Simultaneous Inferences and Other Topics in Regression Analysis

Yang Feng

Simultaneous Inferences

- In chapter 2, we know how to construct confidence interval for β_0 and β_1.
- If we want a confidence level of 95% of both β_0 and β_1
- One could construct a separate confidence interval for β_0 and β_1. BUT, then the probability of both happening is below 95%.
- How to create a joint confidence interval?

Bonferroni Joint Confidence Intervals

- Calculation of Bonferroni joint confidence intervals is a general technique
- We highlight its application in the regression setting
  - Joint confidence intervals for β_0 and β_1
- Intuition
  - Set each statement confidence level to larger than $1 - \alpha$ so that the family coefficient is at least $1 - \alpha$
  - BUT how much larger?

Ordinary Confidence Intervals

- Start with ordinary confidence intervals for β_0 and β_1

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

- And ask what is probability that one or both of these intervals is incorrect

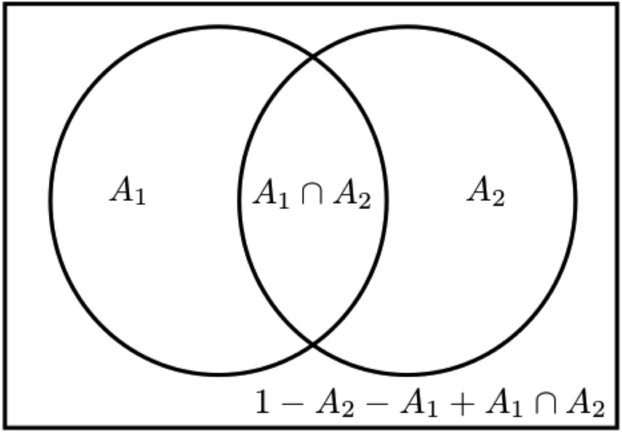
Remember

$$s^2\{b_0\} = MSE\left[\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i-\bar{X})^2}\right]$$
$$s^2\{b_1\} = \frac{MSE}{\sum(X_i-\bar{X})^2}$$

General Procedure

- Let A_1 denote the event that the first confidence interval does not cover β_0, i.e. $P(A_1) = \alpha$
- Let A_2 denote the event that the second confidence interval does not cover β_1, i.e. $P(A_2) = \alpha$
- We want to know the probability that both estimates fall in their respective confidence intervals, i.e. $P(\bar{A}_1 \cap \bar{A}_2)$
- How do we get there from what we know?

Venn Diagram



Bonferroni inequality

- We can see that $P(\bar{A}_1 \cap \bar{A}_2) = 1 - P(A_2) - P(A_1) + P(A_1 \cap A_2)$
  - Size of set is equal to area is equal to probability in a Venn diagram.
- It also is clear that $P(A_1 \cap A_2) \geq 0$
- So,

$$\begin{aligned} P(\bar{A}_1 \cap \bar{A}_2) &\geq 1 - P(A_2) - P(A_1) \\ &= 1 - 2\alpha \end{aligned}$$

Using the Bonferroni inequality cont.

- To achieve a $1-\alpha$ *family* confidence interval for β_0 and β_1 (for example) using the Bonferroni procedure we know that both individual intervals must shrink.
- Returning to our confidence intervals for β_0 and β_1 from before

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

- To achieve a $1-\alpha$ *family* confidence interval these intervals must *widen* to

$$b_0 \pm t(1-\alpha/4; n-2)s\{b_0\}$$
$$b_1 \pm t(1-\alpha/4; n-2)s\{b_1\}$$

- Then $P(\bar{A}_1 \cap \bar{A}_2) \geq 1 - P(A_2) - P(A_1) = 1 - \alpha/2 - \alpha/2 = 1-\alpha$

Using the Bonferroni inequality cont.

- The Bonferroni procedure is very general. To make joint confidence statements about multiple simultaneous predictions remember that

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

- If one is interested in a $1-\alpha$ confidence statement about g predictions then Bonferroni says that the confidence interval for each individual prediction must get wider (for each h in the g predictions)

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

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.

The Toluca Example

- Say, we want to get a 90 percent confidence interval for β_0 and β_1 simultaneously.
- Then we require $B = t(1 - .1/4; 23) = t(.975, 23) = 2.069$
- Then we have the joint confidence interval:

$$b_0 \pm B * s(b_0)$$

and

$$b_1 \pm B * s(b_1)$$

Confidence Band for Regression Line

- Remember in Chapter 2.5, we get the confidence interval for $E\{Y_h\}$ to be

$$\hat{Y}_h \pm t(1-\alpha/2; n-2)s\{\hat{Y}_h\}$$

- Now, we want to get a confidence band for the entire regression line $E\{Y\} = \beta_0 + \beta_1 X$.
- The Working-Hotelling $1-\alpha$ confidence band is

$$\hat{Y}_h \pm W \times s\{\hat{Y}_h\}$$

  here $W^2 = 2F(1-\alpha; 2, n-2)$.
- Same form as before, except the t multiple is replaced with the W multiple.

