

36-309/749
Experimental Design for Behavioral
and Social Sciences

Oct. 13, 2015
Lecture MR: Midterm Review

Schizophrenia Example

- **Scientific background:** People with schizophrenia spend less time focusing on the eyes and mouth than non-schizophrenic people when asked to look at photographs of faces.
- **Scientific Hypotheses:** The people with schizophrenia have a reduced physiological ability to hold their focus vs. they avoid those areas for psychological reasons.
- **Experiment:** Randomize people with schizophrenia to look at faces with different emotions (neutral, happy, angry, disgusted). Record the mean time spent looking at eyes and mouth vs. elsewhere. Repeat over 5 pictures with the same emotion (at 1 minute/picture), so the outcome is an "eyes/mouth" time between 0 and 5 minutes.

2

Example, cont.

- **Operationalization of concepts (concept validity) and variable classification:**
- **Means model:** Define $\mu_N = \text{population}$ (condition) mean parameter for eye/mouth time (EMT) in the neutral, happy, angry, and disgusted conditions
 - Can the means model be wrong?
- **Null and alternate hypotheses:** $H_0: \mu_N = \mu_H = \mu_A = \mu_D$
 $H_1: \text{at least one mean differs.}$

3

Example, cont.

- **Error model:**
 - Is the usual error model plausible? (Also, check using EDA before formal analysis and residual plots at the end.)
- **Making a conclusion:** Plausible parameter values (95%CI) or p-values (compare a statistic to its **null sampling distribution**, e.g., compare $F = MS_B / MS_W$ to the null distribution $F_{k-1, N-k}$).

4

1) Before running the experiment

- **Generalizability (external validity):** If we get “statistically significant” results, will rational critics claim applicability to the real world is overly limited? External validity comes from appropriate (ideally random) selection of subjects from a population (and not-too-rigid or narrow control of experimental conditions).
- **Interpretability:** If we get “statistically significant” results, will rational critics claim that the effects seen may not be due to what we say that are due to?
 - **Internal validity** comes from random assignment of treatment to subjects. A non-randomized study has great potential for **confounding**. If experimental groups differ on **more than one** thing, we also have confounding.
 - Use of a control condition
 - Use of blinding

5

Before running, cont.

- **Power:** (probability of rejecting the null hypothesis when a particular alternative hypothesis is true and for a particular sample size and residual variance) When we see a **non**-significant result, will our critics claim that there really is an effect of treatment, but we just designed our experiment poorly and had insufficient power?
 - Reduce **error (residual; within groups) variance** (σ^2) considering the four sources of variation:
 - Subject-to-subject
 - Environmental
 - Treatment application
 - Measurement of the outcome

6

Power, cont.

- Consider “design measures” to reduce error variance, such as **within-subjects design**, **blocking**, and means modeling of important **factors** and /or **covariates**. All of these improve power by reducing error variance.
- Assure treatment is of appropriate “strength” (compared to control).
- **Calculate** the power for alternative hypotheses that correspond to appropriate substantive significance (e.g., clinical benefit, cost-efficiency, etc.).
- Use enough subjects to get, sufficient power (e.g. $\geq 80\%$) for important alternative hypotheses.

7

Before running, cont.

- **Best experiment possible:** Think about: pilot testing, good experimenter training, good record keeping, valid randomization, effective blinding, good “x” measurement, avoiding inter-subject communication, etc.

8

2) Perform the Experiment

3) Exploratory Data Analysis

- Perform EDA to
 - learn the distribution of each variable
 - find mistakes and outliers
 - make initial check of model assumptions (linearity, additive vs. interaction model, equal variance)
 - make tentative, initial hypothesis testing (mean/median differences, zero vs. non-zero slope, pattern of means).

9

4) Formal Analysis

- Formal analyses produce p-values for choosing between null and alternative hypotheses about scientifically interesting parameters.
- Choose a standard model and corresponding (initial) analysis.

10

Concept of Interaction

- Applies (should be considered) whenever there are **two IVs** in a model (or more than two).
- Terminology: There is an interaction between IV A and IV B in their effects on the DV. **Not** between levels of an IV, e.g., between control group and male.
- Implies: The **effect of** a change in IV A on the mean of the DV **depends on** the level or value of IV B. (And always vice versa.)
- As opposed to: no interaction (additive model) where the effect of a change in IV A on the mean of the DV is the **same** for every level or value of IV B.

11

Overview

- Use the following table for **continuous outcomes** with **independent errors**, especially one measurement per subject (between-subjects design).

12

Explanatory Variables	Model Assumptions	H ₀ vs. H _A	EDA	Formal Analysis
One categorical				
Two categorical				
One quantitative				
Categorical plus Quantitative				

13

Formal Analysis, cont.

