HW2_JY

2023-09-04

*1.20. **Copier maintenance.** 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 below have been collected from 45 recent calls on users to perform routine preventive maintenance service; for each call, X is the number of copiers serviced and Y is the total number of minutes spent by the service person. Assume that first-order regression model (1.1) is appropriate.

i:	1	2	3	...	43	44	45
X_i:	2	4	3	...	2	4	5
Y_i:	20	60	46	...	27	61	77

*2.5. Refer to **Copier maintenance** Problem 1.20.

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.

b. Conduct a t test to determine whether or not there is a linear association between X and Y here; control the α risk at .10. State the alternatives, decision rule, and conclusion. What is the P-value of your test?

c. Are your results in parts (a) and (b) consistent? Explain.

d. The manufacturer has suggested that the mean required time should not increase by more than 14 minutes for each additional copier that is serviced on a service call. Conduct a test to decide whether this standard is being satisfied by Tri-City. Control the risk of a Type I error at .05. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. Does b_0 give any relevant information here about the "start-up" time on calls—i.e., about the time required before service work is begun on the copiers at a customer location?

2.5a. 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.

90% confidence interval (CI) for the expected changes in the mean service time for each 1-unit increase in the number of copiers serviced = 90% CI for β1.

90% CI for β1 = b1 +/- t(1-α/2, df)*sb1*

Based on the outcome by R code:

- *b1 = 15.0352 mins*
- *α = 0.1*
- *df = 45-2 = 43*
- *qt(1-0.1/2, 43) = 1.681071*
- *Sb1 =estimated standard error of β1 =root(Σe^2 /(n-p)) = 0.4831*
- *LL <- b1 - 1.681071*sb1
- UL <- b1 + 1.681071*sb1

90% CI for β1 (by R code)

```
confint(HW2.mod, level = .90)

##                   5 %      95 %
## (Intercept) -5.29378  4.133467
## ser.no      14.22314 15.847352
```

➔ Thus, the estimated mean service time changed by each increase in the number of copier serviced is 15.0352 mins. With 90% confidence, we estimate that the expected mean service time changes by somewhat between 14.22307 and 15.84733 for each one-unit increase in the No. of copier served.

2.5e. Does bo 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?

From above summary,

- b0 = 0.5802
- sb0 = 2.8039
- p.value.b0 = 0.837

➔ Thus, the 'start-up' time on calls (the intercept) is <u>not likely</u> -0.5802 since it is a negative value. Also, it's not statistically significant with a p-value of 0.837.

*2.14. Refer to **Copier maintenance** Problem 1.20.

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

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?

<u>Confidence interval for E{yh}</u>

$$\hat{Y}_h \pm t(1-\alpha/2, df = n-2) \cdot s\{\hat{Y}_h\}$$

$$s\{\hat{Y}_h\} = s\sqrt{\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}}$$

<u>Prediction interval of Yh:</u>

$$\hat{Y}_h \pm t(1-\alpha/2, df = n-2) \cdot s_{\text{pred}}\{\hat{Y}_h\}$$

$$s_{\text{pred}}\{\hat{Y}_h\} = s\sqrt{1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}}$$

5.14a. Obtain a 90 percent confidence interval for the mean service time on calls in which six copiers are serviced. Interpret your confidence interval.

90% CI for β1 = b1 +/- t((1-a)/2,df)*sb1

15.0352 - qt(p=.95, df=43) * 0.4831 # lower limit
15.0352 + qt(p=.95, df=43) * 0.4831 # upper limit

t(0.95,43)= 1.681071

90% CI for β1 (by R code)

```
confint(HW2.mod, level = .90)

##                    5 %      95 %
## (Intercept) -5.29378  4.133467
## ser.no      14.22314 15.847352
```

➔ The 90% CI for the mean service time when 6 copiers are serviced is between 87.28387 and 91.97880. Thus, we are 90% confident that <u>the **mean value** of Y</u> (expected mean service time) for <u>all</u> observations with X=6 (when 6 copiers are serviced) falls between **87.28387 and 91.97880 mins**.

5.14 b. Obtain a 90 percent prediction interval for the service time on the <u>next</u> call in which six copiers are serviced.

```
Y.hat.6 <- data.frame("ser.no"=6)
predict(HW2.mod, interval="confidence", newdata=Y.hat.6, level=.90) #
confidence interval

##        fit      lwr     upr
## 1 89.63133 87.28387 91.9788

predict(HW2.mod, interval="prediction", newdata=Y.hat.6, level=.90) #
prediction interval

##        fit      lwr      upr
## 1 89.63133 74.46433 104.7983
```

➔ The 90% PI for the service time when 6 copiers are serviced is between **74.46433 and 104.79833 mins**. Thus, with 90% confidence, we <u>**predict**</u> that <u>**the value** of Y</u> (service time) for the <u>next</u> observation with X=6 (when 6 copiers are serviced) falls within the interval.

Is your prediction interval wider than the corresponding confidence interval in part (a)? Should it be?

➔ Yes, the **PI is wider than CI** (see the plot below). In general, it is true that PI is wider than CI. The CI provides a range for <u>where we believe **the average value** lies</u>. In contrast, the PI gives a range <u>where **individual data points** are expected to fall</u>. Since the PI accounts for the natural variability among individual observations, it tends to be broader.

