Lecture 16: One-Way ANOVA

Reading Assignment:

- Muller and Fetterman, Chapter 13: "One-Way ANOVA"

We use analysis of variance (ANOVA) to answer questions like the following.

- Do two or more groups differ in mean response?
- Does the new drug reduce symptoms when compared to existing drugs?
- Does group membership differentially predict response?

Now we shift from considering continuous predictors to considering categorical predictors. ANOVA models are a type of GLM used with

categorical predictors.

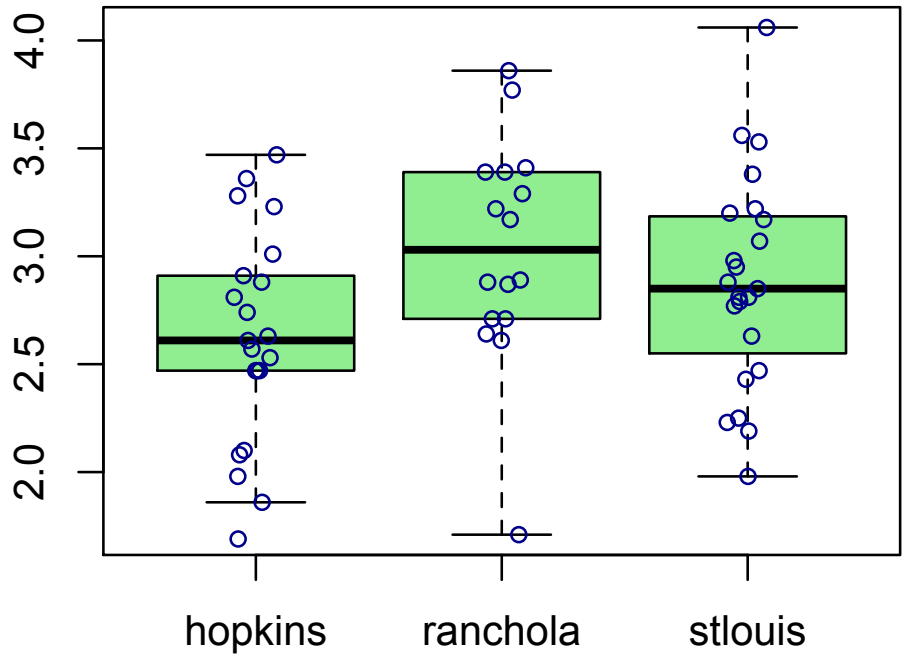
Specification of the Model

We want to determine whether two or more groups differ in location. The Gaussian errors assumption reduces this task to testing equality of means, generalizing Student's T test to three or more groups.

Example: Testing Equality of Cell Means

Pagano and Gauvreau (1993) present lung function data for heart patients at three medical centers: Johns Hopkins University, Rancho Los Amigos, and St. Louis University. For each center, investigators measured lung function as FEV_1, the forced expiratory volume in 1 second in liters.

Below is a plot of the FEV_1 for each subject by center.



The FEV_1 of patients across centers does differ, but there is also a lot of variation within center. While Rancho Los Amigos does have the highest lung function on average, some patients at Johns Hopkins have higher FEV_1 than some of the Rancho Los Amigos patients, for example.

We need a statistical test to determine whether this is a statistically significant difference. In ANOVA (as well as regression more generally), we ask questions about *means* of groups by analyzing their *variances*...how does this work?

GLM Assumptions in ANOVA

Assumption	ANOVA Interpretation
H	within cell / between cell
I	check sampling scheme
L	automatically OK, given design
E	no problem in practice
Gauss	within cell

Violation of the independence assumption may inflate α greatly. We must verify independence with careful thought about the sampling scheme, e.g., one individual may be sampled twice. To evaluate the assumption of Gaussian distribution of residuals, we may look at the Gaussian errors assumption within each group separately if the sample size in each group is large enough (≥ 30 say). Otherwise, we may look at studentized residuals pooled over all the groups.

