BIOS 6611 Homework 10

Due Monday, December 3, 2018 by midnight to Canvas Assignment Basket

The data for this assignment are from a study of demographic, dietary, and other behavioral determinants of plasma levels of retinol, beta-carotene, and other carotenoids. A summary of the study and a description of the variables in the *carotenoids.dat* data file are provided below:

Summary: Observational studies have suggested that low dietary intake or low plasma concentrations of retinol, beta-carotene, or other carotenoids might be associated with increased risk of developing certain types of cancer. However, relatively few studies have investigated the determinants of plasma concentrations of these micronutrients. A cross-sectional study was designed to investigate the relationship between personal characteristics, dietary factors, and plasma concentrations of retinol, beta-carotene and other carotenoids. Study subjects (N = 315) were patients who had an elective surgical procedure during a three-year period to biopsy or remove a lesion of the lung, colon, breast, skin, ovary or uterus that was found to be non-cancerous. The carotenoids.dat data file provides data for retinol and beta-carotene levels.

age	Age (years).
sex	Sex (1=Male, 2=Female).
smoke	Smoking status (1=Never; 2=Former; 3=Current Smoker).
bmi	Body Mass Index (weight/(height^2)).
vitamins	Vitamin Use (1=Yes, regularly; 2=Yes, irregularly; 3=No).
calories	Number of calories consumed per day.
fat	Fat consumed (grams per day).
fiber	Fiber consumed (grams per day).
alcohol	Number of alcoholic drinks consumed per week.
chol	Cholesterol consumed (mg per day).
betadiet	Dietary beta-carotene consumed (mg per day).
retdiet	Dietary retinol consumed (mg per day).
betaplas	Plasma beta-carotene (ng/ml).
retplas	Plasma Retinol (ng/ml).

1) The investigator hypothesizes that plasma beta-carotene levels may differ by smoking status. In this question, you will examine the relationship between plasma beta-carotene (the response) and smoking status (current smokers, former smokers, and never smokers).

A) Obtain the sample size, mean, standard deviation (SD), and standard error of the mean (SE) for plasma beta-carotene levels within each of the three smoking groups.

B) Fit a "reference cell" linear regression model (MODEL 1) regressing plasma beta-carotene levels, *betaplas* (the dependent variable) on smoking status (the independent variable). **Make the never smokers the reference group**. Write down the regression equation.

C) Using MODEL 1, is smoking status significantly associated with plasma beta-carotene levels? Write the null and alternative hypotheses in terms of the appropriate beta coefficient(s) and also in terms of the appropriate means, test the null hypothesis, and state your conclusion.

D) Using MODEL 1, do plasma beta-carotene levels differ between current smokers and never smokers? Write the null and alternative hypotheses in terms of the appropriate beta coefficient(s) and also in terms of the appropriate means, test the null hypothesis, and state your conclusion.

E) Using MODEL 1, do plasma beta-carotene levels differ between former smokers and never smokers? Write the null and alternative hypotheses in terms of the appropriate beta coefficient(s) and also in terms of the appropriate means, test the null hypothesis, and state your conclusion.

F) Using MODEL 1, do plasma beta-carotene levels differ between current smokers and former smokers? Write the null and alternative hypotheses in terms of the appropriate beta coefficient(s) and also in terms of the appropriate means, test the null hypothesis, and state your conclusion. (USE the variance-covariance matrix for the βs to answer this question).

2) For question 2 we will focus on fitting the "cell means" model instead of the "reference cell" model from question 1.

A) Fit a "cell means" linear regression model (MODEL 2) predicting plasma beta-carotene levels from smoking status. Write down the regression equation.

B) Use the cell means model (MODEL 2) to test if smoking status is significantly associated with plasma beta-carotene levels.

C) Use the cell means model (MODEL 2) to test whether plasma beta-carotene levels differ between current smokers and former smokers. Write the null and alternative hypotheses in terms of the appropriate beta coefficient(s) and also in terms of the appropriate means, test the null hypothesis, and state your conclusion.

D) Use the cell means model (MODEL 2) to test whether plasma beta-carotene levels differ between non-smokers (the average of never smokers and former smokers) and current smokers. Write the null and alternative hypotheses in terms of the appropriate beta coefficient(s) and also in terms of the appropriate means, test the null hypothesis, and state your conclusion.

3) Perform an independent samples t-test comparing plasma beta-carotene levels in current smokers versus former smokers. Compare your results to those obtained in parts (1F and 2C) and explain any differences.

HINTS FOR HOMEWORK 10

Overall Comment:

- You do not need to calculate confidence intervals unless explicitly requested or a "brief, but complete" summary of results is requested.

1F) You do not need to use matrix algebra to answer this question (although you could). To test the difference between current and former smokers, use the appropriate beta coefficients for these two groups and then use the appropriate variances and covariance from the Variance-Covariance matrix (which you can request with the SAS COVB option) to perform the statistical test.

2A, B, C) You can use SAS to perform all of these calculations (the TEST statement), but you should also know how to perform them by hand.

2C) For calculating the "average of never smokers and former smokers" you can use any weights for calculating the "average" since it's not specified, although it is probably most appropriate to weight by the sample size.

SAS code for reading in .dat file and providing some labels for categorical values:

```
DATA carotenoids;
     INFILE "~\carotenoids.dat";
     INPUT age sex smoke bmi vitamins calories fat fiber alcohol chol
betadiet retdiet betaplas retplas;
RUN;

PROC FORMAT;
     value sex  1 = 'Male'
                2 = 'Female';
     value smoke      1 = 'Never'
                2 = 'Former'
                3 = 'Current';
     value vitamins   1 = 'Yes, regularly'
                      2 = 'Yes, irregularly'
                      3 = 'No';
RUN;
```