Confidence Band and Simultaneous Confidence Inference in SLR

The Simultaneous(as known as the Family or Joint) Confidence Interval Problem

1. Simultaneously (joint) estimation of **p** parameters, in SLR, $p = 2\ with\ \beta_0\ and\ \beta_1,$ with **all** X values.
   - Provide confidence that the conclusions for both $\beta_0\ and\ \beta_1$ are correct.
2. Simultaneously estimation of **g** mean response $\hat{Y}_h$ with **g** (usually different) X_h values.
   - Only one mean response value is made at one X_h value, $s^2\{\hat{Y}_h\} = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right]$
   - The **g** different mean responses made at different X_h values might not all be correct at the same time, even though all estimates are based on the same fitted regression line, depending on **p** parameters, e.g., $b_0\ and\ b_1, p = 2$ in SLR.
3. Simultaneously estimation of **g** single response $\hat{Y}_h\{new\}$ with **g** (usually different) X_h values
   - Only one single response value is made at one X_h value, $s^2\{pred\} = MSE[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}]$=$s^2 + s^2\{\hat{Y}_h\}$

The Method

1. Simultaneously (joint) estimation of $\boldsymbol{p}$ parameters, in SLR, $p = 2\ with\ \beta_1\ and\ \beta_{2,}$ with **all** X values.
    a) Bonferroni
    b) Working-hoteling
2. Simultaneously estimation of mean response $\hat{Y}_h$ with $\boldsymbol{g}$ different X_h values.
    a) Bonferroni
    b) Working-hoteling
3. Simultaneously estimation of single response $\hat{Y}_h\{new\}$ with $\boldsymbol{g}$ different X_h values
    a) Bonferroni
    b) Schefft

The Method

1. Simultaneously (joint) estimation of $\boldsymbol{p}$ parameters, in SLR, $p = 2\ with\ \beta_1\ and\ \beta_{2,}$ with **all** X values.
   a) Bonferroni
   b) Working-hoteling

The Individual Confidence Interval

- Consider to estimate β_0 and β_1 respectively
- The two individual intervals are

$$b_0 \pm t(1-\alpha/2; n-2)s\{b_0\} \quad = I_0$$
$$b_1 \pm t(1-\alpha/2; n-2)s\{b_1\} \quad = I_1$$

$$The\ event\ I_0^+\ means\ the\ I_0\ sucessfully\ contains\ the\ true\ intercepe, \beta_0$$

$$The\ event\ I_1^+\ means\ the\ I_1\ sucessfully\ contains\ the\ true\ slope, \beta_1$$

$\Pr(I_0^+) = \Pr(I_1^+) = 1 - \alpha_i$, where α_i is used for individual interval

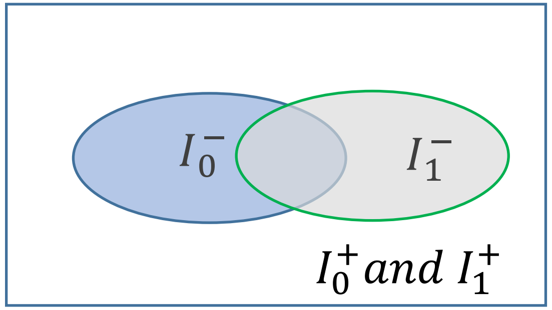
$$The\ event\ I_0^-\ \ and I_1^-\ \ define\ the\ complement\ events, and$$
$$\Pr(I_0^-) = \Pr(I_1^-) = \alpha_i$$

For example,
$\Pr(I_0^+) = \Pr(I_1^+) = 0.95$
$\Pr(I_0^-) = \Pr(I_1^-) = 0.05$

Bonferroni Inequality

When $\Pr(I_0^+) = \Pr(I_1^+) = 0.95$, and $\Pr(I_0^-) = \Pr(I_1^-) = 0.05$

Q: what of the following is true about the probability that both individual intervals are correct, i.e., $= P(I_0^+ \, and \, I_1^+)$.



A) 0.95(0.95)
B) 1-2(0.05)
C) 1-2(0.05)+(0.05)(0.05)
D) $\geq 1 - 2(0.05)$

The answer is D). Note that A) or C) is true only when the individual estimates are independent.

$$P(I_0^+ \, and \, I_1^+) = 1 - [P(I_0^-) + P(I_1^-) - P(I_0^- \, and \, I_1^-)] \geq 1 - (P(I_0^-) + P(I_1^-)) = 1 - 2\alpha_i$$

$$\begin{aligned} P(I_0^+ \, and \, I_1^+) &\geq 1 - 2\alpha_i \\ &= 1 - 2(0.05) \\ &= 1 - 0.1 = 0.9 \end{aligned}$$

