

Linear & Generalized Linear Models/ Advanced Regression for Independent Data

Andrew J. Spieker, Ph.D.

Associate Professor of Biostatistics
Vanderbilt University

Set 3: Hypothesis testing; ANOVA

Version: 09/23/2026

TABLE OF CONTENTS

- 1 Motivation for hypothesis testing
- 2 Testable hypotheses and key lemmas
- 3 The F -statistic and its distribution under the null
- 4 The distribution of the F -statistic under the alternative
- 5 Example: One-way ANOVA
- 6 Large-sample theory
- 7 More general form for the null hypothesis
- 8 ANOVA: A little more formally
- 9 One-way ANOVA
- 10 Two-way ANOVA
- 11 Coefficient of determination

Hypothesis testing:

- We seek to evaluate statistical strength of evidence. Example null hypotheses that I suspect you'll be familiar with (examples provided in parentheses):
 - ▶ $H_0 : \beta_1 = 0$ (association).
 - ▶ $H_0 : \beta_1 = \beta_2 = 0$ (overall/omnibus association).
 - ▶ $H_0 : \beta_2 - \beta_1 = 0$ (between-strata comparison).
 - ▶ $H_0 : \beta_0 + \beta_1 = 0$ (stratum-specific mean).
- How do we perform these tests?
- More generally, can we test hypotheses of the form $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$?
- Even more generally, can we test hypotheses of the form $\mathbf{C}\boldsymbol{\beta} = \mathbf{c}$?

MOTIVATION FOR HYPOTHESIS TESTING

Hypothesis testing:

- To begin with, suppose we want to test $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, where $\boldsymbol{\beta}$ is of length K and $\mathbf{C}$ is a $Q \times K$ matrix (i.e., Q linear hypotheses).
- Intuitively, we want a test that rejects H_0 if $\mathbf{C}\hat{\boldsymbol{\beta}}$ is “very different” from $\mathbf{0}$, and otherwise fails to reject.
- A Wald test statistic takes the form: $(\mathbf{C}\hat{\boldsymbol{\beta}})^\top [\widehat{\text{Cov}}[\mathbf{C}\hat{\boldsymbol{\beta}}]]^{-1} \mathbf{C}\hat{\boldsymbol{\beta}}$.
- If $\mathbf{X}$ is of full rank and we use OLS under the assumption that $\text{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}$, then a Wald test would reject H_0 for large values of:

$$(\mathbf{C}\hat{\boldsymbol{\beta}})^\top [\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} \mathbf{C}\hat{\boldsymbol{\beta}} / S^2,$$

where $S^2 = \text{RSS} / (N - K)$.

Assumptions and road map:

- We will derive a sensible test statistic similar to the Wald test described on the previous slide.
- We will then—under certain parametric assumptions—derive its exact distribution under the null and alternative hypotheses.
- We will be sure to talk a bit more about justification for the Wald test when parametric assumptions are not met.

TABLE OF CONTENTS

- 1 Motivation for hypothesis testing
- 2 Testable hypotheses and key lemmas
- 3 The F -statistic and its distribution under the null
- 4 The distribution of the F -statistic under the alternative
- 5 Example: One-way ANOVA
- 6 Large-sample theory
- 7 More general form for the null hypothesis
- 8 ANOVA: A little more formally
- 9 One-way ANOVA
- 10 Two-way ANOVA
- 11 Coefficient of determination

Setup:

- As has generally been the case thus far, the model is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.
- Assume $\text{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}$ until further notice.
- We will present the results assuming $\mathbf{X}$ is of full rank.
 - ▶ Results generalize to the cases of $\text{rank}(\mathbf{X}) = R < K$ (you would see more pseudo-inverses, and “ K ” would generally be replaced with “ R ” throughout the notes).

Testable hypotheses:

- Suppose we want to test the hypothesis:

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0} \quad \text{or} \quad H_0 : \mathbf{c}_1^T \boldsymbol{\beta} = \dots = \mathbf{c}_Q^T \boldsymbol{\beta} = 0.$$

- We say the hypothesis is *testable* if $\mathbf{c}_i^T = \mathbf{d}_i^T \mathbf{X}$ for $i = 1, \dots, Q$.
 - ▶ This means $\mathbf{c}_1^T \boldsymbol{\beta}$, $\dots$, and $\mathbf{c}_Q^T \boldsymbol{\beta}$ are estimable functions.
 - ▶ This occurs precisely when the rows of $\mathbf{C}$ are linear combinations of the rows of $\mathbf{X}$.
 - ▶ In the case of full rank, any row of $\mathbf{C}$ will be a linear combination of the rows of $\mathbf{X}$ (all linear hypotheses are testable). When $\mathbf{X}$ is not of full rank, not all hypotheses are testable, but we are not concerning ourselves with this case.
- In these notes, we will assume $Q \leq K$ and $\text{rank}(\mathbf{C}) = Q$.
 - ▶ That is, no redundancy.

Unrestricted and restricted models:

- Hypothesis: $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$.
- The premise of the hypothesis test that we're about to develop is based on a fit to an unrestricted model and a fit to the restricted model (under H_0). Let's talk about a couple of examples of what this means, since it's so central to the concepts.

Example 3.1: Testing an overall effect of a categorical variable

- Unrestricted model:

$$Y = \beta_0 + \beta_1 1(X = 1) + \beta_2 1(X = 2) + \beta_3 1(X = 3) + \epsilon.$$

- Hypothesis: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$, or:

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}.$$

- Restricted model (under H_0):

$$Y = \beta_0^* + \epsilon^*.$$

Example 3.2: Testing an overall effect in a two-way interaction model

- Unrestricted model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon.$$

- Hypothesis: $H_0 : \beta_1 = \beta_3 = 0$, or:

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{0}.$$

- Restricted model (under H_0):

$$Y = \beta_0^* + \beta_2^* X_2 + \epsilon^*.$$

Example 3.3: Testing a stratum-specific effect an interaction model

- Unrestricted model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon.$$

- Hypothesis: $H_0 : \beta_1 + 2\beta_3 = 0$, or:

$$H_0 : \mathbf{C}\boldsymbol{\beta} = [0 \quad 1 \quad 0 \quad 2] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = 0.$$

- Restricted model (under H_0):

$$Y = \beta_0^* + \beta_1^*(X_1 - X_1 X_2 / 2) + \beta_2^* X_2 + \epsilon^*.$$

Notation and intuition:

- RSS_H denotes the residual sum of squares under the null, H_0 .
- RSS_H never smaller than RSS (i.e., under the unrestricted model).
 - ▶ I'll need to supply justification for this. Coming up!
- If H_0 is true, we expect RSS_H to be “only slightly greater” than RSS .
- If H_0 is false, we expect RSS_H to be “much greater” than RSS .

