Simultaneous Inference and Other Topics in Regression Analysis

- Simultaneous inference of \((\beta_0, \beta_1)\) and their linear functions
- Regression through the origin
- Effects of measurement error
- Inverse prediction
- Choice of \(X\) levels
Simultaneous Inference of $\beta_0$ and $\beta_1$

We would like to get $1 - \alpha$ confidence regions that the conclusions for both $\beta_0$ and $\beta_1$ are correct. We call the set of estimates (or tests) of interest the family of estimates (or tests).

A statement confidence coefficient is a probability statement about one parameter, which indicates the proportion of correct estimates that are obtained for repeated samples. A family confidence coefficient indicates the proportion of families of estimates that are entirely correct for repeated samples.
Single confidence intervals

\[
\Pr \left( \frac{|b_j - \beta_j|}{s(b_j)} \leq t(1 - \alpha/2, n - 2) \right) = 1 - \alpha, \; j = 0, 1. \quad (1)
\]

Generally

\[
\Pr \left( \frac{|b_j - \beta_j|}{s(b_j)} \leq t(1 - \alpha/2, n - 2), j = 0, 1 \right) \quad (2)
\]

\[
\neq \prod_{j=0}^{1} \Pr \left( \frac{|b_j - \beta_j|}{s(b_j)} \leq t(1 - \alpha/2, n - 2) \right),
\]

unless they are independent.
Bonferroni Joint Confidence Intervals

In theory we can work out the joint probability (2), which involves a complicated two-dimension integral. But we can easily get an lower bound using the following Bonferroni inequality

\[ \Pr \left( \bigcup_{k=1}^{m} A_k \right) \leq \sum_{k=1}^{m} \Pr(A_k), \]  

(3)

and, hence

\[ \Pr \left( \bigcap_{k=1}^{m} A_k^c \right) \geq 1 - \sum_{k=1}^{m} \Pr(A_k). \]  

(4)

Therefore

(2) \geq 1 - 2\alpha.
Generally for \( m \) (correlated) tests, we can use \( \alpha/m \) significance level for individual test to guarantee an overall Type-I error \( \alpha \).
Simultaneous Estimation of Mean Responses

Goal: estimate the mean response at $X$ levels $\{X_h, h = 1, \cdots, m\}$.

**Working-Hotelling Procedure**

The Working-Hotelling confidence band covers the entire regression line. So for confidence interval

$$
\hat{Y}_h \pm W s(\hat{Y}_h), \quad W = \sqrt{2F(1 - \alpha, 2, n - 2)}
$$

the family confidence coefficient for these simultaneous estimates will be at least $1 - \alpha$.

**Bonferroni Procedure**

With Bonferroni procedure, we just need to use $\alpha/m$ as the individual Type-I
error for each $X_h$.

$$\hat{Y}_h \pm B \sigma(\hat{Y}_h), \quad B = t \left(1 - \frac{\alpha / m}{2}, 2, n - 2 \right)$$
Simultaneous Prediction Intervals for New Observations

Goal: predict new observations at \( X \) levels \( \{X_h, h = 1, \cdots, m\} \).

**Scheffe Procedure**
Use the following confidence interval

\[
\hat{Y}_h \pm S_s(\text{pred}), \quad S = \sqrt{mF(1 - \alpha, m, n - 2)}.
\]

**Bonferroni Procedure**
With Bonferroni procedure, we just need to use \( \alpha/m \) as the individual Type-I error for each \( X_h \).

\[
\hat{Y}_h \pm B_s(\text{pred}), \quad B = t \left(1 - \frac{\alpha/m}{2}, 2, n - 2\right)
\]
Multiple Test Comparisons

Critical Values (m=3)
Sheffe
Working–Hotelling
Bonferroni

Critical Values (m=6)
Sheffe
Bonferroni
Working–Hotelling
Joint Intervals for $\beta_0$ and $\beta_1$:

Point estimation

$$b_1 = \frac{\sum_{i=1}^{25} (X_i - \bar{X})Y_i}{\sum_{i=1}^{25} (X_i - \bar{X})^2} = 3.57, \quad b_0 = \bar{Y} - b_1\bar{X} = 62.37,$$

with variance estimated as

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{25} (Y_i - b_0 - b_1X_i)^2}{23}} = 48.82,$$

$$s^2(b_0) = \left\{ \frac{1}{25} + \frac{\bar{X}^2}{\sum_{i=1}^{25} (X_i - \bar{X})^2} \right\} \hat{\sigma}^2, \quad s(b_0) = 26.18,$$
\[ s^2(b_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{25} (X_i - \bar{X})^2}, \quad s(b_1) = 0.347, \]

With overall confidence coefficient being \( \alpha = 0.05 \), using Bonferroni procedure we will use

\[ qt(1 - \alpha/4, 23) = 2.398 \]

as the critical values.

