BMEG 802 – Advanced Biomedical Experimental Design and Analysis

Two Way (Between) Analysis of Variance (ANOVA)

Joshua G. A. Cashaback, PhD

Recap

- One-Way (Between) ANOVA
  - Linear Model Approach
  - Test normality and sphericity assumption
  - Multiple mean comparisons

Today

- Two-Way ANOVA
  - linear model approach
  - interpret main effects and interactions
  - follow up mean comparisons
- n-Way ANOVA
  - general concepts
  - limitations

Two Factor Design

In the example above we have two factors

- Factor A (e.g., Drug) with 2 levels (e.g., drug vs. no drug)
- Factor B (e.g., Biofeedback) with 2 levels (e.g., biofeedback vs. no biofeedback)

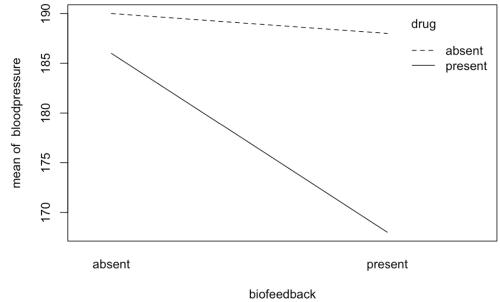
Fully crossed design

- every level of factor A is tested with every level of factor B
- total # groups (cells) is a x b

We will see how to formulate in terms of model comparisons:

- main effect of A
- main effect of B
- interaction effect A x B

2-Way ANOVA



Same approach as before

1. write the equation for the full and restricted models
2. derive the equations for model error $E_{restricted}$ and E_{full}
3. derive the expressions for degrees of freedom $df_{restricted}$ and df_{full}
4. end up with an equation for the F ratio

The Full Model

$$Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha \cdot \beta)_{jk} + \epsilon_{ijk}$$

- Y_{ijk} is an individual score in the jth level of factor A and the kth level of factor B (i indexes subjects within each (j,k) cell)
- μ is the overall mean of all cells
- α_j is the effect of the jth level of factor A
- β_k is the effect of the kth level of factor B
- $(\alpha \cdot \beta)_{jk}$ is the interaction effect of level j of A and level k of B

Hypothesis testing using Restricted Models

Two-Factor (A x B) design: 3 null hypotheses to be tested:

- main effect of A
- main effect of B
- interaction effect of A x B

We will formulate a separate restricted model for each hypothesis test

- each test will involve the same full model
- we will use the usual F test

$$F = \frac{(E_{restricted} - E_{full})/(df_{restricted} - df_{full})}{(E_{full}/df_{full})}$$

Main Effect of A

Full Model: $Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha \cdot \beta)_{jk} + \epsilon_{ijk}$

null hypothesis is that A main effect is zero.

- $H_0 : \alpha_1 = \alpha_2 = ... = \alpha_n = 0$

Restricted Model: $Y_{ijk} = \mu + \beta_k + (\alpha \cdot \beta)_{jk} + \epsilon_{ijk}$

F-Statistic for Main Effect A

$$E_{full} = \sum_{j=1}^{a} \sum_{k=1}^{b} \sum_{i=1}^{n} (Y_{ijk} - \bar{Y}_{jk})^2$$

$$df_{full = a \cdot b(n-1)}$$

$$E_{restricted} - E_{full} = nb \sum_{j=1}^{a} (\bar{Y}_j - \bar{Y})^2$$

$$df_{restricted} - df_{full} = a - 1$$

see Maxwell Delaney, Kelley (Chapter 7) for derivations

Now we can do our F-test!

$$F = \frac{(E_{restricted} - E_{full})/(df_{restricted} - df_{full})}{(E_{full}/df_{full})}$$

Main Effect of B

Full Model: $Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha \cdot \beta)_{jk} + \epsilon_{ijk}$

null hypothesis is that B main effect is zero.