- The α value for the joint confidence interval is at most 0.1, and the joint confidence level is at least 0.9
- This feature is called the Bonferroni Inequality.
- Bonferroni Inequality extends to k groups

$$P(I_1^+ \, and \, I_2^+ \, \ldots . \, and \, I_k^+) \geq 1 - k(\alpha_i)$$

The Bonferroni Joint Confidence Interval

$$P(I_0^+ and\ I_1^+) \geq 1 - 2\alpha_i = 1 - \alpha, \qquad where\ \alpha_i = \frac{\alpha}{2}$$

Thus, if we want the joint confidence interval $P(I_0^+ and\ I_1^+)$ to be at least $1-\alpha$, set the individual alpha value $\alpha_i = \frac{\alpha}{2}$, note that the individual confidence level is now at least $1 - \alpha_i = 1 - \frac{\alpha}{2}$

For MLR with p paramters, if we want $P(I_0^+ and\ I_2^+ \ldots.and\ I_{p-1}^+)$ to be at least $1-\alpha$, set the individual alpha value $\alpha_i = \frac{\alpha}{p}$ such that the joint confidence level

$$P(I_0^+ and\ I_2^+ \ldots.and\ I_{p-1}^+) \geq 1 - p(\alpha_i) = 1 - \alpha\ .$$

This estimation is a very conservative because it overestimates the actual confidence level.

The Bonferroni Critical Value

The $\mathbf{1-\alpha}$ joint confidence interval is done by estimating β_0 and β_1 $(p=2)$ each with the individual confidence level of at least $\mathbf{1-\frac{\alpha}{2}}$, or the alpha value, $\alpha_i = \frac{\alpha}{2}$

We can now adjust the individual confidence interval to the Bonferroni Joint Confidence Interval:

$$b_0 \pm t\left(1-\frac{\alpha_i}{2}, dfE\right)s\{b_0\} = b_0 \pm t\left(1-\frac{\frac{\alpha}{2}}{2}, dfE\right)s\{b_0\} = b_0 \pm t(1-\alpha/4, dfE)s\{b_0\}$$

$$b_1 \pm t\left(1-\frac{\alpha_i}{2}, dfE\right)s\{b_1\} = b_1 \pm t\left(1-\frac{\frac{\alpha}{2}}{2}, dfE\right)s\{b_1\} = b_1 \pm t(1-\alpha/4, dfE)s\{b_1\}$$

In general, the Bonferroni Joint Confidence Interval for p parameters consists of p intervals,

$$b_k \pm t\left(1-\frac{\alpha_i}{2}, dfE\right)s\{b_k\} = b_k \pm t\left(1-\frac{\frac{\alpha}{p}}{2}, dfE\right)s\{b_k\} = b_k \pm t(1-\alpha/2p, dfE)s\{b_k\},$$ where k ranges from 0 to $p-1$.

The Bonferroni critical value is defined as

$$B = t(1-\alpha/2p, dfE),\ or\ t(1-\frac{\alpha}{4}; n-2)$$ in SLR.

Example: in the Toluca company case, compute the 90% family confidence interval for (β_0, β_1)

$$B = t\left(1 - \frac{\alpha}{4}; n-2\right) = \quad t(0.975, 23) = 2.069$$

$$b_0 \pm Bs\{b_0\} = 62.366 \pm 2.069 * 26.177 = (8.2, 116.5)$$

$$b_1 \pm Bs\{b_1\} = 3.570 \pm 2.069 * 0.347 = (2.85, 4.29)$$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   62.366     26.177   2.382   0.0259 *
size           3.570      0.347  10.290 4.45e-10 ***

