

Analysis of variance methods for a one-factor completely randomized design

STA305 Spring 2014

See last slide for copyright information

Optional Background Reading

- Chapter 3 of *Data analysis with SAS*
- *Photocopy 1* from an old textbook: See link on course website.

Topics

- Dummy variable coding schemes
- Contrasts
- Multiple comparisons

Indicator dummy variables with intercept

- $x_1 = 1$ if Drug A, Zero otherwise
- $x_2 = 1$ if Drug B, Zero otherwise
- $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$

Group	x_1	x_2	$\beta_0 + \beta_1 x_1 + \beta_2 x_2$
A	1	0	$\mu_1 = \beta_0 + \beta_1$
B	0	1	$\mu_2 = \beta_0 + \beta_2$
Placebo	0	0	$\mu_3 = \beta_0$

$$\beta_1 = \Delta_1 \text{ and } \beta_2 = \Delta_2$$

Recall the interpretation

- β_0 is the expected response if all members of the population had been exposed to the control condition.
- $\beta_1 = \Delta_1$ is the constant that is added to the response by Treatment 1.
- (Assumption of unit-treatment additivity.)
- $\beta_2 = \Delta_2$ is the constant that is added to the response by Treatment 2.
- β_0 , β_1 and β_2 are 1-1 with μ_1 , μ_2 and μ_3 .
- Use routine regression methods for estimation and testing.

Just one treatment group

Let Y_1, \dots, Y_n be a random sample (i.i.d.) from a $N(\mu, \sigma^2)$ distribution.
Then

- $\sum_{i=1}^n (Y_i - \bar{Y})^2$ and \bar{Y} are independent.
- $\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sigma^2} \sim \chi^2(n - 1)$

To show this, write as a regression model.

- What is $\hat{\beta}$?
- What is \hat{Y} ?
- What is SSE?

Cell means coding: An indicator for each treatment, and no intercept

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$$

Drug	x_1	x_2	x_3	$\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
A	1	0	0	$\mu_1 = \beta_1$
B	0	1	0	$\mu_2 = \beta_2$
Placebo	0	0	1	$\mu_3 = \beta_3$

This model is equivalent to the one with 3 indicators and an intercept.

Cell means coding can be very convenient

- β values are treatment means (expected values).
- $\hat{\beta}$ values are treatment *sample* means.
 - What is the \mathbf{X} matrix?
 - What is $\mathbf{X}'\mathbf{X}$?
 - What is $(\mathbf{X}'\mathbf{X})^{-1}$?
 - What is $\mathbf{X}'\mathbf{Y}$?
 - What is $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$?
 - What is the distribution of $\hat{\beta}$?
- More distribution theory is readily available. Just use regression results.

Re-write the model

Independently for $j = 1, \dots, p$ and $i = 1, \dots, n_j$,
 $Y_{ij} \sim N(\mu_j, \sigma^2)$

- $SST = SSR + SSE$
- $SST = \sum_{j=1}^p \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y})^2$
- $SSR = \sum_{j=1}^p n_j (\bar{Y}_j - \bar{Y})^2$
- $SSE = \sum_{j=1}^p \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2$

Test $H_0: \mu_1 = \dots = \mu_p$

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Treatments	$p - 1$	SSR	$MSR = SSR/(p - 1)$	$F = \frac{MSR}{MSE}$	p -value
Error	$n - p$	SSE	$MSE = SSE/(n - p)$		
Total	$n - 1$	SST			

$$SST = SSR + SSE$$

$$SST = \sum_{j=1}^p \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y})^2$$

$$SSR = \sum_{j=1}^p n_j (\bar{Y}_j - \bar{Y})^2$$

$$SSE = \sum_{j=1}^p \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2$$

Effect coding

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

Group	x_1	x_2	$\beta_0 + \beta_1 x_1 + \beta_2 x_2$
1	1	0	$\mu_1 = \beta_0 + \beta_1$
2	0	1	$\mu_2 = \beta_0 + \beta_2$
3	-1	-1	$\mu_3 = \beta_0 - \beta_1 - \beta_2$

- Just like indicator coding with intercept, except last category gets -1
- β_0 is the grand mean.
- Other β_j are deviations from the grand mean.

