1-way ANOVA

overview

- Comparisons of nested linear models
- Interpretations of ANOVA tables
- Mean Squares are estimates of variance
- Effect size & association strength
- Assumptions of ANOVA & alternatives
**Linear Models**

- ANOVA fits & compares several nested, linear models
  - \( Y_{ij} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + e_{ij} \)
  - \( Y_{ij} \) is score “i” in group “j”
  - \( X_i \) are a predictor variables (e.g., representing groups/conditions)
  - \( \beta_i \) are coefficients adjusted to minimize \( \sum e_{ij}^2 \)

**Linear Models**

- Linear models are a very broad class of models
- can be used to characterize curvilinear associations between \( Y \) and \( X \)'s
  - \( Y_{ij} = \beta_0 + \beta_1 X_1 + \beta_2 X_2^2 + \cdots + e_{ij} \)
  - \( Y_{ij} = \beta_0 + \beta_1 \exp(X_1) + \beta_2 \log(X_2) + \cdots + e_{ij} \)

**Linear Models**

include linear regression & multiple regression

\[
Y_i = B_0 + (B_1 \times X) + (B_2 \times X^2)
\]

**Linear Models**

Least-squares estimate of model parameters

\[
Y_{ij} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + e_{ij}
\]

\( e_{ij} = Y_{ij} - \hat{Y}_{ij} \) residual = difference between observed & predicted scores

\( \Sigma e_{ij}^2 \) Least-squares measure of goodness-of-fit: sum of squared-residuals

ANOVA models are fit using the criterion of least-squares
Comparing Nested Linear Models

- Compare nested linear models
  - models vary in complexity and how well they fit the data
  - select model that provides best fit with fewest parameters
- Define “complexity” as number of parameters (i.e., predictor variables)
- Define “goodness-of-fit” as sum of squared residuals
  - measure effect of removing free parameters on goodness-of-fit
  - if change is small, keep simpler, reduced model
  - if change is large, do not remove parameters & keep full model

Best-fitting (least-squares) parameters

( for a 1-way design )

\[
\text{Nested Linear Models}
\]

- full model: \( Y_{ij} = \mu + \alpha_j + e_{ij} \)
- reduced model: \( Y_{ij} = \mu + e_{ij} \)

\[\sum_{j} \alpha_j = 0\]

- Group effects (alphas)
  - Defined as difference between group mean & overall mean
  - with this definition, alphas MUST sum to zero
  - often called "sigma" or "sum-to-zero" definition of alphas
  - There are other definitions of group effects & intercept that yield the same, minimum value of the sum of squared residuals.

• Estimate parameters of models from sample
• which values of \( \mu \) and \( \alpha_j \) minimize \( \Sigma e_{ij}^2 \)?

\[\sum_{j} \alpha_j = 0\]
Comparing Full & Reduced Models

- We expect full model to fit the data better
- Why?
  - because it has more free parameters (is more complex)
  - adjusting values of alpha is expected to reduce $\Sigma e_j^2$
  - even when true values of alphas in population are zero
- $H_0: \alpha_j = 0$ (for all j’s)  $H_1: \alpha_j \neq 0$ (for at least 1 group, j)
  - when $H_0$ is true, all groups are selected from populations with same mean $\mu$
  - variation among $Y_j$ is due to sampling variation
  - reduced model should provide a reasonably good fit to data
  - difference between $\Sigma e_j^2$ for full & reduced model should not be unusually large
- Question: is change in $\Sigma e_j^2$ unusually large assuming $H_0$ is true (and true alphas are 0)?

Measure of relative goodness-of-fit

$$F = \frac{(E_R - E_F)/ (df_R - df_F)}{E_F/(df_F)}$$

- $E_R$: $\Sigma e_j^2$ for reduced model
- $E_F$: $\Sigma e_j^2$ for full model
- $df_R$: degrees-of-freedom reduced model
- $df_F$: degrees-of-freedom full model

F distribution when $H_0$ is true

$$F = \frac{(E_R - E_F)/(df_R - df_F)}{E_F/(df_F)}$$

$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$
$H_1: \alpha_j \neq 0$

When $H_0$ is true, F statistic follows F distribution with $(df_R - df_F)$ & $(df_F)$ degrees of freedom