Residual standard error: 48.82 on 23 degrees of freedom
Multiple R-squared:  0.8215,    Adjusted R-squared:  0.8138
F-statistic: 105.9 on 1 and 23 DF,  p-value: 4.449e-10
```

- This is essentially setting the two individual confidence level as 95% for both β_0 and β_1, to ensure the family have at least a 90% confidence level.

Some Example on the Bonferroni Critical Value Notation

$\alpha = 0.1, \qquad g = 2$

$$B = t\left(1 - \frac{\alpha}{2g}; n - p\right) = t(1 - 0.025; n - p) = t(0.975, n - p)$$

$\alpha = 0.05, \qquad g = 2$

$$B = t\left(1 - \frac{\alpha}{2g}; n - p\right) = t(1 - 0.0125; n - p) = t(0.9875, n - p)$$

$\alpha = 0.15, \qquad g = 3$

$$B = t\left(1 - \frac{\alpha}{2g}; n - p\right) = t(1 - 0.025; n - p) = t(0.975, n - p)$$

$\alpha = 0.1, \qquad g = 3$

$$B = t\left(1 - \frac{\alpha}{2g}; n - p\right) = t(1 - 0.0167; n - p) = t(0.983, n - p)$$

Working-hoteling Joint confidence interval (confidence band)

$$F = (\mathbf{b} - \boldsymbol{\beta}^*)'(\mathbf{X}'\mathbf{X})(\mathbf{b} - \boldsymbol{\beta}^*)/p\text{MSE},$$

$$F \sim F(p, n-p).$$

$F \leq F_\alpha,$ Where $F_\alpha = F(1-\alpha; p, n-p)$ is the $(1-\alpha)100th$ percentile of the $F - distribution\ (p, n-p)$

$$(b_j - \beta_j)^2 \leq \mathrm{pF}_\alpha \mathrm{MSE} = W^2 \mathrm{MSE} \qquad Where\ W^2 = pF_\alpha(p, n-p)$$

which gives $b_j - Ws(b_j) \leq \beta_j \leq b_j + Ws(b_j)$ $where\ j = 1, 2, ..p$

In SLR, $p = 2$ $W^2 = pF_\alpha(p, n-p) = 2F_\alpha(2, n-p)$

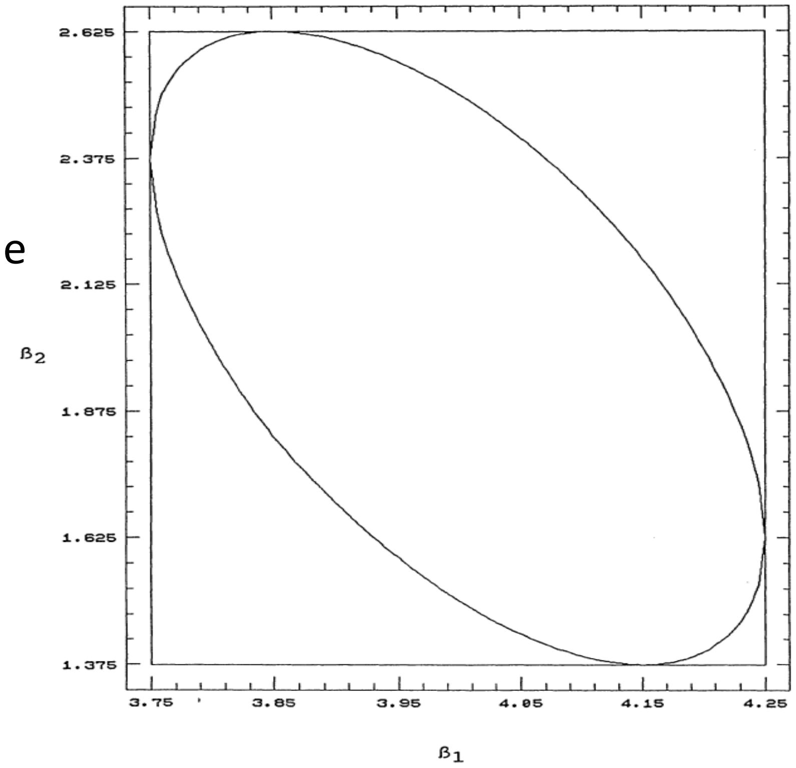


Figure 1. Exact Versus Conservative Confidence Regions. The plot compares a typical exact confidence ellipse with a conservative confidence rectangle for the case $p = 2$; that is, the case $Y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$. The projection of the ellipse onto the β_1 axis is the interval 3.75–4.25, whereas the projection of the ellipse onto the β_2 axis is 1.375–2.625.

Reference: David M. Nickerson, Construction of a conservative confidence region from projections of an exact confidence region in Multiple Linear Regression, The American Statistician, Vol. 48, No.2 (May, 1994)

Working hoteling confidence band on estimating $(\boldsymbol{\beta_0}, \boldsymbol{\beta_1})$

- Working hoteling method is based on a F distribution: $W^2 = pF_\alpha(p, n-p)$

  $\beta_j \pm Ws\{\beta_j\}$ In SLR, $j = 0 \text{ or } 1, and \ \mathrm{W}^2 = 2\, F(1-\alpha; 2, n-2),$

- In the Toluca case, find the joint 90% confidence interval for the parameters (β_0, β_1).

$$W = \sqrt{2\, F(1-\alpha; 2, n-2)} = 2.258$$

```
sqrt(2*qf(0.9,2,23))
```

$$b_0 \pm Ws\{b_0\} = 62.37 \pm W(26.18) = 62.3 \pm 59.11$$

$$b_1 \pm Ws\{b_1\} = 3.57 \pm W(0.347) = 3.57 \pm 0.78$$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   62.366     26.177   2.382   0.0259 *
size           3.570      0.347  10.290 4.45e-10 ***

 Residual standard error: 48.82 on 23 degrees of freedom
 Multiple R-squared:  0.8215,	Adjusted R-squared:  0.8138
 F-statistic: 105.9 on 1 and 23 DF,  p-value: 4.449e-10
```

In this case of $n = 25, p = 2,$ the Bonferroni procedure is better because the critical value is smaller.