All the dummy variable coding schemes are equivalent

- β vectors are connected by 1-1 (and onto) functions.
- $\hat{\beta}$ vectors are connected by those same functions.
- Follows from the *Invariance principle of maximum likelihood estimation*,
- Which basically says that the MLE of such a function is that function of the MLE.

Common ways to state the models

(Note Y_{ij} are observed data)

- Cell means model: $Y_{ij} = \mu_j + \varepsilon_{ij}$
- Effects model: $Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$
where $\tau_1 + \dots + \tau_p = 0$

The effects model is very popular, even when presenting randomization tests. Everything is relative.

Contrasts

$$c = a_1\mu_1 + a_2\mu_2 + \cdots + a_p\mu_p$$

$$\hat{c} = a_1\bar{Y}_1 + a_2\bar{Y}_2 + \cdots + a_p\bar{Y}_p$$

where $a_1 + a_2 + \cdots + a_p = 0$

Overall F-test is a test of $p-1$ contrasts

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

a_1	a_2	a_3	a_4
1	-1	0	0
0	1	-1	0
0	0	1	-1

$$c = a_1\mu_1 + a_2\mu_2 + \cdots + a_p\mu_p$$

Sample Question

Give a table showing the contrasts you would use to test

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

There is one row for each contrast.

a_1	a_2	a_3	a_4
1	-1	0	0
0	1	-1	0
0	0	1	-1

(This is a good format.)

In a one-factor design

- Mostly, what you want are tests of contrasts,
- Or collections of contrasts.
- You could do it with any dummy variable coding scheme.
- Cell means coding is often most convenient.
- With $\boldsymbol{\beta}=\boldsymbol{\mu}$, test $H_0: \mathbf{C}\boldsymbol{\beta}=\mathbf{t}$ using a general linear test.

- And you know how to get a confidence interval for any single contrast.

Orthogonal contrasts

Two contrasts $c_1 = \mathbf{a}'_1 \boldsymbol{\mu}$ and $c_2 = \mathbf{a}'_2 \boldsymbol{\mu}$ are said to be *orthogonal* if $\mathbf{a}'_1 \mathbf{a}_2 = \mathbf{0}$.

If the contrasts c_1 and c_2 are orthogonal and sample sizes are equal, then the estimated contrasts \hat{c}_1 and \hat{c}_2 have zero covariance, and are independent under normality.

Multiple Comparisons

- Most hypothesis tests are designed to be carried out in isolation
- But if you do a lot of tests and all the null hypotheses are true, the chance of rejecting at least one of them can be a lot more than α . This is **inflation of the Type I error probability**.
- Otherwise known as the curse of a thousand t-tests.
- Multiple comparisons (sometimes called follow-up tests, post hoc tests, probing) try to offer a solution.

Multiple Comparisons

- Protect a *family* of tests against Type I error at some *joint* significance level α
- If all the null hypotheses are true, the probability of rejecting at least one is no more than α

Multiple comparison tests of contrasts in a one-factor design

- Usual null hypothesis is $\mu_1 = \dots = \mu_p$.
- Usually do further tests after rejecting the initial null hypothesis with an ordinary F test.
- The big three are
 - Bonferroni
 - Tukey
 - Scheffé

Bonferroni

- Based on Bonferroni's inequality

$$Pr \left\{ \bigcup_{j=1}^k A_j \right\} \leq \sum_{j=1}^k Pr \{ A_j \}$$

- Applies to *any* collection of k tests
- Assume all k null hypotheses are true
- Event A_j is that null hypothesis j is rejected.
- Do the tests as usual
- Reject each H_0 if $p < 0.05/k$
- Or, adjust the p-values. Multiply them by k, and reject if $pk < 0.05$

Bonferroni

- Advantage: Flexible – Applies to *any* collection of hypothesis tests.
- Advantage: Easy to do.
- Disadvantage: Must know what all the tests are before seeing the data.
- Disadvantage: A little conservative; the true joint significance level is *less* than α .

Tukey (HSD)

- Based on the distribution of the largest mean minus the smallest.
- Applies only to pairwise comparisons of means.
- If sample sizes are equal, it's most powerful, period.
- If sample sizes are not equal, it's a bit conservative, meaning $P(\text{Reject at least one}) < \alpha$.

Statistical power

- Power is the probability of rejecting the null hypothesis when the null hypothesis is wrong.
- It is a function of the parameters, the sample size and the design.
- Power is good (by this definition).
- Sample size can be chosen to yield a desired power.
- More on this later.

