Linear Combinations of Means

Design of Experiments - Montgomery
Section 3-5
Linear Combinations of Means

\[ y_{ij} = \mu + \tau_i + \epsilon_{ij} \]

\[ = \mu_i + \epsilon_{ij} \]

• Often study goal more than just testing \( H_0 : \) all \( \mu_i \) equal
• Goal more likely a series of tests \( H_0 : L = \sum c_i \mu_i = L_0 \)
  – Pairwise comparisons (\( a = 3 \), test \( \mu_1 = \mu_2 \), \( \mu_2 = \mu_3 \), \( \mu_1 = \mu_3 \))
  – Treatments vs control (\( a = 4 \), test \( \mu_1 = \mu_C \), \( \mu_2 = \mu_C \), \( \mu_3 = \mu_C \))
  – Comparing combinations of treatments (e.g., \( 2\mu_1 = \mu_2 \))
  – Assessing (curvi)linear relationship between means and ordered trts

• Hypotheses may be planned or “after the fact”
• Very important to make the distinction
Linear Combination of Means

Can use our linear statistical model to construct \( t \)-test (\( F \)-test) of any linear combination

\[
\hat{L} = \sum c_i \bar{y}_i.
\]

\[
V(\hat{L}) = V(\sum c_i \bar{y}_i.)
\]

\[
= \sum c_i^2 V(\bar{y}_i.)
\]

\[
= MS_E \sum (c_i^2 / n_i)
\]

\[
t_o = \frac{\hat{L} - L_0}{\sqrt{V(\hat{L})}}
\]

Under \( H_0 \): \( t_o \sim t_{N-a} \)
Why Linear Combinations?

• These tests specifically address hypotheses of interest

• Overall $F$-test is “jointly” testing all possible alternatives

• Why include alternatives of no interest in your tests?
  - If overall error rate equal to $\alpha$, error rate for single alternative will be $< \alpha$
  - $F$ test will reduce (“water down”) power of specific test
  - Can find individual tests significant while $F$ not!!

• Issues with linear combinations
  - Multiple tests inflate overall Type I error rate
  - Not all tests of interest are independent of each other

• Will discuss procedures that control error rates
Contrasts

• Linear combinations with coeff’s that sum to zero

\[ \Gamma = \sum c_i \mu_i = 0 \text{ with } \sum c_i = 0 \]

\[ H_0 : \mu_1 = \mu_2 \rightarrow c_i = \{1, -1, 0, \ldots, 0\} \]

\[ H_0 : 2\mu_1 = \mu_2 + \mu_3 \rightarrow c_i = \{2, -1, -1, 0, \ldots, 0\} \]

\[ H_0 : 2\mu_1 = \mu_2 + \mu_3 \rightarrow c_i = \{-1, 0.5, 0.5, 0, \ldots, 0\} \]

• Will estimate each \( \Gamma \) using \( C = \sum c_i \bar{y}_i \).

• Recall under \( H_0 \): \( t_o = C / \sqrt{\text{Var}(C)} \sim t_{N-a} \) (Slide 3)
Contrasts

• Also $t^2_{N-a} = F_{1,N-a}$ so could consider $F$ test

• Contrasts often presented in terms of Sum of Squares

$$SS_C = \left( \sum c_i \bar{y}_i \right)^2 / \sum \left( c_i^2 / n_i \right)$$

   If divide by $\text{MS}_E$, simply $t^2_o$

   Can then compare to $F_{1,N-a}$
title1 'Contrast Comparisons';

data one;
  infile 'c:saswork\data\tensile.dat';
  input percent strength time;

proc glm data=one;
  class percent;
  model strength=percent;
  contrast 'C1' percent 0 0 0 1 -1;
  contrast 'C2' percent 1 0 1 -1 -1;
  contrast 'C3' percent 1 0 -1 0 0;
  contrast 'C4' percent 1 -4 1 1 1;
**SAS Output**

