STA 108 Discussion 7: Simultaneous Intervals

Yuanyuan Li

Reference: Textbook Chapter 4.1-4.3.

1. Joint estimation of β_0 and β_1

Let I_0 be the C.I. for β_0, I_1 be the C.I. for β_1. The goal of simultaneous inference is to construct I_0 and I_1 such that

$$P(\{\beta_0 \in I_0\} \cap \{\beta_1 \in I_1\}) \geq 1 - \alpha$$

or

$$P(\{\beta_0 \notin I_0\} \cup \{\beta_1 \notin I_1\}) \leq \alpha.$$

Now we consider to construct the individual intervals with a higher confidence level such that the family confidence coefficient could be at least $1 - \alpha$. This is the idea of **Bonferroni Procedure**. The procedure is a general one that can be applied in many cases, as we shall see, not just for the joint estimation of β_0 and β_1.

Let I_0 be $1 - c$ confidence interval for β_0, I_1 be $1 - c$ confidence interval for β_1, i.e.,

$$P(\{\beta_0 \notin I_0\}) \leq c, \ P(\{\beta_1 \notin I_1\}) \leq c.$$

Then we get

$$\begin{aligned} &P(\{\beta_0 \notin I_0\} \cup \{\beta_1 \notin I_1\}) \leq P(\{\beta_0 \notin I_0\}) + P(\{\beta_1 \notin I_1\}) \leq 2c, \\ \Rightarrow &P(\{\beta_0 \in I_0\} \cap \{\beta_1 \in I_1\}) \geq 1 - 2c. \end{aligned}$$

Therefore, in order to get joint confidence intervals with $1 - \alpha$ family confidence coefficient, we just need to constrcut $1 - \alpha/2$ confidence interval for β_0 and β_1 seperately. Thus, the $1 - \alpha$ family confidence limits for β_0 and β_1 for regression model (2.1) by the Bonferroni procedure are:

$$\hat{\beta}_k \pm B \ s.e.(\hat{\beta}_k), \ k = 0, 1,$$

$$B = t_{1-\alpha/4; \ n-2}.$$

```r
#Read data of Problem 1.21 in textbook
setwd("~/books/108s21/UCDSTA108-master/datasets")#set working directory to "datasets" folder
data1 = read.table("airfreight+breakage.txt")
#replace the file path in (" ") with your own file path for this data set
Y = data1[,1]
X = data1[,2]
n = length(X)
#Get least square estimates:
fit=lm(Y~X)
```

```
b0hat = fit$coefficients[[1]]
b1hat = fit$coefficients[[2]]
mse= summary(fit)$sigma^2
se.b0hat= sqrt(mse*(1/n+ mean(X)^2/sum((X - mean(X))^2)))
se.b1hat = sqrt(mse/sum((X-mean(X))^2))
#Get 95% simultaneous intervals:
alpha=0.05
B= qt(1-alpha/4,n-2)
c(b0hat-B*se.b0hat, b0hat+B*se.b0hat)#c.i. for beta_0
```

```
## [1]  8.374846 12.025154
```

```
c(b1hat-B*se.b1hat, b1hat+B*se.b1hat)#c.i. for beta_1
```

```
## [1] 2.709421 5.290579
```

Interpretation: We conclude that β_0 is between 8.37 and 12.03 *and* β_1 is between 2.71 and 5.29 with 95% family confidence coefficient.

2. Simultaneous Estimation of Mean Response

Often the mean responses at a number of X levels need to be estimated from the same sample data. The separate interval estimates of $E(Y_h)$ at the different X_h levels need not all be correct or all be incorrect, which means we need to construct simultaneous intervals for a number of different mean responses with a family confidence coefficient.

