Stat E-150 Section #4 (but really our first section!)

Neha Dhawan

dhawan@g.harvard.edu

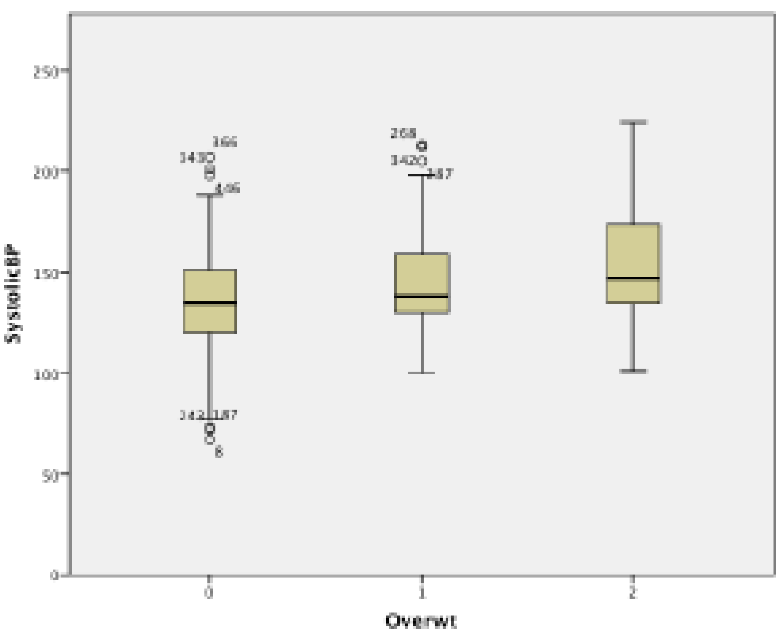
Office hours by request for one-one help

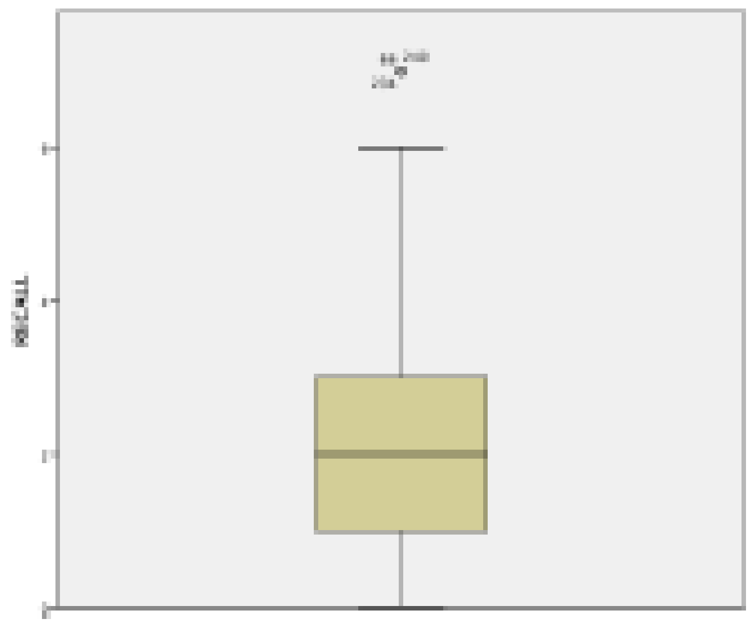
Logistics

- Section is for you!
  - Email dhawan@g.harvard.edu by **Saturday midnight** for questions you want to ask or topics you want to cover in depth
- Section focus
  - Lecture material + SPSS run through intertwined
  - Homework Questions after
- Learn using examples

ANOVA and groups

- ANOVA compare means- more than 2 levels
- Wanna know if one big ol' mean is fine or differentiating into groups helpful
- Error is: predicted value – its group mean





Cereal

Let’s say you are a dietician trying to figure out the best way to predict the calories in Cereal. You are curious about the difference in calories amongst cereals with different main components.

