BIOS 145 Lab #10 10/29/97

1st hour: Computer Exercise #2

2nd hour: questions

Other:

1. Overheads will be on web page & in HSL

2. Computer problem for test #2 will be on web page and on disk at the SPH computer lab tomorrow

3. PostScript printing at SPH lab

4. Thurs. session? (Ch11, anything else)

5. Incorrect answers to HWs on web page

Computer Exercise #2:

In the dataset PONE, define a zero-one variable for SEX, e.g. SEX10 = SEX - 1. Explore the sex differences in HDL (mg/dl), controlling for alcohol consumption (ALC-g/wk), AGE (yr), and BMI {BMI=WT/(HT/100)**2 (kg/m**2)}.

1. Use a regression model to get (unadjusted) differences in mean HDL between men and women.

   Here, Y = HDL, X = SEX10 $= \begin{cases} 1, & \text{if female} \\ 0, & \text{if male} \end{cases}$

   Model:

   $$Y = \beta_0 + \beta_1 X + \epsilon$$

   For males, $\mathrm{E}(Y|X=0) = \beta_0$
   females, $\mathrm{E}(Y|X=1) = \beta_0 + \beta_1$

   Then, $\beta_1 = \mathrm{E}(Y|X=1) - \mathrm{E}(Y|X=0)$
   $=$ true mean difference in HDL between females and males

   L.S. estimates (MLEs) give

   $$\widehat{\beta}_0 = \overline{Y}_{\mathrm{M}}$$
   $$\widehat{\beta}_0 + \widehat{\beta}_1 = \overline{Y}_{\mathrm{F}}$$
   $$\text{so } \widehat{\beta}_1 = \overline{Y}_{\mathrm{F}} - \overline{Y}_{\mathrm{M}}$$

   SAS attachment 1 gives $\overline{Y}_{\mathrm{F}} - \overline{Y}_{\mathrm{M}}$ (unadj.) $= \widehat{\beta}_1 \approx 13.23$

2. Compute means of AGE and ALC by SEX.

Means of AGE and ALC by SEX:

SEX10	N Obs	Variable	Mean
Male	96	AGE	53.3125000
		ALC	13.4612500
Female	83	AGE	53.3012048
		ALC	6.3249742

Compute correlation coefficients between HDL and AGE and between HDL and ALC.

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 179

	HDL
AGE	0.03937
	0.6008
ALC	0.08555
	0.2548

Another way?

How do these results relate to the question of interest, comparing HDL between men and women?

3. Model HDL as a function of AGE, BMI, ALC, SEX, and the ALC*SEX interaction.

Model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_3 X_4 + \epsilon$$

where

$X_1 =$ AGE, $X_2 =$ BMI, $X_3 =$ ALC, $X_4 =$ SEX;

Test the interaction:

Method 1: Fit the following model

Full: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_3 X_4 + \epsilon$
Reduced: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$

and calculate by hand the partial-F (or t) for H_0: $\beta_5 = 0$

or,

Method 2: Let SAS do it for you, i.e. use PROC GLM to get the partial-F (from Type I SS) or t.

From PROC GLM (SAS attachment 2), partial-$F = 3.99$ with p-value $= .0474$ so we reject H_0: $\beta_5 = 0$ at the $\alpha = .05$ level.

What is the sex difference in mean HDL for non-drinkers?

Need to compare the models for male and female non-drinkers. (Note, non-drinkers have $X_3 = 0$, males $\Rightarrow X_4 = 0$, females $\Rightarrow X_4 = 1$)

$$\begin{aligned} \text{male, n/d:}\ & Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \\ \text{female, n/d:}\ & Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 + \epsilon \end{aligned}$$

So,

$$\begin{aligned} \text{males:}\ & \mathrm{E}(Y|X_1, X_2, X_3 = 0, X_4 = 0) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \\ \text{females:}\ & \mathrm{E}(Y|X_1, X_2, X_3 = 0, X_4 = 1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_4 \end{aligned}$$

Thus,

$$\beta_4 = \mathrm{E}(Y|X_1, X_2, X_3 = 0, X_4 = 1) - \mathrm{E}(Y|X_1, X_2, X_3 = 0, X_4 = 0)$$

is the true difference in mean HDL between sexes among non-drinkers

From PROC GLM (SAS attachment 2), $\widehat{\beta}_4 \approx 12.24$

What is it for drinkers of 20 g/day alcohol?

This time we compare the models:

$$\begin{aligned} \text{males:}\ & Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + 20\beta_3 + \epsilon \\ \text{females:}\ & Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + 20\beta_3 + \beta_4 + 20\beta_5 + \epsilon \end{aligned}$$

so $(\beta_4 + 20\beta_5)$ is the true difference in mean HDL between sexes among drinkers of 20 g/day. From PROC GLM (SAS attachment 2), $\widehat{\beta}_4 + 20\widehat{\beta}_5 \approx 12.24 + 20(0.33) = 18.84$

4. We could also consider the question of how the ALC effect on HDL varies by SEX.

So what were we doing in #3??

Assess the "effect" of 20 g/wk alcohol consumption on HDL levels for each sex.

Here, we compare a group with X_3 g/day alcohol consumption with a group with 20 g/day more consumption ($X_3 + 20$).

For *males*, compare the models:

$$\begin{aligned} \underline{X_3 \text{ g/day}}&: \; Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \\ \underline{(X_3 + 20) \text{ g/day}}&: \; Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_3 + 20) + \epsilon \end{aligned}$$

so that $20\beta_3 =$ true mean difference in HDL between $(X_3 + 20)$ g/day male drinkers and X_3 g/day male drinkers.

From PROC GLM (SAS attachment 2),

$$20\widehat{\beta}_3 \approx 20(.168) = 3.36$$

Explain.

For *females*, compare the models:

$$\begin{aligned} \underline{X_3 \text{ g/day}}&: \; Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 + \beta_5 X_3 + \epsilon \\ \underline{(X_3 + 20) \text{ g/day}}&: \; Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_3 + 20) + \beta_4 + \\ & \qquad \beta_5 (X_3 + 20) + \epsilon \end{aligned}$$

so that $20(\beta_3 + \beta_5) =$ true mean difference in HDL between

20 g/day male drinkers and male non-drinkers

From PROC GLM (SAS attachment 2),

$$20(\widehat{\beta}_3 + \widehat{\beta}_5) \approx 20(.168 + .33) \approx 10$$

Explain. Compare to females.

5. Restrict to men only, so drop the sex terms from the HDL model, keeping AGE, BMI, and ALC.

Model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$

What is the partial correlation coefficient between HDL and ALC, controlling for AGE and BMI?

From PROC GLM (SAS attachment 3), we have

$$\begin{aligned} R_{Y,X_3|X_1,X_2} &= \text{sgn}(\widehat{\beta}_3)\sqrt{R^2_{Y,X_3|X_1,X_2}} \\ &= (1)\sqrt{\frac{\text{SSR}(X_3|X_1,X_2)}{\text{SSE}(X_1,X_2)}} \\ &= \sqrt{\frac{\text{SSR}(X_3|X_1,X_2)}{\text{SST} - \text{SSR}(X_1,X_2)}} \\ &\approx \sqrt{\frac{849.86}{13401.625 - (3.678 + 763.451)}} \\ &\approx .2594 \end{aligned}$$

From PROC REG with the PCORR2 option (SAS attachment 4), $R^2_{Y,X_3|X_1,X_2} \approx .0673$ so $R_{Y,X_3|X_1,X_2} \approx .2594$

What is the unadjusted correlation coefficient? $.2128$