2.1 Bonferroni procedure

Using the Bonferrnoni procedure to construct simultaneous intervals for mean responses at different X levels, we calculate in each instance the usual confidence limits for a single mean response $E(Y_h)$, adjusting the statement confidence coefficient to yield the specified family confidence coefficient. When $E(Y_h)$ is to be estimated for g levels X_h with family confidence coefficient $1-\alpha$, the Bonferroni confidence limits for regression model are:

$$\hat{Y}_h \pm B\ s.e.(\hat{Y}_h),$$

$$B = t_{1-\alpha/2g;\ n-2}.$$

```
xh = c(1,2,3)
g=length(xh)
yhat = b0hat + b1hat*xh#
se.yhat = sqrt(mse*(1/n+ (xh - mean(X))^2/sum((X - mean(X))^2)))
B= qt(1-alpha/(2*g), n-2)
library(knitr)
CI.b = data.frame(xh, yhat, lower=yhat-B*se.yhat,upper=yhat+B*se.yhat)
kable(CI.b,caption = "Bonferroni Confidence intervals for mean response")
```

Table 1: Bonferroni Confidence intervals for mean response

xh	yhat	lower	upper
1	14.2	12.78548	15.61452
2	18.2	16.19957	20.20043
3	22.2	19.03704	25.36296

Interpretation: With family confidence coefficient 95%, we conclude that the *mean* number of broken ampules is between 12.79 and 15.61 for 1 transfer, between 16.20 and 20.20 for 2 transfers, between 19.04 and 25.36 for 3 transfers.

2.2 Working-Hotelling procedure

The WoIking-Hotelling procedure is based on the confidence band for the regression line. The confidence band contains the entire regression line and therefore contains the mean responses at all X levels. Hence, we can use the boundary values of the confidence band at selected X levels as simultaneous estimates of the mean responses at these X levels. The family confidence coefficient for these simultaneous estimates will be at least $1-\alpha$ because the confidence coefficient that the entire confidence band for the regression line is correct is $1-\alpha$. The simultaneous confidence limits for g mean responses $E(Y_h)$ for regression model with the Working-HoteIling procedure therefore are:

$$\hat{Y}_h \pm W s.e.(\hat{Y}_h),$$

where

$$W = \sqrt{2F_{1-\alpha;\ 2,n-2}}.$$

```
W = sqrt(2*qf(1-alpha,2,n-2))
CI.w = data.frame(xh, yhat, lower=yhat-W*se.yhat, upper=yhat+W*se.yhat)
kable(CI.w,caption = "W-H Confidence intervals for mean response")
```

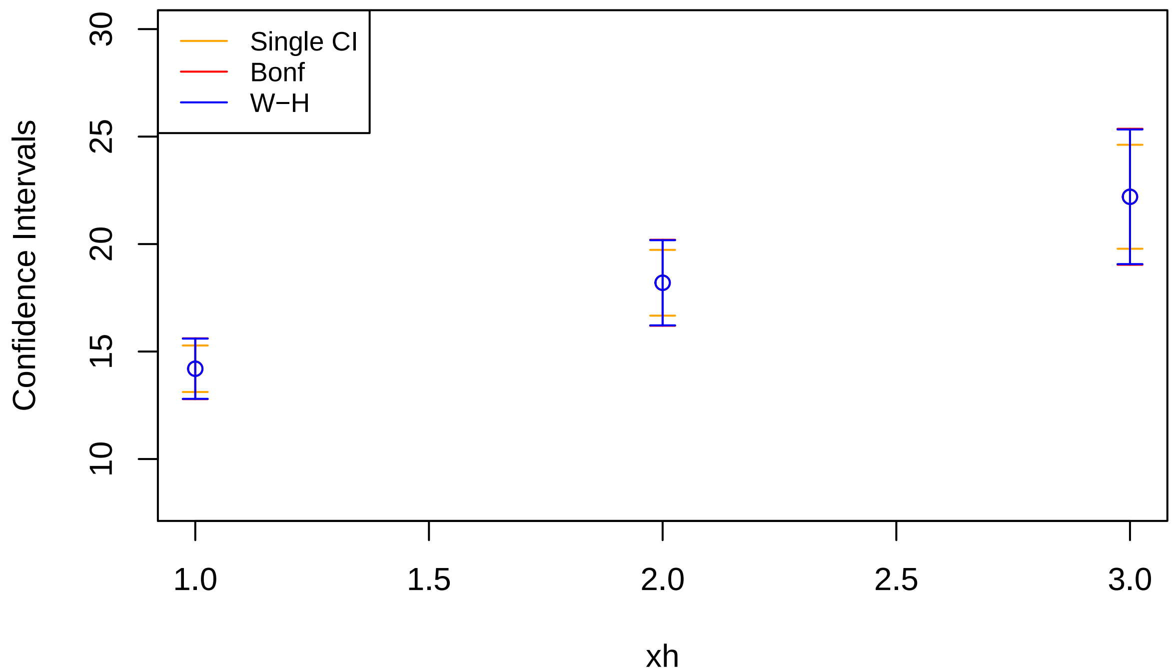
Table 2: W-H Confidence intervals for mean response