Helping to visualize the hypothesis test:

- Under the full model, $\mathbf{X}\boldsymbol{\beta} \in \mathcal{C}(\mathbf{X}) =: \Omega$.
- Since $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is testable, we may write $\mathbf{C} = \mathbf{D}\mathbf{X}$ for some $\mathbf{D}$.
 - ▶ The rows of $\mathbf{C}$ are linear combinations of the rows of $\mathbf{X}$.
 - ▶ The columns of $\mathbf{C}$ are linear combinations of the columns of $\mathbf{D}$.
- Under H_0 , $\mathbf{C}\boldsymbol{\beta} = \mathbf{D}\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$ such that $\mathbf{X}\boldsymbol{\beta} \in \mathcal{N}(\mathbf{D})$.
- Therefore, under the restricted model, $\mathbf{X}\boldsymbol{\beta} \in \mathcal{C}(\mathbf{X}) \cap \mathcal{N}(\mathbf{D}) =: \omega \subset \Omega$.
- Let $\hat{\mathbf{y}} = \mathbf{P}_\Omega \mathbf{y}$, denote the fitted value under the unrestricted model.
- Let $\hat{\mathbf{y}}_H = \mathbf{P}_\omega \mathbf{y}$ denote the fitted value under the restricted model.

TESTABLE HYPOTHESES AND KEY LEMMAS

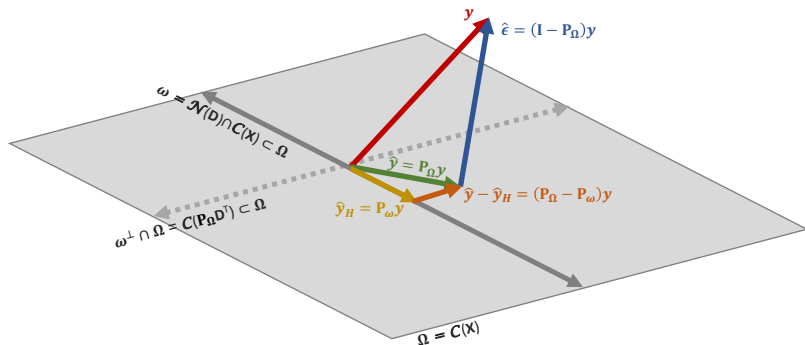


Figure 3.1: Visualizing hypothesis testing.



Understanding RSS:

- For the full model, we have:

$$\text{RSS} = (\mathbf{y} - \hat{\mathbf{y}})^T(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{y}^T(\mathbf{I} - \mathbf{P}_\Omega)\mathbf{y}.$$

- For the restricted model, we have:

$$\text{RSS}_H = (\mathbf{y} - \hat{\mathbf{y}}_H)^T(\mathbf{y} - \hat{\mathbf{y}}_H) = \mathbf{y}^T(\mathbf{I} - \mathbf{P}_\omega)\mathbf{y}.$$

- It follows, then, that:

$$\text{RSS}_H - \text{RSS} = \mathbf{y}^T(\mathbf{P}_\Omega - \mathbf{P}_\omega)\mathbf{y}.$$

- What have we learned about nested linear models?

Lemma 3.1: Difference between $\mathbf{P}_\Omega$ and $\mathbf{P}_\omega$

$$\mathbf{P}_\Omega - \mathbf{P}_\omega = \mathbf{P}_{\omega^\perp \cap \Omega}, \text{ and } \omega^\perp \cap \Omega = \mathcal{C}(\mathbf{P}_\Omega \mathbf{D}^\top).$$

- See an advanced regression textbook (e.g., Seber & Lee) for a formal proof.
- Intuition: $\mathcal{R}(\mathbf{D}) = \mathcal{C}(\mathbf{D}^\top)$ is the orthogonal complement of $\mathcal{N}(\mathbf{D})$. To keep everything within Ω , project onto $\mathcal{C}(\mathbf{X})$.
- This is important; it implies that $\text{RSS}_H - \text{RSS} \geq 0$ (specifically, $\mathbf{P}_\Omega - \mathbf{P}_\omega \succcurlyeq 0$ because it is a projection matrix as we previously learned from [Lemma 1.4](#)).

Understanding RSS:

- We already know that $\mathbf{X}\boldsymbol{\beta} \in \Omega$, such that $(\mathbf{I} - \mathbf{P}_\Omega)\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$. Therefore,

$$\text{RSS} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{I} - \mathbf{P}_\Omega) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

- Under the restricted model, $\mathbf{X}\boldsymbol{\beta} \in \omega$, such that $(\mathbf{I} - \mathbf{P}_\omega)\mathbf{X}\boldsymbol{\beta} = \mathbf{0}$.
Therefore,

$$\text{RSS}_H = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{I} - \mathbf{P}_\omega) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

- It follows, then, that **under the restricted model**, we have:

$$\text{RSS}_H - \text{RSS} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{P}_\Omega - \mathbf{P}_\omega) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

- It may not be clear yet why I want to write RSS and RSS_H in terms of the (unknown) parameter $\boldsymbol{\beta}$; I will later invoke theorems/lemmas from Set 1 that demand I do so. This slide is extremely useful.

Lemma 3.2: Difference between $\mathbf{P}_\Omega$ and $\mathbf{P}_\omega$ (more explicitly)

If $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is a testable hypothesis, then:

$$\mathbf{P}_\Omega - \mathbf{P}_\omega = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top (\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top)^{-1} \mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top.$$

- To prove this, note that $\mathbf{P}_\Omega - \mathbf{P}_\omega$ is a projection matrix that, by [Lemma 3.1](#), projects onto the columns of $\mathbf{P}_\Omega \mathbf{D}^\top$. Expand out the following quantity:

$$\mathbf{P}_\Omega \mathbf{D}^\top ((\mathbf{P}_\Omega \mathbf{D}^\top)^\top (\mathbf{P}_\Omega \mathbf{D}^\top))^{-1} (\mathbf{P}_\Omega \mathbf{D}^\top)^\top,$$

and recognize that $\mathbf{P}_\Omega = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ and $\mathbf{C} = \mathbf{D}\mathbf{X}$ (by the fact that $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is testable).

Theorem 3.1: Expression for $RSS_H - RSS$

If $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is a testable hypothesis, then:

$$RSS_H - RSS = (\mathbf{C}\hat{\boldsymbol{\beta}})^T (\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T)^{-1} \mathbf{C}\hat{\boldsymbol{\beta}},$$

where $\hat{\boldsymbol{\beta}}$ is the OLS estimate from the unrestricted model.

- Using the fact that $RSS_H - RSS = \mathbf{y}^T(\mathbf{P}_\Omega - \mathbf{P}_\omega)\mathbf{y}$, this follows directly from [Lemma 3.2](#) (use the fact that $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$).

Lemma 3.3: Properties involving full-rank $\mathbf{C}$

If $\mathbf{C}$ is of full (row) rank and corresponds to a testable hypothesis $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, then $\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top \succ 0$.