- $H_0 : \beta_1 = \beta_2 = ... = \beta_n = 0$

Restricted Model: $Y_{ijk} = \mu + \alpha_j + (\alpha \cdot \beta)_{jk} + \epsilon_{ijk}$

F-Statistic for Main Effect B

$$E_{full} = \sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{n}(Y_{ijk} - \bar{Y}_{jk})^2$$

$$df_{full=a\cdot b(n-1)}$$

$$E_{restricted} - E_{full} = nb\sum_{k=1}^{b}(\bar{Y}_k - \bar{Y})^2$$

$$df_{restricted} - df_{full} = a - 1$$

Now we can do our F-test!

$$F = \frac{(E_{restricted} - E_{full})/(df_{restricted} - df_{full})}{(E_{full}/df_{full})}$$

- note: denominator of F-test is the same as mean-square within from ANOVA table.

Interaction Effect of A x B

Full Model: $Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha \cdot \beta)_{jk} + \epsilon_{ijk}$

Restricted Model: $Y_{ijk} = \mu + \alpha_j + \beta_k + \epsilon_{ijk}$

F-Statistic for A x B Interaction

$$E_{full} = \sum_{j=1}^{a}\sum_{k=1}^{b}\sum_{i=1}^{n}(Y_{ijk} - \bar{Y}_{jk})^2$$

$$df_{full = a \cdot b(n-1)}$$

$$E_{restricted} - E_{full} = n\sum_{j=1}^{a}\sum_{k=1}^{b}(\bar{Y}_{jk} - \bar{Y}_{j} - \bar{Y}_{k} + \bar{Y})^2$$

$$df_{restricted} - df_{full} = (a-1)(b-1)$$

Now we can do our F-test!

$$F = \frac{(E_{restricted} - E_{full})/(df_{restricted} - df_{full})}{(E_{full}/df_{full})}$$

- don't worry, I won't make you do this by hand...

2-Way ANOVA Example

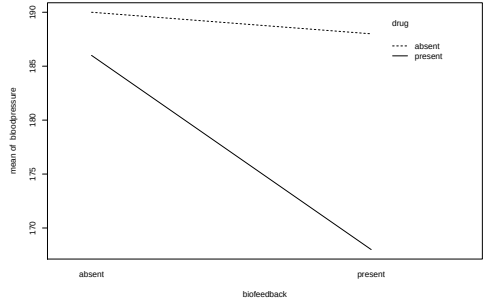
Hypothetical study: explore effects of biofeedback and drug therapy on blood pressure

- two independent variables: drug therapy and biofeedback

Biofeedback + Drug	Biofeedback, no Drug	no Biofeedback, Drug	no Biofeedback, no Drug
158	188	186	185
163	183	191	190
173	198	196	195
178	178	181	200
168	193	176	180
mean = 168	mean = 188	mean = 186	mean = 190
sd = 7.91	sd = 7.91	sd = 7.91	sd = 7.91

2-Way ANOVA Example

```
bloodpressure <- c(158,163,173,178,168,188,183,198,178,193,
    186,191,196,181,176,185,190,195,200,180)
biofeedback <- factor(c(rep("present",10),rep("absent",10)))
drug <- factor(rep(c(rep("present",5),rep("absent",5)),2))
bpdata <- data.frame(bloodpressure, biofeedback, drug)
interaction.plot(biofeedback, drug, bloodpressure)
```



2-Way ANOVA Example

```
myanova <- aov(bloodpressure ~ biofeedback*drug)
summary(myanova)
```

```
##                   Df Sum Sq Mean Sq F value  Pr(>F)
## biofeedback        1    500   500.0    8.00 0.01211 *
## drug               1    720   720.0   11.52 0.00371 **
## biofeedback:drug   1    320   320.0    5.12 0.03792 *
## Residuals         16   1000    62.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Significant interaction of drug, biofeedback, main effect of drug, main effect of biofeedback! How do we interpret? How do we perform followup mean comparisons?

Testing Normality

Before we get to interpretation and followup mean comparisons, lets check normality and homogenous of variance.

