Lecture 5

- Other topics
- Simple linear regression in a matrix form

Joint Estimation of β_0 and β_1

- Confidence intervals are used for a single parameter, confidence regions for a two or more parameters
- The region for (β_0, β_1) defines a set of lines
- Since β_0 and β_1 are (jointly) normal, the *natural* confidence region is an ellipse
- We can also do rectangles

Bonferroni

- We want the probability that *both* intervals are correct to be (at least) .95
- Basic idea is an *error budget* (α =.05)
- Spend half on β_0 (.025) and half on β_1 (.025)
- We use α =.025 for the β_0 CI (97.5% CI)
- and α =.025 for the β_1 CI (97.5% CI)

Bonferroni (2)

- So we use
- $b_1 \pm t^* s(b_1)$
- $b_0 \pm t^* s(b_0)$
- where $t^* = t(.9875, n-2)$
- .9875 = 1 – (.05)/(2*2)

Bonferroni (3)

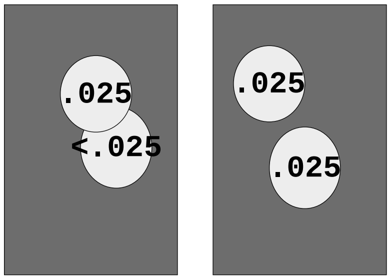
- Note we start with a 5% error budget and we have two intervals so we give
- 2.5% to each
- Each interval has two *ends* so we again divide by 2
- So, .9875 = 1 – (.05)/(2*2)

Bonferroni Inequality

- Let the two intervals be I_1 and I_2
- We will use inc if the interval does not contain the true parameter value

Bonferroni Inequality (2)

- P(both cor)=1-P(at least one inc)
- P(at least one inc)
- = P(I_1 inc)+ P(I_2 inc)-P(both inc)
- $\leq$ P(I_1 inc)+ P(I_2 inc)
- So if we use .05/2 for each interval
- P(at least one inc)≤0.05



Mean Response CIs

- Simultaneous estimation for *all* X_h, use Working-Hotelling
- $\hat{\mu}_h \pm Ws(\hat{\mu}_h)$ where $W^2=2F(1-\alpha; 2, n-2)$
- For simultaneous estimation for *a few* (g) X_h, use Bonferroni $\hat{\mu}_h \pm Bs(\hat{\mu}_h)$
- where $B=t(1-\alpha/(2g), n-2)$
- B: 2.100922 2.445006 2.639145 2.774529
- W: 2.666292

Simultaneous PIs

- Simultaneous prediction for *a few* (g) X_h,
- use Bonferroni
- $\hat{\mu}_h \pm Bs(\text{pred})$
- where $B=t(1-\alpha/(2g), n-2)$

Regression through the Origin

- $Y_i = \beta_1 X_i + \xi_i$
- Generally *not* a good idea
- Problems with r^2 and other statistics

Measurement Error

- For Y, this is usually not a problem
- For X, we can get biased estimators of our regression parameters

Inverse Predictions

- Sometimes called calibration
- Given Y_h, predict the corresponding value of X, $\hat{X}_h$
- Solve the fitted equation for X_h
- $\hat{X}_h = (Y_h - b_0)/b_1$, $b_1 \neq 0$
- Approximate CI can be given

Choice of X Values (Levels)

- Look at the formulas for the variances of the estimators of interest
- Usually we find $\Sigma(X_i - \bar{X})^2$ in a denominator
- So we want to spread out the values of X

The Model in Scalar Form

- $Y_i = \beta_0 + \beta_1 X_i + \xi_i$
- ξ_i are independent normally distributed random errors with mean 0 and variance σ^2

The Model in Matrix Form

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \cdots \\ \beta_0 + \beta_1 X_n \end{pmatrix} + \begin{pmatrix} \xi_1 \\ \xi_2 \\ \cdots \\ \xi_n \end{pmatrix}$$

