**Stat 400 Project 2 Fall 2019**

*Purpose:* To analyze large sets of “real” data in the form of a spread sheet and to use inferential statistics to determine any association between heart disease and the risk factors age, sex, blood pressure, and cholesterol level.

*Introduction*: The Framingham Heart Study was initiated in 1948 in Framingham, Mass., to investigate the relationship between the incidence of coronary heart disease and certain risk factors such as age, sex, serum cholesterol, serum glucose, body-mass index, systolic blood pressure, and cigarette smoking.

*Confidence Intervals*: Consult section 7.2 of the text.

*Hypothesis Tests*: Consult section 9.1 for tests between two means and section 9.4 for tests between two proportions.

*Procedure*: Access the Microsoft Excel file “2.20.Framingham” on ELMS under “Files” which represent a partial listing of the data. The data key is:

# sex : Sex (1 = male; 2 = female)

# sbp : Systolic Blood Pressure

# dbp : Diastolic [Blood Pressure](http://www.math.montana.edu/shancock/courses/stat539/r/Framingham-Feb7.R)

# scl : [Serum](http://www.math.montana.edu/shancock/courses/stat539/r/Framingham-Feb7.R) Cholesterol

# chdfate : Coronary Heart Disease(1 = no disease; 2 = disease)

# followup : Follow-up in Days

# age : Age in Years

# bmi : Body Mass Index (wt (kg) / h^2 (m)

# month : Study Month of Baseline Exam

# id : Subject ID

1. a) From the study data, determine a 95% confidence interval for the true US population proportion of women who develop coronary heart disease. Also, determine a 95% confidence interval for men. What can you say about the actual proportions of heart disease in the US population?

b) Perform a hypothesis test for the difference between two proportions to determine whether there is a significant difference in the population proportions for women and men. Report the p-value and interpret the result in the context of this test.

2. a) From the study data, determine a 95% confidence interval for the true US population dbp mean for those with chd and a 95% confidence interval for those in the US population without chd. What can you say about the actual means in these two cases in the US population?

b) Perform a hypothesis test for the difference between two means to determine whether there is a significant difference in the population means. Report the p-value and interpret the result in the context of this test.

3. a) From the study data, determine a 95% confidence interval for the true US population scl mean for those with chd and a 95% confidence interval for those in the US population without chd. What can you say about the actual means in these two cases in the US population?

b) Perform a hypothesis test for the difference between two means to determine whether there is a significant difference in the population means. Report the p-value and interpret the result in the context of this test.