Dependent Variable: STRENGTH

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>4</td>
<td>475.76000</td>
<td>118.94000</td>
<td>14.76</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>20</td>
<td>161.20000</td>
<td>8.06000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>24</td>
<td>636.96000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square: 0.746923, C.V.: 18.87642, Root MSE: 2.8390, STRENGTH Mean: 15.040

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERCENT</td>
<td>4</td>
<td>475.76000</td>
<td>118.94000</td>
<td>14.76</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contrast</th>
<th>DF</th>
<th>Contrast SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>1</td>
<td>291.60000</td>
<td>291.60000</td>
<td>36.18</td>
<td>0.0001</td>
</tr>
<tr>
<td>C2</td>
<td>1</td>
<td>31.25000</td>
<td>31.25000</td>
<td>3.88</td>
<td>0.0630</td>
</tr>
<tr>
<td>C3</td>
<td>1</td>
<td>152.10000</td>
<td>152.10000</td>
<td>18.87</td>
<td>0.0003</td>
</tr>
<tr>
<td>C4</td>
<td>1</td>
<td>0.81000</td>
<td>0.81000</td>
<td>0.10</td>
<td>0.7545</td>
</tr>
</tbody>
</table>

Linear Combinations
Orthogonal Contrasts

- Suppose you have two contrasts \( \{c_i\} \) and \( \{d_i\} \)
- **Orthogonal** if \( \sum c_i d_i = 0 \) (when \( n_i \) constant)

- Can divide up SS\(_{Trt}\) into \( a - 1 \) orthogonal contrasts
- By Cochran’s Thm → comparisons independent
- Thus, previous four contrasts are independent

Linear Combinations
Orthogonal Polynomial Contrasts

- Orthogonal contrasts used to study trend
- Should only be used if treatments quantitative (ordered)
- If equally-spaced and $n_i = n$, the $c_i$ in Table IX
- Use software to determine coefficients in other situations
- Results in breakdown of polynomial regression

$$
\mu_i = \beta_0 + \beta_1 i + \ldots + \beta_{a-1} i^{a-1}
$$
Using SAS to determine $c_i$

- Often the levels of the trt are not equally spaced

- Can use Proc IML to determine coeffs

```sas
proc iml;
levels={1 2 5 10 20}; /* Consider these 5 levels */
print levels;
coef=ORPOL(levels,3); /* Gives coeffs up through cubic */
coef=t(coef); /* Puts coeffs in rows instead of cols */
coef=coef[2:4,]; /* Eliminates the first row of coef matrix*/
print coef; /* 1st row linear, 2nd quadratic, 3rd cubic */
run;
```

