Notes for Statlab section, 11/11/13

1. We obtain estimates for the model of SAT,

$$sat = \alpha + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 female + \beta_4 black + \beta_5 female \cdot black + u$$

	male	female
non-black	-	β_3
black	β_4	$\beta_3 + \beta_4 + \beta_5$

   - The omitted group is white males
   - β_3 is the difference in SAT score between non-black males and females
   - β_4 is the effect of being male and black relative to being a white male.
     - The p-value that Stata reports for $\hat{\beta}_4$ tests exactly the null hypothesis that there is no difference between nonblack males and black males, against the alternative that there is a difference.
   - $\beta_3 + \beta_5$ is the effect of being female, conditional on being black.
   - $\beta_3 + \beta_4 + \beta_5$ is the total effect of being black and female, so $\beta_4 + \beta_5$ is the difference in SAT score between black and white females
     - What if we want to estimate the effect of being black for females ($\beta_4 + \beta_5$)?
     - Run a slightly different regression:

$$sat = a + b_1 hsize + b_2 hsize^2 + b_3 male + b_4 black + b_5 male \cdot black + u$$

	male	female
non-black	b_3	-
black	$b_3 + b_4 + b_5$	b_4

2. Basics of interpreting coefficients

$$y = a + bx + u$$

   - y level - x level: Increasing x by one unit changes y by b units

$$y = a + blog(x) + u$$

   - y level - x log: Increasing x by one percent changes y by $b/100$ units

$$log(y) = a + bx + u$$

   - y log - x level: Increasing x by one unit changes y by $100 * b$ percent

$$log(y) = a + blog(x) + u$$

   - y log - x log: Increasing x by one percent changes y by b percent