

1 **Does a very statistically significant F value imply a high R<sup>2</sup>?**

- Problem 3.6, G&S page 105 [p 108/9 in 2nd Ed]

2 **Simple vs. Multiple Regression**

- Problem 3.8, G&S page 107 [p 110 in 2nd Ed]

3 **Smaller errors in prediction by adding predictors?**

- Problem 3.12, G&S page 108 [p 111 in 2nd Ed]

4 **Location! Location! Location!** (or did kkmn make up these data?)

[*"house price" data on course web page*]

a Test the improvement brought about by adding the *two indicator variables* for location to the model that already contains the number of rooms and the area of the property. Calculate the (multiple) partial F test manually. [*As I mention in the class notes, chapter 7 of nknw is clear on this -- I summarize that chapter in my web page for course 697. Also, if the improvement (and the sums of squares seem counterintuitively small, it's maybe the numbers -- you are probably still on the right track)*]

b There is a way to get INSIGHT to automatically do the same calculations for you: in the "spreadsheet" window of INSIGHT, make sure that the role of the location variable (last column in dataset) shows as "nominal" (i.e. categorical) rather than "interval". Now test the improvement brought about by adding the *categorical variable* location to the model that already contains the number of rooms and the area of the property.

PROC REG (via the Program Editor) does not handle categorical variables with 3 levels (there is no problem if just 2 levels, just leave it as an interval variable); instead (i) make indicator variables in PROC REG (they were provided here so there is no problem) or (ii) -- if you are too lazy to make them via programming statements -- you can use PROC GLM instead, as below. GLM uses the CLASS statement to specify a categorical variable.

```
proc glm data=sasuser.housepri;  
class location;  
model price = rooms area location; run;
```

c Why do you think this way of using location provides so little extra information?

5 **Is there gain without pain? Is more better?** [*"bodyfat data (large)" on course web page*]

Predict Y = percent body fat using Brozek's equation from X = the Adiposity index (=Weight / Height<sup>2</sup>. Then test the addition of the block of 9 "circumference" measurements (i.e. Neck Chest Hip Thigh Knee Ankle biceps Forearm Wrist ) to this equation.