Homework 2

STAT448 - Advanced Data Analysis

Due: Thursday, February 15, 2018 at 5:00 pm

The data sets are provided in the **Program_HW2_Data_Spring2018.sas** file on the course website. For Exercises 1 and 2 use the **haireyes** data set, and for Exercises 3 and 4 use the **heartbpchol** data set.

The **haireyes** data set is a subset of the "Hair Colour and Eye Colour" data set described and given on page 367 of the textbook. The **heartbpchol** data set contains the cholesterol, blood pressure status, and cholesterol status for alive patients from the sashelp.heart data set.

To retain the order in which categorical values appear in a data set, you can use **order=data** *as an option in the* **proc freq** *statement.*

1. (2 parts)

   (a) For the **haireyes** data as given, construct a contingency table and comment on any apparent associations between hair color and eye color.

   (b) Perform and comment on appropriate tests of association for the table, and interpret the results.

2. (3 parts) Now consider association between the two darker eye colors and hair colors.

   SAS Studio (including SAS OnDemand and University Edition) seems to not correctly process logical connectors (e.g. and, or) in **where** *statements, so you may need to create an additional data set to get the correct subset if you are using SAS Studio.*

   (a) Construct a table with the **medium** values omitted. For the remaining 2x2 table of **light** and **dark** eye color and **fair** and **dark** hair color, comment on any apparent associations between these eye and hair color groups.

   (b) Perform and comment on appropriate tests of association for the table, and interpret the results. Compare the results and conclusions for this 2x2 table to those for

the full 3x3 table analyzed in part (a).

(c) Consider risks for having **fair** hair. Test and comment on whether those with **light** eye color are significantly more likely to have **fair** hair than are those with **dark** eye color.

3. (2 parts)

(a) For the **heartbpchol** data set, construct a contingency table for comparing blood pressure status and cholesterol status and comment on any apparent relationships between blood pressure and cholesterol statuses.

(b) Perform and comment on appropriate tests of association for the table, and interpret the results. What can be said about cholesterol statuses of patients with higher blood pressure statuses?

4. (2 parts) For the **heartbpchol** data set, consider a one-way ANOVA model to identify differences between group **cholesterol** means. The normality assumption is reasonable, so you can proceed without testing normality.

(a) Perform a one-way ANOVA for **Cholesterol** level with **BP_Status** as the categorical predictor, test any assumptions of the model that should be tested (aside from normality, which you do not need to test), comment on the significance of the model, and the variation described by the model.

(b) Comment on any significantly different cholesterol means as determined by the best test for comparing all pairwise differences. Explain what that tells us about differences in cholesterol levels across blood pressure status groups, and comment on how these results compare with the results from part (b) of Exercise 3.