Working hoteling confidence band on estimating ($\boldsymbol{\beta_0}, \boldsymbol{\beta_1}$)

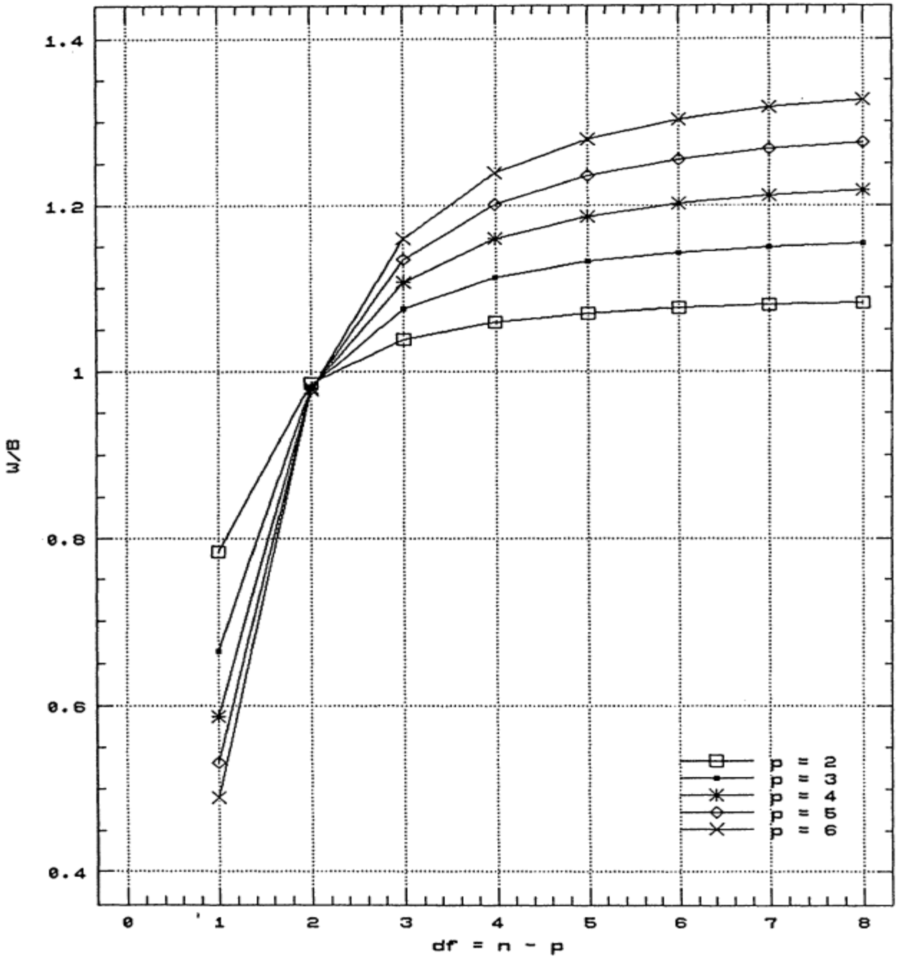


Figure 2. Working–Hotelling–Scheffè/Bonferroni Versus Degrees of Freedom and Dimension. This compares the ratio of the Working–Hotelling–Scheffè multiplier, W, with the Bonferroni multiplier, B, across values of df = $(n - p)$ *for* $p = 2, 3, 4, 5, 6$ *at* $\alpha = 0.10$.

- For $\mathrm{n} - p \geq 2$, the Bonferroni procedure is better because
$$\frac{W}{B} > 1$$

- This conclusion holds true when we try to estimate the **coefficients**, or the **linear impacts ($\boldsymbol{\beta}$) simultaneously in a linear model.**

The method

1. Simultaneously (joint) estimation of $\boldsymbol{p}$ parameters, in SLR, $p = 2\ with\ \beta_1\ and\ \beta_2,$ with **all** X values.
   a) Bonferroni
   b) Working-hoteling
2. Simultaneously estimation of mean response $\hat{Y}_h$ with $\boldsymbol{g}$ different X_h values.
   a) Bonferroni
   b) Working-hoteling
3. Simultaneously estimation of single response $\hat{Y}_h\{new\}$ with $\boldsymbol{g}$ different X_h values
   a) Bonferroni
   b) Scheffe

Bonferroni Joint (or Family) Confidence Interval to predict the mean response, $\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$ ($\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$ given X_1, $\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$ given X_2, …, $\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$ given X_g)

The Bonferroni procedure is very general. To make joint confidence interval for multiple (g) simultaneous prediction

For each mean response $\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$ for a given X_h

$$\hat{Y}_h \pm t\left(1-\frac{\alpha}{2}; n-2\right) s\{\hat{Y}_h\} \qquad s^2\{\hat{Y}_h\} = s^2\left[\frac{1}{n} + \frac{(X_h-\bar{X})^2}{\Sigma(X_i-\bar{X})^2}\right]$$

Then for the $1-\alpha$ joint CI for g predictions, change the confidence level of each individual CI to be $1-\alpha/g$

$$\hat{Y}_h \pm t\left(1-\frac{\alpha}{2\boldsymbol{g}}; n-2\right) s\{\hat{Y}_h\}$$

Note: if a sufficiently large number of simultaneous predictions are made, the width of the individual confidence Intervals may become so wide that they are no longer useful.

