

## Comments on Biostatistics 208 Lab #3 1/21/10

The comments below refer to the accompanying Stata log for the lab on the course website. A do-file for the lab is posted as well.

From the results for the first, unadjusted model, the `lincom` commands run using the do-loop or the `margins` command (or, `adjust` in Stata 10) reproduce what we got using `tab, sum`. There is strong evidence for heterogeneity in mean BMI across levels of physical activity ( $F(4, 546) = 11.41, p < 0.00005$ ), either from the first `testparm` result or the model  $F$ -test results (since the initial model includes no covariates). The pattern in the group-specific means (29.2  $kg/m^2$  among the women who describe themselves as “much less active” than their peers, somewhat higher in the next most active group, but then progressively lower across the remaining groups) constitutes a statistically significant linear trend ( $F(1, 546) = 18.34, p < 0.00005$  from the `test` command result). However, the fact that the observed decrease occurs only at the higher levels of activity is also consistent with a nonlinear trend, possibly quadratic. A method to evaluate evidence for a departure from linearity is discussed in Section 4.3.5 of the book. Recall that the results for `physact` as a score in the second model are not directly interpretable.

Although age and smoking are also statistically significant predictors of BMI, the adjusted results for `physact` are fairly similar to the unadjusted results, suggesting that these covariates are not strongly associated with physical activity in this sample.

- The intercept is an estimate of mean BMI for “much less active” women of age zero who do not smoke or drink – not directly interpretable, but also not of much interest. The coefficient for the 3<sup>rd</sup> level of `physact` estimates the adjusted difference in mean BMI between the “about as active” and “much less active” (reference) groups, holding the covariates constant. The interpretation of the coefficients for the other levels is similar. The coefficient for `age10` estimates the increase in mean BMI for every 10 years increase in age, while that for `drinkany` the difference in mean BMI between women who use any alcohol and those who do not – in both cases, accounting for the effects of the other variables.
- We still find the non-monotone pattern in the adjusted means, and strong evidence for heterogeneity and trend. Note that the adjusted means obtained by the `margins` command (or `adjust` in Stata 10) refer to women at the sample average age; they also represent an average of the means for smokers and non-smokers, and those who do and do not drink, weighted by the proportions in the various groups.
- Even with the adjustment in this model, you can easily verify that the coefficients for the four included `physact` indicator variables are exactly equal to the differences in the adjusted means between the “much less active” reference group and the other four physical activity groups.