Scheffé

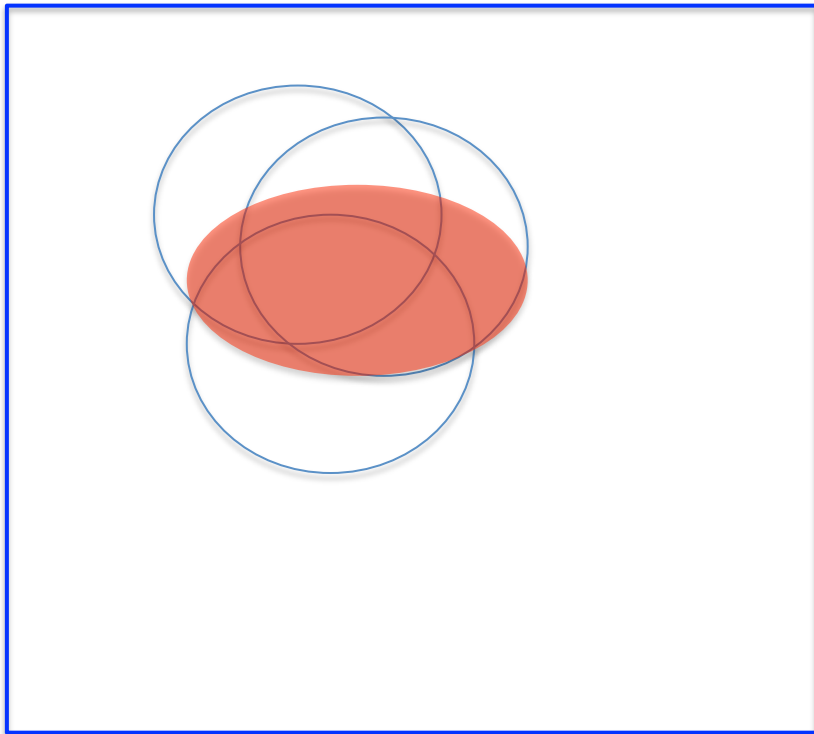
- Find the usual critical value for the initial test. Multiply by $p-1$. This is the Scheffé critical value for the test of a single contrast.
- Carry out the test as usual, comparing F to the Scheffé critical value.
- Family includes *all* contrasts: Infinitely many!
- You don't need to specify them in advance.
- Based on the union-intersection principle.

General principle of union-intersection multiple comparisons

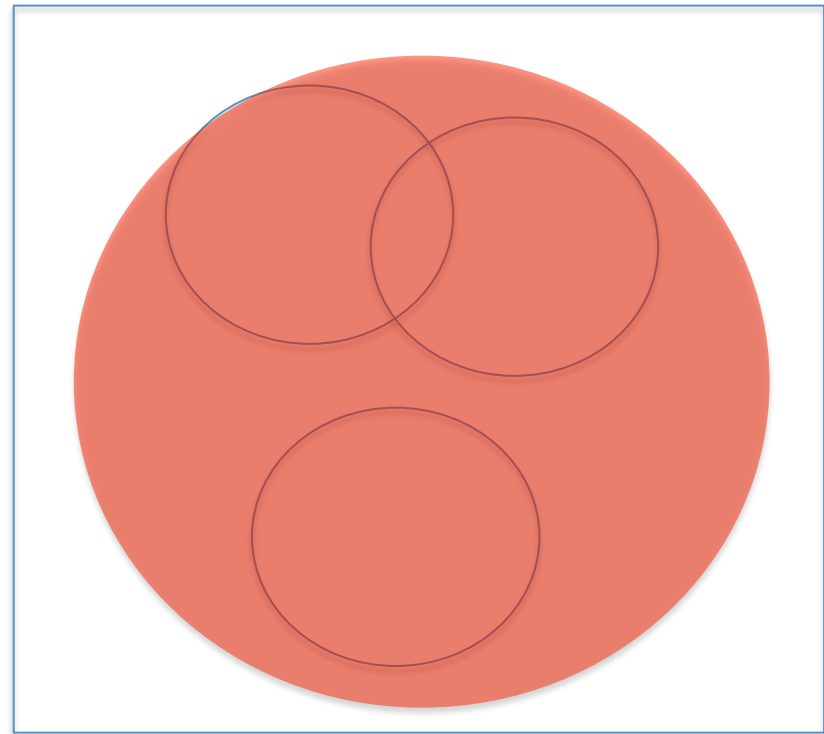
- The intersection of the null hypothesis regions of the tests in the family must be contained in the null hypothesis region of the overall (initial) test, so that if all the null hypotheses in the family are true, then the null hypothesis of the overall test is true.
- The union of critical regions of tests in the family must be contained in the critical region of the overall (initial) test, so if any test in the family rejects H_0 , then the overall test does too.
- In this case the probability that at least one test in the family will wrongly reject H_0 is $\leq \alpha$.

