ONE-FACTOR ANALYSIS: PART 3

Chapter 3-5, 8

LEARNING OBJECTIVES

- Explain F-test for ANOVA using expected values of msT and msE
- Explain difference between central and non-central F
- Explain connection between F and T-statistic

BREAD YEAST AND TIME TO RISE

- Compare three different rise times to see if there is a significant difference in dough height
  - 35, 40, and 45 minutes (quantitative or categorical?)
- Large batch of dough partitioned into *N* loaf pans of equal size
- Determine if any difference between treatment means
- If there is a difference, which are different?

TEST STATISTICS

- For t=2 used two-sample t-test
- With t > 2? Multiple two-sample tests?
  - t=3 there would be 3 such tests
  - t=4, we'd have 6
  - t=5.... 20
  - t=10…45
- Multiple hypothesis testing dramatically increases Type I error rate if we aren't careful
  - Each test has its own Type 1 error rate
- Goal: one test statistic that takes on certain values when the null is true and others when null is false

TESTING FOR TREATMENT DIFFERENCES NULL HYPOTHESIS

- Write this as a testable hypothesis in terms of our model parameters and construct test statistics
  - Must involve estimable functions!

<u>Cell Means</u>

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_t$$

PARTITIONING SOURCES OF VARIANCE

- If H_0 true, then variance estimator should be

$$\hat{\sigma}^2 = \frac{\sum_{i,j}(Y_{ij} - \bar{Y}_{..})^2}{N-1} = \frac{ssTotal}{N-1}$$

- If wrong then this will be an inflated estimator
- Partition ssTotal with identity

$$\sum_{i,j}(Y_{ij} - \bar{Y}_{i.} + \bar{Y}_{i.} - \bar{Y}_{..})^2$$

$$= \sum_{i,j}(Y_{ij} - \bar{Y}_{i.})^2 + \underbrace{\sum_i r_i(\bar{Y}_{i.} - \bar{Y}_{..})^2}_{\text{Source of inflation!}}$$

ANALYSIS OF VARIANCE
SSE AND SST

- $ssE = \sum_{i,j}(Y_{ij} - \bar{Y}_{i.})^2$ $\qquad ssT = \sum_i r_i(\bar{Y}_{i.} - \bar{Y}_{..})^2$

- Look at expected values

$$E(ssE) = (N - t)\sigma^2$$

$$E(ssT) = (t - 1)\sigma^2 + \sum_i r_i(\tau_i - \bar{\tau}_.)^2$$

- Comments about $\sum_i r_i(\tau_i - \bar{\tau}_.)^2$
  - $\bar{\tau}_. = \sum_i r_i \tau_i / N$
  - Equals 0 under null hypothesis

- Use these statistics to create our test statistic

ANALYSIS OF VARIANCE
MEAN SQUARES AND F-STATISTIC

- Convert ssE and ssT to mean sum-of-squares by dividing by their degrees-of-freedom (df)

$$msE = ssE/(N-t) \qquad E(msE) = \sigma^2$$

$$msT = ssT/(t-1) \qquad E(msT) = \sigma^2 + \frac{\sum_i r_i(\tau_i - \bar{\tau}_.)^2}{t-1}$$

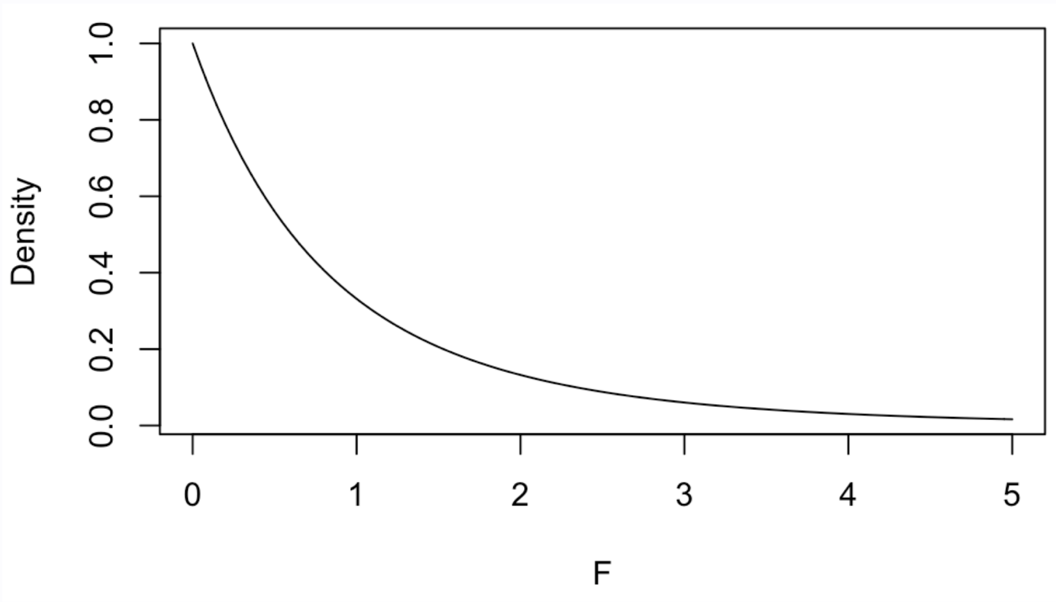
- If msT >> msE then we have evidence null is wrong
- Compare size of msT to msE with *F*-statistic

