1 "**De-confounding" the passive smoking-breast cancer association (removing the part of the association that could be an"effect" of animal fat)**

- Problem 3.17 part A, G&S page 109 [p 112 in 2nd edition]

2 **Effect of maternal smoking on birthweight** *["birthweight data" on course web page]*

Assume for the purposes of this exercise that any effect of maternal smoking is **direct** i.e. on birthweight (via slower intrauterine growth) itself rather than via gestational age i.e., assume that maternal smoking does not affect length of gestation [this assumption has been the subject of some debate in the class in previous years, but on balance it seems from the literature that it is more or less correct]

i How much lower is the mean birthweight of infants of smokers than non-smokers? [designate "smoke" as a "group" variable, so that when you request the distribution of gestational age, it will be reported separately for each level of smoke]

ii How much higher is the mean gestational age of infants of smokers than non-smokers?

iii (a) Is the difference in age "statistically" significant?

you can carry out a t-test on the age difference by regressing age on smoke --after you remember to turn off the "group" designation for the variable smoke!

(b) is statistical significance a non-issue here anyway?

iv If babies grow on average of 100 g/week, how much of a birthweight "advantage" would the infants of smokers have as a result of their longer gestational age?

v On this basis, how much lower would the mean birthweight of infants of smokers be than the mean of infants of non-smokers if "adjusted" for the difference in age?

vi Instead of using the "out of the air" value of 100 g/week, use multiple regression to simultaneously estimate the g/week and g/smoke coefficients i.e., fit the model

$$\text{average[weight} \mid \text{age smoke]} = b_0 + b_{age} \times \text{age} + b_{smoke} \times \text{smoke}$$

vii Verify that if you correct/adjust the comparison in (i) using the fitted b_{age} found in (vi) and the age difference seen in (ii), you arrive at the b_{smoke} obtained in (vi).

We will come back to the appropriateness of this "adjustment model" again.

3 **Sharper/Fairer: Effect of sexual activity on longevity** *["fruitfly longevity" 622 web page]*

• Restrict attention to fruitflies with 1 partner [the difference is obvious in those with 8]

One way: highlight the 50 rows for those with partner=1, then using the menu under the triangle in the top left corner of the spreadsheet, extract these into a separate datasheet.

Early on, one student argued that thorax size cannot be used as a predictor or explanatory variable for the longevity outcome since fruitflies who die young may not be fully grown, i.e., it is also an "intermediate" variable. Later, another epi student who had studied entomology assured us that fruitflies do not grow longer after birth; i.e., thorax length is not time- (age)-dependent!

i How much lower is the mean longevity of sexually active than sexually inactive flies? [designate "type" as a "group" variable, so that when you request the distribution of longevity, it will be reported separately for each level of type]

Is this difference statistically significant?

you can carry out a t-test on the longevity difference by regressing thorax on type -- after you remember to turn off the "group" designation for the variable type!

Is this difference "clinically" significant

ii How much different is the mean thorax length of active and inactive flies?

iii (a) Is the difference in mean thorax length "statistically" significant?

you can carry out a t-test on the thorax difference by regressing thorax on type --after you remember to turn off the "group" designation for the variable type!

(b) is statistical significance a non-issue here anyway?

iv If -- all other things being equal -- flies who are 0.01 mm larger live on average of 1 day longer than ones who are 0.01mm shorter, how much of a longevity "advantage" would the active flies have as a result of their larger average thorax size?

v On this basis, how much lower is the mean longevity of sexually active than sexually inactive flies if "adjusted" for the difference in thorax size?

vi Instead of using the "out of the air" value of 1 day/ 0.01 mm, use multiple regression to simultaneously estimate the days/mm and days/type coefficients i.e., fit the model

$$\text{average[longevity} \mid \text{thorax type]} = b_0 + b_{\text{thorax}} \times \text{thorax} + b_{\text{type}} \times \text{type}$$

vii Verify that if you correct/adjust the comparison in (i) using the fitted b_{thorax} found in (vi) and the thorax difference seen in (ii), you arrive at the b_{type} obtained in (vi).

viii Compare the p-value obtained in (i) with that obtained in (vi). Part of the reason for the difference is that the adjusted estimate of the effect (*numerator* of t-test) is a small bit *larger* ; another is the *smaller* SE of the estimated effect (*denominator* of t-test).

Why is the SE smaller? [hint: compare components of G&S formulae 2.8 & 3.15/16]

See also pages 8-11 of "Intro to multi-variable analysis (incl. SHARPER and FAIRER e.g.'s) from 1995" under "Resources" on the 678 web page.

4 **Analysis of Rates of Fatal Crashes on rural interstate highways in New Mexico in the 5 years 1982-1986 (55 mph limit) and in 1987 (65 mph limit)**.

[see earlier howework for a description of how the authors did the analysis; if you did not save the dataset , you will need to recreate it]

"**Another way to skin a cat**"

It will be helpful to change the values of the variable "limit" to 0(=55) and 1(=65)

and the values of "year" to 0(=1982), 1(=1983), ... , 5(=1987) ...

- (before fitting) Sketch the lines that correspond to the models

  (i) average[rate | limit year] $= \beta_0 + \beta_{limit} \times limit$

  (ii) average[rate | limit year] $= \beta_0 + \beta_{year} \times year + \beta_{limit} \times limit$

- What is the meaning of each β in models(i) and (ii)?

- Fit these two models to the 6 datapoints; choose the more appropriate model of the two and obtain a point and interval estimate for the parameter of prime interest. Is the parameter estimate statistically significant if tested against the null value of zero?

  See pp 4-7 in "Intro to multi-variable analysis (incl. SHARPER and FAIRER e.g.'s) from 1995" under Resources.