BIOS 7345: Advanced Regression for Independent Data

Andrew J. Spieker, Ph.D.

Associate Professor of Biostatistics
Vanderbilt University

Set 6: Confidence and prediction intervals

Version: 09/19/2025

Table of Contents

Confidence intervals for regression parameters

Recall:

- We spent a meaningful amount of time developing the theory for inference regarding regression parameters.
- For these notes, we'll focus on "inverting" the tests in order to create confidence regions/intervals.
- Model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}_{N\times K}$ is of full rank (for simplicity of presentation), $\mathrm{E}[\boldsymbol{\epsilon}] = \mathbf{0}$, and $\mathrm{Cov}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{I}$.
- I will make the formal assumption that $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$, so that the results presented are exact. However, the results surrounding confidence intervals are otherwise approximate for sufficiently large samples.
- Throughout, let $\widehat{\boldsymbol{\beta}}$ denote the OLS estimator.

Special case of the *F*-statistic:

- Throughout, we've written the F-statistic as:

$$F = \frac{(\text{RSS}_H - \text{RSS})/Q}{\text{RSS}/(N-K)} = \frac{(\mathbf{C}\widehat{\boldsymbol{\beta}})^\intercal(\mathbf{C}(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{C}^\intercal)^{-1}\mathbf{C}\widehat{\boldsymbol{\beta}}/Q}{S^2}$$

- When $\mathbf{C} = \mathbf{c}^\intercal$ (one row; $Q = 1$), we can write the F-statistic as:

$$F = \frac{(\mathbf{c}^\intercal\widehat{\boldsymbol{\beta}})^2}{S^2\mathbf{c}^\intercal(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{c}}.$$

- Under $H_0 : \mathbf{c}^\intercal\boldsymbol{\beta} = 0$, this will follow an $\mathcal{F}_{1,N-K}$ distribution.

Lemma 6.1: Relationship between the t- and F-distribution

Let $Z \sim \mathcal{N}(0, 1)$ and let $U \sim \chi^2_K$, with $U \perp\!\!\!\perp Z$. Further, let

$$T = \frac{Z}{\sqrt{U/K}}.$$

Then, $T \sim t_K$ and $T^2 \sim \mathcal{F}_{1,K}$.

Justification for the *t*-statistic:

- Rather than using the F-statistic:

$$F = \frac{(\mathbf{c}^\intercal \widehat{\boldsymbol{\beta}})^2}{S^2 \mathbf{c}^\intercal (\mathbf{X}^\intercal \mathbf{X})^{-1} \mathbf{c}},$$

  note that Lemma 6.1 gives justification instead for the t-statistic:

$$t = \frac{\mathbf{c}^\intercal \widehat{\boldsymbol{\beta}}}{S \sqrt{\mathbf{c}^\intercal (\mathbf{X}^\intercal \mathbf{X})^{-1} \mathbf{c}}}.$$

- Under $H_0 : \mathbf{c}^\intercal \boldsymbol{\beta} = 0$, this will follow an t_{N-K} distribution.

Important note:

- Suppose that the value of $\mathbf{c}^\intercal \boldsymbol{\beta}$ is not zero. Then,

$$\frac{\mathbf{c}^\intercal \widehat{\boldsymbol{\beta}} - \mathbf{c}^\intercal \boldsymbol{\beta}}{S\sqrt{\mathbf{c}^\intercal (\mathbf{X}^\intercal \mathbf{X})^{-1} \mathbf{c}}} \sim t_{N-K}$$

- Note that the left-hand side is not a statistic, but a just a pivotal quantity (one for which the distribution does not depend upon unknown parameters).

A confidence interval for $\boldsymbol{\beta_k}$:

- Suppose we seek to develop a confidence interval for β_k.
- Letting $V_k = [\text{diag}((\mathbf{X}^\top\mathbf{X})^{-1})]_{k,k}$, we have that

$$\frac{\widehat{\beta}_k - \beta_k}{S\sqrt{V_k}} \sim t_{N-K}.$$

- This suggests the following $100(1-\alpha)\%$ confidence interval for β_k:

$$\widehat{\beta}_k \pm t_{1-\alpha/2,N-K} S\sqrt{V_k}.$$

- Two interpretations:
  - $100(1-\alpha)\%$ of intervals generated in this way contain β_j.
  - This interval contains the set of all null hypotheses β_j^* such that $H_0 : \beta_j = \beta_j^*$ cannot be rejected at the nominal level α (two-sided).