The Primacy of Cell Means

Cell mean coding will provide our default choice, reflecting our primary interest in comparing means across groups. Recall that for cell mean coding with G groups, we create G indicator variables and do not include an intercept in the model, which is given by

$$\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times G}\,\boldsymbol{\beta}_{G\times 1} + \boldsymbol{\varepsilon}_{n\times 1}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_{n_1} \\ y_{n_1+1} \\ \vdots \\ y_{n_1+n_2} \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ 1 & 0 & \cdots & \cdots & \cdots & \vdots \\ 0 & 1 & 0 & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & & & \vdots \\ 0 & 1 & 0 & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \vdots \\ \vdots \\ \beta_{G-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

where $\mathbf{X}$ is full rank G. Matrix $\mathbf{X}$ can be summarized with the essence matrix, the matrix created by deleting any duplicate rows from the design matrix. The essence matrix helps us to determine rank and the

relationships between parameter definitions. What is the essence matrix of $\mathbf{X}$? (answer: $\text{Es}(\mathbf{X}) = \mathbf{I}$).

Often, double indices are used to denote y's by group as follows.

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{Gn_G} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & & & \vdots \\ 1 & 0 & \cdots & \cdots & \cdots & \vdots \\ 0 & 1 & 0 & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & & & \vdots \\ 0 & 1 & 0 & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \vdots \\ \vdots \\ \beta_{G-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Parameters in other codings are linear transformations of cell means.

Consider cell mean coding for G groups, with $n = \sum_{g=1}^{G} n_g$. Then

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} n_1 & & & 0 \\ & n_2 & & \\ & & \ddots & \\ 0 & & & n_G \end{bmatrix}_{G \times G}$$

$$\boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \sum_{n=1}^{n_2} y_{2j} \\ \vdots \\ \sum_{n=1}^{n_G} y_{Gj} \end{bmatrix}_{G \times 1}$$

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} \overline{y}_{1\cdot} \\ \overline{y}_{2\cdot} \\ \vdots \\ \overline{y}_{G\cdot} \end{bmatrix}_{G\times 1}$$

Note that in *cell mean* coding, our estimates are the sample means for each group! In other coding schemes, estimates are functions of cell means.

$$\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \mathbf{J}\overline{y}_{1\cdot} \\ \mathbf{J}\overline{y}_{2\cdot} \\ \vdots \\ \mathbf{J}\overline{y}_{G\cdot} \end{bmatrix}_{n\times 1} \qquad \widehat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \begin{bmatrix} y_{11} - \overline{y}_{1\cdot} \\ \vdots \\ y_{1n_1} - \overline{y}_{1\cdot} \\ y_{21} - \overline{y}_{2\cdot} \\ \vdots \\ y_{Gn_G} - \overline{y}_{G\cdot} \end{bmatrix}_{n\times 1}$$

$$\widehat{\sigma}^2 = \frac{\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}}{(n-G)} = \frac{SSE}{(n-G)} = \frac{\sum_{i=1}^{G}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_{i\cdot})^2}{n-G}$$

(Usual) Overall Test

We wish to test the hypothesis that the means are the same for each group. For three groups, we have

$$\begin{aligned} H_0: &\quad \beta_0 = \beta_1 = \beta_2 && \text{(cell mean)} \\ &\quad \Leftrightarrow \\ H_0: &\quad \beta_1 = \beta_2 = 0 && \text{(reference cell)} \\ &\quad \Leftrightarrow \\ H_0: &\quad \beta_1 = \beta_2 = 0 && \text{(effect)} \end{aligned}$$

The overall test involves equality of G means, which implies $G-1$ constraints so that $\boldsymbol{\theta}$ is $(G-1)\times 1$ and $a = G-1$.

We are testing H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

We state $\boldsymbol{C}$ (unique to coding scheme), use $\boldsymbol{\theta}_0 = \mathbf{0}$, and then compute $SSH = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \boldsymbol{M}^{-1} (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ and $f_{obs} = (SSH/a)/\widehat{\sigma}^2$.

Reference Cell Coding

Recall that for reference cell coding with three groups, we have the model

$$\mathbf{y} = \mathbf{X} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \boldsymbol{\varepsilon},$$

where $\text{Es}(\mathbf{X}) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$. To test the hypothesis of equal cell means, we have $\mathbf{C} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ implying that $\boldsymbol{\theta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$. We use $\boldsymbol{\theta}_0 = \mathbf{0}$.

Effect Coding

Recall that for effect coding with three groups, we have the model

$$\mathbf{y} = \mathbf{X} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \boldsymbol{\varepsilon},$$

where $\text{Es}(\mathbf{X}) = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$. To test the hypothesis of equal cell means, we have $\mathbf{C} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ implying that $\boldsymbol{\theta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$. We use $\boldsymbol{\theta}_0 = \mathbf{0}$.