$$F = \frac{msT}{msE}$$

- Under H_0, follows central *F*-distribution with *t*-1 and *N*-t numerator and denominator df

F-DISTRIBUTION UNDER NULL DISTRIBUTION

- Bread example has t=3, r=4 (N=12)
- F-distribution with 2 num df and 9 denom df



- 0.05 critical value = 4.26
- P-value expression: $P(F_{2,9} > F_{obs})$ (one-sided test)

NON-CENTRAL F-DISTRIBUTION

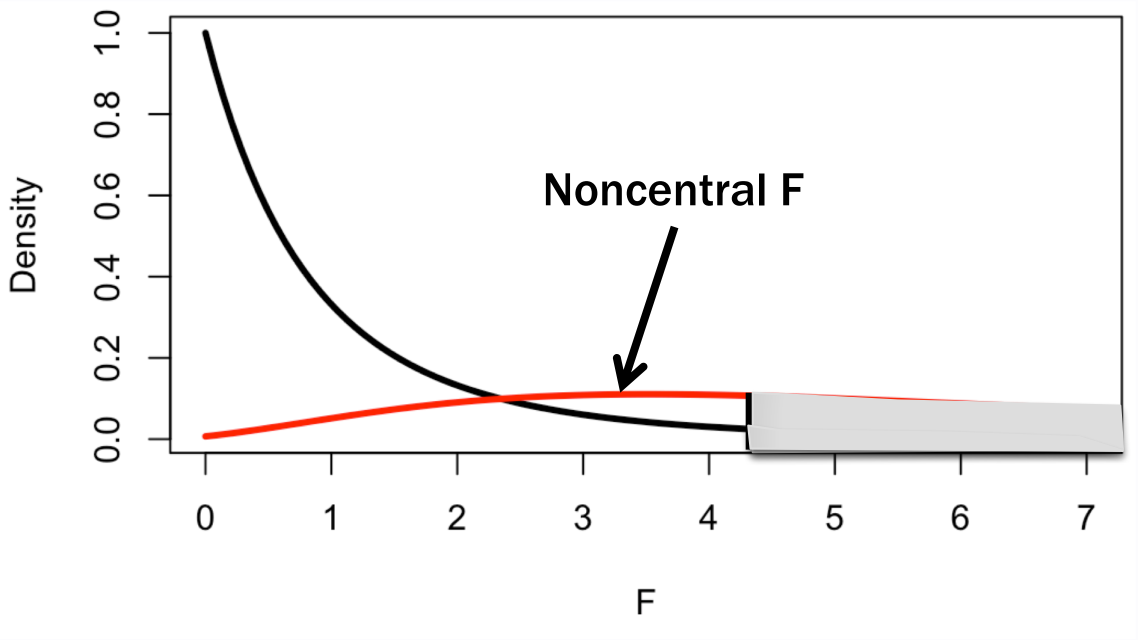
- *F* statistic follows central *F*-distribution under null
- If alternative is true then it follows non-central *F*-distribution with noncentrality parameter

$$\lambda = \frac{n}{\sigma^2} \sum_i \frac{r_i}{n} (\tau_i - \bar{\tau}.)^2 = \frac{n}{\sigma^2} \delta^2$$

- Shifts probability density to larger values, meaning larger *F* value are more likely
- Use to calculate power and sample size