Example: toluca company

- Say we want to estimate the boundary value for the band at $X_h = 30, 65, 100$.
- We have

X_h	$\hat{Y}_h$	$s\{\hat{Y}_h\}$
30	169.5	16.97
65	294.4	9.918
100	419.4	14.27

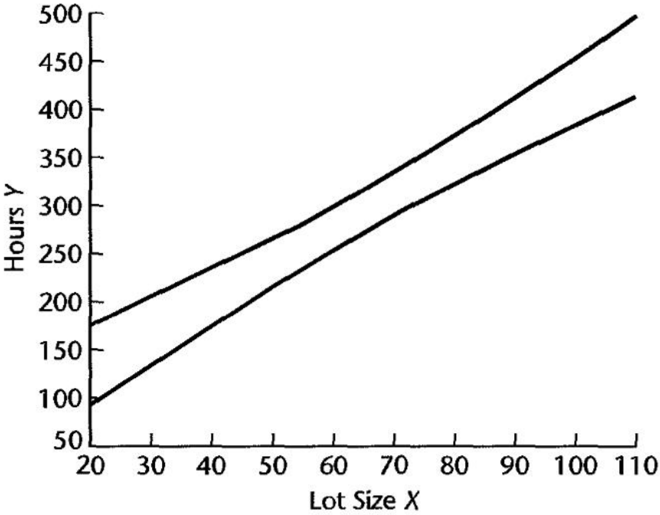
- Looking up the table, $W^2 = 2F(1-\alpha; 2, n-2) = 2F(.9; 2, 23) = 5.098$.
  R code:

```
w2 = 2 * qf(1-0.1,2,23)
```

Now we have the confidence band for the three points are

$$131.2 = 169.5 - 2.258(16.97) \leq E\{Y_h\} \leq 169.5 + 2.258(16.97) = 207.8$$
$$272.0 = 294.4 - 2.258(9.918) \leq E\{Y_h\} \leq 294.4 + 2.258(9.918) = 316.8$$
$$387.2 = 419.4 - 2.258(14.27) \leq E\{Y_h\} \leq 419.4 + 2.258(14.27) = 451.6$$

Example confidence band



Compare with Bonferroni Procedure

- Say we want to simultaneously estimate response for $X_h = 30, 65, 100$.
- Then the simultaneous confidence intervals are

$$\hat{Y}_h \pm t(1-\alpha/(2g); n-2)s\{\hat{Y}_h\}$$

- We have $B = t(1-\alpha/(2g); n-2) = t(1-.1/(2*3), 23) = 2.263$, the confidence intervals are

$$131.1 = 169.5 - 2.263(16.97) \leq E\{Y_h\} \leq 169.5 + 2.263(16.97) = 207.9$$
$$272.0 = 294.4 - 2.263(9.918) \leq E\{Y_h\} \leq 294.4 + 2.263(9.918) = 316.8$$
$$387.1 = 419.4 - 2.263(14.27) \leq E\{Y_h\} \leq 419.4 + 2.263(14.27) = 451.7$$

Bonferroni v.s. Working-Hotelling

- This instance, working-hotelling confidence limits are slighter tighter(better) than bonferroni limits
- However, in larger families (more X) to be considered simultaneously, working-hotelling is always tighter, since W stays the same for any number of statements but B becomres larger.
- The levels of predictor variables are sometimes not known in advance. In such cases, it is better to use Working-Hotelling procedure since the family encompasses all possible levels of X.

$$131.1 = 169.5 - 2.263(16.97) \le E\{Y_h\} \le 169.5 + 2.263(16.97) = 207.9$$
$$272.0 = 294.4 - 2.263(9.918) \le E\{Y_h\} \le 294.4 + 2.263(9.918) = 316.8$$
$$387.1 = 419.4 - 2.263(14.27) \le E\{Y_h\} \le 419.4 + 2.263(14.27) = 451.7$$

Simultaneous Prediction Intervals for g New Observations

1. Scheffe procedure

$$\hat{Y}_h \pm Ss\{pred\}, \tag{1}$$

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

2. Bonferroni procedure

$$\hat{Y}_h \pm Bs\{pred\}, \tag{2}$$

where $B = t(1-\alpha/(2g); n-2)$.

Regression through the origin

Model

$$Y_i = \beta_1 X_i + \epsilon_i$$

- Sometimes it is known that the regression function is linear and that it *must* go through the origin.
- β_1 is parameter
- X_i are known constants
- ϵ_i are i.i.d $N(0, \sigma^2)$.
- The least squares and maximum likelihood estimators for β_1 coincide as before, the estimator is $b_1 = \frac{\sum X_i Y_i}{\sum X_i^2}$

Regression through the origin, Cont

- In regression through the origin there is only one free parameter (β_1) so the number of degrees of freedom of the MSE

$$s^2 = MSE = \frac{\sum e_i^2}{n-1} = \frac{\sum(Y_i - \hat{Y}_i)^2}{n-1}$$

is increased by one.

- This is because this is a "reduced" model in the general linear test sense and because the number of parameters estimated from the data is less by one.

Estimate of	Estimated Variance	Confidence Limits	
β_1	$s^2\{b_1\} = \frac{MSE}{\sum X_i^2}$	$b_1 \pm ts\{b_1\}$	(4.18)
$E\{Y_h\}$	$s^2\{\hat{Y}_h\} = \frac{X_h^2 MSE}{\sum X_i^2}$	$\hat{Y}_h \pm ts\{\hat{Y}_h\}$	(4.19)
$Y_{h(new)}$	$s^2\{\text{pred}\} = MSE\left(1 + \frac{X_h^2}{\sum X_i^2}\right)$	$\hat{Y}_h \pm ts\{\text{pred}\}$	(4.20)
		where: $t = t(1 - \alpha/2; n-1)$	

A few notes on regression through the origin

- $\sum e_i \neq 0$ in general now. Only constraint is $\sum X_i e_i = 0$.
- SSE may exceed the total sum of squares SSTO. In the case of a curvilinear pattern or linear pattern with a intercept away from the origin.
- Therefore, $R^2 = 1 - SSE/SSTO$ may be negative!
- Generally, it is safer to use the original model opposed with regression-through-the-origin model.
- Otherwise, it is the wrong model to start with!