R example (mood induction experiment)
mood induction experiment

```r
> library(sciplot)
> par(mfrow=(1,1), cex=1.5)
> with(data=mood.data, 
  bargraph.CI(x.factor=group, 
  response=mood, 
  main="Mood Data") )
```

```r
> summary(mood.data)
group         mood
neutral   :10   Min.   :1.00
pleasant  :10   1st Qu.:3.25
unpleasant:10   Median :4.00
Mean   :4.33
3rd Qu.:5.00
Max.   :7.00

# grand mean
> mean(mood.data$mood)
[1] 4.3333

# group means
> with(mood.data,tapply(mood,group,mean))
group      neutral pleasant unpleasant
          4        6        3

# group SD
> with(mood.data,tapply(mood,group,sd))
group      neutral pleasant unpleasant
          0.6667    1.1547    1.0541
```

Comparison of full & restricted models

```r
# load data:
> load(file=url("http://pnb.mcmaster.ca/bennett/psy710/datasets/mood_data.rda"))

# fit models:
> mood.full <- lm(mood~1+group,data=mood.data)
> mood.restricted <- lm(mood~1,data=mood.data)
```

```r
Comparison of full & restricted models

# extract residuals:
> (E.full<-sum(residuals(mood.full)^2))
[1] 26
> (E.restricted<-sum(residuals(mood.restricted)^2))
[1] 72.667
> (E.restricted-E.full)
[1] 46.667
> (df.full<-mood.full$df.residual)
[1] 27
> (df.restricted<-mood.restricted$df.residual)
[1] 29

# compare SSresids with F test:
> (F <- ( (E.restricted-E.full)/(df.restricted-df.full)/(E.full/df.full) ) )
[1] 24.231
> (p.value <- 1-pf(F,df1=(df.restricted-df.full),df2=df.full) )
```

Comparison of full & restricted models

```
Comparison of full & restricted models

# # load data:
> load(file=url("http://pnb.mcmaster.ca/bennett/psy710/datasets/mood_data.rda"))

# # fit models:
> mood.full <- lm(mood~1+group,data=mood.data)
> mood.restricted <- lm(mood~1, data=mood.data)
```

ANOVA tables

```
ANNOVA tables

# # extract residuals:
> (E.full<-sum(residuals(mood.full)^2))
[1] 26
> (E.restricted<-sum(residuals(mood.restricted)^2))
[1] 72.667
> (E.restricted-E.full)
[1] 46.667
> (df.full<-mood.full$df.residual)
[1] 27
> (df.restricted<-mood.restricted$df.residual)
[1] 29

# # compare SSresids with F test:
> (F <- ( (E.restricted-E.full)/(df.restricted-df.full)/(E.full/df.full) ) )
[1] 24.231
> (p.value <- 1-pf(F,df1=(df.restricted-df.full),df2=df.full) )
```
ANOVA tables

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>a - 1</td>
<td>SS_B</td>
<td>MS_B</td>
<td>MS_B/MS_W</td>
<td>p</td>
</tr>
<tr>
<td>Residuals</td>
<td>a(n - 1)</td>
<td>SS_W</td>
<td>MS_W</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: A standard ANOVA table for a one-way design.

ANOVA tables

lm()

options(contrasts=c("contr.sum","contr.poly")) # IMPORTANT!
> # use lm()
> mood.lm <- lm(mood~1+group,data=mood.data)

> class(mood.lm)
[1] "lm"

> anova(mood.lm)

Analysis of Variance Table
Response: mood
     Df Sum Sq Mean Sq F value  Pr(>F)
group   2 46.7 23.33 24.2 9.4e-07 ***
Residuals 27 26.0   0.96

ANOVA tables

interpret as changes in goodness of fit in nested models

> anova(mood.lm)

Analysis of Variance Table
Response: mood
     Df Sum Sq Mean Sq F value Pr(>F)
group   2 46.7 23.33 24.2 9.4e-07 ***
Residuals 27 26.0   0.96

Table 1: A standard ANOVA table for a one-way design.

From Equation 14 it is possible to show that Squared Error (i.e., SS values for Residuals often are referred to as group subjects (i.e., freedom. The total degrees of freedom is equal to one less than the total number of other item listed in the Source column, Residuals, has in Table 1. The independent variable, Group, has the same number (3.3.3 Relation to ANOVA

Imagine an experiment in which we measure some aspect of behaviour on

The elements of Equation 13 are closely tied to various components of a standard

distribution, we can calculate the probability of obtaining a value of

F distribution, we can calculate the probability of obtaining a value of

the

F

distribution, we can calculate the probability of obtaining a value of

is change in goodness-of-fit when alphas are set to zero

Is this change in goodness-of-fit unusual when H0 is true?