If any $G - 1$ pairwise differences are zero, then all pairwise differences are zero, so the traditional overall test necessarily has $G - 1 = a$ degrees of freedom. All the above coding schemes give the same

p-value, F, and R^2 because they are full rank transformations of the rows; e.g.,$\boldsymbol{C}_1 = \boldsymbol{T}\boldsymbol{C}_2$.

Traditional Source Table

Traditional ANOVA Source Table

Source	$\boldsymbol{df}$	$\boldsymbol{SS}$	Mean Square	$\boldsymbol{F_{obs}}$	$\boldsymbol{p}$
Between	$G-1$		SSH/a		
Within (Error)	$n-G$		$\widehat{\sigma}^2$		
Total (Corrected)	$n-1$				

Estimating and Testing Cell Means in One-Way ANOVA

Using the essence matrix as the contrast matrix defines the cell means. For cell mean coding, the parameter estimates are the cell means. For other coding schemes, we use the parameter estimates provided to calculate the cell means.

Once you determine the cell means for each coding scheme, it is straightforward to test contrasts in the cell means.

Which Means are Different?

If we reject the null hypothesis that all means are equal, naturally we next want to know which means are in fact different.

There are $\binom{G}{2} = \frac{G(G-1)}{2}$ pairwise comparisons, and infinitely many others (if three or more means are involved).

An overall hypothesis that all means are equal tests all linear combinations of means and not just pairwise comparisons.

For the overall test with cell mean coding,

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

creates a matrix of secondary parameters which are mean differences. Each row of $\boldsymbol{C}$ defines a contrast matrix, a GLH, and an F test. Any

$\boldsymbol{C}$ with a single row creates an F test with one numerator df and hence corresponds to a T.

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \qquad H_0\colon \ \beta_0 = \beta_1$$
$$\boldsymbol{C} = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix} \qquad H_0\colon \ \beta_0 = \beta_2$$
$$\boldsymbol{C} = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix} \qquad H_0\colon \ \beta_1 = \beta_2$$

Contrasts

Describe a linear combination of cell means, $\sum_{i=1}^{G} c_i \mu_i$, as a **contrast** if $\sum_{i=1}^{G} c_i = 0$. Pairwise contrasts (pairwise comparisons) use coefficients such as $\boldsymbol{C} = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix}$ or $\boldsymbol{C} = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$.

Complex contrasts, such as $\boldsymbol{C} =$

$$\begin{bmatrix} -1 & \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

involve 3 or more means.

The null and alternative hypotheses in this case are

$$H_0 : \quad \sum_{i=1}^{G} c_i \mu_i = 0$$

$$H_A : \quad \sum_{i=1}^{G} c_i \mu_i \neq 0.$$

A planned (*a priori*) contrast requires choosing the number of contrasts and specifying the associated $\boldsymbol{C}$ matrices before seeing any data. Unplanned (*a posteriori*) contrasts include all others not specified in advance.

- For **balanced** data, a set of contrasts is **orthogonal** if $\boldsymbol{CC'}$ is diagonal (rows orthogonal).

Example: FEV_1 Data

The following SAS code creates the design matrix for cell mean coding and conducts the overall test of equality of group means for the lung function data.

```
data fev2;
set fev;
if center="hopkins" then h_ind=1; else h_ind=0;
if center="ranchola" then r_ind=1; else r_ind=0;
if center="stlouis" then s_ind=1; else s_ind=0;
run;

proc glm data=fev2;
model fev=h_ind r_ind s_ind/noint;
contrast Usual Overall Test h_ind 1 r_ind -1 s_ind  0,
                            h_ind 1 r_ind 0  s_ind -1;
run;
```

The results of the test are provided below.

```
                          The GLM Procedure

                       Dependent Variable: fev
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	482.5697439	160.8565813	633.19	<.0001
Error	57	14.4802561	0.2540396		
Uncorrected Total	60	497.0500000			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
h_ind	1	144.8344048	144.8344048	570.13	<.0001
r_ind	1	147.1369000	147.1369000	579.19	<.0001
s_ind	1	190.5984391	190.5984391	750.27	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
h_ind	1	144.8344048	144.8344048	570.13	<.0001
r_ind	1	147.1369000	147.1369000	579.19	<.0001
s_ind	1	190.5984391	190.5984391	750.27	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F