---

<table>
<thead>
<tr>
<th>LEVELS</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>COEF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.424967</td>
<td>-0.360578</td>
<td>-0.167411</td>
<td>0.1545335</td>
<td>0.798423</td>
</tr>
<tr>
<td></td>
<td>0.4348974</td>
<td>0.2072899</td>
<td>-0.325207</td>
<td>-0.711616</td>
<td>0.3946361</td>
</tr>
<tr>
<td></td>
<td>-0.433125</td>
<td>0.1365799</td>
<td>0.7252914</td>
<td>-0.510844</td>
<td>0.0820972</td>
</tr>
</tbody>
</table>
```

Linear Combinations
Testing Multiple Contrasts

- If $m$ orthogonal (independent) contrasts
  
  \[ P(\text{at least one type I error}) = 1 - (1 - \alpha')^m \]

  Can control overall error rate exactly

  **Example:** Consider $m = 5$ independent tests and want overall Type I error rate to be 0.05. Using equation above, each test must use
  
  \[ \alpha' = 1 - (1 - .05)^{1/5} = 0.0102 \]

- Scheffé’s Method

  Set up $1 - \alpha$ simultaneous CI for all contrasts

  Protects for unplanned comparisons

  Overall error rate at most $\alpha \leftrightarrow$ low power

  Scheffé suggested using larger $\alpha$ level
Testing Multiple Contrasts

- Scheffé’s Method

Compare $|C|$ to $s_C \sqrt{(a - 1)F_{\alpha,a-1,N-a}}$

Scheffé showed experimentwise error rate no more than $\alpha$

Relationship with overall $F$ test

If the $P$-value of $F$-test is $\gamma$, then at the $\gamma$ significance level, Scheffé method will find one contrast significant.

$$c_i = n_i(y_{i.} - \bar{y}_{..}) = n_i \hat{\tau}_i$$
Testing Multiple Contrasts

- Bonferroni’s Method

Recall $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(\text{at least one type I error in } m \text{ tests}) \leq m\alpha'$

For each test use $\alpha' = \alpha/m$

Designed for planned comparisons only

Looks only at subset of contrasts

Extremely conservative if $m$ is large
Comparison of Means

- Often only interested in pairwise comparisons
- Can be expressed as contrasts
- If comparing trt $j$ and trt $k$

$$c_i = \begin{cases} 
1 & \text{if } i = j \\
-1 & \text{if } i = k \\
0 & \text{otherwise}
\end{cases}$$
Comparison of Means

• Pairwise comparisons a subset of all contrasts

• Want test that considers only subset

• Trade-off between power and prob(Type I error)

• No one “best” test

• Comparison methods vary in protection
  – Experimentwise error (overall Type I)
  – Comparisonwise error (individual Type I)
Pairwise Comparison Methods

- Least Significant Difference (LSD)
  - Use $\text{MS}_E$ and its df to perform usual $t$-test
    \[
    \frac{\bar{y}_i. - \bar{y}_j.}{\sqrt{\text{MSE} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}} \sim t_{N-a}
    \]
  - Does not control experimentwise error, controls $\alpha'$

- Protected Least Significant Difference (LSD)
  - Only do LSD comparisons if $F$-test significant
Pairwise Comparison Methods

- **Tukey’s (Tukey-Kramer) Method**
  - Uses studentized range distribution \( q \) instead of \( t \)
  - Distribution based on \( \bar{y}_{\text{MAX}} - \bar{y}_{\text{MIN}} \) in \( a \) trts
  - Accounts for any possible pair being selected
  - Controls overall experimentwise error rate \( \alpha \)
  - Reject if
    \[
    |\bar{y}_i. - \bar{y}_j.| > \frac{q_{\alpha,a,N-a}}{\sqrt{2}} \sqrt{\frac{\text{MSE} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}}
    \]
    - The value \( q \) available in Table VII
Pairwise Comparison Methods

• Newman-Keuls Test
  – Performs Tukey’s test for varying number of trts
  – Start with $p = a$, then $p = a - 1$, etc

$$K_p = q_{\alpha}(p, f) \sqrt{MS_E/n}$$

  – Stop whenever difference not found significant
  – Controls experimentwise error for all tests with $m$ means
  – When unequal sample size $n = a/\sum 1/n_i$
Pairwise Comparison Methods

- **Duncan’s Multiple Range Test**
  - Similar testing approach as Newman-Keuls
  - Based on least significant range
  - Powerful $\leftrightarrow$ does not control overall Type I

- **Dunnett’s Test**