Calculate F statistic...
ANOVA tables

\[ F = \frac{(E_R - E_F)/(\text{df}_R - \text{df}_F)}{E_F/\text{df}_F} \]

options(contrasts=c("contr.sum","contr.poly")) # IMPORTANT!
> # use lm()
> mood.lm <- lm(mood~1+group,data=mood.data)

> class(mood.lm)
[1] "lm"

> anova(mood.lm)

Analysis of Variance Table
Response: mood
   Df  Sum Sq Mean Sq  F value  Pr(>F)
group  2   46.7  23.33    24.2 9.4e-07 ***
Residuals 27  26.0   0.96

ANOVA tables

\[ F = \frac{(E_R - E_F)/(\text{df}_R - \text{df}_F)}{E_F/\text{df}_F} \]

options(contrasts=c("contr.sum","contr.poly")) # IMPORTANT!
> # use lm()
> mood.lm <- lm(mood~1+group,data=mood.data)

> class(mood.lm)
[1] "lm"

> anova(mood.lm)

Analysis of Variance Table
Response: mood
   Df  Sum Sq Mean Sq  F value  Pr(>F)
group  2   46.7  23.33    24.2 9.4e-07 ***
Residuals 27  26.0   0.96

ANOVA tables

\[ F = \frac{(E_R - E_F)/(\text{df}_R - \text{df}_F)}{E_F/\text{df}_F} \]

options(contrasts=c("contr.sum","contr.poly")) # IMPORTANT!
> # use lm()
> mood.lm <- lm(mood~1+group,data=mood.data)

> class(mood.lm)
[1] "lm"

> anova(mood.lm)

Analysis of Variance Table
Response: mood
   Df  Sum Sq Mean Sq  F value  Pr(>F)
group  2   46.7  23.33    24.2 9.4e-07 ***
Residuals 27  26.0   0.96

• Our observed F is unusual when H0 is true.
• Reject H0 in favor of H1 \( \alpha_j \neq 0 \) for at least one group.
ANOVA tables

$$F = \frac{(E_R - E_F)/(df_R - df_F)}{E_F/(df_F)}$$

options(contrasts=c("contr.sum","contr.poly"))  # IMPORTANT!
> # use aov()
> mood.aov <- aov(mood~1+group,data=mood.data)
> class(mood.aov)
[1] "aov" "lm"
> anova(mood.aov)

# IMPORTANT!

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>a - 1</td>
<td>SS_B</td>
<td>MS_B</td>
<td>MS_B/MS_W</td>
<td>p</td>
</tr>
<tr>
<td>Residuals</td>
<td>a(n - 1)</td>
<td>SS_W</td>
<td>MS_W</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: A standard ANOVA table for a one-way design.

\[ E_R = SS_B \]

\[ E_F = SS_W \]

\( E_f \) and \( E_R \) are \( \sum e_i^2 \) for full & reduced models

ANOVA tables

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>a - 1</td>
<td>SS_B</td>
<td>MS_B</td>
<td>MS_B/MS_W</td>
<td>p</td>
</tr>
<tr>
<td>Residuals</td>
<td>a(n - 1)</td>
<td>SS_W</td>
<td>MS_W</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: A standard ANOVA table for a one-way design.

\[ E_R = SS_B + SS_W = SS_{Total} \]

\[ E_F = SS_W \]

\( E_f \) and \( E_R \) are \( \sum e_i^2 \) for full & reduced models
ANOVA tables

\[ E_R - E_F = SS_{total} - SS_W = SS_B \]

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>( a - 1 )</td>
<td>( SS_B )</td>
<td>MS_B</td>
<td>MS_B/MS_W</td>
<td>( p )</td>
</tr>
<tr>
<td>Residuals</td>
<td>( a(n - 1) )</td>
<td>( SS_W )</td>
<td>MS_W</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: A standard ANOVA table for a one-way design.

\[ E_R = SS_B + SS_W = SS_{total} \]

\[ E_F = SS_W \]

\[ F = \frac{(E_R - E_F)/(df_R - df_F)}{E_F/df_F} = \frac{MS_B}{MS_W} \]

ANOVA tables

\[ E_R - E_F = SS_{total} - SS_W = SS_B \]

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>( a - 1 )</td>
<td>( SS_B )</td>
<td>MS_B</td>
<td>MS_B/MS_W</td>
<td>( p )</td>
</tr>
<tr>
<td>Residuals</td>
<td>( a(n - 1) )</td>
<td>( SS_W )</td>
<td>MS_W</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: A standard ANOVA table for a one-way design.

\[ E_R = SS_B + SS_W = SS_{total} \]

\[ E_F = SS_W \]

\[ MS = \frac{SS}{df} \]

MS values are estimates of variance
We fit the following two models to the data:

- Full model: $Y_{ij} = \mu + \alpha_j + e_{ij}$
- Reduced model: $Y_{ij} = \mu + e_{ij}$

Here, $\alpha_j = \mu_j - \mu_G$ represents the group effect, $\mu$ is the overall mean, and $\mu_j$ is the group mean.