Toluca Company Example: Bonferroni Joint (or Family) Confidence Interval on Mean Response $\hat{Y}_h$

What is the simultaneous estimates for the mean number of work hours for $X_h = 30, 65\ and\ 100$ (i.e., $g = 3$) with family confidence level **0.9** ($\alpha = 0.1$). Suppose $\bar{X} = 70,\ SSX = 19800, s = 48.82, \mathrm{n} = 25$

$$\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2g}; n-2\right) s\{\hat{Y}_h\}\ X_h = 30, 65\ and\ 100\ \ (\hat{Y}_h = 169.5, 294.4, and\ 419.4\ respectively)$$

- Set the **confidence level** for each of the g individual estimate to be $1 - \frac{\alpha}{g} = 1 - \frac{0.1}{3} = 96.7\%$
- Then the **confidence level** for the family estimate is at least $1 - g * \frac{\alpha}{g} = 1 - 0.1 = 90\%$

Toluca Company Example: Bonferroni Joint (or Family) Confidence Interval on Mean Response $\hat{Y}_h$

What is the simultaneous estimates for the mean number of work hours for $X_h = 30, 65 \ and \ 100$ (i.e., $g = 3$) with family confidence level **0.9** $(\alpha = 0.1)$. Suppose $\bar{X} = 70,\ SSX = 19800, s = 48.82, n = 25$

$$\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2g}; n-2\right) s\{\hat{Y}_h\} \quad X_h = 30, 65 \ and \ 100 \ \ (\hat{Y}_h = 169.5, 294.4, and \ 419.4 \ respectively)$$

1. $X_h = 30$ $\quad s^2\{\hat{Y}_h\} = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right] = 48.82^2\left[\frac{1}{25} + \frac{(30-70)^2}{s^2(n-1)}\right] = 48.82^2\left[\frac{1}{25} + \frac{(30-70)^2}{19800}\right] = 288.169$

$$\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2g}; n-2\right) s\{\hat{Y}_h\} = 169.5 \pm t\left(1 - \frac{0.1}{6}; 23\right)\sqrt{288.169} = 169.5 \pm 2.263(16.97) = (131.1, \quad 207.9)$$

2. $X_h = 65$ $\quad \hat{Y}_h \pm t\left(1 - \frac{\alpha}{2g}; n-2\right) s\{\hat{Y}_h\} = 294.4 \pm 2.263\ (9.92) \ = (272, 316.8)$

3. $X_h = 100$ $\quad \hat{Y}_h \pm t\left(1 - \frac{\alpha}{2g}; n-2\right) s\{\hat{Y}_h\} = 419.4 \pm 2.263(14.27) \ = (387.1, 451.7)$

```
d<-data.frame(c(30,65,100))
ci.reg(tol.mod, d, type='b',alpha=0.1)
```

	size <dbl>	Fit <dbl>	Lower.Band <dbl>	Upper.Band <dbl>
1	30	169.4719	131.0570	207.8868
2	65	294.4290	271.9783	316.8797
3	100	419.3861	387.0774	451.6947

Working-Hoteling Joint Confidence Interval to predict the mean response, $\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$

($\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$ given X_1,$\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$ given X_2, …, $\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$ given X_g)

- The WH procedure estimate $\beta_0 \ and\ \beta_1$ with a F distribution using the whole scale of X.
- The estimation of true mean $\mu_h = \beta_0 + \beta_1 X_h$ depends on the estimation of $\beta_0 \ and\ \beta_1$ and a given constant X_h
- The WH procedure for estimating mean for what ever X_h level is through a conservative estimation of $\beta_0 \ and\ \beta_1$

- For SLR, $\hat{Y}_h \pm W\ s\{\hat{Y}_h\}$ $\mathrm{W}^2 = 2F(1-\alpha; 2, n-2)$

- For MLR, $\hat{Y}_h \pm W\ s\{\hat{Y}_h\}$ $\mathrm{W}^2 = pF(1-\alpha; p, n-p)$, where p is the number of parameters, not the number of predictions to make, i.e., g.

- Note that for MLR, the $s^2\{\hat{Y}_h\} = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right]$ formula should be modified.