```
Usual Overall Test  2  1.58283723      0.79141861       3.12    0.0520
```

Parameter	Estimate	Standard Error	t Value	Pr > |t|
h_ind	2.626190476	0.10998692	23.88	<.0001
r_ind	3.032500000	0.12600585	24.07	<.0001
s_ind	2.878695652	0.10509614	27.39	<.0001

We see that there is a marginally significant difference among the cell means ($F = 3.12$ with 2 numerator and 57 denominator degrees of freedom yields $p = 0.05$). At least one center has average FEV that is different from the others.

What is the interpretation of the overall F test and the nine tests for the individual groups?

Step-Down Testing

Describe a test logically subsumed by another test as a **step-down test**. For example, having conducted an (overall) test for equality among a set of means, one may wish to step down to test equality of two particular means. In the population, the truth of the null hypothesis of the parent implies the truth of the null hypothesis of the child, and the truth of the alternative hypothesis for the child implies the truth of the alternative hypothesis for the parent.

Any linear combination of the rows of $\boldsymbol{C}$ represents a step-down test.

Properties of Cell Mean Estimates

If $\boldsymbol{C}$ and $\boldsymbol{X}$ are both orthogonal, then the contrast estimates $\widehat{\theta}_k$ are uncorrelated and statistically independent (under HILE Gauss). This occurs when we estimate cell means with cell mean coding.

There are at most $G-1$ mutually orthogonal contrasts (of mean differences) but infinitely many sets of contrasts are possible. There can be G mutually orthogonal contrasts if the set spans the grand mean.Recall that

$$\begin{aligned}\widehat{\boldsymbol{\theta}} &\sim \mathcal{N}_a[\boldsymbol{\theta},\, \sigma^2\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'] \\ &\sim \mathcal{N}_a[\boldsymbol{\theta},\, \sigma^2\boldsymbol{M} \qquad\qquad\quad]\end{aligned}$$

Each row of $\boldsymbol{C}$ defines a θ_k, with $\widehat{\theta}_k \sim \mathcal{N}(\theta_k, \sigma^2 m_{kk})$ and

$$\frac{\widehat{\theta}_k - \theta_0}{\sqrt{\widehat{\sigma}^2 m_{kk}}} \sim T\left(n-r\right),$$

where r is the rank of $\mathbf{X}$.

For cell mean coding the $\widehat{\theta}_k$'s are independent for orthogonal $\boldsymbol{C}$.

$\widehat{\theta}_k$ tests are never exactly independent (for finite n) because they all use $\widehat{\sigma}^2$.

We invert the test to create confidence intervals:

$$\widehat{\theta}_k \pm \ (\widehat{\sigma} \cdot \sqrt{m_{kk}}) \cdot t_{crit}$$

with $t_{crit} = F_T^{-1}(1 - \alpha/2; n - r)$

Conducting Multiple Comparisons

Choosing An Error Rate

The type I error rate for a collection of tests may be higher than α.

K independent tests with a nominal type I error rate of α lead to a probability of one or more false positives of $1-(1-\alpha)^K$. Using $\alpha = .05$ and $K = 10$ implies $1-(1-\alpha)^K \approx 0.40 >> 0.05$.

For dependent tests, the Bonferroni inequality provides an upper bound of $K\alpha$, which equals 0.50 in this example. For the Bonferroni approach, test each comparison at $\alpha_k = \alpha/K = 0.05/10 = 0.005$.

Overview of Methods for Multiple Comparisons

Only a few of many techniques for controlling the error rate will be highlighted here.

The Bonferroni approach is completely general for planned (*a priori*) comparisons.

The development of alternative methods arose from a desire for

1. increased statistical power,
2. the ability to control unplanned comparisons, or
3. increased robustness to violation of GLM assumptions.

For a small number of planned contrasts, use the Bonferroni correction.

With an overall test size of .01 and four contrasts, $\alpha_k = \alpha/K$ implies a nominal size for each test of $.01/4 = .0025$.

Some authors recommend not using any correction in the special case of a set of planned and orthogonal contrasts because a significant overall test implies a trend of some sort, and the step-down tests merely need to identify the order of the trend. This approach reduces the chances of having a significant overall test with no particular trend significant.

Dunnett's (1955, 1964) test was designed to test each of $G-1$ treatments against a control. Considering any other comparison with Dunnett's correction inflates test size.

Tukey's HSD (honest significant difference) method and generalizations of it (Tukey, 1953; Kramer, 1956) allow testing all pairwise comparisons.

Scheffé's (1953, 1959) test controls test size with any contrast, including complex contrasts (non-pairwise), whether planned or not. This test also protects against testing infinitely possible contrasts. It

has the desirable property of *coherence*. That is, the failure of the overall test ensures no significant step-down tests (e.g.,mean comparison tests) exist, while a significant overall test ensures at least one significant step-down test.

For a small number of comparisons, a Bonferroni correction suffices. For a larger number or for unplanned comparisons, Scheffé's test is preferred.

Example Formulas

In general, $\widehat{\boldsymbol{\theta}} \sim \mathcal{N}_a(\boldsymbol{\theta}, \sigma^2 \boldsymbol{M})$, with $\boldsymbol{M} = \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'$.

For a cell mean coding with G groups, $(\boldsymbol{X}'\boldsymbol{X})$ =Dg$(n_1, n_2, \ldots n_G)$, which implies $(\boldsymbol{X}'\boldsymbol{X})^{-}$ =Dg$(1/n_1, 1/n_2, \ldots 1/n_G)$. For $\boldsymbol{C}_k$ =row$_k(\boldsymbol{C})$ and *cell mean coding*,

$$m_{kk} = \boldsymbol{C}_k(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}_k' = \sum_{g=1}^{G} \frac{c_{kg}^2}{n_g}.$$

A Bonferroni correction for planned comparisons may be implemented with any linear models program merely by proper specification of the test size.

For other corrections, most of the work involves selecting the proper critical value based on determining the set of comparisons.

Scheffé's method for unplanned comparisons provides the most general and sturdy technique. Its downside is a lack of power.

Example: Testing Combinations of Means for FEV Data

We use the following SAS code to test all possible differences among cell means for the FEV data, using Scheffé's correction.

