V. Analysis of Variance

A. One-Way Analysis of Variance

1. Models for Response Variable Differences by Group
   a. Suppose that response values $Y_i$ come with
      i. Explanatory variables $x_{i1}, \ldots, x_{ip}$
      ii. Group membership $G_i$
   b. How do we
      i. Model dependence on all of this explanatory information
      ii. Estimate parameters associated with each?
      iii. Test interesting hypotheses?

2. Easiest Case: $p = 0$.
   a. Suppose
      i. there are $K$ groups indicated by $G_i$
      ii. and the number of observations in group $k \in \{1, \ldots, K\}$ is $n_k$
      iii. Hence $\sum_{k=1}^{K} n_k = n$.
   b. Conditions (as before):
      i. All of the observations in a given group have the same
expectation.

ii. All of the observations are independent.

iii. All of the observations have the same variance.

iv. All of the observations are approximately normal.

c. Sometimes write \((Y_1, \ldots, Y_n)\) as \((W_{11}, W_{12}, \ldots, W_{1n_1}, W_{21}, \ldots, W_{2n_2}, \ldots, W_{K1}, \ldots, W_{Kn_K})\)

i. \(W\) is not a matrix, because there could be different numbers of observations per group.

ii. Consider the case in which these are independent.

d. Parameters of interest are \(\mu_k = E\left[W_{kj}\right]\).

e. Plausible estimators are \(\hat{\mu}_k = \sum_{j=1}^{n_k} W_{kj} / n_k\).

i. In this case, \(\hat{\mu}_k\) satisfies \(\text{argmin}(\sum_{i=1}^{n_k} (W_{kj} - \mu_k)^2)\)

ii. Hence \(\mu\) minimizes \(\sum_{k=1}^{K} \sum_{i=1}^{n_k} (W_{kj} - \mu_k)^2\)

iii. Hence the group-wise means are least squares estimates.

f. Interesting null and alternative hypotheses are \(H_0 : \mu_k = \mu_1 \forall k\)

and \(H_A : \) there exist \(k, \ell\) such that \(\mu_k \neq \mu_\ell\) respectively.

g. If the data are approximately normal, one can test the null
hypothesis $H_0$ vs. the alternative $H_A$ by calculating

$$F = \frac{\sum_{k=1}^{K} n_k (\bar{W}_k - \bar{W})^2 / (K - 1)}{\sum_{k=1}^{K} \sum_{j=1}^{n_k} (W_{kj} - \bar{W}_k)^2 / (n - K)}$$

and comparing it to an $F_{K-1,n-K}$ distribution.

3. Equivalent approach using multiple regression:

a. For each $k$ create a new covariate $V_k$

   i. $V_{ki} = 1$ if observation $i$ is in group $k$ (that is, if $G_i = k$),
      and $V_{ki} = 0$ otherwise.

b. Then previous model is $E[Y_i] = \mu_{G_i} = \sum_{k=1}^{K} V_{ki} \mu_k$.

c. Hence we can replicate a means model with a regression model.

   i. Least-squares estimates are means.

   ii. Note the lack of an intercept.

   iii. Adding an intercept gives model $\hat{Y}_i = \beta_0 + \sum_{k=1}^{K} \beta_k V_{ki}$.

   iv. An intercept term would be co-linear, since $\sum_{k=1}^{K} V_{ki} = 1 \forall i$.

d. Often remove induced co-linearity by dropping one of the
   indicator variables.

   i. Suppose that the removed parameter is $\beta_1$

   ii. $\hat{\beta}_1 = 0$.

   iii. $\bar{W}_1 = \hat{\beta}_0 + \hat{\beta}_1 = \hat{\beta}_0$.

   iv. $\bar{W}_k = \hat{\beta}_0 + \hat{\beta}_k$ and so $\hat{\beta}_k = \bar{W}_k - \bar{W}_1$ for $k > 1$. 
v. Group with parameter set to zero is called reference group.

e. Can calculate matrix inverse in regression fit in closed form.

i. Take first group as reference.

\[
X = \begin{pmatrix}
1 & 0 & 0 & \ldots & 0 & 0 \\
1 & 0 & 0 & \ldots & 0 & 0 \\
1 & 1 & 0 & \ldots & 0 & 0 \\
1 & 1 & 0 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
1 & 1 & 0 & \ldots & 0 & 0 \\
1 & 0 & 1 & \ldots & 0 & 0 \\
1 & 0 & 1 & \ldots & 0 & 0 \\
1 & 0 & 1 & \ldots & 0 & 0 \\
1 & 0 & 0 & \ldots & 1 & 1 \\
\end{pmatrix}
\]

\[
X^\top X = \begin{pmatrix}
n & n_2 & n_3 & \ldots & n_k \\
n_2 & n_2 & 0 & \ldots & 0 \\
n_3 & 0 & n_3 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\end{pmatrix}
\]

ii. Then

\[
(X^\top X)^{-1} = \begin{pmatrix}
\frac{1}{n_1} & -\frac{1}{n_1} & -\frac{1}{n_1} & \ldots & -\frac{1}{n_1} \\
-\frac{1}{n_1} & \frac{1}{n_1} + \frac{1}{n_2} & \frac{1}{n_1} & \ldots & \frac{1}{n_1} \\
-\frac{1}{n_1} & \frac{1}{n_1} & \frac{1}{n_1} + \frac{1}{n_3} & \ldots & \frac{1}{n_1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\end{pmatrix}
\]

f. Matrix inverse can be expressed in closed form.