```
shapiro.test(bloodpressure[1:5])$p.value # group 1
```

```
## [1] 0.9671739
```

```
shapiro.test(bloodpressure[6:10])$p.value # group 2
```

```
## [1] 0.9671739
```

```
shapiro.test(bloodpressure[11:15])$p.value # group 3
```

```
## [1] 0.9671739
```

```
shapiro.test(bloodpressure[16:20])$p.value # group 4
```

```
## [1] 0.9671739
```

No violations of normality

Homogenous of Variance (Sphericity)

Test for sphericity (homogeneous of variance) in each main effect and interaction

```
bartlett.test(bloodpressure ~ interaction(drug, biofeedback))$p.value
```

```
## [1] 1
```

```
bartlett.test(bloodpressure ~ drug)$p.value
```

```
## [1] 0.1760607
```

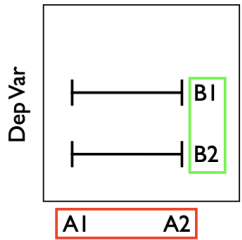
```
bartlett.test(bloodpressure ~ biofeedback)$p.value
```

```
## [1] 0.1440305
```

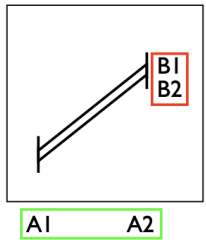
No violations of sphericity

Can use Anova() in the Car package and use Greenhouse-Geisser corrections (adjusts df on F-tests) if violations.

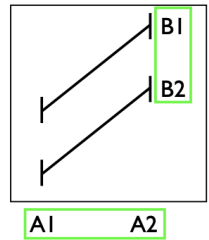
Interpretating Main Effects



Main effect of B



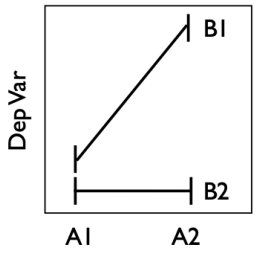
Main effect of A



Main effects of A and B

in all 3 cases: **no A x B interaction effect**

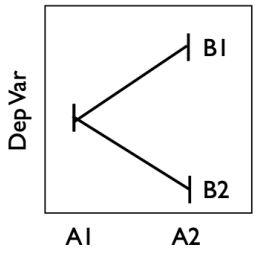
Interpretating Interactions



A:

B:

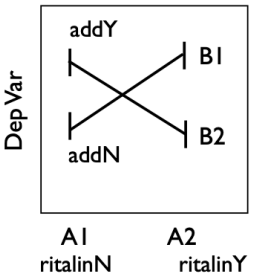
AxB:



A:

B:

AxB:

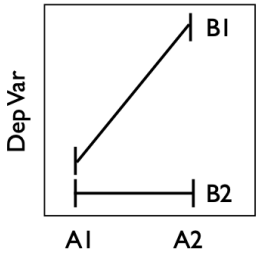


A:

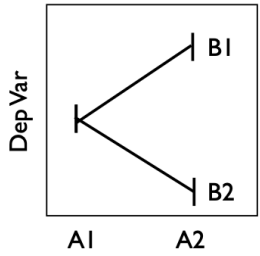
B:

AxB:

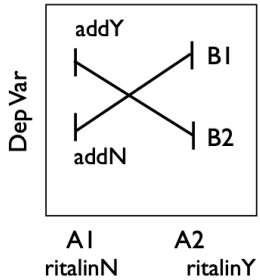
Interpretating Interactions



A:
B:
AxB:



A:
B:
AxB:



A:
B:
AxB:

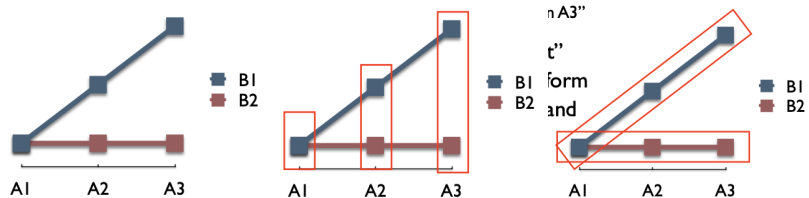
Rule of Thumb: parallel lines = main effect, non-parallel lines = interaction effect

Mean Comparison Procedure

First look at the interaction effect

IF interaction effect is significant,

- perform mean comparisons (e.g., t-tests)
- i.e. (investigate the nature of the interaction)
- pick what you care about more (e.g., **drug** vs. biofeedback)
  - e.g., no drug vs drug (biofeedback) and no drug vs. drug (no biofeedback)
- DON'T bother looking at involved main effects (not informative)



Mean Comparison Procedure

ELSE if interaction effect is not significant,

- perform mean comparisons within each significant main effect to understand the nature of the differences
- e.g., for a main effect of drug, 'Collapse across' biofeedback and compare drug (n = 10) vs. no drug (n = 10) in our example. Likewise for a main effect of biofeedback.

note: for three way ANOVA, can have a significant two way interaction but can look at the non-involved main effect (assuming no three-way interaction)

Correcting for Multiple Mean Comparisons

- Done for **each 'Family'** (e.g., main effect only, interaction only)
  - i.e., you don't have to correct for multiple comparisons *across* each family
- e.g., Bonferroni-Holm or others (e.g., Tukey HSD)

Interaction Effect, Mean Comparisons

Let's say we are more interested in differences in drugs

