Categorical Data Analysis
Lab material #1

Consider Table 1. The placebo group contains 15 patients, and the test education group contains 28 patients. The columns contain the information concerning the categorical response measure: 8 patients in the Test group had a favorable response to the medication, 20 subjects did not. The outcome in this example is dichotomous, and the analysis investigates the relationship between the response and the treatment.

Table 1: Respiratory Outcomes

Treatment	Favorable	Unfavorable	Total
Placebo	5	10	15
Test	8	20	28

Consider the following SAS Data step that inputs the data displayed in Table 1.

```
* example1_1.sas;
data respire;
  input treat $ outcome $ count;
  cards;
  placebo f 5
  placebo u 10
  test f 8
  test u 20
  ;
run;
proc freq;
  weight count;
  tables treat*outcome;
run;
```

The data set RESPIRE contains three variables: TREAT is a character variable containing values for treatment, OUTCOME is a character variable containing values for the outcome (‘f’ for favorable and ‘u’ for unfavorable), and COUNT contains the number of observations that have the respective TREAT and OUTCOME values. Thus, COUNT effectively takes values corresponding to the cells of Table 1. The PROC FREQ statements request that a table be constructed using TREAT as the row variable and OUTCOME as the column variable. By default, PROC FREQ orders the values of the rows (column) in alphanumeric order. The WEIGHT statement is necessary to tell the procedure that the data are count data, or frequency data; the variable listed in the WEIGHT statement contains the values of the count variable.

Output contains the resulting frequency table.

Output 1 Frequency Table

```
The FREQ Procedure
```

Table of treat by outcome

treat	outcome		
Frequency / Percent / Row Pct / Col Pct	f	u	Total
placebo	5 11.63 33.33 38.46	10 23.26 66.67 33.33	15 34.88
test	8 18.60 28.57 61.54	20 46.51 71.43 66.67	28 65.12
Total	13 30.23	30 69.77	43 100.00

These data may be stored in case record form, which means that each individual is represented by a single observation. You can also use this type of input with the FREQ procedure. The only difference is that the WEIGHT statement is not required.

The following statements create a SAS data set for these data and invoke PROC FREQ for case record data. The @@ symbol in the INPUT statement means that the data lines contain multiple observations.

```
* example1_2.sas;
data respire;
  input treat $ outcome $ @@;
  cards;
  placebo f placebo f placebo f
  placebo f placebo f
  placebo u placebo u placebo u
  placebo u placebo u placebo u
  placebo u placebo u placebo u
  placebo u
  test f test f test f
  test f test f test f
  test f test f
  test u test u test u
  test u test u test u
  test u test u test u
  test u test u test u
  test u test u test u
  test u test u test u
  test u test u
  ;
proc freq;
  tables treat*outcome;
run;
```

This SAS statement basically gives the same table. In this course, the data are generally presented in count form.