Quiz 3: Nutr 6101-Nutrition Research Methods

**Your Name: Kelsey Higgins**

[Please use this file to type your answers. Add your last name to the file name when you save your responses (i.e., 'Save As…"). If your name is Smith, the file name will be "Nutr6101-Quiz3_Jane Smith". (Please use the name used on the official class roll.) Send the file with your answers to me **by no later than Sunday, November 16th @ 5:00pm.]** NOTE: This quiz is to be taken without the assistance/discussion/cooperation of anyone else, and without the use of notes/books/internet or other materials. Treat the quiz as if you were taking it silently in class by yourself at your desk.]

Q. You have assessed: education level achieved (i.e., H.S., B.S., M.S., PhD/MD, etc), current income, age, race, height, weight, body fat percent, LDL cholesterol, HDL cholesterol, and total cholesterol in all the active players on 3 NFL teams (60 x 3).

1. You wish to determine the variables most associated with: LDL, HDL, and Total Cholesterol in this sample. Describe the statistical procedure(s) you would use to make this determination. (Not just the name of the statistical test, but also the procedure you would follow leading up to the test that includes identifying the dependent/independent variables.)
    a) In order to determine the variables most associated with LDL, HDL, and Total Cholesterol in this sample you would have to do some sort of correlation. These are continuous variables, so Pearson's correlation coefficient would be suitable. The correlation coefficient measures the degree of association and leads to a standard *r,* which takes a value between -1 to +1. *r* measures the degree of straight-line association between the values of two variables. If the points on the scatter plot lie in a wonderful and perfect straight line, a value of +1 or -1 is obtained. +1 (positive correlation) if higher values of one variable associate with higher values from the other variable. -1 (negative correlation) if one variable decreases while the other increases (an inverse relationship). A correlation near zero means the variables are uncorrelated. Furthermore, a correlation >0.75 represents a good to excellent relationship.
    b) To prep for performing the Pearson correlation coefficient, you need to see which variables are independent or dependent, continuous or categorical, interval or ratio. (Nominal data would not be used in a correlation, more so in Chi Square)
    c) First, you would create a scatter plot on SPSS. Next, you can run the correlation tests on SPSS by clicking analyze, correlate, bivariate. This will create a chart of the variables selected. You can run the non-parametric test (Spearman) at the same time as the parametric test (Pearson). One example for this question would be to see the association between LDL and weight. LDL is the dependent variable (y-axis) and weight is the independent variable (x-axis). The scatter plot will visually clarify the shape and strength of the correlation as well as allow you to see any

outliers, which would need to be removed. The charts created from the correlation test will allow you to determine the statistical significance of the correlation if any.

Dan Benardot 11/20/14 7:38 AM
**Comment:** OK

2. Assuming age and total cholesterol are found be strongly related, please indicate all the different ways you can think of for handling the 'age' and 'total cholesterol' values for doing multiple *different* Pearson correlations.
    a) One way would be to split the age groups up even further. Instead of general age, you could analyze it by ages 20-25, 25-30, and so on by 5-year intervals. You could do this by 10-year intervals or endless different ways really. You could also break it down by intervals of total cholesterol to age. Maybe one interval is <200 total cholesterol and one interval is >200 total cholesterol and you see the correlation between those groups and age. Using 200 because <200 is considered a healthy cholesterol level and >200 is seen as borderline to high levels. The list goes on, but those are some ways I would look at correlation between age and total cholesterol.

Dan Benardot 11/20/14 7:38 AM
**Comment:** OK

3. You wish to see if different age players have different Total Cholesterol values, and you are limited to doing a t-test. *Describe* how you would do this.
    a) A t-test compares two means (averages) of independent groups. A paired samples t-test requires two separate variables measuring the same thing on each subject. An independent group t-test compares a single variable between two groups. For more than two groups you would use ANOVA.
    b) Total Cholesterol (dependent) and Age (independent)
    c) In order to use a t-test you must have data that is normally distributed. Using a histogram before performing the t-test on SPSS allows you to see any outliers (which you would remove from the data set), and to ensure the data is normally distributed, before moving on to the t-test. In order to compare these two groups, you could break down age by players <27 years old, and players >27 years old. Because there is a single variable (total cholesterol) being compared between two groups (<27 yo and >27 yo) you would use the independent samples t-test. (Using 27 yo because 26.7 was the average age for all NFL players according to footballoutsiders.com in 2013)

