STP 530: Applied Regression Analysis

Name : **Sai Swaroop Reddy Vennapusa**

Homework 2

Instructor : **Yi Zheng**

Due Date : 5th Sep 2023, 10:30AM

Question 2.5 (a) : Estimate the change in the mean service time when the number of copiers serviced increases by one. Use a 90 percent confidence interval. Interpret your confidence interval.

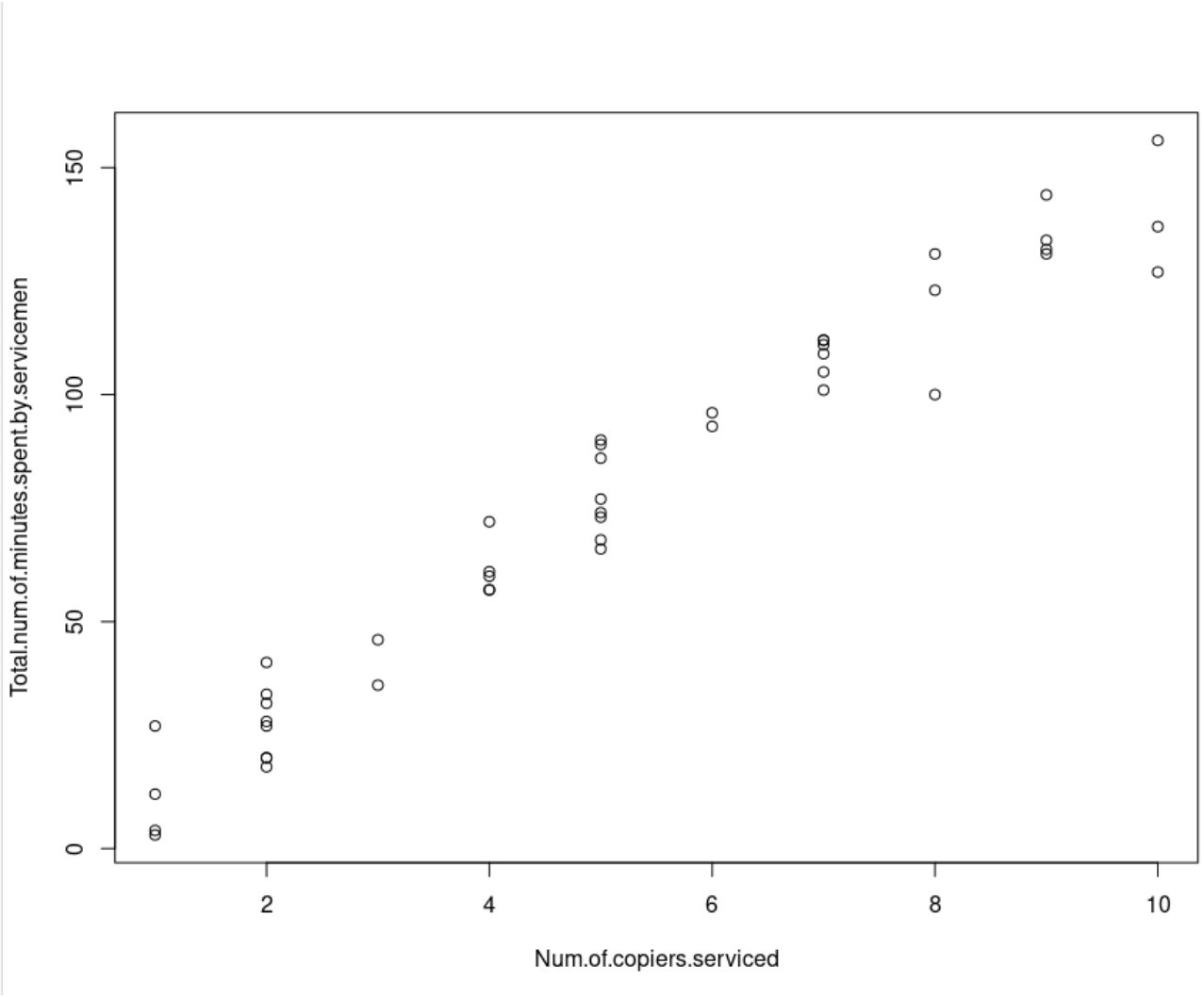
R Code:

```
mydata <- read.table("~/Downloads/Assignments/STP530/HW2/CH01PR20.txt", quote="\"", comment.char="")
head(mydata)

colnames(mydata) <- c("Total.num.of.minutes.spent.by.servicemen", "Num.of.copiers.serviced") # Rename the columns
head(mydata)

plot(Total.num.of.minutes.spent.by.servicemen ~ Num.of.copiers.serviced, data=mydata)

m <- lm(Total.num.of.minutes.spent.by.servicemen ~ Num.of.copiers.serviced, data=mydata)
summary(m)


# Question 2.5
confint(m, level=0.9)
```