Toluca Company Example: Working-Hoteling Confidence band on Mean Response $\hat{Y}_h$

What is the simultaneous estimates for the mean number of work hours for $X_h = 30, 65\ and\ 100$ (i.e., $g = 3$) with family confidence level **0.9** ($\alpha = 0.1$). Suppose $\bar{X} = 70,\ SSX = 19800, s = 48.82, n = 25$

$X_h = 30, 65\ and\ 100,$
$\hat{Y}_h = 169.5, 294.4, and\ 419.4\ respectively,$

$$s^2\{\hat{Y}_h\} = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right] = 288.169, 98.41\ and\ 203.63\ \ respectively,$$

$Wc = 2.258\ \ (for\ any\ given\ X_h)$

```
sqrt(2*qf(0.9,2,23))
x<-data.frame(size=c(30,65,100))
ci.reg(toluca.mod, x, type='w',alpha=0.1)  # Working hotelling
```

Lower.Band <dbl>	Upper.Band <dbl>
131.1542	207.7897
272.0351	316.8229
387.1591	451.6130

1. $X_h = 30$ $\quad \hat{Y}_h \pm W\ s\{\hat{Y}_h\}$ =(131,15, 207.79)

2. $X_h = 65$ $\quad \hat{Y}_h \pm W\ s\{\hat{Y}_h\}$ =(272.04, 316.82)

3. $X_h = 100$ $\quad \hat{Y}_h \pm W\ s\{\hat{Y}_h\}$ =(387.16, 465.61)

Bonferroni simultaneous CI vs Working hoteling confidence band on estimating mean $\hat{Y}_h$

Working hoteling procedure: $\hat{Y}_h \pm W s\{\hat{Y}_h\}$ where $W^2 = p\,F(1-\alpha; p, n-p)$

Bonferroni procedure: $\hat{Y}_h \pm t\left(1-\frac{\alpha}{2g}; n-2\right) s\{\hat{Y}_h\}$

Compare the critical value with 90% confidence level, $dfE = 23$

$W = 2.258\ (for\ all\ g)$

$W = \sqrt{pF(1-\alpha, p, n-p}$
$= \sqrt{2F(0.9, 2, 23)} = 2.258$

B=$t\left(1-\frac{\alpha}{2g}; n-2\right) = 1.714\ (g=1)$

B=$t(0.975, 23) = 2.069\ \ (g=2)$

B=$t(0.9833, 23) = 2.263\ (g=3)$

B=$t(0.99, 23) = 2.5\ \ \ (g=5)$

B=$t(0.995, 23) = 2.807\ (g=10)$

- WH is better than Bonferroni when $g \geq 3$

Bonferroni simultaneous CI vs Working hoteling confidence band on estimating mean $\hat{Y}_h$

Comments:

- Both the WH and Bonferroni procedures provide wider bounds to the actual family confidence level.
- For larger families (g), the WH confidence band will be narrower since W stays the same for all g, while B gets larger.
- The levels for the predictor variable (X_h) are sometimes not known in advance. In such case, it is better to use the WH procedure because the family confidence interval encompasses all possible levels of X.
- The estimation is the better when X_h is closer to the mean, since $\boldsymbol{s^2_{\{\hat{Y}_h\}} = s^2 \left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right]}$