xh	yhat	lower	upper
1	14.2	12.79930	15.60070
2	18.2	16.21912	20.18088
3	22.2	19.06795	25.33205

Simultaneous intervals for g mean responses at different X levels are always wider than g confidence intervals for single mean response (in discussion 3). The B multiple is larger than the original t multiple($t_{1-\alpha/2}$) since B has higher confidence coefficient($B = t_{1-\alpha/2g}$). The W multiple will be larger than the original t multiple because the confidence band must encompass the entire regression line, whereas the confidence limits for $E(Y_h)$ at X_h apply only at the single level X_h. When g is larger(larger family), the Working-Hotelling confidence limits will always be tighter than Bonferroni confidence limits, since W stays the same for any number of statements (g) in the family whereas B becomes larger as the g increases. In practice, once the family confidence coefficient has been decided upon, one can calculate the W and B multiples to determine which procedure leads to tighter confidence limits.

```
#CI for single mean reponse at a given Xh(same with c.i. in discussion 3)
cv= qt(1-alpha/2, n-2)
```

```r
CI= data.frame(yhat, lower=yhat-cv*se.yhat, upper=yhat+cv*se.yhat)
library(plotrix)
plotCI(xh, CI$yhat, ui=CI$upper, li=CI$lower, col = "orange",ylim=c(8,30)
       ,ylab="Confidence Intervals")
plotCI(xh, CI.b$yhat, ui=CI.b$upper, li=CI.b$lower, col = "red", add = TRUE)
plotCI(xh, CI.w$yhat, ui=CI.w$upper, li=CI.w$lower, col = "blue",add = TRUE)
legend("topleft",legend=c("Single CI","Bonf","W-H"),
       col = c("orange", "red","blue"),lty=1, cex=0.8)
```



3. Simultaneous Prediction Intervals for New Observations

Now we consider the simultaneous predictions of g new observations on Y in g independent trials at g different levels of X.

3.1 Bonferroni procedure

$$\hat{Y}_h \pm B \text{ p.s.e}(\hat{Y}_h),$$
$$B = t_{1-\alpha/2g;\ n-2}.$$

The B multiple is the same with that in part 2, while p.s.e$(\hat{Y}_h)$ is the prediction standard error(in discussion 3).

```r
pse.yhat = sqrt(mse*(1+1/n+ (xh - mean(X))^2/sum((X - mean(X))^2)))
PI.b = data.frame(xh, yhat, lower=yhat-B*pse.yhat,upper=yhat+B*pse.yhat)
kable(PI.b, caption = "Bonferroni Prediction intervals for new observations")
```

Table 3: Bonferroni Prediction intervals for new observations

xh	yhat	lower	upper
1	14.2	9.508576	18.89142
2	18.2	13.299967	23.10003
3	22.2	16.721597	27.67840

Interpretation: With family confidence coefficient 95%, we conclude that the number of broken ampules is between 9.51 and 18.89 for 1 transfer, between 13.30 and 23.10 for 2 transfers, between 16.72 and 27.68 for 3 transfers.