For the full model, it can be shown that the sum of squared residuals for Equation 8 is minimized by setting $\alpha_j = \mu_j - \mu_G$. For the reduced model, even after taking into account its greater complexity, the question of interest is whether the full model provides a better fit to the data than the reduced model. This is often assessed using model selection criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC).

### ANOVA tables

**Estimates of Population Error Variance**

$\text{MS}_W = \frac{E_F}{df_F} = \hat{\sigma}_e^2$

$\text{MS}_W$ is the weighted average of within-group variances.

### ANOVA tables

**MS_{void}** = AVG(within-group variance)

```r
> anova(mood.lm)
Analysis of Variance Table
Response: mood
  Df  Sum Sq Mean Sq  F value Pr(>F)
group  2   46.7  23.33    24.2  <.001
Residuals 27  26.0   0.96

> # mean of group variances:
> mean(with(mood.data, tapply(mood, group, var)))
[1] 0.96296
```

**MS_{group}** = n x (between-group variance)

```r
> anova(mood.lm)
Analysis of Variance Table
Response: mood
  Df  Sum Sq Mean Sq  F value Pr(>F)
group  2   46.7  23.33    24.2  <.001
Residuals 27  26.0   0.96

> n <- 10
> ( groupMeans <- with(mood.data, tapply(mood, group, mean)) )
neutral   pleasant unpleasant
4  6  3

> n * var(groupMeans)
[1] 23.33333
```
Estimates of Population Error Variance
when H0 is TRUE variation among group means is pure SAMPLING VARIANCE

- between-group variance has 2 sources
  - real differences among µ values
  - sampling variation
- if H0 is true: αj = 0
- between-group variance is caused only by sampling variation

\[ \sigma_i^2 = \frac{\sigma^2}{n} \]

Estimates of Population Error Variance
when H0 is FALSE variation among groups caused by SAMPLING VARIANCE plus group effects

- if H0 is false: αj ≠ 0 (for at least group)
- between-group variance is caused by sampling variation & alpha’s
  - \[ \sigma_i^2 = \frac{\sigma^2}{n} + \frac{\sum \alpha_j^2}{a - 1} \]
  - E(MSb) = n × \sigma^2 + \frac{n \sum \alpha_j^2}{a - 1}
- and we expect MSb > MSw as \( \Sigma \alpha_j^2 \) increases, and \( F \gg 1 \)
- evaluate with 1-tailed tests... look for unusually large values of F

When H0 is true & \( \alpha_j = 0 \)
\[ E(\text{MSb}) = E(\text{MSw}) \]
\[ F \approx 1 \]

When H0 is false & \( \alpha_j ≠ 0 \)
\[ E(\text{MSb}) > E(\text{MSw}) \]
\[ F > 1 \]
ANOVA tables

MS values are estimates of variance

> anova(mood.lm)
Analysis of Variance Table
Response: mood
  Df  Sum Sq Mean Sq F value Pr(>F)
group  2   46.7   23.33  24.2 9.4e-07 ***
Residuals 27   26.0    0.96

When H0 is true:
MS\text{resid} & MS\text{group} are independent estimates of VAR(error)
F = MS\text{group}/MS\text{resid} = 1 (distributed as F(2,27))

When H1 is true:
MS\text{resid} = estimate of VAR(error)
MS\text{group} = estimate of VAR(error) + [positive number that depends on \alpha’s]
F = MS\text{group}/MS\text{resid} >> 1

ANOVA tables

Using F to test null hypothesis

When H0 is true:
MS\text{resid} & MS\text{group} are independent estimates of VAR(error)
F = MS\text{group}/MS\text{resid} = 1 (distributed as F(2,27))

When H1 is true:
MS\text{resid} = estimate of VAR(error)
MS\text{group} = estimate of VAR(error) + [positive number that depends on \alpha’s]
F = MS\text{group}/MS\text{resid} >> 1

ANOVA tables

interpretations of H0 & H1

> anova(mood.lm)
Analysis of Variance Table
Response: mood
  Df  Sum Sq Mean Sq F value Pr(>F)
group  2   46.7   23.33  24.2 9.4e-07 ***
Residuals 27   26.0    0.96

H0: all group effects (alpha’s) are zero
H1: not all group effects (alpha’s) are zero

H0: all populations have the same mean
H1: not all populations have the same mean

Assuming all of the effects are zero, is the change in goodness of fit (SS error) unusually large?