0. Ask your research question: What is it you want to know?

Does the main ingredient in cereal affect the number of calories it has?

1. **Choose a form for the model: Identify the variables and their types**; Examine graphs to help identify the appropriate model

Calories- response- quantitative

Main component- predictor- categorical

- In ANOVA predictor variables are called ‘factors’
- Each factor has ‘levels’
  - Component: 3 levels= Oat, wheat, and bran

- 1 categorical predictor; 1 quantitative response variables

One-way ANOVA!!

2. **Fit the model to the data**: Use the sample data to **create statistics**, which estimate the values of the model parameters

Q. Is there a difference between the groups?

H_0: $\mu_{Wheat} = \mu_{Bran} = \mu_{Oats}$

H_a: the means are not all equal

Sine $p<0.001$, we can reject the null hypothesis, and conclude that the **mean** amount of calories significantly differ between cereal types.

But this doesn't tell us **how** they differ!

3. Assess how well the model fits the data: **Verify assumptions,** examine the residuals, compute significance, refine model as needed

- **Random**- random sample
- Distribution of errors
  - **Zero Mean-** always true due to analysis
  - **Constant Variance**- same standard deviation <u>for each group</u>
    - Residual plot
    - Rule of thumb: largest std devi/ smallest.
    - Levene's test of equal variance
  - **Normality** - NPP
  - **Independence**- errors independent of each other; presume with random assumption

2/3 variance tests should agree
Try transforming variable if all fail

Rule of thumb

- Get a value <2 OK
- Value >2 concern
  - Can accept >2 if sample size per group is small or if deign is balanced

Descriptive Statistics

Dependent Variable: Calories

Main_Component	Mean	Std. Deviation	N
Bran	30.111	21.5954	9
Oats	107.083	6.8418	12
Wheat	54.600	14.5101	15
Total	65.972	34.2215	36

$$\frac{\text{Largest StDevi}}{\text{Smallest StDevi}} = \frac{21.595}{6.841} = 3.157$$ **>2 OHOH!!**

Levene's test

$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$

H_a: Not all variances are equal

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Calories

F	df1	df2	Sig.
3.435	2	33	.044

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Main_Component

Since p-value <0.05, we reject the null hypothesis that conclude all the variances are NOT equal

- So we know our groups differ but which ones differ and how?
  - Post hoc tests!
    - When two or more groups; with two can see difference
  - Liberal LSD vs. middle ground Tukey’s vs. conservative Bonefferroni
    - Main diff is the distribution used
    - Usually use Tukey
  - Never run post hocs if main effect is not significant, and if we only have two levels

Why can't we use multiple two sample t-tests to see differences between means of the groups?

- If we do three (# of groups) tests, we need to concern ourselves with the family-wise error rate as well as the individual error rate. If we perform three two-sample tests (necessary to compare three groups) and each has a 5% individual error rate, there is a higher risk that at least one of the three tests is wrong. Family-wise error rate is the probability of making type I errors among all the hypotheses when performing multiple hypotheses tests.

Post hoc test

H_0: $\mu_i = \mu_j$

H_a: $\mu_i \neq \mu_j$

Based on the output we can see that cereal with Oats as the main component has significantly more calories than Bran or Wheat , and Wheat has significantly more calories than Bran (all $p<0.001$)

Research Methods!

- Linear Regression usually for observational study
  - Correlation ≠ Causation
- ANOVA usually comparative study
  - Control group
  - Manipulation of treatment groups
  - Randomization
  - Cause and effect

Terminology

- **Random assignment**- randomly assigning people in your sample to different treatment groups
  - Comparative studies
- **Random selection:** randomly selecting people from your population to be in your study
  - Observational studies

Terminology

- **Blocking** is the grouping of sources of variability (like gender, age, race etc.) into 'blocks' (i.e. group of similarities).
- **Factorial Crossing:** In ANOVA, two factors are crossed if all combinations of levels of two factors appear in the design
- **Balanced design** is if the factors are fully crossed and there are equal numbers of observations per cell(equal #s of experimental units)

Our example

Medical Researchers are interested in knowing which sort of intervention works best for patients with depression (measured using the Beck Depression Inventory). They had a sample of 90 men and women (1:1 ratio) and randomly assigned subjects to three groups: control, Cognitive Behavioural Therapy, and medications.

