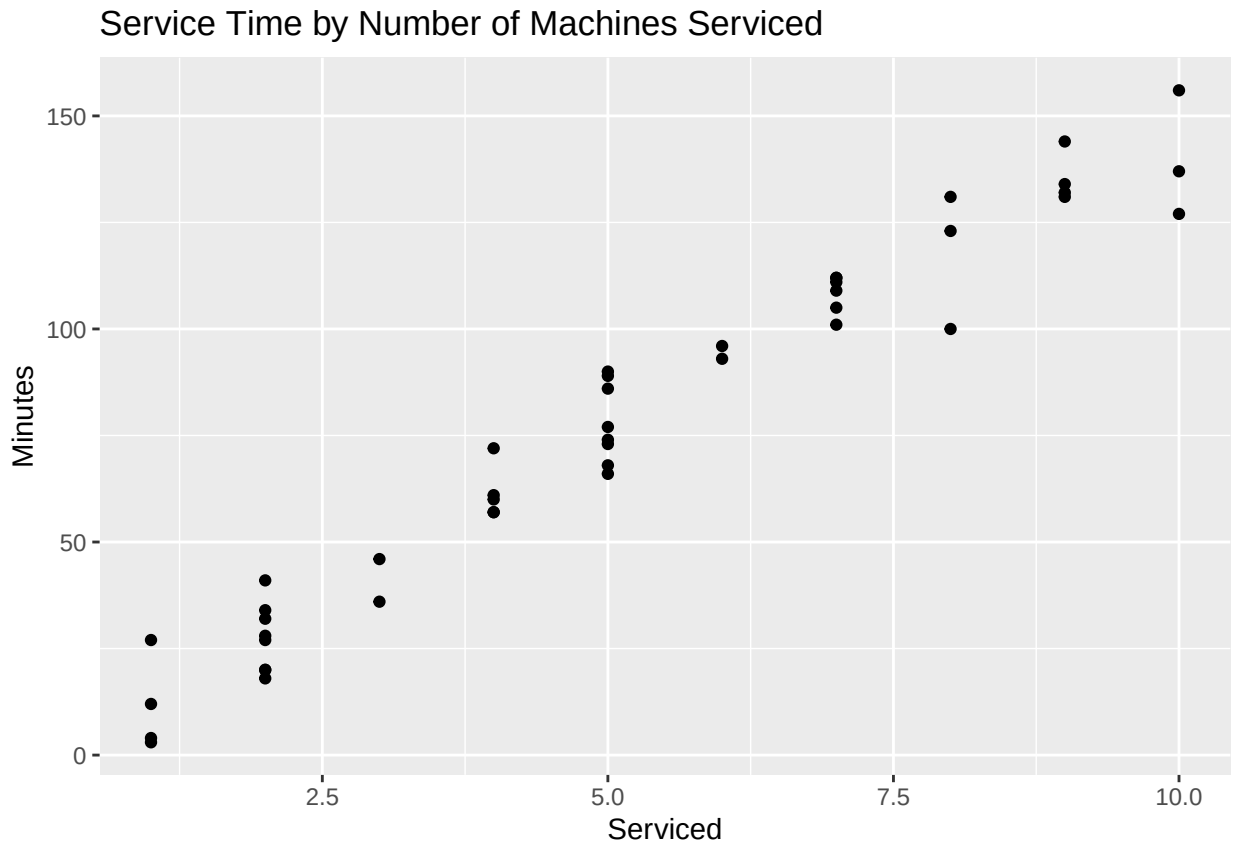
The response variable is *Minutes*, the total time taken by the service person, and the predictor is *Serviced*, the number of copiers serviced.

Question (b)

```
library(tidyverse)

data <- read.table("Copier.txt", header=TRUE ,sep="")

ggplot(data, aes(x=Serviced,y=Minutes))+
  geom_point()+
  labs(x="Serviced", y="Minutes", title="Service Time by Number of Machines Serviced")
```



We can see there is a strong positive linear association between the total time taken by the service person and the number of copiers serviced.

Question (c)

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

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

The correlation is 0.978517. This indicates a strong positive linear relationship.

Question (d)

Since the scatterplot shows a linear relationship, we can reliably interpret the correlation as a measure of linear relation.

Question (e)

```
result<-lm(Minutes~Serviced, data=data)
```

```
summary(result)
```

```
##
## Call:
## lm(formula = Minutes ~ Serviced, data = data)
##
## 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
```

- $\hat{\beta}_1 = 15.0352$
- $\hat{\beta}_0 = -0.5802$
- $R^2 = 0.9575$
- $\hat{\sigma}^2 = 8.914^2 = 79.4594$

Question (f)

For each additional copier serviced, the predicted service time increases by 15.0352 minutes.

When the number of copiers serviced is 0, the predicted service time is -0.5802 minutes. The intercept makes no sense in context because service time cannot be negative. (This is a by product of extrapolation)

Question (g)

```
anova(result)
```

```
## Analysis of Variance Table
##
## Response: Minutes
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Serviced   1  76960   76960  968.66 < 2.2e-16 ***
## Residuals 43   3416      79
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA F statistic is 968.66. The hypotheses are $H_0 : \beta_1 = 0, H_a : \beta_1 \neq 0$. Since the p-value is small, we reject the null hypothesis. The data supports the claim that there is a linear association between the total service time and the number of copiers serviced.

Alternatively, using the critical value approach. Critical value is 4.07 (using qf(0.95,1,43) in R). Since the F-stat is greater than the critical value, we reject the null hypothesis. The data supports the claim that there is a linear association between the total service time and the number of copiers serviced.