NON-CENTRAL F-DISTRIBUTION

- Bread example has 2, 9 df
- Say $\lambda = 10$
- Critical value to reject at 0.05 level = 4.26



$P(F_{2,9} > 4.26) = 0.05$ $\quad$ $P(F_{2,9,10} > 4.26) = 0.660$

POWER CALCULATION

- To calculate power, we have to specify a single alternative hypothesis through λ
  - Replicate numbers
  - Treatment effect parameters**
  - Variance

$$\lambda = \frac{n}{\sigma^2}\sum_i \frac{r_i}{n}(\tau_i - \bar{\tau}.)^2 = \frac{n}{\sigma^2}\delta^2$$

- **Need to relax this unrealistic requirement

POWER
CONSERVATIVE APPROACH

- Unrealistic to give treatment effects, that's what we're trying to figure out!
- Instead, choose the smallest practically significant paired difference to detect, Δ, assume equal reps
- For bread example, perhaps you'd like to detect a difference of $\Delta = 3$ inches with high probability
- Fact: $\sum_i (\tau_i - \bar{\tau}_.)^2 \geq \frac{\Delta^2}{2}$
- Smallest λ under alternative is then $\frac{n}{\sigma^2}\frac{\Delta^2}{2t}$

POST-HOC CONTRAST ANALYSIS

- If we reject H_0, conclude treatment effects differ…but which ones?
- Post-hoc contrast analysis estimates differences
- Recall, contrast is a comparison of only the τ_i

$$\sum_{i=1}^{t} c_i \tau_i \qquad\qquad \sum_{i=1}^{t} c_i = 0$$

- $\sum_{i=1}^{t} c_i \tau_i = 0$ → some equality of treatment effects

- Three testing scenarios to consider:
  1. A single contrast
  2. Whether set of contrasts are simultaneously 0 or not
  3. One-by-one analysis of a set of contrasts (multiple testing)

SINGLE CONTRAST SAMPLING DISTRIBUTION

- For $\sum_i c_i \tau_i$ the sampling distribution is

$$\sum_i c_i \bar{Y}_{i.} \sim N\left(\sum_i c_i \tau_i,\ \sigma^2 \sum_i \frac{c_i^2}{r_i}\right)$$

- Don't know σ^2 so we use the T-statistic

$$T = \frac{\sum_i c_i \bar{Y}_{i.} - \sum_i c_i \tau_i}{\sqrt{msE \sum_i \frac{c_i^2}{r_i}}} \sim t_{N-t}$$

Standard error

Degrees-of-freedom from MSE

SINGLE CONTRAST HYPOTHESIS TESTING

- Don't know $\sum_i c_i\tau_i$, so we test it

$$H_0 : \sum_i c_i\tau_i = 0 \qquad\qquad H_A : \sum_i c_i\tau_i \neq 0$$

- If H_0 true then test statistic is

$$T = \frac{\sum_i c_i \bar{Y}_{i.}}{\sqrt{msE \sum_i \frac{c_i^2}{r_i}}} \sim t_{N-t}$$

- P-value (two-sided):

$$2 \times P(t_{N-t} > |T|)$$

TESTING WITH FULL AND REDUCED MODELS

- Null and alternative distributions implicitly assume all remaining contrasts are potentially nonzero
  - Didn’t have to think about this for two-sample problem

- Referred to as full vs reduced model approach
  - Full model: all contrasts are significant (model DF = t-1)
  - Reduced model: one contrast is insignificant (model DF = t-2)

- Important when simultaneously testing multiple contrasts and testing polynomial regression models
  - Also comes up with analysis of factorial experiments

SINGLE CONTRAST CONFIDENCE INTERVAL

- All a hypothesis tells us is what contrast isn't
- Common to use a confidence interval

$$\sum_i c_i \bar{Y}_{i.} \pm t_{N-t,\alpha/2} \times \sqrt{msE \sum_i \frac{c_i^2}{r_i}}$$

- Similar to the two-sample confidence interval

SINGLE CONTRAST SUM-OF-SQUARES

- Alternative test statistic for contrasts is based on the *F*-distribution with 1 numerator df

$$F = T^2 = \frac{\boxed{\left(\frac{(\sum_i c_i \bar{Y}_{i.})^2}{\sum_i c_i^2 / r_i}\right)}}{msE}$$

Referred to as contrast sum-of-squares (SS)