```
# nobiofeed, nodrug vs nobiofeed, drug
pval_nbnd_v_nbd = t.test(bloodpressure[1:5], bloodpressure[6:10], alternative = "two.sided")$p.value
# biofeed, nodrug vs biofeed, drug
pval_bnd_v_bd = t.test(bloodpressure[11:15], bloodpressure[16:20], alternative = "two.sided")$p.value
pvals = c(pval_nbnd_v_nbd, pval_bnd_v_bd)
p.adjust(pvals, method = "holm", n = length(pvals))
```

```
## [1] 0.007899546 0.446813334
```

With biofeedback present, there is significantly lower blood pressure when taking the drug compared to not taking the drug ($p = 0.008$).

Main Effect, Mean Comparisons

Typically we would stop after a significant two-way interaction in a 2-way ANOVA. But, lets carry out the mean comparison's for the two main effects so that you know how to do it.

```
# drug vs no drug --- group all data as with or without drug
drug_main = c(bloodpressure[1:5],bloodpressure[11:15])
nodrug_main = c(bloodpressure[6:10],bloodpressure[16:20])
pval_d_v_nd = t.test(drug_main, nodrug_main, alternative = "two.sided")$p.value
# biofeedback vs no biofeedback
biof_main = c(bloodpressure[1:10])
nobiof_main = c(bloodpressure[11:20])
pval_bf_v_nbf = t.test(biof_main, nobiof_main, alternative = "two.sided")$p.value
pval_d_v_nd
```

```
## [1] 0.01746623
```

```
pval_bf_v_nbf
```

```
## [1] 0.05333271
```

The drug leads to significantly lower blood pressure ($p = 0.017$). Biofeedback did not lead to significantly lower blood pressure ($p = 0.053$). REMEMBER, we were just doing this as an example of followup mean comparisons for a main effects (even though there was a significant interaction, which makes these main effects not meaningful).

Effect Sizes

install.packages("effectsize") install.packages("effsize")

```
library(effectsize)
library(effsize)
omega_squared(myanova)
```

```
## Parameter         | Omega2 (partial) |       90% CI
## ------------------------------------------------------
## biofeedback       |             0.26 | [0.02, 0.51]
## drug              |             0.34 | [0.06, 0.58]
## biofeedback:drug  |             0.17 | [0.00, 0.44]
```