➤ Interpreting ANOVA tables

- Rows: corrected model; between group / individual factor main effects; interaction; error / residual / within groups; (corrected) total
- Columns: SS(D); df; MS=SS/df;
 - F=MS_{between} / MS_{error}; p-values (“sig”)
 - “D” is a specific deviation from a sample mean
 - Degrees of freedom is a measure of the “effective” number pieces of information in an SS or MS or F.
- MS_{within} or MS_{error} or MS_{residual} is $\hat{\sigma}^2$: 95% of **individual** values fall within $\pm 2 \sigma$ of the **group mean**.
- Other MS values estimate σ^2 + some treatment effect

14

Formal Analysis, cont.

➤ Interpreting ANOVA tables, cont.

- Null & alternative hypotheses and p-values
 - **Corrected model:** H₀: no IVs have any effect on the DV
 - **Interaction (A*B):** Used for model selection.
H₀: additive model is good enough (or interaction is not needed) H_A: Interaction model is needed.
 - **Main effects (A, B, separately):** H₀: $\mu_1 = \mu_2 = \dots = \mu_k$ (ignoring other factors). It is silly to ignore other factors if the effect of the factor of interest depends on other factors.

15

Formal Analysis, cont.

➤ Interpreting ANOVA, cont.

- Interpreting 2-way ANOVA results
 - If 1/A p-value ≤ 0.05 , conclude “both factors affect the DV in a complicated way, i.e., the effect of a level change of factor A on the DV depends on (varies with) the level of factor B”. [Contrast tests are needed for any additional conclusions.]
 - If 1/A p-value > 0.05 , choose the additive model (parsimony). Read additive model main effects to test the individual null hypothesis about each factor ignoring the other. Four cases: A, B, both, neither affect the DV.
 - Compare means if k=2 (which is better?). [Do planned contrasts and/or post-hoc testing when k>2.]

16

Formal Analysis, cont.

- Interpreting regression results (incl. ANCOVA)
 - Rows: Intercept (“Constant”) and each IV, possibly including indicator variables and interactions (product variables).
 - Columns: “B”, i.e., unstandardized parameter estimates; SE(B) ; [standardized estimates]; $T=B/SE(B)$; p-value (“sig”)
 - In SPSS, the “B” column shows estimates of parameters, e.g., b_0 or $\widehat{\beta}_0$, not parameter values, e.g., β_0 , because the latter is a secret of nature.
 - All p-value are for $H_0: \beta_{\text{something}}=0$ vs. $H_A: \beta_{\text{something}} \neq 0$.

17

Formal Analysis, cont.

- Interpretation of regression results
 - Simple regression means model:
 - $E(Y|x) = \beta_0 + \beta_1 x$
 - β_0 is the population mean of Y when $x=0$
 - Interpretation as “mean of Y when $x=0$ ” requires $x=0$ makes sense and there is data near $x=0$.
 - Interpretation of $H_0: \beta_0=0$ is meaningful only if both H_0 and H_A were scientifically possible before running the experiment.
 - β_1 or β_x is the change in population mean of Y when x goes up by one unit.

18

Formal Analysis, cont.

- **Setup** for ANCOVA (in the regression format)
 - For any k-level factor (categorical IV), choose a baseline, code all other levels as indicator variables, and use **only** the k-1 indicators to represent the factor in the model.
 - Create interaction variables, as products of the IVs. Specifically for one factor and one covariate (quantitative IV), multiply each indicator by the covariate to create k-1 interaction variables.
 - Means models (e.g., indicators B,C, covariate x):
 - Additive: $E(Y) = \beta_0 + \beta_B B + \beta_C C + \beta_x x$
 - I/A: $E(Y) = \beta_0 + \beta_B B + \beta_C C + \beta_x x + \beta_{B*x} Bx + \beta_{C*x} Cx$
 - Note: Baseline level, e.g., “A”, is “invisible”.

19

Formal Analysis, cont.

- Interpretation of ANCOVA (in the regression format)
 - For ANCOVA in SPSS, use the single model-2 F-change statistic and p-value to choose which model to interpret.
 - ● Know how to write the overall model equation plus the simplified equations for each individual factor level (including the baseline). Model equations use $E(Y)$ and β 's. Prediction (fit) equations use \hat{Y} and b 's.
 - Interpret using EDA plot, coefficient table and simplified equations all together. Make detailed interpretations (as opposed to 2-way ANOVA).

20

Formal Analysis, cont.

