STA 610L: Module 1.5

Review of one way ANOVA (step-down tests and matrix formulation)

Dr. Olanrewaju Michael Akande

STA 610L

WHERE IS THE DIFFERENCE?

After an *overall F test* is rejected, the next obvious questions is "Why?"

It could be the case that each group is different from all other groups, or maybe only one group differs from the other.

Many types of step-down tests are available with varying degrees of protection of the type I error rate.

STEP-DOWN TESTS

The phrasing step-down is a reminder that these tests are not conducted unless the overall F test is rejected. Some options include the following.

- The Bonferroni comparison using the significance level $\frac{\alpha}{\text{number of tests}}$ is generally the most conservative and is not guaranteed to find a signficant pairwise comparison even if the overall F test is rejected.

- Fisher's LSD test is generally the least conservative, as it involves regular pairwise t-tests with no correction to α, with the type I error rate protection in the sense of not carrying out the pairwise tests if the overall F test is not rejected.

- Many methods lie in between these extremes, with some additional protection of the type I error rate after rejection of the overall null hypothesis, including Tukey's method, Scheffe's method, ...

STEP-DOWN TESTS

```
TukeyHSD(aov.res)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = `passing distance` ~ kerb, data = PsychBikeData)
##
## $kerb
##                  diff         lwr         upr      p adj
## 0.5-0.25   -0.10758034 -0.16590883 -0.04925185 0.0000051
## 0.75-0.25  -0.19253456 -0.25993094 -0.12513817 0.0000000
## 1-0.25     -0.20746951 -0.26834834 -0.14659068 0.0000000
## 1.25-0.25  -0.28524048 -0.35310701 -0.21737394 0.0000000
## 0.75-0.5   -0.08495422 -0.15489916 -0.01500928 0.0082429
## 1-0.5      -0.09988917 -0.16357790 -0.03620045 0.0001874
## 1.25-0.5   -0.17766014 -0.24805821 -0.10726207 0.0000000
## 1-0.75     -0.01493495 -0.08702041  0.05715051 0.9799835
## 1.25-0.75  -0.09270592 -0.17078247 -0.01462937 0.0105846
## 1.25-1     -0.07777097 -0.15029619 -0.00524575 0.0284527
```

Because we're curious, we'll look for where the difference appears to be with Tukey's HSD test, which is basically a regular t-test that controls the family-wide error rate.

Looks like the passing distance is similar for curb distances of 0.75m and 1m but different for other combinations. Contrary to the hypothesis, the passing distance was greater when bikers rode closer to the curb.

STA 610L

ADVANTAGES AND DRAWBACKS OF CLASSICAL APPROACH

- This approach controls type I error rate (compared to testing group differences separately)

- Easy implementation and reporting

- Rejecting H_0 does not mean there are no *similarities* across groups; in fact, $\overline{y}_j$ can be a pretty inefficient estimate of μ_j

MATRIX FORMULATION

As we move towards more complex models, it will be a lot easier to work with the model in *matrix* form rather than in *scalar* form.

The *general linear model* is written in matrix form as $y = X\beta + \varepsilon$. Consider the treatment means model $y_{ij} = \mu_j + \varepsilon_{ij}$. We can represent this model in matrix form as follows.

$$y = X\mu + \varepsilon$$

$$
\begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n_1 1} \\ y_{12} \\ y_{22} \\ \vdots \\ y_{n_J J} \end{bmatrix}
=
\begin{bmatrix}
1 & 0 & \cdots & 0 \\
1 & 0 & \cdots & 0 \\
\vdots & & & \vdots \\
1 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots \\
0 & 1 & 0 & \cdots \\
\vdots & \ddots & & \vdots \\
0 & \cdots & & 1
\end{bmatrix}_{(\sum_j n_j) \times J}
\begin{bmatrix} \mu_1 \\ \vdots \\ \mu_J \end{bmatrix}
+
\begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{n_1 1} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{n_J J} \end{bmatrix}
$$

STA 610L

ASIDE: ESSENCE MATRIX

In ANOVA, the X matrix has a special form, which is sometimes summarized by the *essence matrix*, which shows the unique rows of the matrix.