  - Specifically designed for trts vs control situation
  - Distribution based on $\text{MAX}((\bar{y}_\text{Control} - \bar{y}_\text{Ttrt})$ for $a - 1$ trts
  - Similar in approach to Tukey’s test
  - Controls experimentwise error rate
SAS Procedures

title1 'Means Comparison';

data one;
   infile 'c:saswork\data\tensile.dat';
   input percent strength time;

proc glm data=one;
   class percent;
   model strength=percent;
   means percent / alpha=.05 lines bon snk tukey;
   lsmeans percent / lines adjust=tukey;
   means percent / lines duncan lsd scheffe;
   means percent /dunnett;
   means percent / lsd clm;
run;
T tests (LSD) for variable: STRENGTH

NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.

Alpha= 0.05  df= 20  MSE= 8.06
Critical Value of T= 2.09
Least Significant Difference= 3.7455

Means with the same letter are not significantly different.

<table>
<thead>
<tr>
<th>T Grouping</th>
<th>Mean</th>
<th>N</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>21.600</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>B</td>
<td>17.600</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>C</td>
<td>10.800</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>B</td>
<td>15.400</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>C</td>
<td>9.800</td>
<td>5</td>
<td>15</td>
</tr>
</tbody>
</table>
Bonferroni (Dunn) T tests for variable: STRENGTH

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

Alpha= 0.05   df= 20   MSE= 8.06
Critical Value of T= 3.15
Minimum Significant Difference= 5.6621

Means with the same letter are not significantly different.

<table>
<thead>
<tr>
<th>Bon Grouping</th>
<th>Mean</th>
<th>N</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>21.600</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>A</td>
<td>21.600</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>B A</td>
<td>17.600</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>B</td>
<td>17.600</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>B C</td>
<td>15.400</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>C</td>
<td>15.400</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>C</td>
<td>10.800</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>C</td>
<td>10.800</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>C</td>
<td>9.8000</td>
<td>5</td>
<td>15</td>
</tr>
</tbody>
</table>
Scheffe’s test for variable: STRENGTH

NOTE: This test controls the type I experimentwise error rate but generally has a higher type II error rate than REGWF for all pairwise comparisons

Alpha = 0.05  df = 20  MSE = 8.06
Critical Value of F = 2.86608
Minimum Significant Difference = 6.0796

Means with the same letter are not significantly different.

<table>
<thead>
<tr>
<th>Scheffe Grouping</th>
<th>Mean</th>
<th>N</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>21.600</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B A</td>
<td>17.600</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B C</td>
<td>15.400</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>10.800</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>9.800</td>
<td>5</td>
<td>15</td>
</tr>
</tbody>
</table>
Tukey’s Studentized Range (HSD) Test for variable: STRENGTH

NOTE: This test controls the type I experimentwise error rate, but generally has a higher type II error rate than REGWQ.

\[ \text{Alpha} = 0.05 \quad \text{df} = 20 \quad \text{MSE} = 8.06 \]

Critical Value of Studentized Range = 4.232

Minimum Significant Difference = 5.373

Means with the same letter are not significantly different.

<table>
<thead>
<tr>
<th>Tukey Grouping</th>
<th>Mean</th>
<th>N</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>21.600</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>B A</td>
<td>17.600</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B C</td>
<td>15.400</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D C</td>
<td>10.800</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>9.800</td>
