Problem set 9: week 13

2./3. A t-test of $H_0 : \mu_{male} - \mu_{female} = 0$ against the two-sided alternative is equivalent to the testing $H_0 : \beta = 0$ against the two sided alternative in the regression $Y_i = \alpha + \beta X_i + \epsilon_i$ where Y is wages and X_i is 1 if male and zero otherwise. To see example of this see do file for this problem set.

4. see do file.

5.

- $\hat{\beta}_4 = 0.37$: males in the south earn 37% more than females in the south, on average, holding education and experience fixed.
- $\hat{\beta}_5 = -0.10$: wages in the north are 10% lower than wages in the south, on average, holding education and experience fixed.
- $\hat{\beta}_6 = 0.04$: the gender wage gap in the north is 4% higher than the gender wage gap in the south, on average, holding education and experience fixed.
- $H_0 : \beta_4 = 0$ v $H_1 : \beta_4 \neq 0$. $t = 0.038/0.042 = 0.91$. Fail to reject. Limited statistical evidence of difference in gender wage gap between north and south.

6.

I suggest running this regression for males only. The rationale behind this could be that man labour markets are sufficiently different from female labour markets to warrant separate consideration of each. Here we look at the male labour market. Statistically you could test this using a chow test as follows:

- Restricted model: $ln(wage_i) = \beta_0 + \beta_1 school_i + \beta_2 exper_i + \beta_3 exper_i^2 + \beta_4 male_i + \epsilon_i$.
- Unrestricted model: $ln(wage_i) = \beta_0 + \beta_1 school_i + \beta_2 exper_i + \beta_3 exper_i^2 + \beta_4 male_i + \delta_0(school_i * male_i) + \delta_2(exper_i * male_i) + \delta_3(exper_i^2 * male_i) + epsilon_i$.
- $H_0 : \delta_1 = \delta_2 = \delta_3 = 0$ v. H_1 at least one equality doesn't hold.
- $RSS_R = 480.73$, $RSS_{UR} = 465.89$, $q = 3$, $dof_{UR} = 2294$.
- $F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/dof_{UR}} = \frac{(480.73 - 465.89)/3}{465.89/2994} = 24.4$, reject at common levels of significance.
- There is significant evidence to suggest that a gender dummy variable is not enough to capture the difference in wage functions between men and women.

7.

- $H_0 : \delta_2 = \delta_2 = ... = \delta_8 = 0$ v. H_1 : at least one of the coefficients is non-zero.
- This is a multiple restriction test, therefore we need to use an F-test.
- $F = \frac{(R^2_{UR} - R^2_R)/q}{(1-R^2_{UR})/(dog_{UR})} \sim F_{q,dof_{UR}}$; where q is the number of restrictions in the null hypothesis (the number of equality signs); R^2_{UR} is the r-squared from the unrestricted model; R^2_R is the r-squared from the restricted model; and $dof_{UR} = n - k - 1$ is the degrees of freedom in the unrestricted model, which is always the number of observations minus the number of parameters being estimated.
- Our restricted model is $ln(wage) = \alpha + \beta_1 school + \beta_2 exper + \beta_3 exper^2 + \epsilon_i$. Estimating we get an $R^2_R = 0.2972$.
- Our unrestricted model is $ln(wage) = \alpha + \beta_1 school + \beta_2 exper + \beta_3 exper^2 + \sum_{j=2}^{8} \delta_j region_j + \epsilon_i$. Estimating we get an $R^2_{UR} = 0.3304$.
- $q = 7$, $n - k - 1 = 1374 - 10 - 1 = 1363$.
- $F = \frac{(0.3304-0.2972)/7}{(1-0.3304)/(1363)} = 10.2$. While $F^{0.05}_{7,1363} = 2.01$. Therefore, we strongly reject H_0 in favour of H_1.
- The regional dummies are jointly significant, which implies there is statistically significant regional variation in wages.