```
proc glm data=fev2;
class center;
model fev=center/noint;
lsmeans center/ pdiff adjust=scheffe;
run;
```

The LSMEANS statement produces the least squares estimates of CLASS variable means, where CLASS is used to indicate to SAS which variables are categorical. You may user the ORDER= option in the proc glm statement to alter how SAS chooses the reference level in proc glm (several options). In one-way ANOVA, these are just the usual means. In higher-way ANOVA models, the least squares means are simply the means of group means (thus if the design is not *balanced*, LSMEANS are not simple means).

This code yields the following additional output.

```
             Least Squares Means
  Adjustment for Multiple Comparisons: Scheffe

                                    LSMEAN
    center          fev LSMEAN      Number
   hopkins        2.62619048           1
   ranchola       3.03250000           2
   stlouis        2.87869565           3

     Least Squares Means for effect center
      Pr > |t| for H0: LSMean(i)=LSMean(j)

             Dependent Variable: fev

i/j              1             2             3
  1                         0.0603        0.2605
  2          0.0603                       0.6466
  3          0.2605        0.6466
```

We see some evidence that the mean FEV_1 at Rancho Los Amigos (3.03 liters) is higher than the mean FEV_1 at Johns Hopkins (2.63 liters), but the other means do not appear to differ significantly.

Trend Tests

Usually, orthogonal sets of tests are unappealing because $G-1$ tests do not cover all interesting mean comparisons.

For example, if the categorical predictor can be interpreted as an interval scale variable (e.g.,low, medium, or high dose), then the $G-1$ orthogonal trend tests are appealing.

Assume we are interested in one species of tree, grown at 200, 400, or 600m, with corresponding hypothesis H_0: $\mu_{200} = \mu_{400} = \mu_{600}$.
Having rejected H_0, what is trend across elevation?

First, we can plot the observed treatment means as a function of the factor levels. This plot may provide a general idea of any pattern that is present.

G levels imply $G-1$ trends can be tested. With equal spacing and equal cell sizes, we may use the orthogonal polynomial coefficients in

Table B-10. Without these conditions, tests are more complicated.

Using the contrast matrices described here with cell mean coding allows computing exactly the same tests as available in polynomial coding, which equal tests in polynomial regression.

For *cell mean* coding, consider

$$\boldsymbol{C}_{\mathsf{O}} = \begin{bmatrix} -1 & 0 & 1 \\ 1 & -2 & 1 \end{bmatrix} \begin{matrix} \text{linear} \\ \text{quadratic} \end{matrix}$$

This spans the matrix commonly used for the overall test, namely

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

in that each equals a full rank transformation of the rows of the other ($\boldsymbol{C}_{\mathsf{O}} = \boldsymbol{T}\boldsymbol{C}$, with $\boldsymbol{T}$ a full rank matrix of dimension $a \times a$).

The GLH p-value and test statistic do not change with such a transformation.

Plotting the elements of a row of $\boldsymbol{C}_{\mathsf{O}}$ against the column $\#$ displays the shape indicated by the labeling.

Separate one df tests provide a set of $G-1$ orthogonal trend tests.

It seems reasonable to ignore any Bonferroni or other sort of correction in this setting if the trend tests represent the only planned comparisons.

Homogeneity