- To prove this, note that $\mathbf{C} = \mathbf{D}\mathbf{X}$; therefore, $\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top = \mathbf{D}\mathbf{P}_\Omega \mathbf{D}^\top$.
- $\mathbf{P}_\Omega \succeq 0 \Rightarrow \mathbf{D}\mathbf{P}_\Omega \mathbf{D}^\top \succeq 0$. To strengthen to positive definiteness, we need to argue that $\mathbf{v}^\top \mathbf{D}\mathbf{P}_\Omega \mathbf{D}^\top \mathbf{v} = 0 \Rightarrow \mathbf{v} = \mathbf{0}$ for $\mathbf{v} \in \mathbb{R}^Q$. Set $\mathbf{d} = \mathbf{D}^\top \mathbf{v}$.
- $\mathbf{v}^\top \mathbf{D}\mathbf{P}_\Omega \mathbf{D}^\top \mathbf{v} = 0 \Leftrightarrow \mathbf{d} \in \mathcal{N}(\mathbf{X}^\top)$. But $\mathbf{X}^\top \mathbf{d} = \mathbf{X}^\top \mathbf{D}^\top \mathbf{v} = (\mathbf{D}\mathbf{X})^\top \mathbf{v} = \mathbf{C}^\top \mathbf{v}$.
- For $\mathbf{C}^\top \mathbf{v} = \mathbf{0}$ to hold without $\mathbf{v} = \mathbf{0}$ would contradict the assumption that $\mathbf{C}$ is of full row rank. Therefore, the result follows.

Theorem 3.2: Expectation of $RSS_H - RSS$

If $\text{rank}(\mathbf{C}) = Q$, then:

$$E[RSS_H - RSS] = \sigma^2 Q + \boldsymbol{\beta}^T \mathbf{C}^T (\mathbf{C}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T)^{-1} \mathbf{C} \boldsymbol{\beta}.$$

- To prove this, recall that $RSS_H = RSS = \mathbf{y}^T (\mathbf{P}_\Omega - \mathbf{P}_\omega) \mathbf{y}$, employ the formula for the expectation of a quadratic form, and use [Lemma 3.2](#).
- By [Lemma 3.3](#), $\boldsymbol{\beta}^T \mathbf{C}^T (\mathbf{C}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T)^{-1} \mathbf{C} \boldsymbol{\beta} = 0$ only when $H_0 : \mathbf{C} \boldsymbol{\beta} = \mathbf{0}$ is true, and is otherwise positive (which is *exactly* the property that we would want).

Motivating a reasonable test statistic:

- We've argued that $RSS_H - RSS$ will be (in expectation) equal to $Q\sigma^2$ under H_0 , and will be larger when H_0 is false.
- This implies that the following may be reasonable to test H_0 :

$$F = \frac{(RSS_H - RSS)/Q}{RSS/(N - K)}.$$

- If F is large, we want to reject H_0 .
- Thus far, we have only assumed $E[\epsilon] = \mathbf{0}$ and $\text{Cov}[\epsilon] = \sigma^2 \mathbf{I}$.
- To obtain an exact form for the distribution under the null and under the alternative, we have to make a distributional assumption on ϵ .
- Certainly, the fact that we're calling it the F -statistic is suggestive of its distribution! :)

TABLE OF CONTENTS

- 1 Motivation for hypothesis testing
- 2 Testable hypotheses and key lemmas
- 3 The F -statistic and its distribution under the null
- 4 The distribution of the F -statistic under the alternative
- 5 Example: One-way ANOVA
- 6 Large-sample theory
- 7 More general form for the null hypothesis
- 8 ANOVA: A little more formally
- 9 One-way ANOVA
- 10 Two-way ANOVA
- 11 Coefficient of determination

THE DISTRIBUTION OF F UNDER THE NULL

Theorem 3.3: Distribution of $\text{RSS}_H - \text{RSS}$ under H_0

Suppose $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $\text{rank}(\mathbf{X}) = K$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. If $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is a testable hypothesis with $\text{rank}(\mathbf{C}_{Q \times K}) = Q$, then under the null hypothesis:

$$F = \frac{(\text{RSS}_H - \text{RSS})/Q}{\text{RSS}/(N - K)} \sim \mathcal{F}_{Q, N-K}.$$

THE DISTRIBUTION OF F UNDER THE NULL

Theorem 3.3: Proof

- Note that $\text{rank}(\mathbf{P}_\Omega) = K$ and $\text{rank}(\mathbf{P}_\omega) = K - Q$.
- Recall that $\mathbf{y}^\top(\mathbf{I} - \mathbf{P}_\Omega)\mathbf{y} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{I} - \mathbf{P}_\Omega)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.
- Therefore, applying Theorem 1.3:
 - ▶ $\text{RSS}/\sigma^2 = \mathbf{y}^\top(\mathbf{I} - \mathbf{P}_\Omega)\mathbf{y}/\sigma^2 \sim \chi_{N-K}^2$.
- By Theorem 1.5 (together with Lemma 1.4):
 - ▶ $\text{RSS}_H - \text{RSS}$ is independent of RSS .
- By Theorem 1.6:
 - ▶ $(\text{RSS}_H - \text{RSS})/\sigma^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{P}_\Omega - \mathbf{P}_\omega)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/\sigma^2 \sim \chi_Q^2$ under H_0 .
- The result therefore follows immediately by the definition of the F -distribution.

TABLE OF CONTENTS

- 1 Motivation for hypothesis testing
- 2 Testable hypotheses and key lemmas
- 3 The F -statistic and its distribution under the null
- 4 The distribution of the F -statistic under the alternative
- 5 Example: One-way ANOVA
- 6 Large-sample theory
- 7 More general form for the null hypothesis
- 8 ANOVA: A little more formally
- 9 One-way ANOVA
- 10 Two-way ANOVA
- 11 Coefficient of determination

THE DISTRIBUTION OF F UNDER THE ALTERNATIVE

Motivation:

- Knowing the distribution of F under the null is what allows us to conduct the hypothesis test of interest having a specific level.
- However, finding a test that achieves the nominal level is not the hard part (if you don't believe me, construct a 20-sided fair die and reject the null hypothesis if and only if it comes up as a 17).
- We also want to know that the distribution of F is meaningfully different from its null distribution when in fact the null is false.
- Ideally, the “more different” $\mathbf{C}\boldsymbol{\beta}$ is from $\mathbf{0}$, the “more different” we would want the distribution of F to be from its null distribution. Given [Theorem 3.2](#), we should feel pretty optimistic that this will be the case!

THE DISTRIBUTION OF F UNDER THE ALTERNATIVE

Theorem 3.4: Distribution of $RSS_H - RSS$

Suppose $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $\text{rank}(\mathbf{X}) = K$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. If $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is a testable hypothesis with $\text{rank}(\mathbf{C}_{Q \times K}) = Q$, then:

$$F = \frac{(RSS_H - RSS)/Q}{RSS/(N - K)} \sim \mathcal{F}_{Q, N-K}(\lambda), \text{ where}$$

$$\lambda = \boldsymbol{\beta}^T \mathbf{X}^T (\mathbf{P}_\Omega - \mathbf{P}_\omega) \mathbf{X} \boldsymbol{\beta} / (2\sigma^2) = (\mathbf{C}\boldsymbol{\beta})^T (\mathbf{C}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{C}^T)^{-1} \mathbf{C}\boldsymbol{\beta} / (2\sigma^2).$$

Note: You can use this information to determine, for instance:

- 1 Sample size needed to achieve a certain power for a specified effect.
- 2 The power to detect a specified effect under a specified sample size.
- 3 The minimum effect at which a certain power will be achieved under a specified sample size.

