

## Homework 4: due Tuesday, 25 October

**Take-home midterm will be handed out on Thursday, 27 October, and due in class on Tuesday, 1 November.**

1. **OGTT categories.** In HW 2, you classified individuals into categories according to the results of their OGTT (oral glucose tolerance test). In this problem, we will examine whether these categories are associated with body mass index (BMI).

Download the spreadsheet OGTT2.xls containing data from adults with cystic fibrosis. OGTT categories are given in several ways: OGTT\_category has 7 categories, as text strings; the numeric variable OGTTclass = 1 (Normal), 2 (Normal-Indeterminate), 3 (IGT), or 4 (CFRD); and IFG is an indicator for IFG status.

- a. Delete all patients with OGTTclass=2. Make a table classifying patients according to IFG status and OGTTclass. This is called the *design space*.
- b. Fit the model  $BMI = OGTTclass \text{ IFG } OGTTclass*IFG$ ; and report the type III ANOVA table. Make a plot of studentized residuals from this model. *Hint*: to add a reference line  $y = 0$  to a plot:

```
Proc SGplot;
  scatter x=pred y=ehat;
  refline 0 / axis=y;
```

- c. In your residual plot from (b), why are the points in vertical strips?  
If `yhat` is the name of the variable of predicted values from (b), define a new variable:

```
yhat_jitter = yhat+0.3*ranuni(477567)-0.15;
```

Make another residual plot using this new variable in place of `yhat`. Look up `ranuni` and explain what this new variable is.

- d. Use `birthdate` and `OGTT_date` to compute each patient's age in years at the time of OGTT. Calculate an indicator variable for *underweight* patients: women with BMI < 21 and men with BMI < 22. Make a table comparing the 3 OGTTclasses with respect to mean age and proportion underweight.
- e. Plot BMI against age (years) and add a fitted regression line.
- f. Why should we adjust the ANOVA in (b) for age, but not for underweight status?
- g. Fit the model  $BMI = age | OGTTclass | IFG$  ; and report the type III ANOVA table. The vertical bars tell Proc GLM to make all possible interactions. Do we need to keep any of the interactions with age in the model?
- h. Fit the model  $BMI = age \text{ } OGTTclass \text{ } IFG \text{ } OGTTclass*IFG$ ; report the type III ANOVA table, and make a plot of studentized residuals from this model. Why are these residuals *not* in vertical strips?
- i. Make two interaction plots for the model in (h): in one plot, use `sliceby=IFG` and in the other use `sliceby=OGTTclass`.
- j. Were there any differences in BMI associated with IFG status? Give a brief summary with appropriate means, standard errors, and *p*-values, and refer to only one of the plots from (i) in your explanation.

2. The spreadsheet `Collection Dates.xls` gives medical record number (MRN), date of OGTT collection, OGTT values, and subject's birthdate.

Calculate the age of each participant at collection data. Print MRN (an 8-digit string), age in years, birth date and collection date. Dates should be printed in the format 23JAN2009.