B. Model can contain an additional explanatory variable.
1. Analysis of Covariance model

a. Structure that is more simple than general multivariate regression model
   i. Computationally easier to fit.
   ii. Solutions can be expressed without an arbitrary matrix inverse.
   iii. Simplification is seldom needed with modern computing.

b. Model is of form $Y_i = \beta_0 + \sum_{k=1}^{K} \beta_k V_{ki} + \gamma X_i + \epsilon_i$,
   i. Holding group fixed,
      - the relation is linear
      - with an intercept depending on group,
      - with a slope not depending on group.
   ii. Hence fit is a set of parallel lines.

c. You might want to consider an interaction:
   i. Resulting fits are no longer parallel.

2. Model is problematic if the grouping variable and the linear covariate are related.

a. Problem does not arise if group assignment is randomized.

b. This problem is common to all linear regression models.

3. Testing the model via multivariate Wald test:
a. Recall: to test $\beta_j = \beta_j^0$ for $j = k - m + 1, \ldots, k$

i. Partition $\beta$ as $(\lambda, \gamma)$.

- Consider case with $\beta_j^0 = 0$ for $j = k - m + 1, \ldots, k$.

b. Let $W$ be the part of matrix $(X^\top X)^{-1}$.

c. Then $[(\hat{\gamma} - \gamma^0)^\top W^{-1}(\hat{\gamma} - \gamma^0)/m]/\hat{\sigma}^2 \sim F_{m,n-k}$.

d. Express design matrix as $(X_a, X_b)$

e. Simple case: design matrix is orthonormal

i. Then $\hat{\gamma} = X_b^\top Y$, $W = I$,

ii. and so $(\hat{\gamma} - \gamma^0)^\top W^{-1}(\hat{\gamma} - \gamma^0) = Y^\top X_b X_b^\top Y = \check{Y}^\top \check{Y}$ for $\check{Y}$ the fitted values for regression with just $X_b$.

iii. Note that $\hat{Y} = \check{Y} + \tilde{Y}$, where $\hat{Y} = (X_a, X_b)^\top Y$ and $\check{Y} = X_a^\top Y$.

iv. $\check{Y}$ and $\tilde{Y}$ are orthogonal

v. $\check{Y}^\top \check{Y} = \check{Y}^\top \check{Y} + \tilde{Y}^\top \tilde{Y}$

vi. $(\hat{\gamma} - \gamma^0)^\top W^{-1}(\hat{\gamma} - \gamma^0) = \check{Y}^\top \check{Y} - \check{Y}^\top \tilde{Y}$.

f. Outside simple orthonormal case:

i. Suppose $\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$ makes $X$ orthonormal.

ii. Inverse of this matrix is $\begin{pmatrix} A^{-1} & -A^{-1}BC^{-1} \\ 0 & C^{-1} \end{pmatrix}$.
iii. Hence \( X\beta = X \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} \left( \begin{pmatrix} A^{-1} & -A^{-1}BC^{-1} \\ 0 & C^{-1} \end{pmatrix} \right) \beta \)

iv. New design matrix is \( X \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} \), parameter is 
\( \left( \begin{pmatrix} A^{-1} & -A^{-1}BC^{-1} \\ 0 & C^{-1} \end{pmatrix} \right) \beta \)

v. Hence \( \hat{\beta} \) transformed to \( C^{-1}\hat{\beta} \), and \( W \) transforms to \( C^{-1\top}WC^{-1} \).

vi. Hence Wald statistic numerator is unchanged, and so are the difference in fitted values.

g. Hence test can be represented as a difference in sum of regression sum of squares.

4. Test is built from differences in regression sums of squares.

a. Differences are for more complex and less complex models.

b. Differences can be represented in a table of sums of squares:

c. Summarized by Analysis of Variance (ANOVA) table:

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
<th>Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Group</td>
<td>( SS_R )</td>
<td>( m )</td>
<td>( MS_R )</td>
<td>( F )</td>
<td></td>
</tr>
<tr>
<td>Within Group</td>
<td>( SS_{Res} )</td>
<td>( n - 1 - m )</td>
<td>( MS_{Res} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( SS_t )</td>
<td>( n - 1 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

d. Test applies only to last alternative hypothesis under
5. Treatment of a covariate as group indicator is different from treating it as a numeric variable.
   a. Construction of indicator variables does not take ordering into account.
   b. If the groups are ordered, one might regress on numeric labels to take ordering into account.
   c. This is appropriate only if the distance between categories reflects difference in expectation of group mean.

