Biometry 700 – Introduction to Clinical Biostatistics
Lab 4 – Descriptive Statistics II

I. PROC Chart:
- Used to create many types of charts
- Vertical Bar Chart (Histogram)
  - Use the keyword VBAR in the code
- Horizontal Bar Charts
  - Use the keyword HBAR in the code
- Basic Code and Examples:

```
PROC CHART DATA = <data set name>;
   BY <list variables>;
   VBAR <list variables> <options>;
   HBAR <list variables> <options>;
RUN;
```

- See Examples

II. PROC Means:
- Displays basic descriptive statistics such as: (highlighted statistics = the default output)
  - Sample Size (n)
  - Mean (mean)
  - Standard Deviation (std)
  - Standard Error (stderr)
  - Minimum and Maximum Values (min ; max)
  - Number of Observations with Missing Values (nmiss)
  - Range of Values (range)
  - Sum (sum)
  - Variance (var)
  - T – Test (t)
  - P-value of a T-Test (prt)
- Basic Code:

```
PROC MEANS DATA=<data set name> <statistics>;
   BY <variable list>;
   VAR <variable list>;
   OUTPUT out=<output data set name> <STATS=new variable names>;
RUN;
```

- Example:

```
PROC MEANS DATA=utility mean sum std;
   VAR fuel phone electric total;
RUN;
```

Find the total and the average amount spent on phone, fuel, and electricity costs.
III. PROC Freq:
- Generates tables for categorical data.
- Uses a TABLES statement (instead of the VAR statement) to indicate the variables to be included in the table.
- Reads each observation with a frequency of one (as a default).
- BUT, allows you to used grouped data instead of only reading each observation as an individual.
- Can request more than one table at a time.
- Basic Code:

```plaintext
PROC FREQ DATA=<data set name> ;
   BY <variable list> ;
   TABLES <vertical variable>*<horizontal variable> / <options> ;
   WEIGHT <variable> ;
RUN ;
```

- OPTIONS:
  - Chi Square Test for Independence or Homogeneity (chisq)
  - Fisher’s Exact Test (exact)
  - Expected Values Cell Counts (expected)
  - You can suppress some of the output
    - Column Percents (nocol)
    - Row Percents (norow)
    - Cumulative Frequencies (nocum)
    - Cell Frequencies (nofreq)
    - All Percentages (nopercent)

- Example

```plaintext
Find the number of students (in data set Grades.dat) at each letter grade, also how many are in each class-sex group.

PROC FREQ DATA=grades;
   TABLES course class*sex / CHISQ EXPECTED;
RUN;
```

- Example when to use the WEIGHT statement

```plaintext
DATA = Agegroups
Agegroup Gender Number of People
1     M      10
1     F      15
2     M      20
2     F      18
3     M      19
3     F      22
```

```plaintext
PROC FREQ DATA=agegroups;
   TABLES Agegroup*gender;
   WEIGHT Number;
RUN;
```