Assuming all of the effects are zero, is the variation among group means unusually large?

effect size & association strength
**effect size (Cohen’s f)**

\[ d_i = \frac{(\mu_1 - \mu_3)}{\sigma_e} \]

\[ f = \sqrt{\frac{\sum_{i=1}^{3} d_i^2}{3}} \]

```r
> library(effectsize)
> coens_f(mood.lm)
Parameter | Cohen's f | 90% CI
--- | --- | ---
group | 1.34 | [0.88, 1.75]
```

**association strength**
proportion of variance accounted for by group

\[ \tilde{R}^2 \]

\[ \omega^2 = \frac{\sigma^2_{\theta}}{\sigma_{\theta}^2 + \sigma^2_{e}} \]

\( \omega^2 = 0.01 \) is a small association

\( \omega^2 = 0.06 \) is a medium association

\( \omega^2 = 0.14 \) or larger is a large association

**association strength**
proportion of variation that is accounted for by group

```r
> summary(mood.lm)
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.333  0.179 24.190  < 2e-16
group1   -0.333  0.253  -1.313    0.19
group2    1.667  0.253   6.580  < 2e-10
Residual standard error: 0.981 on 27 degrees of freedom
Multiple R-squared: 0.642, Adjusted R-squared: 0.616
F: 24.2 on 2 and 27 DF, p-value: 9.42e-07
```

```r
> eta_squared(mood.lm)
Parameter | Eta2 | 90% CI
--- | --- | ---
group | 0.64 | [0.43, 0.75]
```

\[ \eta^2 = R^2 \]

```r
> omega_squared(mood.lm)
Parameter | Omega2 | 90% CI
--- | --- | ---
group | 0.61 | [0.39, 0.73]
```

\[ \omega^2 \approx \tilde{R}^2 \]

**assumptions**

- VAR(residuals) = 0.896
- VAR(original scores) = 2.506
- so we accounted 64%, of VAR(original scores)
Assumptions of ANOVA

- The scores must be statistically independent of each other
- Population of scores, Y, distributed normally within each group
  - equivalent to assuming that error follows normal distribution
- Error variance is constant across groups
- If assumptions are met, F statistic follows F distribution
  - if they are not met, F statistic does not follow F distribution

Non-normality & non-constant variance

- ANOVA reasonably robust to deviations from normality
  - if deviations are similar in all groups
  - less robust to deviations that differ across groups
  - e.g., positive skew in 1 group and negative skew in others
  - robustness also declines if n is not equal across groups
- ANOVA is reasonably robust to 3-4 fold differences in variances
  - if Y's normally distributed and equal n per group

Tests for non-normality

- Kolmogorov-Smirnov test: ks.test()
- Shapiro-Wilk’s test: shapiro.test()
- Both tests have low power, though shapiro.test is better
  - to compensate, could use tests with alpha = 0.10

```r
shapiro.test(residuals(mood.full))
# H0: residuals are distributed normally
```

Bartlett.test

test for homogeneity of variance

```r
bartlett.test(mood~group, data=mood.data)
bartlett.test(mood.data$mood, mood.data$group)
# Bartlett test of homogeneity of variances
# data: mood.data$mood and mood.data$group
# Bartlett's K-squared = 2.6, df = 2, p-value = 0.3
# H0: variance is constant across groups
( also see leveneTest in car package )
```
Alternative Analyses

- perform ANOVA on transformed data
  - square-root, log, & inverse-sine transformations common
  - conclusions apply to transformed data
- Welch correction for non-constant variance
  - oneway.test () [N.B. Assumes normality]

```r
oneway.test(mood~group, data=mood.data)
```

```r
## One-way analysis of means (not assuming equal variances)
## data: mood and group
## F = 18, num df = 2, denom df = 17, p-value = 6e-05
```

Alternative Analyses

- Kruskal-Wallis test for group differences [kruskal.test()]
  - does not assume normality or constant variance
  - HO: the means of ranked data are the same in each group
  - if distributions for each group have same shape (not necessarily normal), then KW test evaluates null hypothesis that group MEDIANS are equal

```r
kruskal.test(mood~group, data=mood.data)
```

```r
## Kruskal-Wallis rank sum test
## data: mood by group
## Kruskal-Wallis chi-squared = 19, df = 2, p-value = 7e-05
```