R functions `pF` and `qF()` are useful!

THE DISTRIBUTION OF F UNDER THE ALTERNATIVE

Theorem 3.4: Proof

- As we've seen, $\text{rank}(\mathbf{P}_\Omega) = K$ and $\text{rank}(\mathbf{P}_\omega) = K - Q$.
- Recall that $\mathbf{y}^\top(\mathbf{I} - \mathbf{P}_\Omega)\mathbf{y} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top(\mathbf{I} - \mathbf{P}_\Omega)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.
- Therefore, applying Theorem 1.3:
 - ▶ $\text{RSS}/\sigma^2 = \mathbf{y}^\top(\mathbf{I} - \mathbf{P}_\Omega)\mathbf{y}/\sigma^2 \sim \chi^2_{N-K}$.
- By Theorem 1.5 (together with Lemma 1.4):
 - ▶ $\text{RSS}_H - \text{RSS}$ is independent of RSS .
- By Theorem 1.7:
 - ▶ $(\text{RSS}_H - \text{RSS})/\sigma^2 = \mathbf{y}^\top(\mathbf{P}_\Omega - \mathbf{P}_\omega)\mathbf{y}/\sigma^2 \sim \chi^2_Q((\mathbf{X}\boldsymbol{\beta})^\top(\mathbf{P}_\Omega - \mathbf{P}_\omega)\mathbf{X}\boldsymbol{\beta}/2\sigma^2)$.
- The result therefore follows immediately by the definition of the non-central F -distribution.

More about the non-centrality parameter:

- Recall that [Lemma 3.2](#) presented an expression for $\mathbf{P}_\Omega - \mathbf{P}_\omega$, which we can leverage to write down the non-centrality parameter more specifically as:

$$\lambda = \frac{1}{2\sigma^2} \boldsymbol{\beta}^\top \mathbf{C}^\top (\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top)^{-1} \mathbf{C} \boldsymbol{\beta}.$$

- It is straightforward to show that:

$$E[F] = \left(1 + \frac{2\lambda}{Q}\right) \frac{N - K}{N - K - 2}.$$

TABLE OF CONTENTS

- 1 Motivation for hypothesis testing
- 2 Testable hypotheses and key lemmas
- 3 The F -statistic and its distribution under the null
- 4 The distribution of the F -statistic under the alternative
- 5 Example: One-way ANOVA**
- 6 Large-sample theory
- 7 More general form for the null hypothesis
- 8 ANOVA: A little more formally
- 9 One-way ANOVA
- 10 Two-way ANOVA
- 11 Coefficient of determination

EXAMPLE: ONE-WAY ANOVA

Setup:

- Consider the following model:

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1N_1} \\ Y_{21} \\ \vdots \\ Y_{2N_2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1N_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2N_2} \end{bmatrix}.$$

- To-do:
 - Argue that $H_0 : \mu_1 = \mu_2$ is testable.
 - Write out H_0 in the form $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$.
 - Determine the F statistic and its distribution under H_0 assuming the errors are normally distributed.

EXAMPLE: ONE-WAY ANOVA

Arguing testability:

- Model:

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1N_1} \\ Y_{21} \\ \vdots \\ Y_{2N_2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1N_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2N_2} \end{bmatrix}.$$

- $H_0 : \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0$ can be written in the form $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, with

$$\mathbf{C} = \begin{bmatrix} 1 & -1 \end{bmatrix},$$

which is testable because the (single) row of $\mathbf{C}$ is a linear combination of the rows of $\mathbf{X}$.

EXAMPLE: ONE-WAY ANOVA

Deriving the F -statistic:

- By [Theorem 3.1](#) (expression for $RSS_H - RSS$) and [Lemma 3.3](#) ($\mathbf{C}$ is of full row-rank), $RSS_H - RSS = (\mathbf{C}\hat{\boldsymbol{\beta}})^\top (\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top)^{-1} \mathbf{C}\hat{\boldsymbol{\beta}}$.
- Easiest to take this in in pieces:

$$\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix} \Rightarrow (\mathbf{X}^\top \mathbf{X})^{-1} = \begin{bmatrix} 1/N_1 & 0 \\ 0 & 1/N_2 \end{bmatrix}$$

$$\begin{aligned} \Rightarrow \mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top &= \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1/N_1 & 0 \\ 0 & 1/N_2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &= 1/N_1 + 1/N_2. \end{aligned}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix}$$

$$\Rightarrow (\mathbf{C}\hat{\boldsymbol{\beta}})^\top (\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top)^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}}) = \frac{(\bar{Y}_1 - \bar{Y}_2)^2}{1/N_1 + 1/N_2}.$$

EXAMPLE: ONE-WAY ANOVA

Deriving the F -statistic:

- Additionally, $RSS = \sum_{i=1}^{N_1} (Y_{1i} - \bar{Y}_1)^2 + \sum_{i=1}^{N_2} (Y_{2i} - \bar{Y}_2)^2$.
- In this example, $N = N_1 + N_2$, $Q = 1$ and $K = 2$. It follows, then, that

$$F = \frac{(RSS_H - RSS)/Q}{RSS/(N_1 + N_2 - 2)} = \frac{(\bar{Y}_1 - \bar{Y}_2)^2}{s_p^2(1/N_1 + 1/N_2)}.$$

- You may recognize this as the square of the t -statistic from a two-sample t -test that assumes equal variances between groups.
- Under H_0 , **Theorem 3.3** says that $F \sim \mathcal{F}_{1, N_1 + N_2 - 2}$.
 - ▶ Note that if $T \sim t_{df}$, then $T^2 \sim \mathcal{F}_{1, df}$, which agrees with our result.

EXAMPLE: ONE-WAY ANOVA

Simulation: Ideas

- It would be nice to see that this theoretical result holds up in a simulation setting.
- To check that the theory holds up, we want to generate data under the null, compute the F -statistic (repeatedly), and compare to, say, randomly generated data from F distribution with the degrees of freedom that the statistic should have (“simulation studies 101”).
 - ▶ In this example, we want to check to see that the “whole distribution” matches. How do we do this?
 - ★ Histograms.
 - ★ QQ-plot (plot quantiles).
 - ★ Descriptive statistics (e.g., mean, standard deviation, quantiles).