0. Ask your research question: What is it you want to know?

Do depression scores differ between the three intervention groups?

1. Choose a form for the model: Identify the variables and their types; Examine graphs to help identify the appropriate model

   – Intervention group; categorical– factor
   – Depression score; quantitative – response

We could either randomly assign all participants to treatment groups or factor out by Gender!

Create new categorical factor

- Choose model:
  - ANOVA: categorical predictors
  - Factoring out by Gender so Two-way ANOVA
  - Factorial Crossing!

- Observational study or experiment?
  - Comparative study/experiment
  - Manipulating variable/ treatment groups

- Random?
  - Random assignment – comparing groups

Blocking and Factorial crossing

Factorial Crossing

Factor A: Depression intervention **Treatments**

Factor B: Gender

	Control	CBT	Meds
Male	Male/ control 15	Male/CBT 15	Male/ meds 15
Female	Female / control 15	Female/CBT 15	Female/ meds 15

Blocks
-All levels of treatment; but have to randomize order!

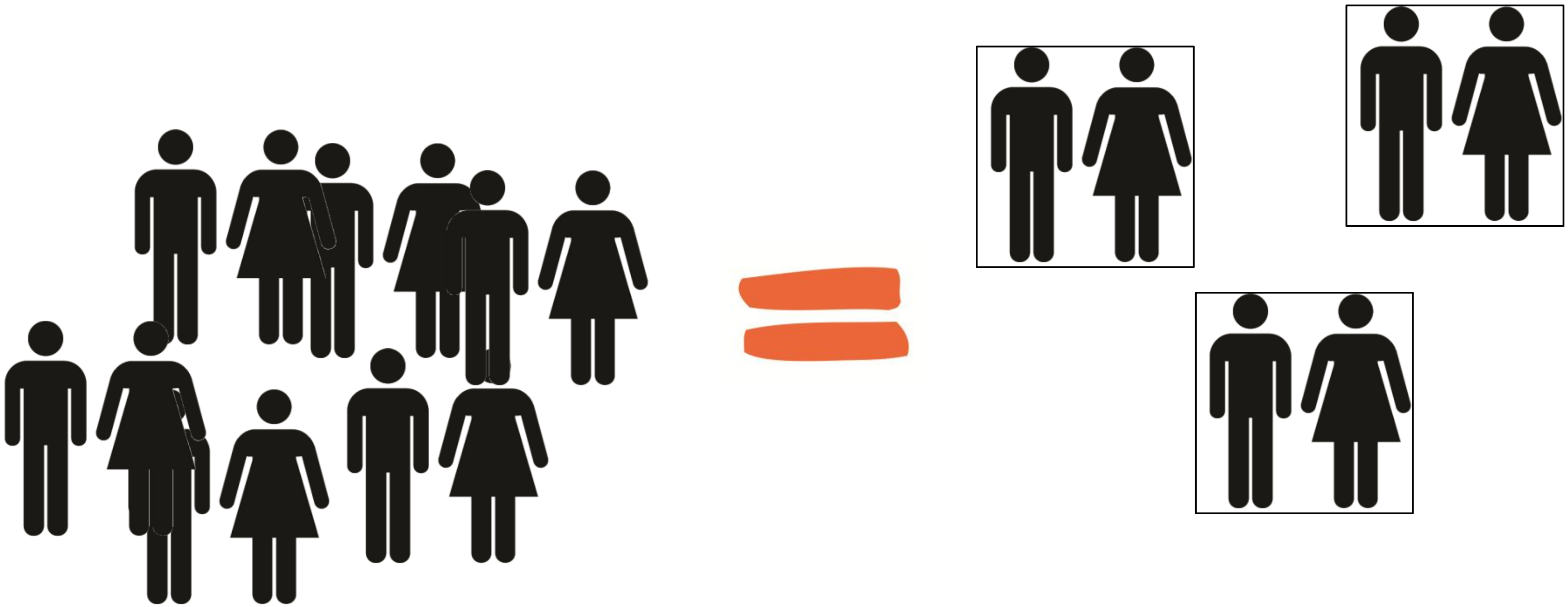
Balanced Design

Blocking as a within-subjects design

- Repeated measures ANOVA
- Each subject gets all treatments, each treatment is given once to all subjects, and the order of the conditions is randomized
- Benefit is that each person is their own control
- Block by sub-dividing or by sub-grouping

Blocking reduces variability

- Individuals can be blocks
- Blocking= is to categorize



Why do we block or factorial cross?

- A **nuisance** factor is a factor that has some effect on the response, but is of no interest to the experimenter; however, the variability it transmits to the response needs to be minimized or explained. Blocking deals with nuisance factors so it better explain the outcome.

- Factorial cross because we want to see the interaction of the two factors

Test time!!

A study reported in 1994 compared different psychological therapies for teenaged girls with anorexia. Each girl's weight was measured before and after a period of therapy designed in the experiment. One group used a cognitive-behavioral treatment, a second group received family therapy, and the third group was a control group which received no therapy. The subjects in this study were randomly assigned to these groups.

a) What kind of study is this? What analysis would you use for this data and why?

Tests of Between-Subjects Effects

Dependent Variable: **Weight Difference**

