Writing about Numeric Data

One of the challenges of using numeric data is choosing how and what to write when reporting results. This handout provides a number of examples and practical tips for the most common types of statistical analyses. Practices may vary by discipline, so ask your adviser or instructor if your write-up should differ from our recommendations. See the "Information and Assistance" section at the end of this document for further resources for writing about data.

General Tips

One important consideration when reporting your findings in writing is *how much information to present*. Generally, a good rule of thumb is to present all of the results from your statistical analyses in *tables*. In the *text*, report only the most relevant findings. This strategy draws the reader's attention to the most important conclusions without omitting any useful information.

You need to make a number of decisions about the *format* of your paper. What will your tables look like and what kind of information will they report? How long will your results section be, and how much detail should you go into for each type of analysis? No general rules apply for such decisions; they must be made on a case-by-case basis. We suggest that you look at journal articles in your discipline that conduct similar analyses to get ideas about formatting. If applicable, you can also read papers or theses from previous years and apply similar formatting.

Describing the Data

Whether you used data from another source or collected your own, you should provide a basic description in the text. For *secondary data*, or data collected by someone else, name the source of the data and provide a citation. The proper citation information is usually at the beginning of the User's Guide. For *primary data* that you collected yourself, include acknowledgements for any funding you received.

You should spend some time describing various aspects of your **sample**, even if it was not chosen randomly. If you are using secondary data, this information will be in the User's Guide. The sampling frame details how the sample was chosen. Include the size of the sample, both before and after missing data were eliminated. State how much information was missing, and provide an explanation if necessary. Also make sure to discuss any restrictions you placed on the original sample to create a smaller subsample (i.e., analyzing only people between the ages of 18 and 30).

Describing the Variables

Once readers have become familiar with your data, you can begin describing the *variables* in your analysis. The data may contain many more variables than those you chose to use. You should focus only on the variables included in your analysis.

You should clearly identify the **dependent variable(s)**, the variables whose values you are trying to predict. The other variables are **independent variables**. You may also distinguish between *explanatory variables*, the variables of interest to you theoretically, and *control variables*, which are included in the analysis but are not your primary concern.

For each variable, you can present *univariate statistics* to describe it numerically. You may want to report these in a table, in the text, or both. Always specify the *units* of measurement. For variables that are *continuous* (i.e., that have a range of possible values), the *mean* and *standard deviation* are good descriptions. For example, to describe an AGE variable, you could write, "the mean age in this sample is 35 years, with a standard deviation of 15."

For *discrete* variables that have a limited number of specific values, you can present a *frequency table* or report frequencies in the text. The mean may also be useful in some cases. For example, to describe a variable reporting RESIDENCE TYPE, you could write, "53 percent of respondents own their residence, 35 percent are renters, and 12 percent live in someone else's household." A mean would not be useful in this case.

Testing Hypotheses

Statistical analysis is usually driven by a number of *research questions* or *hypotheses*. For example, you may be interested in the relationship between age and residence type, so you might hypothesize that as age increases, the probability of owning your residence increases.

Tests of significance are useful in evaluating hypotheses. For a description of significance tests and the statistical analyses used to provide them, please see the "Information and Assistance" section below. When reporting analyses that test hypotheses, you should state both the direction of the relationship and its level of significance. This rule holds whether you report findings in a table or in the text. You must also decide a cut-off value for considering relationships significant. In these examples, our cut-off is p=.05.

Four types of bivariate analyses are useful for testing hypotheses. For more information on how to conduct these tests, see the "Information and Assistance" section. We list the tests and provide example sentences reporting findings.

To evaluate a hypothesis involving two continuous variables, a **correlation** is suitable. For example, the correlation coefficient between AGE and PHYSICAL HEALTH CONDITION is -.33, and the p-value is smaller than .05. To report this finding in the text, you could write, "As hypothesized, there is a significant negative relationship between age and physical health. The correlation coefficient is -.33 ($p<.05$). This means that for every one-year increase in age, the physical health score decreases by .33 points."

For testing relationships between two discrete variables, a **crosstabulation** with a chi-square significance test is appropriate. For example, a crosstabulation of GENDER and HUNTING LICENSE variables has a significant chi-square test with a p-value lower than .05. You might write, "Gender and possession of a hunting license are significantly related ($p<.05$), with a disproportionate number of all licenses issued to men (84 percent)."

To test a hypothesis about the relationship between a continuous and a *dichotomous* variable (i.e., a variable with two discrete values), you may use a **t-test** for differences between means. For example, a t-test for differences by GENDER between mean values of WAGES reveals that men's average wages are $42,000, while the women's mean is $31,000, and the p-value is less than .05. You could write, "A t-test for differences between means reveals that men earn significantly more than women at an average of $42,000 and $31,000 respectively ($p<.05$). This finding supports the hypothesis."

For evaluating relationships between a continuous and a discrete variable, a one-way **ANOVA** is appropriate. This test is useful for analyzing the hypothesis about AGE and RESIDENCE TYPE that was stated at the beginning of this section. This time there is no significant result; the p-level is higher than .05. You might write, "The relationship between age and residence type is not significant in this analysis ($p>.05$). The average age for homeowners is 34 years, with means of 36 years for renters and 35 for those who live in someone else's household. These results do not support the hypothesis."

You may also have run *multivariate analyses*, such as multiple regression, MANOVA, or logit. For interpretations of these results, please refer to the "Information and Assistance" section below or journal articles in your discipline.

For More Information and Assistance

SSDS Software Services at Stanford

The software consultants at Social Science Data and Software (SSDS) provide technical support for users of statistical applications at Stanford. While our service does not help individual clients learn statistics or interpret or write about data, we can help you run your chosen analyses and refer you to statistical and writing resources.

Users can send questions to SSDS Software Services by sending email via our website, http://ssds.stanford.edu, or they can come to the Velma Denning Room (120F) of the Social Sciences Resource Center in the Bing Wing of Green Library during walk-in hours; no appointment is necessary. Current walk-in hours are listed on the website.

Manuals and Publications

The Social Science Data and Software website, http://ssds.stanford.edu/, has a complete list of books and manuals in our non-circulating collection that may be useful for writing about numeric data. Some recommendations are: *Your Statistical Consultant: Answers to your data analysis questions* (Newton and Rudestam), *The Cartoon Guide to Statistics* (Gonick and Smith), and

Statistics with Stata 8 (Hamilton). These publications can be found in the Velma Denning Room (Green Library Bing Wing, Room 120F).

For help getting started with your statistical analyses, refer to the "Getting Started" documents on our website. Documents for SPSS for Macintosh/Windows/UNIX, Stata, and SAS are available online. You may also request an appointment with us from the website.

Other Resources at Stanford

Stanford has a number of other resources that may be helpful to you. If you need help with selecting or understanding statistical analyses or interpreting your results, the Statistics Department's consulting service may be of use. Their website is: http://www-stat.stanford.edu/consulting/index.html.

For help with writing, you can contact the Stanford Writing Center. Their website is: http://swc.stanford.edu. Honors thesis writers or advanced undergraduate writers can contact Hilton Obenzinger, Associate Director for Honors Writing, Undergraduate Research Programs, at hobnzngr@stanford.edu.

Social Science Data and Software
Document revised: August, 2005