- The essence matrix of a treatment means formulation of ANOVA is the $J \times J$ identity matrix, with each row repeated n_j times: if $j = 3$ we have
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

ASIDE: ESSENCE MATRIX

- The essence matrix in the treatment effects model is a column of ones

  appended to the left of the identity matrix: $\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$

- The essence matrix in reference cell coding is given by $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$
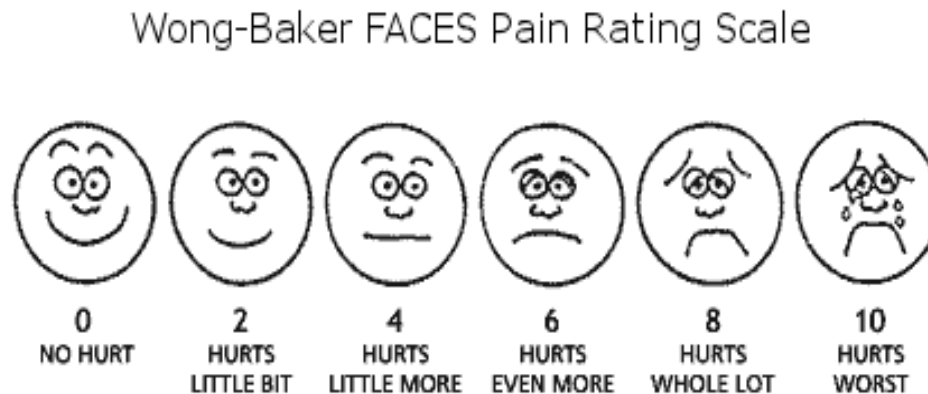
  (assuming without loss of generality that the last group is the reference)

In each case, the row corresponding to group j is repeated n_j times so that the full X matrix has $\sum_j n_j$ rows.

STA 610L

Simple example

Consider a study in which we wish to compare pain levels among children recovering from surgery who were randomized to one of three groups: audio books (chosen by each child), music (the child could select the playlist), or control (noise cancelling headphones).

After 30 minutes of one of the three conditions, each child rated his or her pain status using the following scale.

Wong-Baker FACES Pain Rating Scale

0	2	4	6	8	10
NO HURT	HURTS LITTLE BIT	HURTS LITTLE MORE	HURTS EVEN MORE	HURTS WHOLE LOT	HURTS WORST

From Wong D.L., Hockenberry-Eaton M., Wilson D., Winkelstein M.L., Schwartz P.: Wong's Essentials of Pediatric Nursing, ed. 6, St. Louis, 2001, p. 1301. Copyrighted by Mosby, Inc. Reprinted by permission.

SIMPLE EXAMPLE

The study is described here, but we will consider similar data from a hypothetical smaller study (n=15 total).

Suppose the data from the three groups is as follows.

- Audio books: 5,6,7,2,6

- Music: 5,4,4,7,6

- Control: 4,8,7,6,10

Write each element of $y = X\beta + \varepsilon$ in matrix or vector form using a group means coding scheme.

Once you've finished, repeat the exercise using reference cell coding.

LEAST SQUARES ESTIMATION

The least squares estimate minimizes the sum of squared residuals given by

$$\text{SSE}(\mu) = (y - X\mu)'(y - X\mu)$$
$$= y'y - 2\mu'X'y + \mu'X'X\mu$$

To find μ that minimizes the $\text{SSE} = y'y - 2\mu'X'y + \mu'X'X\mu$, take derivatives:

$$\frac{\partial SSE}{\partial \mu} = 0 - 2X'y + 2X'X\mu$$

and then set to 0:

$$0 = -2X'y + 2X'X\mu$$

to get the *normal equations*

$$(X'X)_{p \times p}\widehat{\mu} = X'y.$$

Read more here if you're rusty on matrix differentiation.

STA 610L

LEAST SQUARES ESTIMATION

When X has rank p, we solve the normal equations

$$(X'X)\widehat{\mu} = X'y$$
$$(X'X)^{-1}(X'X)\widehat{\mu} = (X'X)^{-1}X'y$$
$$\widehat{\mu} = (X'X)^{-1}X'y$$

Our least squares estimate in this case is unique, the best linear unbiased estimate, and if our errors are Gaussian, $\widehat{\mu}$ is the MLE and minimum variance unbiased estimator.

When X has rank $< p$, we can use a generalized inverse, but $\widehat{\mu}$ is not unique, though the predicted values and residuals are unique.

WHAT'S NEXT?

MOVE ON TO THE READINGS FOR THE NEXT MODULE!