Dan Benardot 11/20/14 7:39 AM
**Comment:** OK

4. Same as question "3", but for education level instead of age.
    a) When looking at total cholesterol and education level, the steps of how to do the t-test do not change. Instead of breaking up our independent variable (age) in the previous question to two groups (<30 yo and >30 yo), we will break up the new independent variable (education level) to two groups. If it were more than two

groups, we would run ANOVA. Since we are limited to a t-test, we will break down education level to pre-professional (H.S., and B.S) and professional (M.S. and PhD/MD).

b) In order to use a t-test you must have data that is normally distributed. Using a histogram before performing the t-test on SPSS allows you to see any outliers (which you would remove from the data set), and to ensure the data is normally distributed, before moving on to the t-test.

Dan Benardot 11/20/14 7:39 AM
**Comment:** OK

5. Same as question "3", but for income instead of age.
   a) When looking at total cholesterol and income, the steps of how to do the t-test do not change. Instead of breaking up our independent variable (age) in question 3 to two groups (<30 yo and >30 yo), we will break up the new independent variable (income) to <$2 million per year and >$2 million per year (which is the 2013 average salary for NFL players according to businessinsider.com)
   b) In order to use a t-test you must have data that is normally distributed. Using a histogram before performing the t-test on SPSS allows you to see any outliers (which you would remove from the data set), and to ensure the data is normally distributed, before moving on to the t-test.

Dan Benardot 11/20/14 7:39 AM
**Comment:** OK

6. You wish to predict total cholesterol from the variables you have collected. Please describe the statistical procedure you will follow to do this, and indicate the values you will look for to determine the strength of your prediction.
   a) In order to predict total cholesterol from the variables you have collected, you would run a regression analysis. Regression analysis predicts the value of a dependent variable based on the value of at least one independent variable. It also explains the impact of changes in an independent variable on the dependent variable.
   b) The dependent variable is the variable we wish to explain, and the independent variable is the variable used to explain the dependent variable. In this case, the dependent variable is total cholesterol.
   c) Simple linear regression is with only one independent variable, and the relationship between the independent variable and dependent variable are described by a linear function. Because we have three independent variables we have looked at thus far in relationship to total cholesterol (age, income, and education level) we would run multiple regressions. Regression is also normally the next step after correlation and is used as a statistical approach for explaining and predicting quantifiable clinical outcomes.

d) To predict total cholesterol from our continuous independent variables, an equation for the regression line is determined and used to predict values of Y, the outcome variable (Y=a+bX, where a is a constant and b is the slope).
e) Create a scatter plot to identify and remove outliers, as outliers can seriously affect regression results.
f) Analysis of a plot of residuals will determine if the predictions have been met. (Residual = Y – predicted Y)
g) If the plot resembles a horizontal band of points then the predictions are met but if the plot is curvilinear then data may need to be transformed to normalize the distribution or non-linear regression methods used.

Dan Benardot 11/20/14 7:40 AM
**Comment:** OK

7. You wish to see if the players on the 3 different teams have different LDL, HDL, and Total Cholesterol values. Describe at least 1 statistical procedure you would use to do this.
    a) Now that we are not limited to a t-test, and we have more than two groups to compare, this is where ANOVA comes in. ANOVA stands for analysis of variance. Simply stated ANOVA runs multiple t-tests at one time. ANOVA requires a categorical variable and preferably a ratio-level variable. ANOVA will allow us to see if there are any statistical significant differences in the means (averages) of three or more independent groups.
    b) The dependent variables for this problem are (LDL, HDL, and total cholesterol). The independent groups for this problem are the 3 different NFL teams.
    c) Each dependent variable would be assessed according to each of the 3 independent NFL teams. LDL with each team, HDL with each team, and total cholesterol with each team. From there, if ANOVA states there is a difference between at least two of the three teams, you would run the post-hoc test in order to determine exactly where the groups differ.
    d) On SPSS you click analyze, compare means, One-Way ANOVA, Add in the dependent variables to the dependent list, Click Post-Hoc and select Bonferonni, and voila.
    e) Using Bonferroni's Post Hoc test will let you determine where the differences are and obtain the descriptive statistics.

Dan Benardot 11/20/14 7:40 AM
**Comment:** OK

Dan Benardot 11/20/14 7:40 AM
**Formatted:** Normal, No bullets or numbering

Dan Benardot 11/20/14 7:40 AM
**Formatted:** Font:Times New Roman

Kelsey: Best responses. Nice job.