EXAMPLE: ONE-WAY ANOVA

Simulation: Setup under H_0

```
1 ## Set seed for reproducibility
2 set.seed(7345)
3
4 ## Number of simulations
5 nsim <- 50000
6
7 ## Space to store results
8 res <- matrix(0, nrow = nsim, ncol = 1)
9
10 ## Set sample size
11 n1 <- 25
12 n2 <- 35
13
14 ## Set common standard deviation
15 sigma <- 2
16
17 ## Set mean difference (H0 : mean difference = 0)
18 delta <- 0
```

EXAMPLE: ONE-WAY ANOVA

Simulation: Run simulation under H_0

```
1 ## Run simulations
2 for (j in 1:nsim)
3 {
4   ## Generate data
5   Y0 <- rnorm(n1, 0, sigma)
6   Y1 <- rnorm(n2, delta, sigma)
7
8   ## Compute RSS
9   RSS <- sum((Y0 - mean(Y0))^2) + sum((Y1 - mean(Y1))^2)
10
11  ## Compute F statistic
12  Fstat <- ((mean(Y1) - mean(Y0))^2) / ((1/n1 + 1/n2)*RSS/(n1 + n2 - 2))
13  res[j,1] <- Fstat
14 }
15
16 ## Generate data from theoretical distribution
17 f.compare <- rf(20000, df1 = 1, df2 = n1 + n2 - 2)
```

EXAMPLE: ONE-WAY ANOVA

Simulation: Present results under H_0

```
1 ## Histograms overlaid (harder to compare)
2 hist(res, freq = FALSE, col = rgb(1, 0, 0, 0.2), xlab = "F statistic",
3       main = "Histogram (Null distribution)", ylim = c(0,0.8),
4       xlim = c(0,35), breaks = seq(0,30,1))
5 par(new = TRUE)
6 hist(f.compare, freq = FALSE, col = rgb(0, 0, 1, 0.2),
7       xlab = "", ylab = "", main = "", ylim = c(0,0.8),
8       xlim = c(0,35), breaks = seq(0,35,1))
9 legend(10, 0.7, pch = 15, col = c(rgb(1, 0, 0, 0.2),
10   rgb(0, 0, 1, 0.2)), lwd = 1, lty = 0,
11        legend = c("Observed", "Theoretical"))
12
13 ## Quantile-quantile plot (easier to compare, but be cautious)
14 qqplot(res, f.compare, frame.plot = FALSE, cex = 0.8, pch = 20,
15         col = "darkgray", xlim = c(0,30), ylim = c(0,30),
16         xlab = "F statistic quantile",
17         ylab = "F(1, 58) - Random/Theoretical",
18         main = "QQ-plot (Null distribution)")
19 abline(0,1, col = "blue")
```

EXAMPLE: ONE-WAY ANOVA

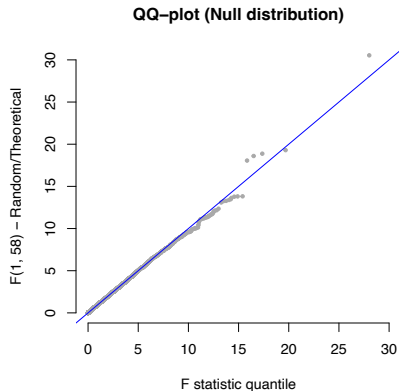
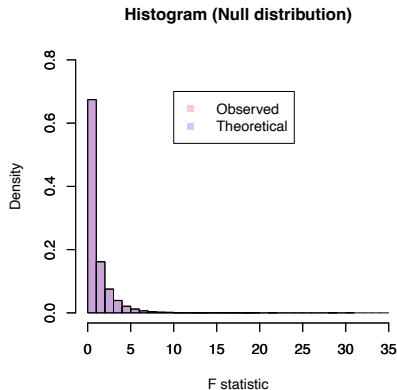


Figure 3.2: Simulation results under H_0 .



EXAMPLE: ONE-WAY ANOVA

Simulation: Setup and run simulation under alternative

```
1 ## Set seed for reproducibility
2 set.seed(7345)
3
4 ## Set mean difference (mean difference = 2)
5 delta <- 2
6
7 ## Space to store results
8 res <- matrix(0, nrow = nsim, ncol = 1)
9
10 ## Run simulations
11 for (j in 1:nsim)
12 {
13   ## Generate data
14   Y0 <- rnorm(n1, 0, sigma)
15   Y1 <- rnorm(n2, delta, sigma)
16
17   ## Compute RSS
18   RSS <- sum((Y0 - mean(Y0))^2) + sum((Y1 - mean(Y1))^2)
19
20   ## Compute F statistic
21   Fstat <- ((mean(Y1) - mean(Y0))^2) / ((1/n1 + 1/n2) * RSS / (n1 + n2 - 2))
22   res[j,1] <- Fstat
23 }
24
25 ## Non-centrality parameter
26 lambda <- (1/(2*sigma^2)) * ((2^2) / (1/n1 + 1/n2))
27
28 ## R parameterizes the non-centrality parameter differently
29 f.compare <- rf(20000, df1 = 1, df2 = n1 + n2 - 2, ncp = 2*lambda)
```

EXAMPLE: ONE-WAY ANOVA

Simulation: Present results under alternative

```
1 ## Histograms overlaid (harder to compare)
2 hist(res, freq = FALSE, col = rgb(1, 0, 0, 0.2), xlab = "F statistic",
3      main = "Histogram (Alternative distribution)", ylim = c(0,0.06),
4      xlim = c(0,100), breaks = seq(0,110,2))
5 par(new = TRUE)
6 hist(f.compare, freq = FALSE, col = rgb(0, 0, 1, 0.2),
7      xlab = "", ylab = "", main = "", ylim = c(0,0.06), xlim = c(0,100)
8      ,
9      breaks = seq(0,110,2))
10 legend(40, 0.05, pch = 15, col = c(rgb(1, 0, 0, 0.2),
11    rgb(0, 0, 1, 0.2)), lwd = 1, lty = 0,
12    legend = c("Observed", "Theoretical"))
13 ## Quantile-quantile plot (easier to compare, but be cautious)
14 qqplot(res, f.compare, frame.plot = FALSE, cex = 0.8, pch = 20,
15        col = "darkgray", xlim = c(0,100), ylim = c(0,100),
16        xlab = "F statistic quantile",
17        ylab = "F(1, 58; lambda) - Random/Theoretical",
18        main = "QQ-plot (Alternative distribution)")
19 abline(0,1, col = "blue")
```

EXAMPLE: ONE-WAY ANOVA

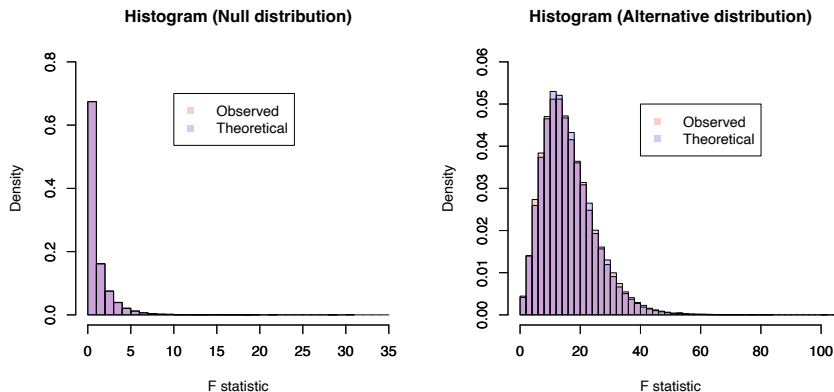


Figure 3.3: Simulation results under H_1 .