6. Can add more categorical variables
   a. Add with indicator variables.
   b. Express as a regression
   c. Generally the $X^\top X$ cannot be inverted in closed form with more than one categorical variable.
      i. Except for the balanced case.
      ii. In balanced case the two factors are orthogonal.
   d. Otherwise the regression approach is required.
      : 3.3

C. Contrasts: Tests for differences within set.
1. Multiple Comparisons Problem
   a. We have $K(K - 1)/2$ chances to find a significant result.
   b. If done at nominal level, this will inflate experiment-wise error rate.
   c. If nominal level is multiplied by number of possible comparisons performed ($\frac{1}{2}K(K - 1)$).
      i. Will get usually a very conservative procedure.
      ii. This is the Bonferroni procedure.

2. Fisher’s Least Significant Difference method:
   a. Do ANOVA.
   b. If $F$ test does not reject $H_0$, do not declare anything significant.
   c. If $F$ test rejects $H_0$, then do test on each pairwise comparison between group means.
   d. This fails to control Type I error rate if $K > 3$.
      i. Suppose groups 1 through $K - 1$ are identical, and group $K$ is offset by $\Delta \neq 0$.
      ii. Then null hypotheses of equality of group expectations $i$ and $j$ are true, for $i, j < K$. 
iii. Can make $\Delta$ so large that $F$ test rejects equality of all
distributions with probability close to 1.

iv. Then at least one of the true null hypotheses is rejected on the
union of the critical regions for the separate $(K - 1)(K - 2)/2$
tests.

v. If $K = 3$, there is only one such test, and so no problem with
multiple comparisons.

3. Tukey’s Honest Significant Difference method:

a. Studentized range distribution:

i. Setup: $\bar{W}_j \sim N(\mu, \sigma^2/n_j)$ for $j \in \{1, \ldots, K\}$,
$\hat{\sigma}^2 \sim \sigma^2 \chi^2_m$, independent, $n_j$ all equal.

ii. If all $n_j$ are equal, $\max_{1 \leq i, j \leq K} (|\bar{W}_j - \bar{W}_i| / (\hat{\sigma} / \sqrt{n_j}))$ has
the studentized range distribution with $K$ and $m$ degrees of
freedom, exactly.

iii. If $n_j$ are not all equal,
$$\max_{1 \leq i, j \leq K} \left( |\bar{W}_j - \bar{W}_i| / (\hat{\sigma} \sqrt{(1/2)(1/n_j + 1/n_k)}) \right)$$
has the studentized range distribution with $K$ and $m$ degrees of
freedom, approximately.

iv. So differences in means greater than $q\hat{\sigma} \sqrt{(1/2)(1/n_i + 1/n_j)}$
4. Mean comparisons generalized

a. $\overline{W}_j - \overline{W}_k = c\overline{W}$ for $c$ a row vector of all zeros except $c_j = 1$, $c_k = -1$.

b. A summary $c\overline{W}$ is a contrast if $\sum_k c_k = 0$.

i. Ex., if $K = 3$, then $c = (1, -2, 1)$ tests to see if the relation between $k$ and $E[\overline{W}_k]$ is linear.

c. Let $\mu = E[\overline{W}]$, $\mu_0$ be expectation of grand mean.

d. Let $\kappa = c\mu$, $\hat{\kappa} = c\overline{W}$,

e. Scheffé Method: $\kappa \in \hat{\kappa} \pm \hat{\sigma} \sqrt{\sum_k c_k^2 / n_k} \sqrt{F_{1-\alpha,K-1,n-K}}$
holds with probability $1 - \alpha$ for all contrasts $c$.

i. Let $c^* = (c_1 / \sqrt{n_1}, \ldots, c_K / \sqrt{n_K})$, $\overline{W}^* =
(\sqrt{n_1}(\overline{W}_1 - \overline{W} - \mu_1 + \mu_0), \ldots, \sqrt{n_K}(\overline{W}_K - \overline{W} - \mu_K + \mu_0))^\top$

ii. Then $\hat{\kappa} - \kappa = c^*\overline{W}^*$.

iii. By Schwartz inequality, $(\hat{\kappa} - \kappa)^2 \leq \sqrt{c^*c^\top} \sqrt{\overline{W}^*\overline{W}^*}$

iv. $\overline{W}^*\overline{W}^* / \sigma^2 \sim \chi^2_{K-1}$, independent of $\hat{\sigma}$.

v. $(K - 1)^{-1} \overline{W}^*\overline{W}^* / \hat{\sigma}^2 \sim F_{K-1,n-K}$. 

v. $q$ is quantile from studentized range distribution with degrees of freedom $n - K$. 

are significant.
vi. Hence $P\left[\left|\hat{\kappa} - \kappa\right| \leq \sqrt{\sum_k \frac{c_k}{n_k^2} / n_k \sqrt{F_{1-\alpha, K-1, n-K\hat{\sigma}}} \right] \geq 1 - \alpha$.

f. Can be extended to sets of regression parameters.