A confidence interval for $\mathbf{c}^\mathsf{T}\boldsymbol{\beta}$:

- Suppose, more generally, that we seek to develop a confidence interval for $\mathbf{c}^\mathsf{T}\boldsymbol{\beta}$. Then,

$$\frac{\mathbf{c}^\mathsf{T}\widehat{\boldsymbol{\beta}} - \mathbf{c}^\mathsf{T}\boldsymbol{\beta}}{\sqrt{S^2\mathbf{c}^\mathsf{T}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{c}}} \sim t_{N-K}.$$

- This suggests the following $100(1-\alpha)\%$ confidence interval for $\mathbf{c}^\mathsf{T}\boldsymbol{\beta}$:

$$\mathbf{c}^\mathsf{T}\widehat{\boldsymbol{\beta}} \pm t_{1-\alpha/2,N-K} S\sqrt{\mathbf{c}^\mathsf{T}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{c}}.$$

- Quantities of the form $\mathbf{c}^\mathsf{T}\boldsymbol{\beta}$ include comparisons of non-reference categories, stratum-specific effects, etc.

Table of Contents

A confidence interval for $\mathbf{x}_0^\mathsf{T}\boldsymbol{\beta}$:

- Parameters of the form $\mathbf{c}^\mathsf{T}\boldsymbol{\beta}$ include stratum-specific means: $\mathrm{E}[\mathbf{y}|\mathbf{x}_0] = \mathbf{x}_0^\mathsf{T}\boldsymbol{\beta}$.
- Indeed, $\mathbf{x}_0$ does not need to mark a stratum that is specifically represented in the data (though it should be "in the range" of the covariate space to avoid extrapolation).
- Letting $\widehat{y}(\mathbf{x}_0) = \mathbf{x}_0^\mathsf{T}\widehat{\boldsymbol{\beta}}$, we have that:

$$\frac{\widehat{y}(\mathbf{x}_0) - \mathbf{x}_0^\mathsf{T}\boldsymbol{\beta}}{\sqrt{S^2\mathbf{x}_0^\mathsf{T}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{x}_0}} \sim t_{N-K}.$$

- This suggests the following $100(1-\alpha)\%$ confidence interval for $\mathbf{c}^\mathsf{T}\boldsymbol{\beta}$:

$$\widehat{y}(\mathbf{x}_0) \pm t_{1-\alpha/2,N-K} S\sqrt{\mathbf{x}_0^\mathsf{T}(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{x}_0}.$$

A confidence interval for $\mathbf{x}_0^\intercal \boldsymbol{\beta}$:

- In the special case of simple linear regression, $\mathbf{x}_0 = \begin{bmatrix} 1 & x_0 \end{bmatrix}^\intercal$, and the following $100(1-\alpha)\%$ confidence interval can be derived:

$$\widehat{y}(x_0) \pm t_{1-\alpha/2, N-2} S \sqrt{\frac{1}{N} + \frac{(x_0 - \overline{x})^2}{\sum_{i=1}^{N} (x_i - \overline{x})^2}}.$$

Table of Contents

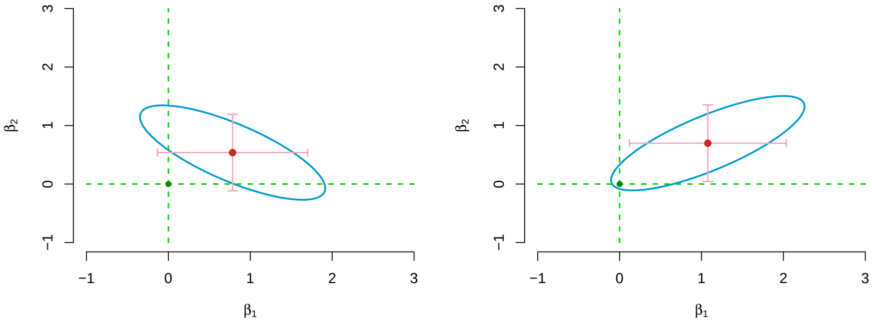
A confidence region for $(\beta_1, \beta_2)^\intercal$:

- A $100(1-\alpha)\%$ joint confidence region for $\boldsymbol{\beta}$ can be formed by finding the set of values of $\mathbf{b}$ for which the following inequality is satisfied:

$$(\widehat{\boldsymbol{\beta}} - \mathbf{b})^\intercal [\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}]]^{-1} (\widehat{\boldsymbol{\beta}} - \mathbf{b}) \leq Q \times \mathcal{F}_{Q,N-K;1-\alpha}$$

- This is an elliptical region (possibly in high dimensions) that inverts the hypothesis test.
  - Comprises all points, $\mathbf{b}$ with which your data are consistent—and would not be ruled out by an α-level F-test with that point as the null.
- In practice, we often use the fact that $Q \times F \overset{\cdot}{\sim} \chi^2_2$.
  - In the case of two dimensions, think of the ellipse as a suitable contour of a multivariate normal distribution.