TABLE OF CONTENTS

- 1 Motivation for hypothesis testing
- 2 Testable hypotheses and key lemmas
- 3 The F -statistic and its distribution under the null
- 4 The distribution of the F -statistic under the alternative
- 5 Example: One-way ANOVA
- 6 Large-sample theory**
- 7 More general form for the null hypothesis
- 8 ANOVA: A little more formally
- 9 One-way ANOVA
- 10 Two-way ANOVA
- 11 Coefficient of determination

Ideas behind asymptotic justification:

- Building off of our previous example, note that key results tell us that the t -test is sensible in large samples even if the errors are not normally distributed.
- The natural question is: can we justify an F -test asymptotically even if our errors are not normally distributed?
 - ▶ Short answer: Yes!
- Without going into much technical detail, the F -test has theoretical justification if the errors are not normally distributed, provided the sample size is sufficiently large.

Note: Connection to χ^2 test

- If $X \sim \mathcal{F}_{d_1, d_2}$, then $Y = d_1 X \dot{\sim} \chi_{d_1}^2$ for large values of d_2 .
- It stands to reason that, under H_0 ,

$$Q \times F \dot{\sim} \chi_Q^2,$$

provided $N - K$ is sufficiently large.

Note: Connection to χ^2 test

- If $X \sim \mathcal{F}_{d_1, d_2}(\lambda)$, then $Y = d_1 X \dot{\sim} \chi_{d_1}^2(\lambda)$ for large values of d_2 .
- It stands to reason that, under alternative hypotheses,

$$Q \times F \dot{\sim} \chi_Q^2 \left(\frac{1}{2\sigma^2} \boldsymbol{\beta}^\top \mathbf{C}^\top (\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top)^{-1} \mathbf{C} \boldsymbol{\beta} \right),$$

provided $N - K$ is sufficiently large.

TABLE OF CONTENTS

- 1 Motivation for hypothesis testing
- 2 Testable hypotheses and key lemmas
- 3 The F -statistic and its distribution under the null
- 4 The distribution of the F -statistic under the alternative
- 5 Example: One-way ANOVA
- 6 Large-sample theory
- 7 More general form for the null hypothesis
- 8 ANOVA: A little more formally
- 9 One-way ANOVA
- 10 Two-way ANOVA
- 11 Coefficient of determination

TESTS OF $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{c}$

Expanding our results:

- Suppose you're interested in a set of the hypothesis $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{c}$.
- Another way to express this is that you're interested in the hypothesis $\mathbf{C}\boldsymbol{\beta} - \mathbf{c} = \mathbf{0}$.
- Results generalize in the expected way. For instance:
 - ▶ $\text{RSS}_H - \text{RSS} = (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{c})^\top (\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top)^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{c})$.

TABLE OF CONTENTS

- 1 Motivation for hypothesis testing
- 2 Testable hypotheses and key lemmas
- 3 The F -statistic and its distribution under the null
- 4 The distribution of the F -statistic under the alternative
- 5 Example: One-way ANOVA
- 6 Large-sample theory
- 7 More general form for the null hypothesis
- 8 ANOVA: A little more formally**
- 9 One-way ANOVA
- 10 Two-way ANOVA
- 11 Coefficient of determination

ANOVA: A LITTLE MORE FORMALLY

Ideas:

- ANOVA is an abbreviation for **analysis of variance**. We will see now why it is so termed.
- Historically, ANOVA has been the approach of choice in the social sciences, laboratory research, etc., allowing one to compare means across discrete strata.

ANOVA: A LITTLE MORE FORMALLY

Ideas:

- Regression: design matrix is front and center:
 - ▶ $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.
- ANOVA: design matrix is an afterthought. Instead, we express the outcome in terms of group-specific means or between-group contrasts.
- Example: $i = 1, \dots, n_j$ observations strata $j = 0, 1$.
 - ▶ Possible model: $Y_{ij} = \mu + \alpha_j + \epsilon_{ij}$.
 - ▶ Possible model: $Y_{ij} = \alpha_j + \epsilon_{ij}$.
- Example: $i = 1, \dots, n_{kj}$ observations in strata defined by $j = 1, \dots, J$ and $k = 1, \dots, K$.
 - ▶ Possible model: $Y_{ijk} = \mu_{jk} + \epsilon_{ijk}$.
 - ▶ Possible model: $Y_{ijk} = \alpha_j + \beta_k + \epsilon_{ijk}$.

TABLE OF CONTENTS

- 1 Motivation for hypothesis testing
- 2 Testable hypotheses and key lemmas
- 3 The F -statistic and its distribution under the null
- 4 The distribution of the F -statistic under the alternative
- 5 Example: One-way ANOVA
- 6 Large-sample theory
- 7 More general form for the null hypothesis
- 8 ANOVA: A little more formally
- 9 One-way ANOVA**
- 10 Two-way ANOVA
- 11 Coefficient of determination

Example 3.4: One-way ANOVA with two groups

- Suppose we have $i = 1, \dots, n_j$ independent observations in strata $j = 1, 2$ (shared outcome variance, σ^2). Let $N = n_1 + n_2$.

	Observations	Mean
Population 1	$y_{11}, \dots, y_{1n_1}$	$\bar{y}_1.$
Population 2	$y_{21}, \dots, y_{2n_2}$	$\bar{y}_2.$
Overall		$\bar{y}_{..}$

- Model: $Y_{ij} = \mu + \alpha_j + \epsilon_{ij}$.
- Suppose we seek to test $H_0 : \alpha_1 = \alpha_2$. We can use the F -statistic:

$$\begin{aligned}
 F &= \frac{(\text{RSS}_H - \text{RSS})/(J - 1)}{\text{RSS}/(N - 2)} \\
 &= \frac{(n_1(\bar{y}_1. - \bar{y}_{..})^2 + n_2(\bar{y}_2. - \bar{y}_{..})^2)/(2 - 1)}{(\sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1.)^2 + \sum_{i=1}^{n_2} (y_{2i} - \bar{y}_2.)^2)/(N - 2)}.
 \end{aligned}$$

Example 3.5: One-way ANOVA with J groups

- Suppose we have $i = 1, \dots, n_j$ independent observations in strata $j = 1, \dots, J$ (shared outcome variance, σ^2). Let $N = \sum_{j=1}^J n_j$.

	Observations	Mean
Population 1	$y_{11}, \dots, y_{1n_1}$	$\bar{y}_1.$
Population 2	$y_{21}, \dots, y_{2n_2}$	$\bar{y}_2.$
$\vdots$	$\vdots$	$\vdots$
Population J	$y_{J1}, \dots, y_{Jn_J}$	$\bar{y}_J.$
Overall		$\bar{y}_{..}$

- Model: $Y_{ij} = \mu + \alpha_j + \epsilon_{ij}$.
- To test $H_0 : \alpha_1 = \dots = \alpha_J$, we can use the F -statistic:

$$F = \frac{(\text{RSS}_H - \text{RSS})/(J-1)}{\text{RSS}/(N-J)} = \frac{(\sum_{j=1}^J n_j (\bar{y}_{j.} - \bar{y}_{..})^2)/(J-1)}{(\sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_{j.})^2)/(N-J)}.$$

Example 3.5: One-way ANOVA with J groups

- The F -statistic has an exact F -distribution under the null hypothesis *if* the errors are normally distributed.
- If normality does not hold, the F -distribution is approximate for large samples.

