

**STATISTICS 210**  
**Exercise #1. Due Lecture #2**

**Introduction**

This exercise concerns analysis of data from a cross-sectional study of the relationship between blood pressure levels, and hypertension, weight-to-height ratio (the Quetelet index), age and smoking history in a sample of men over age 40. The objective is to determine how blood pressure relates to other factors, and how these other factors are related to age. Perform the following analyses using the suggested SAS PROCs (we'll discuss a lot of the statistical features of these later). Write a report summarizing your results as specified in the syllabus.

**Data**

The program **ex1.sas** is on the course website. This file, which is shown below can be the basis for the SAS job to complete this exercise.

\* Program EX1 SAS;

Data EX1; Input SBP Quet Age Smoke;

Label SBP \\ = 'Systolic blood pressure (mm Hg)'

Quet = 'Quetelet index=weight/height (kg/cm)'

Age = 'Age in years'

Smoke = '1 if smoker, 0 if not smoker';

Cards;

```
135 2.876 45 0
122 3.251 41 0
130 3.100 49 0
148 3.768 52 0
146 2.979 54 1
129 2.790 47 1
162 3.668 60 1
160 3.612 48 1
144 2.368 44 1
180 4.637 64 1
166 3.877 59 1
138 4.032 51 1
152 4.116 64 0
138 3.673 56 0
140 3.562 54 1
134 2.998 50 1
145 3.360 49 1
142 3.024 46 1
135 3.171 57 0
142 3.401 56 0
150 3.628 56 1
144 3.751 58 0
```

```
137 3.296 53 0
132 3.210 50 0
149 3.301 54 1
132 3.017 48 1
120 2.789 43 0
126 2.956 43 1
161 3.800 63 0
170 4.132 63 1
152 3.962 62 0
164 4.010 65 0
;
```

```
Title 'Printout of the EX1 data';
Proc Print Data=EX1;
Run;
```

### **Analysis Sequence**

Use the following analysis sequence when solving this exercise.

1. Describe the distributions of the variables Age, SBP and Quet (use PROC UNIVARIATE with options NORMAL and PLOT) separately for smokers and non-smokers. Consider means, medians, standard deviations, quartiles, the stem and leaf plots and any other information you deem relevant. Comment on any differences between smokers and nonsmokers.
2. Separately for smokers and non-smokers use PROC PLOT to generate the following scatterplots (y vs x, as given): SBP vs Age, SBP vs Quet, and Quet vs Age.
3. To accompany the scatterplots, separately for smokers and non-smokers compute the Pearson correlations among Age, SBP and Quet using PROC CORR. Comment on the significance, strengths and nature of the relationships among smokers and non-smokers. Comment on any differences between smokers and non-smokers.
4. Perform two-sided t-tests (PROC TTEST) comparing smokers to nonsmokers for Age, SBP and Quet. Comment.
5. Hypertension is usually defined as a systolic blood pressure greater than 140. Define a new variable HYPERT=1 if SBP> 140, 0 if not. Use PROC FREQ to compute the proportion of subjects with hypertension for smokers and non-smokers in a 2x2 table, and to conduct the contingency chi-square test of independence, or of the differences in proportions with hypertension among smokers and non-smokers. Cite the p-value of the chi-square test. Comment on the differences between smokers and non-smokers.

## Report

Your work should be organized into a report following the structure for exercises described in the syllabus.

1. **Introduction:** Your report should start with an introduction describing the objectives of the study (this can be copied directly from the introduction above).
2. **Methods :** The methods section of the report should then describe the statistics and techniques you use (for example, "A preliminary analysis of the data was performed. Means and medians were used to assess the central tendency of the distributions. Dispersion was evaluated using the standard deviation and etc... Stem and leaf plots were used to ..." and so on.) Note that the methods section does not say what you found, only what you planned to do and did. This would be a summary of the analysis sequence specified in the exercise.
3. **Results:** The results section will contain your relevant findings, i.e. the values, patterns and relationships that you found. This is the place to report such things as distributional oddities ("the distribution of ... was found to be skewed ..."), important differences ("the mean of ... was ... among smokers and ... among non smokers"), and answers to the questions posed in the analysis sequence section above. Tables and Figures should be constructed to present the summary findings. These can be cut from SAS output but should not be just a page of output. What this section does not contain is subject-matters (i.e. medical) conclusions about the results. These will be included in the next and final section.
4. **Discussion:** The discussion section is where you make your conclusions in terms of the original objective of the study: considering everything you have found in this exercise, what can you say about the relationship between blood pressure and the other factors. How do the factors relate to age? Are there caveats to your conclusions? Are there questions that need to be explored further?