- Under the null hypothesis $\sum_i c_i \tau_i = 0$

$$E\left(\frac{(\sum_i c_i \bar{Y}_{i.})^2}{\sum_i c_i^2 / r_i}\right) = \sigma^2$$

- This test statistic follows central *F*-distribution with 1 and *N*-t degrees-of-freedom

ANOVA AND CONTRASTS

- Recall the null hypothesis for ANOVA

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_t$$

- Subtract τ_t from each part of equality and we get

$$H_0 : (\tau_1 - \tau_t) = (\tau_2 - \tau_t) = \cdots = 0$$

- **Conclusion:** ANOVA is simultaneously testing whether *t*-1 contrasts all equal 0

- **Full model:** all possible contrasts are significant
- **Reduced model:** no contrast significant

ANOVA AND CONTRASTS
LINEAR INDEPENDENCE

- Many equivalent expressions for null hypothesis but all of them involve *t*-1 linearly independent contrasts

- Concept from linear algebra but we can get the gist with an example
- Have three contrasts with coefficients c_i, d_i, e_i in bread example
- Write contrast coefficients as a list (or vector)

$$c = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad d = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \qquad e = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

ANOVA AND CONTRASTS
LINEAR INDEPENDENCE

- The last vector can be derived from the first two

$$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

$$\boxed{d - c = e}$$

- Called a linearly dependent set of contrasts
  - Can write one vector as a linear combination of others
- Set of contrasts is linearly independent if no contrast coefficient vector can be written as a linear combination of the others

TESTING MULTIPLE CONTRASTS

- Fact 1: Largest set of linearly independent contrasts has t-1 contrasts
- Fact 2: ANOVA test is equivalent to simultaneous test of t-1 linearly independent contrasts
- Takeaway: We can use F-tests to simultaneously test a set of linearly independent contrasts
- Works for linearly independent sets with less than t-1 contrasts as well!

GENERAL LINEAR HYPOTHESIS TEST

- Have $k < t$-1 linearly independent contrasts and want to test whether all simultaneously equal 0
  - Defines the reduced model
- Software generates required *F*-statistic and calculates *p*-value based on *F*-distribution with k and N-t degrees-of-freedom
- *F*-statistic in general requires matrix computations but can be calculated by hand in special situations

ONE-AT-A-TIME TESTING OF CONTRASTS
MULTIPLE COMPARISONS

- Even with these general contrast hypotheses, still left without knowledge of which individual effects are truly different from 0

- Naïve Approach: Confidence intervals for each

- Issue: Proposed confidence level only applies to a single interval

- We want to be confident that all proposed intervals simultaneously have the proposed confidence level

MULTIPLE COMPARISONS

- Suppose we want $k>1$ confidence intervals
- Then the probability of all confidence intervals capturing the true contrasts simultaneously could be as low as

$$1 - k\alpha$$

- Example: $k=5$ and $\alpha = 0.05$ would give lower bound of 0.75
- Idea: make intervals larger to prevent this…
- …can't make them too large or they won't be useful

MULTIPLE COMPARISON ADJUSTMENTS

- Tukey's HSD (Honestly significant difference):
  - Only works for paired differences
  - Adjusts critical value to be from studentized range distribution

- Dunnett Adjustment:
  - Specific pairwise comparisons to reference treatment
  - Usually control or the "best" treatment

POLYNOMIAL REGRESSION INFERENCE: OVERALL MODEL

- Highest-order polynomial = t-1
  - t=2 levels means simple linear regression
  - t=3 levels means quadratic regression
- Start with overall model test

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0 \qquad H_A: at\ least\ one\ different$$

- **Reduced (null) model:** intercept only
- **Full model:** all parameters included
- Just like with post-hoc contrast testing, want to know which polynomial terms we can drop

BACKWARDS ELIMINATION

- Avoid aliasing issue by starting with high-order polynomial model and test largest order coefficient

$$H_0: \beta_p = 0 \quad H_A: \beta_p \neq 0$$

- Reduced (null) model: polynomial of order *p*-1
- Full model: polynomial of order *p*

- Can test with either *t*-test or *F*-test (equivalent)
- Continue removing terms until largest coefficient becomes significant

LEARNING OBJECTIVES REVIEW

- Explain F-test for ANOVA using expected values of msT and msE
- Explain difference between central and non-central F
- Explain connection between F and T-statistic