Example 3.5: One-way ANOVA with J groups

- Why can the quantity $RSS_H - RSS$ be written in this format?
 - ▶ Note first that $RSS_H = \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_{..})^2$.
 - ▶ Then, take note of the following very useful decomposition:

$$\sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_{..})^2 = \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_{j.})^2 + \sum_{j=1}^J n_j (\bar{y}_{j.} - \bar{y}_{..})^2$$

$$SS_T = SS_W + SS_B.$$

- The total sum of squares can be written as the within-group sum of squares plus the between-group sum of squares.
- The residual sum of squares is none other than the within-group sum of squares.

ONE-WAY ANOVA

Example 3.5: One-way ANOVA with J groups

- Suppose $n_1 = \dots = n_J =: n$, so that $N = n \times J$.
- An ANOVA table typically would have the following information:

Source of variation	df	Sum of squares	Mean sum of squares
Between-group	$J - 1$	SS_B	$SS_B/(J - 1)$
Within-group	$N - J = J(n - 1)$	SS_W	$SS_W/(N - J)$
Total	$N - 1 = nJ - 1$	SS_T	$SS_T/(N - 1)$

- With this, we can see more readily why this is called ANOVA!
- You can think of “between-group” as marking variability explained by the predictor of interest and “within group” as marking everything else (residual).

TABLE OF CONTENTS

- 1 Motivation for hypothesis testing
- 2 Testable hypotheses and key lemmas
- 3 The F -statistic and its distribution under the null
- 4 The distribution of the F -statistic under the alternative
- 5 Example: One-way ANOVA
- 6 Large-sample theory
- 7 More general form for the null hypothesis
- 8 ANOVA: A little more formally
- 9 One-way ANOVA
- 10 Two-way ANOVA**
- 11 Coefficient of determination

Factorial design:

- Suppose we want to compare the effects of J chemotherapy agents together with K different dosage levels. In total, there are $J \times K$ combinations of levels. A total of n_{jk} patients are assigned to each group, with $N = \sum_{1 \leq j \leq J; 1 \leq k \leq K} n_{jk}$.
- ANOVA model: $Y_{ijk} = \mu + \alpha_{jk} + \epsilon_{ijk}$.
 - ▶ What does the design matrix look like?
 - ▶ What is the interpretation of each of the following hypotheses?
 - ★ $H_0 : \alpha_{jk} = 0$ for all $1 \leq j \leq J, 1 \leq k \leq K$.
 - ★ $H_0 : \alpha_{11} = \alpha_{12} = \cdots = \alpha_{1K}$.
 - ★ $H_0 : \alpha_{11} = \alpha_{12} = \cdots = \alpha_{1K}; \cdots ; \alpha_{J1} = \alpha_{J2} = \cdots = \alpha_{JK}$.
 - ★ $H_0 : \alpha_{jk} = \alpha_j + \gamma_k$.

TWO-WAY ANOVA

Two-way ANOVA:

- Suppose $n_{jk} = n$, so that $N = n \times J \times K$.
- A two-way ANOVA table typically would have the following information:

Source of variation	df	Sum of squares	Mean sum of squares
Factor X	$J - 1$	SS_B^X	$SS_B^X / (J - 1)$
Factor Z	$K - 1$	SS_B^Z	$SS_B^Z / (K - 1)$
Interaction XZ	$(J - 1)(K - 1)$	SS_B^{XZ}	$SS_B^{XZ} / ((J - 1)(K - 1))$
Error	$JK(n - 1)$	SS_W	$SS_W / (JK(n - 1))$
Total	$N - 1$	SS_T	$SS_T / (N - 1)$

Example 3.6: Doxorubicin and serum

- Chemotherapy agents are generally preliminarily evaluated in a laboratory setting for activity.
- A sample drawn from a liquid culture of some cell line is exposed to a new drug or combinations of new drugs at varying concentrations.
- The cells are then put in a solid culture medium and incubated. The resulting colonies of cells can be counted using an optical scanner.
- The number of colonies visible at the end of incubation are a fair representation of the number of cells not killed by the treatment.
- As an example, consider a laboratory experiment of doxorubicin (concentrations of 0.05, 0.10, 0.50, and 1.00 $\mu\text{mol/L}$). Further, half the samples are exposed to 10% calf serum.
- Each of the eight defined groups has a sample size of $n = 24$.

EXAMPLE: ONE-WAY ANOVA

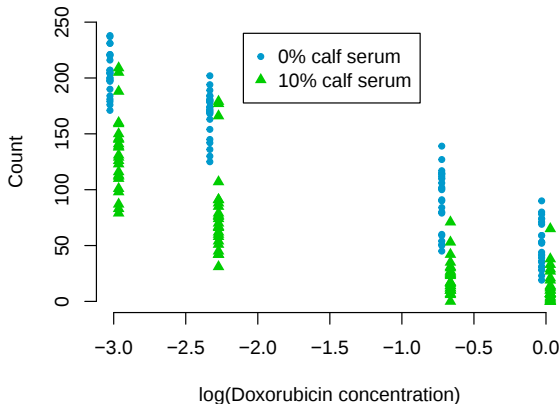


Figure 3.4: Doxorubicin, serum type, and count.



TWO-WAY ANOVA

Example 3.6: Doxorubicin and serum

```
1 ## Run ANOVA
2 zz <- aov(count ~ factor(doxconc) * factor(serum), data = dat)
3
4 > summary(zz)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
factor(doxconc)	3	558897	186299	285.870	< 2e-16	***
factor(serum)	1	214067	214067	328.480	< 2e-16	***
factor(doxconc):factor(serum)	3	16039	5346	8.204	3.74e-05	***
Residuals	184	119911	652			