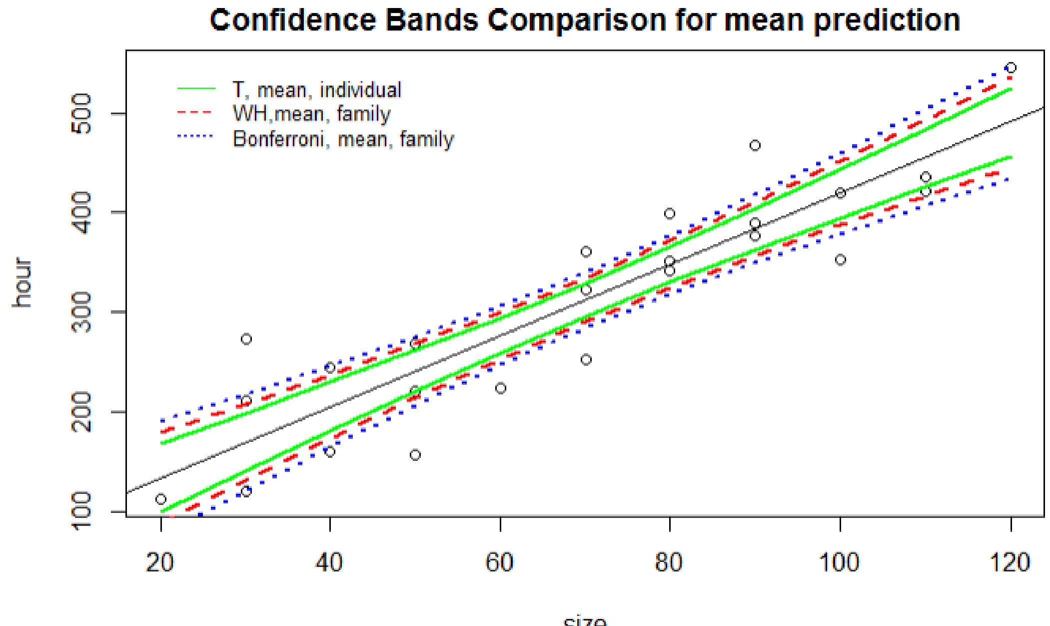
Bonferroni simultaneous CI vs Working hoteling confidence band on estimating mean $\hat{Y}_h$

```
library(ALSM)
tol.mod<-lm(hour~size, toluca)
ox<-data.frame(size=unique(toluca$size))
x<-sort(ox$size)
x
```

```
[1]  20  30  40  50  60  70  80  90 100 110 120
```

```
plot(hour ~ size, toluca, main="Confidence Bands Comparison for mean prediction")
abline(lm(tol.mod))
lines(x, cim$Lower.Band,col="green", lwd=2, lty=1)
lines(x, cim$Upper.Band, col="green", lwd=2, lty=1)
lines(x, ciw$Lower.Band,col="red",lwd=2, lty=2)
lines(x, ciw$Upper.Band, col="red", lwd=2, lty=2)
lines(x, cib$Lower.Band,col="blue", lwd=2, lty=3)
lines(x, cib$Upper.Band, col="blue", lwd=2, lty=3)

legend(x=20, y=550, legend=c("T, mean, individual", "WH,mean, family","Bonferroni, mean, family"), lty=c(1,2,3), col=c("green","red","blue"),cex=0.8,
bty="n")
```



$$t\left(1-\frac{\alpha}{2g};n-2\right) = 1.714\ (g=1)$$

$$W = 2.258\ (for\ all\ X_h)$$

$$t\left(1-\frac{\alpha}{2g};n-2\right) = 2.807\ (g=10)$$

The method

1. Simultaneously (joint) estimation of $\boldsymbol{p}$ parameters, in SLR, $p = 2\ with\ \beta_1\ and\ \beta_{2,}$ with **all** X values.
   a) Bonferroni
   b) Working-hoteling
2. Simultaneously estimation of mean response $\hat{Y}_h$ with $\boldsymbol{g}$ different X_h values.
   a) Bonferroni
   b) Working-hoteling
3. Simultaneously estimation of single response $\hat{Y}_h\{new\}$ with $\boldsymbol{g}$ different X_h values
   a) Bonferroni
   b) Schefft

Bonferroni Joint (or Family) Confidence Interval to predict the g single response, $\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$\{new\} ($\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$ \{new\} given X_1,$\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$ \{new\} given X_2 ,… ,$\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$\{new\} given X_g)

To make joint confidence interval for multiple (g) simultaneous prediction for

For each $\hat{\boldsymbol{Y}}_{\boldsymbol{h}}$

$$\hat{Y}_h \pm t\left(1-\frac{\alpha}{2}; n-2\right) s\{pred\}$$

$s^2\{pred\} = MSE[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}] = s^2 + s^2\{\hat{Y}_h\}$ in SLR

Then for g prediction with the same X_h

$$\hat{Y}_h \pm t\left(1-\frac{\alpha}{2\boldsymbol{g}}; n-2\right) s\{pred\}$$

Note: if a sufficiently large number of simultaneous predictions are made, the width of the individual confidence Intervals may become so wide that they are no longer useful.

Toluca Company Example: Bonferroni Joint (or Family) Confidence Interval on g single Response

What is the simultaneous estimates for the single prediction of number of work hours for $X_h = 30, 65\ and\ 100$ with family confidence level **0.9** ($\alpha = 0.1$). Suppose $\bar{X} = 70,\ SSX = 19800, s = 48.82, n = 25$

$$\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2g}; n-2\right) s\{pred\}\ \ X_h = 30, 65\ and\ 100\ \ (\hat{Y}_h = 169.5, 294.4, and\ 419.4\ respectively)$$

1. $X_h = 30$

$$s^2\{pred\} = s^2 + s^2\{\hat{Y}\} = 48.82^2 + 288.169 = 2671.56$$

$$\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2g}; n-2\right) s\{pred\} = 169.5 \pm t\left(1 - \frac{0.1}{6}; 23\right)\sqrt{2671.56} = 169.5 \pm 2.263(51.69) = (52.5, \quad 286.5)$$

2. $X_h = 65$

$$\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2g}; n-2\right) s\{pred\} = 294.4 \pm 2.263(49.82) = (181.6 \quad 407.2)$$