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	57.600[a]	5	11.520	144.500	.000
Intercept	346.800	1	346.800	.222	.642
Therapies	.533	1	.533	4.800	.004
Error	57.600	24	2.400		
Total	462.000	30			
Corrected Total	115.200	29			

b) What can you conclude from the output?

c) Would factorial crossing be an appropriate study design in this case? Why or why not?

Multiple Comparisons

Tukey HSD

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	2	-3.4569	2.0333	.212	-8.327	1.413
	3	-7.7147*	2.3482	.005	-13.339	-2.090
2	1	3.4569	2.0333	.212	-1.413	8.327
	3	-4.2578	2.2996	.161	-9.766	1.251
3	1	7.7147*	2.3482	.005	2.090	13.339
	2	4.2578	2.2996	.161	-1.251	9.766

*. The mean difference is significant at the 0.05 level.

d) What conclusions can you draw about the pairwise differences between Group 1 and 2 and Group 1 and 3? (group 1 used a cognitive-behavioral treatment, group 2 received family therapy, and third group 3 was a control group which received no therapy)

Solutions

a) This is **a comparative study or an experiment,** since we have **three treatment groups, a control group, and are using random assignment** to compare differences between each group. We can make a cause and effect judgment in this case. We would use **One-way ANOVA** to analyze our data because we have a **categorical predictor/factor (therapy)** with **three levels**: CBT, family and control, and **one quantitative response** variable: **weight difference** (10)

Solutions

b) **H_0: $\mu_1 = \mu_2 = \mu_3$**

H_a: the means are not all equal

Based on the **p-value of .004**, we **reject the null hypothesis**. The data provides **evidence of a difference in the mean weight difference for the three groups**. (5)

c) Factorial crossing would **be inappropriate** in this study since **we only have one factor of interest**. In order to have a factorial crossed design, **we would need two factors of interest and all combinations of the levels of the two factors should appear in the design** (4)

Solutions

d) $\mathbf{H_0: \mu_i = \mu_j}$

$\mathbf{H_a: \mu_i \neq \mu_j}$

The comparison between Group 1 and Group 2 (CBT vs. family) shows a **mean difference of -3.4569, but it is not significant since p = .212**.

The mean difference of -7.7147 between Group 1 **and Group 3 (CBT vs. control) is significant since p =.005.**

This suggest that **the mean weight difference for teenage after CBT was significantly lower the control group.** (6)

The End!
Questions?

Email dhawan@g.harvard.edu for feedback and any future changes