<td>5</td>
<td>15</td>
</tr>
</tbody>
</table>
Student-Newman-Keuls test for variable: STRENGTH

NOTE: This test controls the type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha = 0.05  df = 20  MSE = 8.06

Number of Means  
2  3  4  5
Critical Range  
3.7454541  4.5427098  5.0256318  5.3729606

Means with the same letter are not significantly different.

<table>
<thead>
<tr>
<th>SNK Grouping</th>
<th>Mean</th>
<th>N</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>21.600</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>B</td>
<td>17.600</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>B</td>
<td>15.400</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>C</td>
<td>10.800</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>C</td>
<td>9.800</td>
<td>5</td>
<td>15</td>
</tr>
</tbody>
</table>
Duncan’s Multiple Range Test for variable: STRENGTH
NOTE: This test controls the type I comparisonwise error rate, not the experimentwise error rate

Alpha= 0.05  df= 20  MSE= 8.06

Number of Means  2  3  4  5
Critical Range  3.745 3.931 4.050 4.132

Means with the same letter are not significantly different.

<table>
<thead>
<tr>
<th>Duncan Grouping</th>
<th>Mean</th>
<th>N</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>21.600</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>B</td>
<td>17.600</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>B</td>
<td>15.400</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>C</td>
<td>10.800</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>C</td>
<td>9.800</td>
<td>5</td>
<td>15</td>
</tr>
</tbody>
</table>
Dunnett’s T tests for variable: STRENGTH

NOTE: This tests controls the type I experimentwise error for comparisons of all treatments against a control.

Alpha= 0.05  Confidence= 0.95  df= 20  MSE= 8.06

Critical Value of Dunnett’s T= 2.651
Minimum Significant Difference= 4.7602
Comparisons significant at the 0.05 level are indicated by ‘***’.

<table>
<thead>
<tr>
<th>PERCENT</th>
<th>Comparison</th>
<th>Lower Limit</th>
<th>Difference Between Means</th>
<th>Upper Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>30 - 15</td>
<td>7.040</td>
<td>11.800</td>
<td>16.560</td>
</tr>
<tr>
<td>25</td>
<td>25 - 15</td>
<td>3.040</td>
<td>7.800</td>
<td>12.560</td>
</tr>
<tr>
<td>20</td>
<td>20 - 15</td>
<td>0.840</td>
<td>5.600</td>
<td>10.360</td>
</tr>
<tr>
<td>35</td>
<td>35 - 15</td>
<td>-3.760</td>
<td>1.000</td>
<td>5.760</td>
</tr>
</tbody>
</table>
General Testing Setting

• Consider $m$ tests:

\[ H_0^1 \quad \text{vs} \quad H_a^1 \]
\[ H_0^2 \quad \text{vs} \quad H_a^2 \]
\[ \vdots \]
\[ H_0^m \quad \text{vs} \quad H_a^m \]

• Consider that for $m_0$ of these tests $H_0$ is true and for $m_a$ of these tests $H_a$ true $\rightarrow m_0 + m_a = m$
**General Testing Setting**

- With $m_0$ and $m_a$ fixed but unknown, the given tests will allocate in the following way with $V$ and $T$ random variables.

<table>
<thead>
<tr>
<th></th>
<th>DNR</th>
<th>Reject</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ True</td>
<td>$U$</td>
<td>$V$</td>
<td>$m_0$</td>
</tr>
<tr>
<td>$H_a$ True</td>
<td>$T$</td>
<td>$S$</td>
<td>$m_a$</td>
</tr>
<tr>
<td></td>
<td>$W$</td>
<td>$R$</td>
<td>$m$</td>
</tr>
</tbody>
</table>

- When $m = 1$, test designed to maximize $E(S|m_a = 1)$ given $E(V|m_0 = 1) \leq \alpha$.
General Testing Setting $m > 1$

- Need to determine what to control

<table>
<thead>
<tr>
<th></th>
<th>DNR</th>
<th>Reject</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null True</td>
<td>$U$</td>
<td>$V$</td>
<td>$m_0$</td>
</tr>
<tr>
<td>Alt True</td>
<td>$T$</td>
<td>$S$</td>
<td>$m_a$</td>
</tr>
<tr>
<td></td>
<td>$W$</td>
<td>$R$</td>
<td>$m$</td>
</tr>
</tbody>
</table>

- **Familywise Error Rate** (FWER): $Pr(V > 0)$
  - What Montgomery calls experimentwise error rate
  - Controls the chance of there being a single false positive

- **False Discovery Rate** (FDR): $E\left(\frac{V}{R}\right| R > 0)Pr(R > 0)$