Intersection of null hypotheses regions contained in null hypothesis region.
Union of critical regions contained in critical region

Parameter Space



Sample Space



A very small example

- Consider a 2-sided test, say of $H_0: \beta_3=0$
- Reject if $t > t_{\alpha/2}$ or $t < -t_{\alpha/2}$
- If you reject H_0 , is there a formal basis for deciding whether $\beta_3 > 0$ or $\beta_3 < 0$?
- YES!

A family of 2 tests

- First do the initial 2-sided test of $H_0: \beta_3=0$.
- Reject if $|t| > t_{\alpha/2}$.
- If H_0 is rejected, follow up with 2 one-sided tests:
- One with $H_0: \beta_3 \leq 0$, reject if $t > t_{\alpha/2}$
- The other with $H_0: \beta_3 \geq 0$, reject if $t < -t_{\alpha/2}$
- H_0 will be rejected with one follow-up if and only if the initial test rejects H_0
- And you can draw a directional conclusion.
- **This argument is valuable because it allows you to use common sense.**
- This is a union-intersection family.

Scheffé are union-intersection tests

- Reject H_0 for follow-up test if $F_2 > f^*(p-1)$, where f is the critical value of the initial F test.
- Follow-up tests *cannot* reject H_0 if the initial F-test does not. Not quite true of Bonferroni and Tukey.
- If the initial test (of $p-1$ contrasts) rejects H_0 , there is a contrast for which the Scheffé test will reject H_0 (not necessarily a pairwise comparison).
- Adjusted p-value is the smallest α of the initial test that would make the Scheffé follow-up reject H_0 .
- It's also the tail area above $F_2/(p-1)$ under the null distribution of the initial test.

Which method should you use?

- If the sample sizes are nearly equal and you are only interested in pairwise comparisons, use Tukey because it's most powerful.
- If the sample sizes are not close to equal and you are only interested in pairwise comparisons, there is (amazingly) no harm in applying all three methods and picking the one that gives you the greatest number of significant results. (It's okay because this choice could be determined in advance based on number of treatments, α and the sample sizes.)

- If you are interested in follow-up tests that go beyond pairwise comparisons and you can specify *all* of them before seeing the data, Bonferroni is almost always more powerful than Scheffé. (Tukey is out.)
- If you want lots of special contrasts but you don't know in advance exactly what they all are, Scheffé is the only honest way to go, unless you have a separate replication data set.

How far should you take this?

- Protect all follow-ups to a given test?
- Protect all tests that use a given model?
- Protect all tests reported in a study?
- Protect all tests carried out in an investigator's lifetime?

We will be very modest. If we follow up a test whose null hypothesis has multiple equals signs, we will hold the joint significance level of the follow-up tests to 0.05 somehow.

Scheffé family also contains tests of *multiple* contrasts

- In regression with p regression coefficients, initial test is of $q \leq p$ linear constraints on β . Test statistic is F_1 . Reject if $F_1 > f$
- Follow-up test is test of $s < q$ constraints on β .
- Make sure the null hypothesis of the follow-up test follows logically from that of the initial test.
- Calculate F_2 , test statistic of the ordinary F-test of the follow-up null hypothesis.
- Scheffé test is to reject H_0 of follow-up test if $F_2 > q/s * f$.

Is it a union-intersection test?

- Reject H_0 of follow-up test if $F_2 > q/s * f$.
- Show this implies $F_1 > f$.

$$F_1 = \frac{SSR - SSR_1}{q MSE}$$

$$F_2 = \frac{SSR - SSR_2}{s MSE}$$

Copyright Information

This slide show was prepared by Jerry Brunner, Department of Statistics, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. These Powerpoint slides will be available from the course website:

<http://www.utstat.toronto.edu/brunner/oldclass/305s14>