The Model in Matrix Form (2)

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ & \cdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \xi_1 \\ \xi_2 \\ \cdots \\ \xi_n \end{pmatrix}$$

Design matrix

$$X_{nx2} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ & \cdots \\ 1 & X_n \end{pmatrix}$$

Vector of parameters

$$\beta_{2x1} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Vector of error terms

$$\xi_{nx1} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \cdots \\ \xi_n \end{pmatrix}$$

Vector of response

$$Y_{nx1} = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{pmatrix}$$

Simple Linear Regression in Matrix Form

Y = Xβ + ξ

$$\underset{nx1}{Y} = \underset{nx2}{X}\ \underset{2x1}{\beta} + \underset{nx1}{\xi}$$

Covariance Matrix

$$Cov(Y) = \begin{pmatrix} Var(Y_1) & \mathrm{Cov}(Y_1, Y_2) & \cdots & \mathrm{Cov}(Y_1, Y_n) \\ \mathrm{Cov}(Y_2, Y_1) & \mathrm{Var}(Y_2) & \cdots & \mathrm{Cov}(Y_2, Y_n) \\ \cdots & \cdots & \cdots & \cdots \\ \mathrm{Cov}(Y_n, Y_1) & \cdots & \cdots & \mathrm{Var}(Y_n) \end{pmatrix}$$

Covariance Matrix of ξ

$$Cov\begin{pmatrix} \xi_1 \\ \xi_2 \\ \cdots \\ \xi_n \end{pmatrix} = \sigma^2 \mathbf{I}$$

Covariance Matrix of Y

$$Cov\begin{pmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_n \end{pmatrix} = \sigma^2 \mathbf{I}$$

Distributional Assumptions in Matrix Form

- $\xi \sim N(0, \sigma^2 I)$
- I is an *nxn* identity matrix
- Ones in the diagonal elements specify that the variance of each ξ_i is 1 times σ^2
- Zeros in the off-diagonal elements specify that the covariance between different ξ_i is zero
- This implies that the correlations are zero

Least Squares

- We want to minimize $(Y-X\beta)'(Y-X\beta)$
- We take the derivative with respect to the (vector) β
- This is like a quadratic

Least Squares (2)

- The derivative is 2 times the derivative of $(Y-X\beta)'$ with respect to β which is $-X'$
- times $(Y-X\beta)$
- We set this equal to 0 (a vector)
- So, $-2X'(Y-X\beta) = 0$
- Or, $X'Y = X'X\beta$

Normal Equations

- $X'Y = (X'X)\beta$
- Solving for β gives the least squares solution $b = (b_0, b_1)'$
- $b = (X'X)^{-1}(X'Y)$
- The same approach works for multiple regression

Fitted Values

$$\hat{\mathbf{Y}} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \cdots \\ \hat{Y}_n \end{pmatrix} = \begin{pmatrix} b_0 + b_1 X_1 \\ b_0 + b_1 X_2 \\ \cdots \\ b_0 + b_1 X_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \cdots & \\ 1 & X_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \mathbf{Xb}$$

Hat Matrix

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b}$$

$$\hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}$$

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Estimated Covariance Matrix of b

- We have linear combinations of the elements of Y
- These are normal if Y is normal
- Approximately normal in general

A Useful MultivariateTheorem

- U ~ N(μ, Σ), a multivariate normal vector
- V = c + DU, a linear transformation of U
- c is a vector, D is a matrix
- V ~ N(c+Dμ, DΣD′)

Application to b

- b = $(X'X)^{-1}(X'Y) = ((X'X)^{-1}X')(Y)$
- Y ~ N(Xβ, σ^2I)
- So b ~ N($(X'X)^{-1}X'(X\beta)$, $\sigma^2 ((X'X)^{-1}X')\, I\, ((X'X)^{-1}X')'$
- b ~ N(β, $\sigma^2 (X'X)^{-1}$)