So the joint 95\% confidence intervals for \( \beta_0 \) and \( \beta_1 \) are

\[ \beta_0 : b_0 \pm qt(1 - \alpha/4, 23)s(b_0) \rightarrow (-0.40, 125.14) \]

\[ \beta_1 : b_1 \pm qt(1 - \alpha/4, 23)s(b_1) \rightarrow (2.74, 4.40) \]

Using the joint distribution of \( b_0, b_1 \), we can derive another (elliptical)
confidence regions for $\beta_0, \beta_1$ (see Lecture Note 1)

$$
(b_0 - \beta_0, b_1 - \beta_1) \Sigma X^{-1} (b_0 - \beta_0, b_1 - \beta_1)^T / 2 \leq \frac{\hat{\sigma}^2}{\hat{\sigma}^2} F(1 - \alpha, 2, n - 2).
$$
Confidence Intervals Comparisons

- 95% CR
- 95% Bonferroni CIs

(\(b_0, b_1\))
Simultaneous Mean Estimation at $X_h = 30, 65, 100$:
The mean average estimates are

$$\hat{Y}_h = b_0 + b_1 X_h = 169.47, 294.43, 419.39,$$

with the variance estimates being

$$s^2(\hat{Y}_h) = \hat{\sigma}^2 \left\{ \frac{1}{25} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{25}(X_i - \bar{X})^2} \right\}, \quad s(\hat{Y}_h) = 16.97, 9.92, 14.27.$$

Working-Hotelling Procedure

$$W = \sqrt{2F(1 - \alpha, 2, 23)} = 2.616,$$
and the three confidence intervals are

\[ 125.08 = 169.47 - 2.616(16.97) \leq E(Y_h) \leq 169.47 + 2.616(16.97) = 213.87 \]
\[ 268.48 = 294.43 - 2.616(9.92) \leq E(Y_h) \leq 294.43 + 2.616(9.92) = 320.37 \]
\[ 382.05 = 419.39 - 2.616(14.27) \leq E(Y_h) \leq 419.39 + 2.616(14.27) = 456.72 \]

This 95% family confidence coefficient means that the procedure leads to all correct estimates for three means at least 95% of the time for repeated samples.

**Bonferroni Procedure**

\[ B = qt(1 - \alpha/6, 23) = 2.582, \]
and the three confidence intervals are

$$125.66 = 169.47 - 2.582(16.97) \leq E(Y_h) \leq 169.47 + 2.582(16.97) = 213.28$$

$$268.82 = 294.43 - 2.582(9.92) \leq E(Y_h) \leq 294.43 + 2.582(9.92) = 320.04$$

$$382.53 = 419.39 - 2.582(14.27) \leq E(Y_h) \leq 419.39 + 2.582(14.27) = 456.24$$

So for this dataset and the specified three new $X$ levels, and overall 95% family confidence coefficient, the Bonferroni procedure provides slightly tighter confidence regions than Working-Hotelling procedure.
Simultaneous Predictions for New Observations at $X_h = 30, 65, 100$:

The point estimates are

$$\hat{Y}_h = b_0 + b_1 X_h = 169.47, 294.43, 419.39,$$

with the variance estimates being

$$s^2(\text{pred}) = \hat{\sigma}^2 \left\{ 1 + \frac{1}{25} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{25} (X_i - \bar{X})^2} \right\}, \quad s(\text{pred}) = 51.69, 49.82, 50.87.$$

**Sheffe Procedure**

$$W = \sqrt{3F(1 - \alpha, 3, 23)} = 3.014,$$
and the three confidence intervals are

\[ 13.68 = 169.47 - 3.014(51.69) \leq E(Y_h) \leq 169.47 + 3.014(51.69) = 325.26 \]
\[ 144.27 = 294.43 - 3.014(49.82) \leq E(Y_h) \leq 294.43 + 3.014(49.82) = 444.59 \]
\[ 266.08 = 419.39 - 3.014(50.87) \leq E(Y_h) \leq 419.39 + 3.014(50.87) = 572.70 \]

This 95% family confidence coefficient means that the procedure leads to all correct estimates for three new observations at least 95% of the time for repeated samples.