```
10 ---
```

- How do we interpret each p-value?

Concluding thoughts on ANOVA:

- There is a lot more that can be said about the ANOVA framework.
- The “balanced” case results in particularly nice properties.
- Special considerations are necessary for unbalanced cases.
- The assumption of equal variances is dreadful, and ANOVA doesn't provide a way out of it as easily as regression does.
- Mixed models/random effects models fall under the category of “ANCOVA” (**an**alysis of **cov**ariance). This falls under the umbrella of BIOS 7346/8346.
- I am biased toward the more flexible regression framework. However, ANOVA remains the approach of choice in some fields and you should familiarize yourself with the essential ideas.

TABLE OF CONTENTS

- 1 Motivation for hypothesis testing
- 2 Testable hypotheses and key lemmas
- 3 The F -statistic and its distribution under the null
- 4 The distribution of the F -statistic under the alternative
- 5 Example: One-way ANOVA
- 6 Large-sample theory
- 7 More general form for the null hypothesis
- 8 ANOVA: A little more formally
- 9 One-way ANOVA
- 10 Two-way ANOVA
- 11 Coefficient of determination**

Decomposing variability:

- Another topic that has been notably absent so far is the coefficient of determination (and the related correlation coefficient).
- It's easiest to motivate some of these measures of association once we've wrapped our minds around sources of variation.

COEFFICIENT OF DETERMINATION

Important decomposition:

- Consider the regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $E[\boldsymbol{\epsilon}] = \mathbf{0}$ and $\text{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{I}$. Assume that $\mathbf{X}$ is of full rank.
- Let $\hat{y}$ denote the fitted value based on $\hat{\boldsymbol{\beta}}$, the OLS estimate of $\boldsymbol{\beta}$.
- Then,

$$\begin{aligned} SS_T &= \sum_{i=1}^N (y_i - \bar{y})^2 \stackrel{\text{math}}{=} \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \\ &= SS_R + SS_M. \end{aligned}$$

- SS_T : total sum of squares.
 - ▶ Marks variability of Y about $\bar{Y}$.
- SS_R : residual sum of squares (akin to “within”).
 - ▶ Marks variability of Y about $\hat{Y}$.
- SS_M : model sum of squares (akin to “between”).
 - ▶ Marks variability of $\hat{Y}$ about $\bar{Y}$.

Additional decomposition:

- Note: $SS_R = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{y}^\top\mathbf{y} - \hat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\mathbf{y}$. Why?
- With algebra, $SS_R = \sum_{i=1}^N (y_i - \bar{y})^2 - \hat{\boldsymbol{\beta}}_1^\top (\mathbf{X}^C)^\top \mathbf{y}$, where $\mathbf{X}^C$ denotes the design matrix with predictors centered about respective means (omit column of ones for the intercept) and $\hat{\boldsymbol{\beta}}_1$ denotes coefficients of $\hat{\boldsymbol{\beta}}$ without the intercept.
 - ▶ Hint 1: $SS_T = \sum_{i=1}^N (y_i - \bar{y})^2 = \mathbf{y}^\top\mathbf{y} - N\bar{y}^2$.
 - ▶ Hint 2: Centering X 's alters $\hat{\beta}_0$ (updated to $\bar{y}$) and the remaining estimates are unaffected.
 - ▶ Hint 3: $\hat{\boldsymbol{\beta}}_1^\top (\mathbf{X}^C)^\top \mathbf{y} = \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \mathbf{y} - N\bar{y}^2$.
- Taken together, this implies that $SS_M = \hat{\boldsymbol{\beta}}_1^\top (\mathbf{X}^C)^\top \mathbf{y}$.
- Now, since $\hat{\boldsymbol{\beta}}$ solves the appropriate normal equations, we further find that $SS_M = \hat{\boldsymbol{\beta}}_1^\top [(\mathbf{X}^C)^\top \mathbf{X}^C] \hat{\boldsymbol{\beta}}_1$.

COEFFICIENT OF DETERMINATION

Interpretation:

- The coefficient of determination is given as:

$$R^2 = \frac{SS_T - SS_R}{SS_T} = \frac{SS_M}{SS_T} = \frac{\hat{\beta}_1^T [(\mathbf{X}^C)^T \mathbf{X}^C] \hat{\beta}_1}{\sum_{i=1}^N (y_i - \bar{y})^2}.$$

- Marks the proportion of “total variation” in the outcome explained by the overall model.
- Sometimes described as a measure of model fit, although I find this ambiguous. I want to avoid conflating concepts of correct model specification and a model's predictive capacity.
- Now, SS_T can also be partitioned as:

$$\begin{aligned} SS_T &= \sum_{i=1}^N (y_i - \bar{y})^2 = \mathbf{y}^T \mathbf{y} - N\bar{y}^2 = (\mathbf{y}^T \mathbf{y} - \hat{\beta} \mathbf{X}^T \mathbf{y}) + (\hat{\beta} \mathbf{X}^T \mathbf{y} - N\bar{y}^2) \\ &= SS_R + SS_M. \end{aligned}$$

Properties:

- ① $0 \leq R^2 \leq 1$.
- ② $R = \sqrt{R^2}$ marks the correlation between Y and $\hat{Y}$.
- ③ Adding variables to a model cannot decrease the value of R^2 .

Distribution under the null:

- Consider the hypothesis $H_0 : \beta_1 = \dots = \beta_{K-1} = 0$ (i.e., the predictors are not linearly associated with the outcome).
- The F -statistic associated with this hypothesis can be written as:

$$F = \frac{R^2/(K-1)}{(1-R^2)/(N-K)}.$$

- With a little algebra, we can rearrange, finding:

$$R^2 = \frac{(K-1)F}{(N-K) + (K-1)F} = \frac{\left(\frac{K-1}{N-K}\right) F}{1 + \left(\frac{K-1}{N-K}\right) F}.$$

COEFFICIENT OF DETERMINATION

Distribution under the null: $H_0 : \beta_1 = \cdots = \beta_{K-1} = 0$

- If we are willing to assume that $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, then we know that $F \sim F_{K-1, N-K}$ under H_0 , from which we learn that:

$$R^2 = \frac{\left(\frac{K-1}{N-K}\right) F}{1 + \left(\frac{K-1}{N-K}\right) F} \sim \text{Beta}\left(\frac{K-1}{2}, \frac{N-K}{2}\right).$$

- From this, we also learn that (under H_0):

$$E[R^2] = \frac{K-1}{N-1}.$$

- This is the motivation for the adjusted R^2 , which will have expectation zero under the null.

Adjusted R^2 :

- The adjusted R^2 penalizes the number of parameters in the model:

$$R_{\text{adj}}^2 = 1 - (1 - R^2) \frac{N - 1}{N - K}.$$

- In so doing, we have that (under H_0),

$$E[R_{\text{adj}}^2] = 0.$$

- Trade-off: The adjusted R^2 can take on a negative value.

Coefficient of partial determination: Partial R^2

- Suppose we seek to understand the proportion of variation not explained in a reduced model that is explained by predictors in a full(er) model.
- The partial R^2 is given by:

$$R^2_{\text{Full}|\text{Reduced}} = \frac{SS_R^{\text{Reduced}} - SS_R^{\text{Full}}}{SS_R^{\text{Reduced}}}.$$

- Here, SS_R^{Reduced} is taking the place of SS_T , which you can think of as being based on a model that has been reduced to the maximal extent.

Relationship to more general F -test:

- Similarly, suppose we seek to test (intercept *not* involved):

$$H_0 : \begin{bmatrix} 0 & \mathbf{C}_{Q \times (K-1)} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta}_{-0} \end{bmatrix} = \mathbf{0}.$$

- Then, the F -statistic can be expressed as:

$$F = \frac{(R^2 - R_H^2)/Q}{(1 - R^2)/(N - K)},$$

where R_H^2 is the coefficient of determination under the restricted model. You should be able to verify that this is the more general version of what we just saw.

- Keep in mind that we are assuming $\text{rank}(\mathbf{C}) = Q$ and $\mathbf{X}$ is of full rank.

This unit:

- Testable hypotheses.
- The F -statistic.
- ANOVA.
- The coefficient of determination.

SUMMARY: SO FAR

- Random vectors and matrices; multivariate normal theory.
- Ordinary least squares.
- Hypothesis testing and ANOVA.

SUMMARY: COMING UP

- Weighted least squares.
- Confidence regions and prediction.
- Diagnostics.
- 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.