There are only G distinct groups, so comparing variances for G cells is useful for assessing homogeneity of variances. The Hartley's test involves computing the ratio of the largest group variance, $\max(s_j^2)$ to the smallest group variance, $\min(s_j^2)$. The resulting ratio, $F_{\max}$, is then compared to a critical value from a table of the sampling distribution of $F_{\max}$. If the computed ratio is less than the critical value, the groups are assumed to have similar or equal variances.

Bartlett's test is used to test the null hypothesis, H_0 that all k population variances are equal against the alternative that at least two are different. It is a modification of the normal-theory likelihood ratio test. Bartlett's test is sensitive to departures from normality. That is, if the samples come from non-normal distributions, then Bartlett's test may simply be testing for non-normality.

Newer tests (the Brown and Forsythe test or O'Brien test)(HOVTEST=BF or HOVTEST=OBRIEN in PROC GLM) are much more robust to the underlying distribution. They transform the original values of the dependent variable to derive a dispersion variable and then to perform analysis of variance on this variable.

The significance level for the test of homogeneity of variance is the p-value for the ANOVA F-test on the dispersion variable. All of the homogeneity of variance tests available in PROC GLM except Bartlett's use this approach.

If one of these tests rejects the assumption of homogeneity of variance, you should use Welch's ANOVA (which is a very messy formula. I certainly would never expect anyone to memorize this; but $G = 2$, this test is equivalent to Welch's t-test: $\frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$) instead of the usual ANOVA to test for differences between group means.

```
proc glm data=fev2;
class center;
model fev=center/noint;
means center/hovtest=bf welch;
run;
```

The GLM Procedure

Brown and Forsythes Test for Homogeneity of fev Variance

ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
center	2	0.00818	0.00409	0.04	0.9586
Error	57	5.5105	0.0967		

Welchs ANOVA for fev

Source	DF	F Value	Pr > F
center	2.0000	3.02	0.0614
Error	35.5447		

How do we do this in R?

```
> library(car)
> FEV = read.table("FEV2.dat", header = T)
> head(FEV)
```

```
   center  fev
1 hopkins 3.23
2 hopkins 3.47
3 hopkins 1.86
4 hopkins 2.47
5 hopkins 3.01
6 hopkins 1.69
```

```
> table(FEV$center)
```

```
 hopkins ranchola  stlouis
      21       16       23
```

```
> summary(FEV$fev)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.690   2.515   2.830   2.831   3.220   4.060
```

```
> # cell means predictor matrix
> X_cell = model.matrix(~ factor(center)-1, data = FEV)
```

```
> head(X_cell, 3)

  factor(center)hopkins factor(center)ranchola factor(center)stlouis
1                     1                       0                     0
2                     1                       0                     0
3                     1                       0                     0

> tail(X_cell, 3)

   factor(center)hopkins factor(center)ranchola factor(center)stlouis
58                     0                       0                     1
59                     0                       0                     1
60                     0                       0                     1

> # factor() tells R center is categorical (factor variable),
> # and -1 suppresses the intercept
```

We can specify this cell means coding directly in lm()

```
> fit = lm(fev ~ factor(center)-1, data = FEV)
> summary(fit) # similar to SAS

Call:
lm(formula = fev ~ factor(center) - 1, data = FEV)

Residuals:
     Min       1Q   Median       3Q      Max
-1.32250 -0.32250 -0.02244  0.32630  1.18130