**Bonferroni Procedure**

\[ B = qt(1 - \alpha/6, 23) = 2.582, \]
and the three confidence intervals are

\[
36.01 = 169.47 - 2.582(51.69) \leq E(Y_h) \leq 169.47 + 2.582(51.69) = 302.93 \\
165.79 = 294.43 - 2.582(49.82) \leq E(Y_h) \leq 294.43 + 2.582(49.82) = 423.07 \\
288.05 = 419.39 - 2.582(50.87) \leq E(Y_h) \leq 419.39 + 2.582(50.87) = 550.72
\]

So for this dataset and the specified three new $X$ levels, and overall 95% family confidence coefficient, the Bonferroni procedure provides slightly tighter confidence regions than Sheffe procedure.

R codes for previous analysis are available at [http://www.biostat.umn.edu/~baolin/teaching/ph7405-05F/RFive.html](http://www.biostat.umn.edu/~baolin/teaching/ph7405-05F/RFive.html) and [http://www.biostat.umn.edu/~baolin/teaching/ph7405-05F/SupplementaryNotes.html](http://www.biostat.umn.edu/~baolin/teaching/ph7405-05F/SupplementaryNotes.html).
Regression Through Origin

Model assumption

\[ Y_i = \beta_1 X_i + \epsilon_i, \ i = 1, \ldots, n, \]  

where \( \beta_1 \) is a parameter, \( X_i \) are known constants and \( \epsilon_i \) are independent \( N(0, \sigma^2) \).

The least squares (maximum likelihood) estimation

\[ b_1 = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} \sim N(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2}). \]
The residuals and unbiased variance estimates are

\[ e_i = Y_i - \hat{Y}_i = Y_i - b_1 X_i, \quad \hat{\sigma}^2 = MSE = \frac{\sum_{i=1}^{n} e_i^2}{n - 1}. \]

Similarly as before we can derive the estimated variance for the estimates of \( \beta_1, E(Y_h) \) and \( Y_h \)

\[ s^2(b_1) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n} X_i^2}, \quad s^2(\hat{Y}_h) = \frac{X_h^2 \hat{\sigma}^2}{\sum_{i=1}^{n} X_i^2}, \quad s^2(\text{pred}) = \hat{\sigma}^2 \left( 1 + \frac{X_h^2}{\sum_{i=1}^{n} X_i^2} \right), \]

and based on \( t \)-distribution we have the following confidence intervals

\[ b_1 \pm ts(b_1), \quad \hat{Y}_h \pm ts(\hat{Y}_h), \quad \hat{Y}_h \pm ts(\text{pred}), \quad t = t(1 - \alpha/2, n - 1). \]
Some Cautions When Using Regression Through Origin

• $\sum_{i=1}^{n} X_i e_i = 0$, but usually $\sum_{i=1}^{n} e_i \neq 0$.

• ANOVA decomposition

$$SSTOU = SSRU + SSE,$$

where $SSTOU = \sum_{i=1}^{n} Y_i^2$, $SSRU = \sum_{i=1}^{n} \hat{Y}_i^2$ and $SSE = \sum_{i=1}^{n} (Y_i - b_1 X_i)^2$.

• Simultaneous inference is not a problem anymore since we only have one parameter $\beta_1$. 
• It is generally a safe practice to use the regression model with intercept $\beta_0$. If the regression line does go through the origin, $b_0$ will differ from 0 only by a small sampling error.
Effects of Measurement Errors

When random measurement errors are present in $Y$, no new problems are created if these errors are uncorrelated and not biased. The model error term $\epsilon$ always reflects the composite effects of a large number of factors not considered in the model.

Suppose the observed $X$ levels $X_i^*$ have some measurement errors

$$X_i^* = X_i + \delta_i,$$

and the regression model becomes

$$Y_i = \beta_0 + \beta_1 X_i = \beta_0 + \beta_1 X_i^* + (\epsilon_i - \beta_1 \delta_i),$$
where the predictor variable $X^*$ and the error term $\epsilon_i - \beta_1 \delta_i$ are generally correlated with each other.

If we further assume that

$$E(\delta_i) = E(\delta_i \epsilon_i) = E(\epsilon_i) = 0,$$

we can derive the covariance between predictor and error term

$$Cov(X^*_i, \epsilon_i - \beta_1 \delta_i) = Cov(\delta_i, \epsilon_i - \beta_1 \delta_i)$$
$$= Cov(\delta_i, -\beta_1 \delta_i)$$
$$= -\beta_1 Var(\delta_i).$$

This covariance is not zero whenever there is a linear regression relation between $X$ and $Y$. 
Berkson Model

The model is the same as (7), but now $X_i^*$ are fixed and $X_i$ are random. Under the same assumption (8), we have

$$Cov(X_i^*, \epsilon_i - \beta_1 \delta_i) = 0,$$

since $X_i^*$ are fixed constants. Therefore the problem just transforms into an ordinary linear regression model.
Inverse Predictions

The regression model of $Y$ on $X$

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \ i = 1, \cdots, n,$$

is used to predict the value of $X_h$ which gave rise to a new observation $Y_h$. A natural estimator is

$$\hat{X}_h = \frac{Y_h - b_0}{b_1},$$

where $b_0, b_1$ are least squares estimators.

