HW Module 5

Michael Puchalski

2025-02-17

1. (R required) For this question, you will use the dataset "Copier.txt" for this question.This is the same data set that you used in the last homework. The Tri-City Office Equipment Corporation sells an imported copier on a franchise basis and performs preventive maintenance and repair service on this copier. The data have been collected from 45 recent calls on users to perform routine preventive maintenance service; for each call, Serviced is the number of copiers serviced and Minutes is the total number of minutes spent by the service person.

```
library(tidyverse)
```

It is hypothesized that the total time spent by the service person can be predicted using the number of copiers serviced. Fit an appropriate linear regression and answer the following questions:

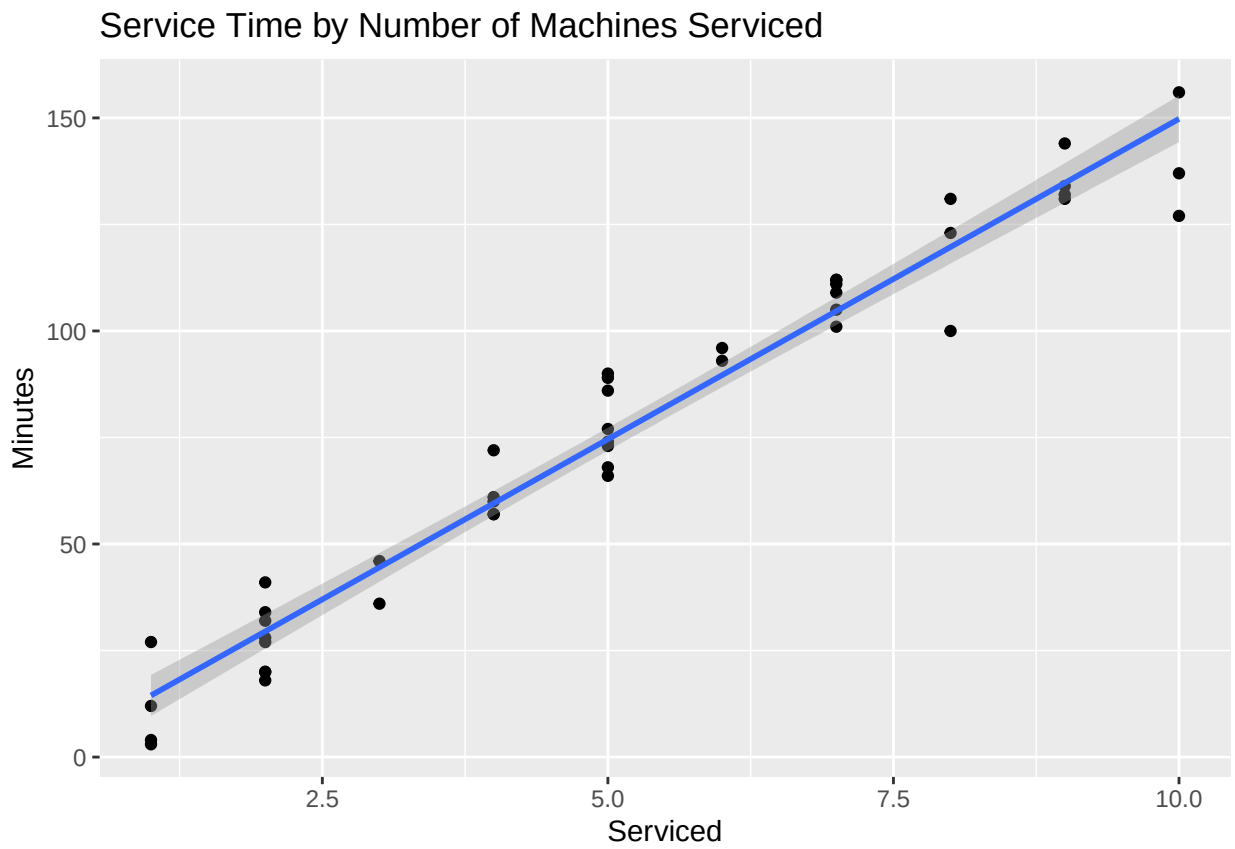
```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

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

mylm <- lm(Minutes~Serviced, data=data)

ggplot(data, aes(x=Serviced,y=Minutes))+
  geom_point()+
  geom_smooth(method = "lm", se = TRUE, level=0.95)+
  labs(x="Serviced", y="Minutes", title="Service Time by Number of Machines Serviced")
```

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



Shown above is a scatterplot fitted with an appropriate linear regression fitted as well as the 95% confidence interval shaded in grey.

```
summary(mylm)
```

(a) Obtain the 95% confidence interval for the slope, B1.

```
##
## 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
```

```r
confint(mylm, level=0.95)
```

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

Here the confidence interval for the slope, indicated in the Serviced row, falls between the lower bound of 14.061 and 16.009 Minutes. This means it is expected that 95% of the time, the average time to fix a copier should be essentially between 14 and 16 minutes with the predicted mean being 15.035 minutes.

```r
new_copiers <- data.frame(Serviced=5)
predict(mylm, new_copiers, level=0.95, interval="confidence")
```

(b) Suppose a service person is sent to service 5 copiers. Obtain an appropriate 95% interval that predicts the total service time spent by the service person.

```
##        fit      lwr      upr
## 1 74.59608 71.91422 77.27794
```

The predicted time is 74.596 minutes, at just under an hour and 15 with a corresponding confidence interval of (71.914, 77.280)mins to be spent on servicing the 5 copiers.

(c) What is the value of the residual for the first observation? Interpret this value contextually. The value of the residual in the first observation is 8.914 on 43 degrees of freedom. This is a low residual value, meaning the extraneous variance not described by the model is low suggesting the model does a fairly good job modeling the relationship between copiers serviced and total service time.