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
factor(center)hopkins    2.6262     0.1100   23.88   <2e-16 ***
factor(center)ranchola   3.0325     0.1260   24.07   <2e-16 ***
factor(center)stlouis    2.8787     0.1051   27.39   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.504 on 57 degrees of freedom
Multiple R-squared:  0.9709,        Adjusted R-squared:  0.9693
F-statistic: 633.2 on 3 and 57 DF,  p-value: < 2.2e-16
```

Now lets check for HOV

```
> leveneTest(y=FEV$fev, group=FEV$center, center=median)

Levenes Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.0423 0.9586
      57

> # if reject the null, Welchs ANOVA
> oneway.test(fev ~ center,
+             data=FEV,
+             var.equal=FALSE)

        One-way analysis of means (not assuming equal variances)

data:  fev and center
F = 3.0211, num df = 2.000, denom df = 35.545, p-value = 0.06141
```

What if we wanted to perform the overall test, where we test for the equality of means? Specify contrast matrix for cell means coding

```
> # overall test
> C = matrix(c(1, -1, 0, 1, 0, -1), nrow = 2, byrow = T)
> print(C)
     [,1] [,2] [,3]
[1,]    1   -1    0
[2,]    1    0   -1
> linearHypothesis(fit, C)
Linear hypothesis test

Hypothesis:
factor(center)hopkins - factor(center)ranchola = 0
factor(center)hopkins - factor(center)stlouis = 0

Model 1: restricted model
Model 2: fev ~ factor(center) - 1

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     59 16.063
```

```
2     57 14.480  2    1.5828 3.1153  0.052 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Now peform pairwise comparison of means with Scheffe p-value correction

```
> library(DescTools)
> ScheffeTest(aov(fev ~ factor(center), data = FEV))

  Posthoc multiple comparisons of means : Scheffe Test
    95% family-wise confidence level

$factor(center)
                        diff      lwr.ci    upr.ci   pval
ranchola-hopkins  0.4063095 -0.01408874 0.8267078 0.0603 .
stlouis-hopkins   0.2525052 -0.12986364 0.6348740 0.2605
stlouis-ranchola -0.1538043 -0.56622280 0.2586141 0.6466

---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

What if we wanted to use reference cell coding?

```
> fit = lm(fev ~ factor(center), data = FEV)
> summary(fit)

Call:
lm(formula = fev ~ factor(center), data = FEV)

Residuals:
     Min       1Q   Median       3Q      Max
-1.32250 -0.32250 -0.02244  0.32630  1.18130

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)              2.6262     0.1100  23.877   <2e-16 ***
factor(center)ranchola   0.4063     0.1673   2.429   0.0183 *
factor(center)stlouis    0.2525     0.1521   1.660   0.1024
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.504 on 57 degrees of freedom
Multiple R-squared:  0.09854,        Adjusted R-squared:  0.06691
F-statistic: 3.115 on 2 and 57 DF,  p-value: 0.052
```

```
> # overall test for regression
> anova(fit)

Analysis of Variance Table


Response: fev
               Df  Sum Sq Mean Sq F value Pr(>F)
factor(center)  2  1.5828 0.79142  3.1153  0.052 .
Residuals      57 14.4803 0.25404
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Note that the overall corrected test for regression here is the same as the overall test in cell means encoding (utilizing the proper contrasts). The overall corrected test of regression in cell means coding is testing what exactly?

If you want to choose a different reference level, you can specify a different reference level in center using relevel()

```
> FEV$center = relevel(FEV$center, ref = "ranchola")
> fit = lm(fev ~ factor(center), data = FEV)
> summary(fit)

Call:
lm(formula = fev ~ factor(center), data = FEV)

Residuals:
     Min       1Q   Median       3Q      Max
-1.32250 -0.32250 -0.02244  0.32630  1.18130

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)             3.0325     0.1260  24.066   <2e-16 ***
factor(center)hopkins  -0.4063     0.1673  -2.429   0.0183 *
factor(center)stlouis  -0.1538     0.1641  -0.937   0.3525
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.504 on 57 degrees of freedom
```

```
Multiple R-squared:  0.09854,        Adjusted R-squared:  0.06691
F-statistic: 3.115 on 2 and 57 DF,  p-value: 0.052
```

Next: Two-way ANOVA

- Muller and Fetterman, Chapter 14: "Complete, Two-Way Factorial ANOVA"