Example 6.1: Joint confidence regions for $(\beta_1, \beta_2)^\intercal$



Example 6.1: Noteworthy items for previous slide

- The confidence region on the left notably *excludes* the origin.
  - However, the respective CIs for β_1 and β_2 do not rule out the individual null hypotheses $H_0 : \beta_1 = 0$ and $H_0 : \beta_2 = 0$.
- The confidence region on the right notably *includes* the origin.
  - However, the respective CIs for β_1 and β_2 rule out the individual null hypotheses $H_0 : \beta_1 = 0$ and $H_0 : \beta_2 = 0$.
- While seemingly paradoxical, these are *not* contradictions: absence of evidence is not evidence of absence. A frequentist must contend with that bitter fact.
- Elliptical area: $\chi^2_{2,0.95} \cdot \pi(\det(\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}]))^{1/2}$.
  - The nonzero covariance serves to rotate the elliptical region.
- The `ellipse()` function in R is helpful for plotting confidence regions of this variety.

Joint regions: Independent tests

- Suppose you seek to form a joint confidence region for $(\theta_1, \theta_2)^\intercal$, whereby you know in advance that $\mathrm{Cov}[\widehat{\theta}_1, \widehat{\theta}_2] = 0$.
- A rectangular confidence region based on the Cartesian product of individual 97.468% confidence intervals for θ_1 and θ_2 should have valid coverage.
- Notice that the individual confidence intervals suggested above have levels of $\alpha = 1 - (1 - 0.05)^{1/2}$ and not $\alpha = 0.05$. This familywise error control method is known as the Šidák correction and is appropriate when dealing with a small number of independent tests.
  - The Bonferroni correction would use $\alpha = 0.05/2$.
- However, a 95% confidence ellipse will generally have a smaller area than the rectangle.

Table of Contents

A confidence interval for σ^2:

- Recall: $\text{RSS}/\sigma^2 = (N-K)S^2/\sigma^2 \sim \chi^2_{N-K}$.
- This suggests the following $100(1-\alpha)\%$ confidence interval for $1/\sigma^2$:

$$\left[\frac{\chi_{\alpha/2,N-K}}{(N-K)S^2}, \frac{\chi_{1-\alpha/2,N-K}}{(N-K)S^2}\right]$$

- This suggests the following $100(1-\alpha)\%$ confidence interval for σ^2:

$$\left[\frac{N-K}{\chi_{\alpha/2,N-K}}S^2, \frac{N-K}{\chi_{1-\alpha/2,N-K}}S^2\right]$$

- This suggests the following $100(1-\alpha)\%$ confidence interval for σ:

$$\left[\sqrt{\frac{N-K}{\chi_{\alpha/2,N-K}}}S, \sqrt{\frac{N-K}{\chi_{1-\alpha/2,N-K}}}S\right]$$

Table of Contents

Prediction intervals

Future/external observations:

- Imagine we want to use a model to establish a range of typical values for outcomes corresponding to a covariate profile $\mathbf{x}_0$.
  - Also termed a prediction interval.
  - Characterizes a "reference range."
- Similar to a confidence interval, we want a certain percentage of intervals we create in this fashion to capture the value of Y_0, a random variable marking a future observation with covariate vector $\mathbf{x}_0$.