3.2 Scheffe procedure

$$\hat{Y}_h \pm S \text{ p.s.e}(\hat{Y}_h),$$
$$S = \sqrt{gF_{1-\alpha;\ g,n-2}}.$$

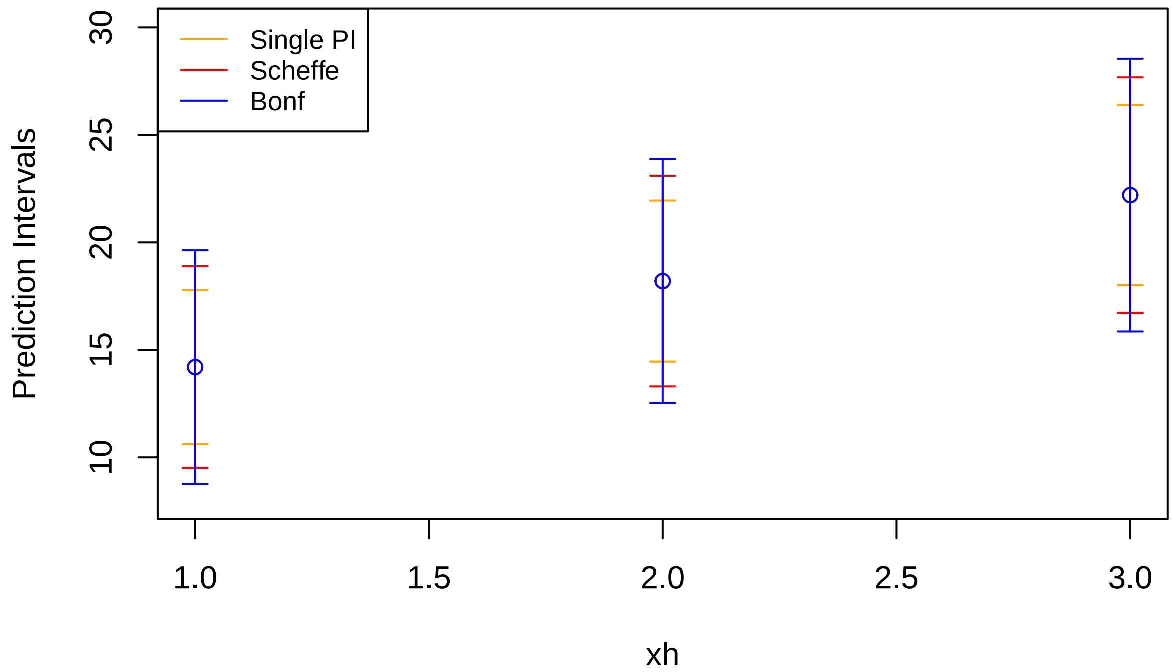
```
S = sqrt(g*qf(1- alpha,g,n-2))
PI.s = data.frame(xh, yhat, lower=yhat-S*pse.yhat,upper=yhat+S*pse.yhat)
kable(PI.s, caption = "Scheffe Prediction intervals for new observations")
```

Table 4: Scheffe Prediction intervals for new observations

xh	yhat	lower	upper
1	14.2	8.766726	19.63327
2	18.2	12.525130	23.87487
3	22.2	15.855302	28.54470

Simultaneous prediction intervals for g new observations on Y at g different levels of X with $1-\alpha$ family confidence coefficient are wider than the corresponding single prediction intervals. Both B and S multiples will be larger as g increase. They can be evaluated in advance to see which procedure provides tighter prediction limits.

```
#PI for single observation at a given Xh(same with p.i. in discussion 3)
PI= data.frame(yhat, lower=yhat-cv*pse.yhat, upper=yhat+cv*pse.yhat)
plotCI(xh, PI$yhat, ui=PI$upper, li=PI$lower,col = "orange",ylim=c(8,30)
       ,ylab="Prediction Intervals")
plotCI(xh, PI.b$yhat, ui=PI.b$upper, li=PI.b$lower,col = "red",add = TRUE)
plotCI(xh, PI.s$yhat, ui=PI.s$upper, li=PI.s$lower,col = "blue",add = TRUE)
legend("topleft",legend=c("Single PI", "Scheffe","Bonf"),
       col = c("orange","red","blue"),lty=1, cex=0.8)
```


Single PI
Scheffe
Bonf
Prediction Intervals
10
15
20
25
30
1.0
1.5
2.0
2.5
3.0
xh