```
d_dvnd = cohen.d(drug_main, nodrug_main, var.equal = False)$estimate
abs(d_dvnd)
```

```
## [1] 1.193388
```

Summary of Example

We found a significant interaction between drug and biofeedback ($p = 0.038$, $\omega_p^2 = 0.17$). With biofeedback present, there is significantly lower blood pressure when taking the drug compared to not taking the drug ($p = 0.008$, $d = 1.193$).

note: we do not consider main effects and their followup mean comparisons since there was a significant interaction.

Power

Let's say we have a two-factor design, with factors A (three levels) and B (two levels). W expect the effect size for the main effect of A to be medium ($f = 0.25$, according to Cohen, 1988).

First, lets figure out $u = df_{restricted} - df_{full} = (a-1)(b-1) = (3-1)(2-1) = 2$ for the interaction term.

Our unknown is $v = df_{full} = a \cdot b(n-1)$, where we want to know n.

Power

install.packages("pwr")

```
library(pwr)
# u = levels, v = dof residuals (df_residuals = a x b x (n-1))
pwr.f2.test(u=2, v=NULL, f2=0.25^2, sig.level=0.05, power=0.80)
```

```
##
##      Multiple regression power calculation
##
##               u = 2
##               v = 154.1898
##              f2 = 0.0625
##       sig.level = 0.05
##           power = 0.8
```

$v = 154.2 = a \cdot b(n-1)$ With simple rearranging, $n = 26.7$. Thus we need 27 participants per group in our 3 x 2 design to be sufficiently powered.

Advantages of a Factorial Design

- factorial design enables us to test for an interaction
- factorial design allows for greater generalizability
- factorial design can produce the same statistical power as 2 single-factor designs using half as many subjects!

N-Way ANOVA

Eg., three-way ANOVA: 2 X 2 X 2 Design

- 3 Main Effects (A, B, and C)
- 3 Two-Way Interactions (AxB, AxC, BxC)
- 1 Three-Way Interaction (AxBxC)

Two-Way Interactions in a 3 Factor Design

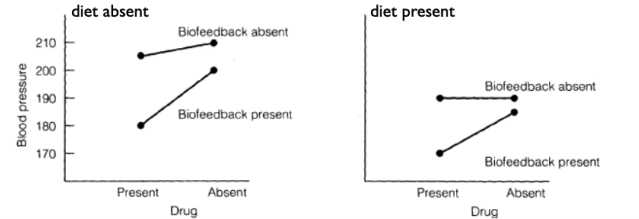
- Main Effects are the same
- Two-way interactions here are slightly trickier
  - AB interaction
    - when averaged over Factor C, the effect of Factor A is different depending on the level of Factor B (or vice versa)
    - same principle applies for AC, BC interactions.

Three-Way Interactions in a 3 Factor Design

Meaning of three-way interaction (AxBxC)

- AB interaction is different depending on the level of C
- AC interaction is different depending on the level of B
- BC interaction is different depending on the level of A

Example: Add diet to our previous example:



3-way Procedure

- look at 3-way interaction first. if significant,
  - perform follow up mean comparisons
  - do not look at 2-way interactions or main effects
- no significant 3-way interaction? Move to two-way interactions
  - if significant two-way (e.g., AxB)
    - peform follow up mean comparisons
    - don't look at involved main effects (main effect of A, main effect of B)
    - Can look at main effect of C if significant AxB, and no significant AxC, BxC interactions.
- no significant 2-way interactions? Move to main effects.

I suggest using the most simple design possible (personally, never greater than a two-level ANOVA). Harder to interpret, articulate, explain or visualize 3-way, 4-way interactions.

Unbalanced Design

Balanced design means that you have the same number of data points per group. If unbalanced, use the Anova() function: install.packages("car"), library(car). Look up R documentation for details.

Next Week

- Within (Repeated Measures) ANOVA
- Friedman Test