

- Review for test
- Final Paper—Writing Your Analysis Plan
- Transcultural approach in research (see handout)
- Dummy variables
- Multiple regression lab

I. Review for test

A. Choosing Statistical Tests

1. Know when to use statistical procedures
2. Know the rules, e.g. knowing type and number of IVs, DVs
3. And for specific research scenarios
4. Test will cover these:
  - a. Independent groups  $t$  test
  - b. Dependent (or Paired) groups  $t$  test
  - c. Multivariate regression (but not logistic regression)

B. Hypothesis Testing--as in the first exam, you will be asked to go through the 8 steps of hypothesis testing for a given scenario

1. What is (are) IV(s)? Type of measure
2. What is DV? Type of measure
3. Identify null hypothesis
4. Identify alternate hypotheses
5. Choose statistical test and alpha
6. Interpret SPSS output
7. Justify decision to reject or not reject null hypothesis
8. Interpret results
  - a. Statistical significance

- b. Direction of the relationships
- c. Meaning and implications of results
- d. Limitations/future studies

## II. Final Paper--Writing Your Analysis Plan

### Example of Proposed Analysis Section

*See the example Analysis Plan on p. 42-43 of the Rubin, Babbie & Lee Supplement. Your Analysis Plan (for the Final Paper) should have a table like the one on p. 43 to describe all univariate, bivariate and multivariate statistical procedures as they apply for your study. They might not all apply, although all papers (even those with qualitative-only designs) should include a plan to analyze univariate statistics of the sample. As an introduction to the table, your papers should also have a very brief narrative summary, as in:*

#### Analysis Plan

The analysis plan for this study is summarized in Table \_\_\_\_\_. Univariate analyses will be conducted on all independent and dependent variables in order to describe the sample. Bivariate analyses will be conducted to explore the relationship among each independent variable (age, quality of family life, social support, gender, and ethnicity) and the dependent variable (life satisfaction). Multivariate analysis will consist of multiple linear regression of the independent variables on the dependent variable to address extent to which age, quality of family life, and social support are related to and predict life satisfaction, controlling for gender and ethnicity.

For the qualitative component, the narrative text (*or other types of data as relevant to your study*) will be coded for themes related to how participants view life satisfaction and its relationship of any of the independent variables in their own lives, as well as other factors not included in the quantitative analyses.

*You can cut and paste this text for your own purposes. Note that this paragraph summarizes but is not redundant with the table.*

## III. Dummy Variables (this material excerpted from class notes on Correlation and Multiple Regression)

**Dummy coding** in multiple linear regression and binomial logistic regression:

- If an independent/control variable is categorical, then **dummy coding** (AKA creating indicator variables) is necessary for proper analysis with multiple regression. This involves creating a separate variable for each category within the categorical variable and using a “baseline” category with which to compare all other categories.

- For instance, race/ethnicity is a very common demographic variable that is included in many multivariate statistical models. We normally think of race/ethnicity as one categorical (nominal) variable with multiple categories within it (i.e. White, African American, Latino, Asian/Pacific Islander, Other). However, to include race/ethnicity in a multivariate model, we need to use a procedure called dummy coding (AKA creating indicator variables) to create several *dichotomous* variables. To do this, the one variable of race/ethnicity is re-coded (in SPSS) into 4 indicator variables:

1. **White: Value labels: 0 = Not White 1 = White**
2. **African American: Value Labels: 0 = Not African American 1 = African American**
3. **Latino: Value labels: 0 = Not Latino, 1 = Latino**
4. **Asian/Pacific Islander: Value labels: 0 = Not API, 1 = API**

Why four and not five? Because we don't need to create dummy variables for all five original attributes. The analysis treats the missing dummy variable as a baseline with which to compare all others. (If you did code all five and tried to run the multivariate analysis, your analysis would be in error.)

- One indicator variable is chosen as the “baseline” to which all other racial/ethnic categories are then compared. For instance, if White is chosen as the baseline, then the statistical output provided by SPSS will indicate a comparison between African Americans and Whites, Latinos and Whites, and APIs and Whites with respect to the dependent variable. How do you chose which variable to exclude as a dummy variable, hence using it as a baseline variable? The decision is based on a combination of theory and standard research practice—often an “Other” category is typically as the baseline. However, sometimes you are interested in comparing all others to White (or another ethnicity), so that variable would be used as the baseline.
- You can also create a dummy variables from “Gender”, where Female = 1 and Male = 2. You would recode “Gender” as a

dichotomous dummy variable called “Female.gender” where 0=not Female and 1=Female. “Male” would be the baseline, and the analysis would then compare females to males. We use “0” and “1” since the interpretation of the results focuses on having either a presence or absence of the variable.

Here’s what it looks like for ethnicity, using “Other” as a baseline. Compare these two alternate ways of coding “ethnicity”:

Data set with one ethnicity variable (having five attributes or categories):

<i>Subject.ID</i>	<i>Ethnicity</i>
	1=White 2=Latino 3=Afr Amer 4=Asian/PI 5=Other
1	3
2	3
3	1
4	4
5	1
6	2
7	2
8	5
9	2
10	2

Recoded data set with *four* ethnicity indicator (dummy) variables (and one, the “Other” is excluded):

<i>Subject.ID</i>	<i>White</i> 0=non-White 1=White	<i>Latino</i> 0=non-Latino 1=Latino	<i>Afr.amer</i> 0=non-African American 1=African American	<i>Asian.pi</i> 0=non-Asian/PI 1=Asian/PI
1	0	0	1	0
2	0	0	1	0
3	1	0	0	0
4	0	0	0	1
5	1	0	0	0
6	0	1	0	0
7	0	1	0	0
8	0	0	0	0
9	0	1	0	0
10	0	1	0	0

\*\*\*Is “Other” really *missing*? Hint—look at subject ID 8.