Future/external observations:

- Let $\widehat{Y}(\mathbf{x}_0)$ denote the fitted value (estimated mean value of Y in the stratum $\mathbf{x}_0$.
- By assumption, the future observation Y_0 is taken to be independent of all observations that contributed to our fitted value, $\widehat{Y}(\mathbf{x}_0)$.
- Therefore, we find that:

$$\begin{aligned} \mathrm{Var}[Y_0 - \widehat{Y}(\mathbf{x}_0)] &= \mathrm{Var}[Y_0] + \mathrm{Var}[\widehat{Y}(\mathbf{x}_0)] \\ &= \sigma^2 + \mathrm{Var}[\mathbf{x}_0^\intercal \widehat{\boldsymbol{\beta}}] \\ &= \sigma^2 + \mathbf{x}_0^\intercal[\sigma^2(\mathbf{X}^\intercal\mathbf{X})^{-1}]\mathbf{x}_0 \\ &= \sigma^2(1 + \mathbf{x}_0^\intercal(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{x}_0), \end{aligned}$$

which can be estimated as:

$$\widehat{\mathrm{Var}}[Y_0 - \widehat{Y}(\mathbf{x}_0)] = S^2(1 + \mathbf{x}_0^\intercal(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{x}_0),$$

Prediction intervals

Future/external observations:

- Since $Y_0 \perp\!\!\!\perp \widehat{Y}(\mathbf{x}_0)$ and $S^2 \perp\!\!\!\perp \widehat{Y}(\mathbf{x}_0)$, we have:

$$\frac{Y_0 - \widehat{Y}(\mathbf{x}_0)}{S\sqrt{(1 + \mathbf{x}_0^\intercal(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{x}_0)}} \sim t_{N-K}.$$

- This suggests the following $100(1-\alpha)\%$ prediction interval for Y_0:

$$\widehat{y}(\mathbf{x}_0) \pm t_{1-\alpha/2,N-K} S\sqrt{1 + \mathbf{x}_0^\intercal(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{x}_0}.$$

Prediction intervals

Future/external observations:

- In the special case of simple linear regression, $\mathbf{x}_0 = \begin{bmatrix} 1 & x_0 \end{bmatrix}^\mathsf{T}$, and the following $100(1-\alpha)\%$ prediction interval can be derived:

$$\widehat{y}(x_0) \pm t_{1-\alpha/2,N-2} S \sqrt{1 + \frac{1}{N} + \frac{(x_0 - \overline{x})^2}{\sum_{i=1}^{N}(x_i - \overline{x})^2}}.$$

Example 6.2: Simulation (setup)

```r
## Set seed for reproducibility
set.seed(7345)

## Sample size
n <- 100

## Number of simulations
nsim <- 5000

## Simulation parameters
beta0 <- 1
beta1 <- 2
sigma <- 2

## Predictor of interest (fixed)
x <- runif(n, 1, 4)
xbar <- mean(x)

## Stratum of interest
x0 <- 4
```

Example 6.2: Simulation (data generation/model fit)

```
## Place to store results
coverage <- matrix(1, nrow = nsim, ncol = 1)

## Conduct simulation
for (j in 1:nsim)
{
  ## External observation
  y0 <- rnorm(1, beta0 + beta1 * x0, sigma)

  ## Generate outcome
  y <- beta0 + beta1*x + rnorm(n, 0, sigma)

  ## OLS (coefficients and error variance)
  zz <- lm(y ~ x)
  bhat <- coef(zz)
  S <- sqrt(sum(zz$residuals^2)/(n - 2))

  ## Critical value
  cv <- qt(0.975, df = n - 2)
```

Example 6.2: Simulation (interval and results)

```
  ## Fitted value
  yhat <- as.numeric(bhat[1] + bhat[2]*x0)

  ## Prediction interval
  pilo <- yhat - cv*S*sqrt(1 + 1/n + (x0 - xbar)^2/((n - 1)*var(x)))
  pihi <- yhat + cv*S*sqrt(1 + 1/n + (x0 - xbar)^2/((n - 1)*var(x)))

  ## Decide whether coverage was achieved
  if (pilo > y0 | pihi < y0) {coverage[j] <- 0}
}

## > mean(coverage)
## [1] 0.9462
```

Summary

This unit:

- Confidence intervals for parameters.
- Confidence intervals for means.
- Joint confidence regions.
- Confidence intervals for the error variance.
- Prediction intervals.

Summary: So far

- Random vectors and matrices; multivariate normal theory.
- Ordinary least squares.
- Hypothesis testing and ANOVA.
- Weighted least squares.
- Misspecification.
- Confidence regions and prediction.

Summary: Coming up

- Diagnostics.
- Regularization.
- Bayesian regression.
- Exponential families.
- Generalized linear models.
- Sandwich and bootstrap.
- Quasi-likelihood.
- Hypothesis testing for GLMs.
- Diagnostics for GLMs.
- Further considerations for binary outcomes.
- Nonlinear least squares.