- Interpreting ANCOVA (in the regression format)
 - Interpret intercept as the true (parameter β_0) or estimated (b_0 , SPSS "(constant)") **mean outcome** for subjects in the **baseline category** with **the covariate equal to zero** (convert to subject matter words, including for indicator variables). Ignore if substantively meaningless or if $x=0$ is a gross extrapolation.
 - Interpret the non-indicator, non-interaction slope coefficient as the **change** (up or down) in **mean outcome** for a **one-unit increase** in the covariate **for the baseline category**. In an additive model this also applies for **all other categories**.

21

Formal Analysis, cont.

- Interpreting ANCOVA
 - Interpret indicator variable coefficients as **changes in intercept** for a given category **relative to the baseline** category (all zero indicators, missing in SPSS tables). For additive models this becomes **at each x value**.
 - Interpret continuous by indicator interaction coefficients as the **change in slope** for a given category **relative to the baseline** category.

22

Formal Analysis, cont.

- Interpreting ANCOVA
 - Mean Squared Error, MSE, is an estimate of σ^2 . (Square root is in the units of the outcome.)
 - The residual for each observation is the observed value minus the expected (predicted) value. This is an estimate of the "error" and represents unexplained variation.
 - Residual plots are used to test assumptions.
 - R-squared is a unitless measure of the closeness of the observations to the prediction "line". Quantitatively it is "percent of variation in the outcome explained by the explanatory variables."

23

5) Model Assessment

- Know that p-values and confidence intervals have the correct meaning only if the assumptions are (approximately) met. (Depends on robustness.)
- Check assumptions with formal tests
 - Less commonly done (often less useful): Levene's test for equal variance, t-test for skew or kurtosis
 - Commonly done: check if interaction is needed (F-change test in ANCOVA, interaction p-value in 2-way ANOVA)
 - Commonly done: Add a square term for a quantitative IV to check for non-linearity
 - [Sometimes done: Durbin Watson test for serial correlation]

24

Model Assessment, cont.

- Check assumptions informally
 - ANOVA or regression (incl. ANCOVA): Check normality of residuals by **quantile-normal (or PP) plot**. (These analyses are robust to moderate non-normality.)
 - ANOVA: Check **residual vs. predicted (fitted) plot**. Non-zero group means suggests non-additivity (interaction). Check for unequal spread (funneling); these analyses are robust to less the 2:1 spread ratio.
 - Regression (incl. ANCOVA): Check residual vs. predicted plot. A consistent pattern of non-zero means (smile or frown) suggests non-linearity (no analyses are robust to non-linearity). Also check for unequal spread.

25

6) If needed, modify the model and try again

- Drop non-significant interaction terms (Occam's razor = parsimony).
- Transform the DV (e.g., log or square root) to see if equal spread, normality of errors, and/or linearity are better on a different scale.
- Transform IV(s) to see if linearity is better on a different scale. (Commonly, **add** the square of a continuous explanatory variable.)

26

7) Report Results (Critical on exam!)

- Give appropriate regression coefficient estimates (best with confidence intervals), prediction equations and/or plots, condition means and/or interaction plots. Usually 3 significant digits are all that are meaningful, so round appropriately.
- Report p-value (**NOT** 0.000!) not just reject/retain. Include the corresponding statistic (t, F, etc.). State which specific H_0/H_A the p-value refers to. (Do not report unimportant p-values, e.g., the intercept in an ANOVA.)
- Report the direction of the effect(s)!!! Put coefficient estimates into words, never just "x" and "y".
- Appropriately explain interaction (plot plus corresponding meaning of p-values).

27

Report Results, cont.

- [Perform appropriate planned or post-hoc comparisons (contrasts).]
- [Un-transform where appropriate.]
- Never say proved! "Provides evidence supporting" is good jargon. Be honest about removal of outliers, assumption violation, limited generalizability, and limited interpretability. Appropriate use "cause" vs. "associated with".
- Consider the possible role of "bad luck" (type-1 error if $p \leq 0.05$; type-2 error if $p > 0.05$).

28

8) Place your experiment in context

- Speculate on the connection between the operationalized world of the experiment and the real world.
- Consider what future experiments are needed or suggested by your work.

29

Possible schizophrenia experiments

- Outcome is EMT (eye-mouth time).
- Explanatory variable is emotion shown (neutral vs. negative vs. positive).

30

Possible schizophrenia experiments

- Outcome is EMT (eye-mouth time).
- Explanatory variables are emotion shown and photo vs. drawing.

31

Possible schizophrenia experiments

- Outcome is EMT (eye-mouth time).
- Explanatory variable is a quantitative emotion measure (anger pictures only).

32

Possible schizophrenia experiments

- Outcome is EMT (eye-mouth time).
- Explanatory variables are emotion shown and a quantitative measure of reaction time for an unrelated task.

33