  - Control the percent of rejected $H_0$’s that are wrong

- When $m$ very large, trend towards controlling FDR
Multiplicity Adjustment

- So far discussed approaches that adjust multiplier to SE
  - Alter $\alpha$ level (e.g., Bonferroni)
  - Use different distribution
- Conservative $\rightarrow$ strong control of overall Type I error - avoids false positives
- Powerful $\rightarrow$ able to pick up differences that exist - avoids false negatives
- All approaches try to strike some sort of balance
Procedures Based on P-values

• Order the P-values from smallest to largest

\[ P_1, P_2, \ldots, P_m \rightarrow P(1), P(2), \ldots, P(m) \]

1. Holm
   - “Refinement” of Bonferroni
   - Instead of using
     \[ \alpha^* = \frac{\alpha}{m} \]
     for all comparisons
   - Continue to reject until \( P(k) > \alpha/(m - k + 1) \)

2. False Discovery Rate (Benjamini and Hochberg)
   - Continue to reject until \( P(k) > k\alpha/m \)

• Both available in Proc Multtest in SAS
False Discovery Rate

- Procedure 1: reject until $P_{(k)} > k\alpha/m$
  - Known as a *step-up* procedure
  - Depends on independence (although works well with some dependence among test statistics)

- Procedure 2: reject until $P_{(k)} > \delta_k$
  \[
  \delta_k = 1 - \left[1 - \min\left(1, \frac{m}{m-k+1}\alpha\right)\right]^{1/(m-k+1)}
  \]
  - Known as a *step-down* procedure
  - More powerful than Procedure 1 when number of tests is small and many hypotheses are “clearly” true
False Discovery Rate

- Procedure 3: reject until $P_{(k)} > \frac{k}{m} \left( \frac{\alpha}{\left( \sum \frac{1}{i} \right)} \right)$
  - Controls FDR under any dependence structure
  - Not as powerful as first procedure

- Procedure 4: reject until $P_{(k)} > \delta_k$

  $$\delta_k = \min \left( 1, \frac{m}{(m-k+1)^2} \alpha \right)$$
  
  - Controls FDR under any dependence structure
  - More powerful than Procedure 3 when $m$ small and most are not true
Some Facts

- $\text{FDR} \leq \text{FWER}$
  
  Equality when $m = m_0$

- Newman-Keuls controls FDR

- There is range test by Ryan-Einot-Gabriel-Welsch (REGWR) which is similar to Holm’s control of FWER
SAS Commands

title1 'Multiple Comparisons';

data one;
  infile 'U:\.www\datasets1\tensile.dat';
  input percent strength time;
proc multtest holm fdr data=one out=new1 noprint;
  class percent;
  contrast '12' 1 -1 0 0 0;  contrast '13' 1 0 -1 0 0;
  contrast '14' 1 0 0 -1 0;  contrast '15' 1 0 0 0 -1;
  contrast '23' 0 1 -1 0 0;  contrast '24' 0 1 0 -1 0;
  contrast '25' 0 1 0 0 -1;  contrast '34' 0 0 1 -1 0;
  contrast '35' 0 0 1 0 -1;  contrast '45' 0 0 0 1 -1;
  test mean(strength);
proc print data=new1;
run;
<table>
<thead>
<tr>
<th>contrast</th>
<th>value</th>
<th>se</th>
<th>raw_p</th>
<th>stpbon_p</th>
<th>fdr_p</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>-140</td>
<td>44.8888</td>
<td>0.00541</td>
<td>0.02704</td>
<td>0.00901</td>
</tr>
<tr>
<td>13</td>
<td>-195</td>
<td>44.8888</td>
<td>0.00031</td>
<td>0.00252</td>
<td>0.00105</td>
</tr>
<tr>
<td>14</td>
<td>-295</td>
<td>44.8888</td>
<td>0.00000</td>
<td>0.00002</td>
<td>0.00002</td>
</tr>
<tr>
<td>15</td>
<td>-25</td>
<td>44.8888</td>
<td>0.58375</td>
<td>0.58375</td>
<td>0.58375</td>
</tr>
<tr>
<td>23</td>
<td>-55</td>
<td>44.8888</td>
<td>0.23471</td>
<td>0.46943</td>
<td>0.26079</td>
</tr>
<tr>
<td>24</td>
<td>-155</td>
<td>44.8888</td>
<td>0.00251</td>
<td>0.01509</td>
<td>0.00503</td>
</tr>
<tr>
<td>25</td>
<td>115</td>
<td>44.8888</td>
<td>0.01859</td>
<td>0.07438</td>
<td>0.02656</td>
</tr>
<tr>
<td>34</td>
<td>-100</td>
<td>44.8888</td>
<td>0.03754</td>
<td>0.11262</td>
<td>0.04693</td>
</tr>
<tr>
<td>35</td>
<td>170</td>
<td>44.8888</td>
<td>0.00116</td>
<td>0.00810</td>
<td>0.00289</td>
</tr>
<tr>
<td>45</td>
<td>270</td>
<td>44.8888</td>
<td>0.00001</td>
<td>0.00006</td>
<td>0.00004</td>
</tr>
</tbody>
</table>

Holm similar to FDR except 2 vs 5 and 3 vs 4

Holm similar to Bonferroni

FDR similar to Newman-Keuls