<u>R code and the outcome:</u>

set work directory

```r
setwd("C:/Users/jyang/OneDrive - Arizona State University/10
Classes_OneDrive/2023_STP530_Regression")
```

import data

```r
HW2.data <- read.table("CH01PR20.txt")
head(HW2.data)
```

```
##    V1 V2
## 1  20  2
## 2  60  4
## 3  46  3
## 4  41  2
## 5  12  1
## 6 137 10
```

```r
colnames(HW2.data) <- c("min", "ser.no")
head(HW2.data)
```

```
##   min ser.no
## 1  20      2
## 2  60      4
## 3  46      3
## 4  41      2
## 5  12      1
## 6 137     10
```

regression model

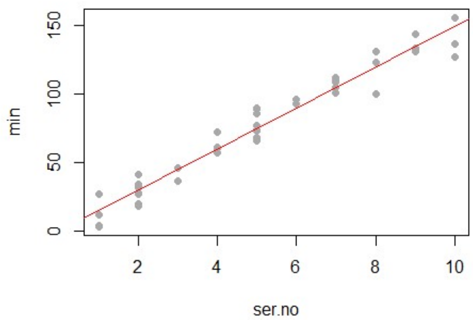
y.hat = b0 + b1X

```r
HW2.mod <- lm(min ~ ser.no, data = HW2.data)
summary(HW2.mod)
```

```
##
## Call:
## lm(formula = min ~ ser.no, data = HW2.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.7723  -3.7371   0.3334   6.3334  15.4039
##
## Coefficients:
##             Estimate Std. Error(sb1)   t value Pr(>|t|)
## (Intercept)  -0.5802     2.8039        -0.207    0.837
## ser.no       15.0352     0.4831         31.123   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error (Se): 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
```

```
plot(min ~ ser.no, data = HW2.data, pch = 16, col = "darkgrey")
abline(coef(HW2.mod), col="red")
```



90% CI for β1

90% confidence interval (CI) for the expected changes in the mean service time for each 1-unit increase in the number of copiers serviced = 90% CI for β1.

90% CI for β1 (by R code)

```
confint(HW2.mod, level = .90)
```

```
##                   5 %      95 %
## (Intercept) -5.29378  4.133467
## ser.no      14.22314 15.847352
```

90% CI for β1 = b1 +/- t((1-a)/2,df)*sb1

15.0352 - qt(p=.95, df=43) * 0.4831 # lower limit

15.0352 + qt(p=.95, df=43) * 0.4831 # upper limit

t(0.95,43)= 1.681071

```
b1 <- 15.0352
df <- 43
qt(1-0.1/2,43)
```

```
## [1] 1.681071
```

```
sb1 <- 0.4831
LL <- b1 - 1.681071*sb1
UL <- b1 + 1.681071*sb1
LL
```

```
## [1] 14.22307
```

```
UL
```

```
## [1] 15.84733
```

95% confidence interval and prediction interval for X=6

```
Y.hat.6 <- data.frame("ser.no"=6)
predict(HW2.mod, interval="confidence", newdata=Y.hat.6, level=.90) #
confidence interval

##         fit      lwr     upr
## 1 89.63133 87.28387 91.9788

predict(HW2.mod, interval="prediction", newdata=Y.hat.6, level=.90) #
prediction interval

##         fit      lwr      upr
## 1 89.63133 74.46433 104.7983
```

plot 95% confidence or prediction interval (CI, PI) for X=6

```
plot(min ~ ser.no, data = HW2.data, pch = 16, col = "darkgrey",
     main = "HW2_Copier maintenance",
     xlab = "No. of copiers serviced",
     ylab = "Mean service time")
min_val <- min(HW2.data$ser.no, na.rm = TRUE)
max_val <- max(HW2.data$ser.no, na.rm = TRUE)
Y.hat.6 <- data.frame("ser.no" = seq(from = min_val, to = max_val, by =
1))
HW2.conf <- predict(HW2.mod, interval="confidence", newdata=Y.hat.6,
level=.90) # confidence interval
head(HW2.conf,6)

##         fit      lwr      upr
## 1 14.45509 10.43812 18.47206
## 2 29.49034 26.11796 32.86272
## 3 44.52559 41.70977 47.34140
## 4 59.56084 57.15175 61.96992
## 5 74.59608 72.36054 76.83162
## 6 89.63133 87.28387 91.97880

lines(HW2.conf[, 1], lwd=2, col = "red")
lines(HW2.conf[, 2], lwd=2, lty = 3, col = "forestgreen" )
lines(HW2.conf[, 3], lwd=2, lty = 3, col = "forestgreen" )
HW2.pred <- predict(HW2.mod, interval="prediction", newdata=Y.hat.6,
level=.90) # prediction interval
head(HW2.pred)

##         fit      lwr       upr
## 1 14.45509 -1.05824  29.96842
## 2 29.49034 14.13129  44.84939
## 3 44.52559 29.27907  59.77210
## 4 59.56084 44.38417  74.73750
## 5 74.59608 59.44600  89.74617
## 6 89.63133 74.46433 104.79833
```

```r
lines(HW2.pred[, 2], lwd=2, lty = 2, col = "blue" )
lines(HW2.pred[, 3], lwd=2, lty = 2, col = "blue" )
legend("topleft", legend = c("regression", "confidence", "prediction"),
       lty=c(1,3,2), lwd = 2, col=c("red", "forestgreen", "blue"),
cex=1)
```