3. $X_h = 100$

$$\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2g}; n-2\right) s\{pred\} = 419.4 \pm 2.263(50.86) = (304.2, \quad 534.5)$$

```
d<-data.frame(c(30,65,100))
ci.reg(tol.mod, d, type='gn',alpha=0.1)
```

size <dbl>	Fit <dbl>	Lower.Band <dbl>	Upper.Band <dbl>
30	169.4719	52.46349	286.4804
65	294.4290	181.64910	407.2089
100	419.3861	304.23781	534.5343

Scheffe procedure of Joint (or Family or simultaneous) Confidence Interval to predict the g single responses

$$\hat{Y}_h \pm S\ s\{pred\}$$

$$S^2 = gF(1-\alpha; g, n-2)$$

$$s^2_{\{pred\}} = s^2_{\{\hat{Y}_h\}} + s^2 = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2} + 1\right] \text{ in SLR.}$$

- For SLR, $\hat{Y}_h \pm S\ s\{pred\}$ $\mathrm{S}^2 = gF(1-\alpha; g, n-2)$
- For MLR, $\hat{Y}_h \pm S\ s\{pred\}$ $\mathrm{S}^2 = gF(1-\alpha; g, n-p)$, where p is the number of parameters, and g is the number of predictions to make.
- In the WH procedure for estimating mean where all mean estimations are done through a conservative estimation of the p parameters, e.g. $\mathrm{E}(\hat{Y}_h) = \beta_0 + \beta_1 X_h$, we only need to concern a combined distribution of p variables, e.g, b_0 and b_1
- On the other hand, in the Schefft procedure, the g single responses each has its own distribution. We need to concern the combined distribution of g variables, $Y_1, Y_2, Y_3, \ldots Y_g$.

Toluca Company Example: Schefft Joint (or Family) Confidence Interval on g single Response $\hat{Y}_h$

What is the simultaneous estimates for the single number of work hours for $X_h = 30, 65\ and\ 100$ (i.e., $g = 3$) with family confidence level **0.9** ($\alpha = 0.1$). Suppose $\bar{X} = 70,\ SSX = 19800, s = 48.82$

$$\hat{Y}_h \pm S\, s\{pred\} \quad X_h = 30, 65\ and\ 100 \quad (\hat{Y}_h = 169.5, 294.4, and\ 419.4\ respectively)$$

1. $X_h = 30$

$s^2\{pred\} = s^2 + s^2\{\hat{Y}_h\} = 48.82^2 + 288.169 = 2671.56$

$S^2 = 3F(1-0.1; 3, n-2) = 3(3.028) = 9.084 \Rightarrow S = 3.01$

$\hat{Y}_h \pm S\, s\{pred\} = 169.5 \pm 3.01(51.69) = (13.68,\quad 325.26)$

2. $X_h = 65$

$\hat{Y}_h \pm S\, s\{pred\} = 294.4 \pm 3.01(49.82) = (144.27,\quad 444.59)$

3. $X_h = 100$

$\hat{Y}_h \pm S\, s\{pred\} = 419.4 \pm 3.01(50.86) = (266.08,\quad 572.7)$

```
d<-data.frame(c(30,65,100))
ci.reg(tol.mod, d, type='s',alpha=0.1)
```

- Unfortunately, the ci.reg function in R is wrong when computing this method.
- You can compute by hand or use the self-defined function on the next page.

	size <dbl>	Fit <dbl>	Lower.Band <dbl>	Upper.Band <dbl>
1	30	169.4719	-2502.216	2841.160
2	65	294.4290	-2187.645	2776.503
3	100	419.3861	-2168.029	3006.801

Family confidence interval, Bonferroni and Scheffe procedure


```
ci.sim <- function(model, newdata, type = c("B", "S"), alpha = 0.05)
{
  g  <- nrow(newdata)
  CI <- predict(model, newdata, se.fit = TRUE)
  M  <- ifelse(match.arg(type) == "B",
          qt(1 - alpha / (2*g), model$df),              # B "Bonferroni""
          sqrt( g * qf( 1 - alpha, g, model$df))        # S "scheffe""

  spred <- sqrt( CI$residual.scale^2 + (CI$se.fit)^2 )
  x <- data.frame(
    "x"      = newdata,
    "credV"  = M,
    "fit"    = CI$fit,
    "lower"  = CI$fit - M * spred,
    "upper"  = CI$fit + M * spred)

  return(x)
}

toluca<-read.table("U:/data/Toluca.txt", header=FALSE)
colnames(toluca)<-c("size","hour")

toluca.mod<-lm(hour~size, data=toluca)

new <- data.frame(size= c(30, 65, 100))
ci.sim(toluca.mod, new, type = "B")
ci.sim(toluca.mod, new, type = "S")
```