Approximate $1 - \alpha$ confidence limits for $X_h$ are

$$\hat{X}_h \pm t(1 - \alpha/2, n - 2)s(predX),$$  \hspace{1cm} (9)
where

\[ s^2(\text{pred}X) = \frac{\hat{\sigma}^2}{b_1^2} \left\{ 1 + \frac{1}{n} + \frac{(\hat{X}_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right\}. \]

Some comments on inverse prediction

1. The inverse prediction problem is also known as a *calibration problem* since it is applicable when inexpensive, quick, and approximate measurements \( Y \) are related to precise, often expensive, and time-consuming measurements \( X \) based on \( n \) observations.

2. The approximate confidence interval (9) is appropriate if

\[ \frac{[t(1 - \alpha/2, n - 2)]^2 \hat{\sigma}^2}{b_1^2 \sum_{i=1}^{n}(X_i - \bar{X})^2} \]

is small, say \( \leq 0.1 \).
3. Simultaneous prediction intervals for \( m \) different new observed measurements \( Y_h \), with a \( 1 - \alpha \) family confidence coefficient, are easily obtained using Bonferroni or Sheffe procedures with the following critical values

\[
B = t \left( 1 - \frac{\alpha/m}{2}, n - 2 \right) \text{ or } S = \sqrt{mF(1 - \alpha, m, n - 2)}.
\]

4. Sometimes it is suggested that the inverse prediction should be made in direct fashion by regressing \( X_0 \) on \( Y \), which is called inverse regression. There are still some controversy in this problem.
Inverse Prediction for Galton’s “regression to the average” Example

Suppose $X_i$ are used to code the father’s heights, and $Y_i$ are used to code the son’s heights.

The “regression to the average” effect refers to the following regression model

$$\frac{Y_i - \bar{Y}}{SD_Y} = r \frac{X_i - \bar{X}}{SD_X}, \quad r \in [-1, 1], \quad (10)$$

where the sample correlation

$$r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{SD_X \cdot SD_Y}.$$
and

\[ SD_X = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}}, \quad SD_Y = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n}}. \]

Notice the symmetry of \( r \) with respect to \( X \) and \( Y \). If we use inverse regression for the purpose of inverse prediction, we will have the following prediction model

\[ \frac{X_i - \bar{X}}{SD_X} = r \frac{Y_i - \bar{Y}}{SD_Y}, \quad r \in [-1, 1], \tag{11} \]

which says that “father’s height tends to go to the average compared to son’s height”, i.e. “son’s height tends to diverge from the average compared to father’s height”. But it is contrary to model (10), which says “son’s height tends to converge to the average compared to father’s height”.
The reason is the difference between *inverse prediction* model

\[
\frac{X_i - \bar{X}}{SD_X} = \frac{1}{r} \frac{Y_i - \bar{Y}}{SD_Y}, \text{ based on model (10)},
\]

and the *inverse regression* model (11)

\[
\frac{X_i - \bar{X}}{SD_X} = r \frac{Y_i - \bar{Y}}{SD_Y}.
\]
Choice of $X$ Levels

For different purpose of experiment design we may choose different $X$ levels in controlled experiment. We can get some guidance from the variance formulas for some parameters

\[
\begin{align*}
Var(b_0) & = \sigma^2 \left\{ \frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \right\} \\
Var(b_1) & = \frac{\sigma^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}
\end{align*}
\]
\[ Var(\hat{Y}_h) = \sigma^2 \left\{ \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right\} \]

\[ Var(pred) = \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right\} \]

If the goal is to estimate \( \beta_1 \), we need to maximize \( \sum_{i=1}^{n}(X_i - \bar{X})^2 \) to get minimum variance (maximum accuracy), which can be achieved by using two extreme values of \( X \), and placing half of the observations at each of the two levels. But of course this will cause some trouble for diagnosis if we are not sure about the linearity of the models.

If the main purpose is to estimate \( \beta_0 \), the variance of \( \beta_0 \) won’t be affected by \( X \) as long as \( \bar{X} = 0 \). On the other hand, to estimate the mean response or to predict a new observation at the level \( X_h \), the variance will be minimized by making \( \bar{X} = X_h \).