R Output:

```
> summary(m)

Call:
lm(formula = Total.num.of.minutes.spent.by.servicemen ~ Num.of.copiers.serviced,
    data = mydata)

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
Num.of.copiers.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

> confint(m, level=0.9)
                             5 %      95 %
(Intercept)             -5.29378  4.133467
Num.of.copiers.serviced 14.22314 15.847352
```

Plot:



Answer:

From the summary(m):
Coefficients of Num.of.copiers.serviced = 15.0352

This coefficient represents the estimated change in the mean service time (in minutes) when the number of copiers serviced increases by one unit. Therefore, for each additional copier serviced, the mean service time increases by approximately 15.0352 minutes.

For a 90% confidence interval:

From confint(m, level=0.9) for Num.of.copiers.serviced: [14.22314, 15.847352]

This means we're 90% confident that the true mean increase in service time for each additional copier is between approximately 14.22 minutes and 15.85 minutes.

Question 2.5 (e) : Does b o give any relevant information here about the "start-up" time on calls-Le., about the time required before service work is begun on the copiers at a customer location?

Answer:

The b_0 coefficient, which in the context of linear regression is commonly known as the intercept.

From the summary output:
b_0 = -0.5802

The intercept, b_0, represents the expected mean response Y when X = 0. In this context, b_0 would be the expected total number of minutes spent by the servicemen when no copiers are serviced.

Given that b_0 is -0.5802, it implies that if a serviceman went to a call and serviced no copiers (which doesn't really make sense in a practical scenario), the expected total number of minutes spent would be a negative value.

Conclusion -
The intercept b_0 value of -0.5802 doesn't provide relevant or realistic information about the "start-up" time on calls. A negative time doesn't make sense in the context of this problem. Ideally, if it were meaningful, the intercept would represent the "start-up" time or the base amount of time it takes for a serviceman before even beginning to work on the copiers, but the negative value suggests that there may be some peculiarities in the data or the linear model might not be the perfect fit for the given data.

Question 2.14(a) : Obtain a 90 percent confidence interval for the mean service time on calls in which six
copiers are serviced. Interpret your confidence interval.

R Code:

```
# Question 2.14(a)
mean_service_time_CI <- predict(m, newdata = data.frame(Num.of.copiers.serviced = 6), interval = "confidence", level = 0.90)
mean_service_time_CI
```

R Output:

```
> mean_service_time_CI
       fit      lwr      upr
1 89.63133 87.28387 91.9788
```

Answer:

The 90% confidence interval for the mean service time when six copiers are serviced is [87.28387,91.9788] minutes.

We are 90% confident that the true mean service time for servicing six copiers lies between approximately 87.28387 minutes and 91.9788 minutes.

Question 2.14(b) : Obtain a 90 percent prediction interval for the service time on the next call in which six copiers are serviced. Is your prediction interval wider than the corresponding confidence interval in part (a)? Should it be?

R Code:

```
# Question 2.14(b)
predict_interval <- predict(m, newdata = data.frame(Num.of.copiers.serviced = 6), interval = "prediction", level = 0.90)
predict_interval
```

R output:

```
> predict_interval
       fit      lwr      upr
1 89.63133 74.46433 104.7983
```

Answer :

The 90% prediction interval for the service time on the next call in which six copiers are serviced is [74.46433,104.7983] minutes.

We are 90% confident that the service time for the next individual call (where six copiers are serviced) will lie between approximately 74.46433 minutes and 104.7983 minutes.

Comparison: The prediction interval [74.46433,104.7983] is wider than the confidence interval [87.28387,91.9788]. This makes sense because prediction intervals account for both the variability in estimating the mean service time (which is what the confidence interval captures) and the inherent variability of individual service times around that mean. Predicting a single observation's outcome is inherently more uncertain than predicting the average outcome, so the prediction interval is wider than the confidence interval.