№1

$$Y = \beta_1 + \beta_2 EDUC + \beta_3 TE + \beta_4 K + \beta_5 MEDUC + \beta_6 FUDUC + \beta_7 weight + \delta \cdot S + \lambda \cdot EDUCS + u$$

~~EDU~~ $S = \begin{cases} 1, & male \\ 0, & female \end{cases}$

$$~~S~~ EDUCS \overset{}{=} \begin{cases} ~~S,~~ EDUC, & male \\ 0, & female \end{cases}$$

$$\text{"} \quad S \cdot EDUC$$

Coef. δ of Dummy variable S ~~can~~ shows ~~how~~ if there is a difference between the hourly earnings of males and females ~~with~~ with (holding other variables zero)

Coef. λ of Slope dummy variable ~~S~~ EDUCS shows if there is a difference in marginal effect of Education (one more year of education) on hourly earning of males and femalles (holding other variables constant)

If these coefs are significant, then there is a significant difference in impact of education on earnings for different genders.

Chow test can be used in this case. We should run three separate regressions ~~to~~ with this model

$$Y = \beta_1 + \beta_2 EDUC + \beta_3 TE + \beta_4 K + \beta_5 MEDUC + \beta_6 FEDUC + \beta_7 weight + u$$

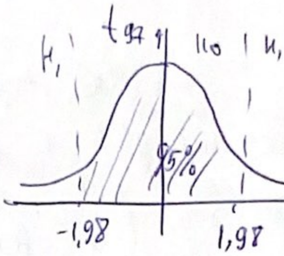
One for females only, one for males only and one for the whole sample and save their RSS

$$F(k, n-2k) = \frac{(RSS_{pooled} - RSS_A - RSS_B)/k}{(RSS_A + RSS_B)/(n-2k)}$$ – then we compute this F-stat, where RSS_A and RSS_B are RSS for models for males and females only, and RSS pooled is for the whole sample. H_0 is that there is no significant difference between coefs in separate models ⇒ we compare F-stat with f-crit and reject or don't reject the H_0.

№2.

$n = 100$

① $$\log \hat{y} = \underset{(-0{,}2)}{\overset{\hat\beta_1}{-0{,}5}} + \underset{(0{,}3)}{\overset{\hat\beta_2}{3{,}1}} \log x_1 + \underset{(0{,}1)}{\overset{\hat\beta_3}{0{,}4}} x_2$$



$\alpha = 0{,}05$ $\quad t_{100-3} = t_{97}$

for $\hat\beta_2$:

$$3{,}1 - 1{,}98 \cdot 0{,}3 \le \beta_2 \le 3{,}1 + 1{,}98 \cdot 0{,}3$$

$$2{,}506 \le \beta_2 \le 3{,}694$$

for $\hat\beta_3$:

$$0{,}4 - 1{,}98 \cdot 0{,}1 \le \beta_1 \le 0{,}4 + 1{,}98 \cdot 0{,}1$$

$$0{,}232 \le \beta_1 \le 0{,}568$$

With $p = 0{,}95$ $\beta_1 \in [0{,}232; 0{,}568] \Rightarrow$ ~~X_3~~ X_2 is a significant variable at 5% signif. level.

② $$\hat y = \overset{\hat\beta_1}{-0{,}5} + \overset{\hat\beta_2}{2}x_1 + \overset{\hat\beta_3}{3{,}1}x_2 + \overset{\delta}{3}D + \overset{\lambda}{2{,}1}x_1 \cdot D$$

for ① if x_1 increases on 1% y increases on $\hat\beta_2$%
($\hat\beta_2$ is elasticity)

if x_2 increases on 1 unit y increases on $\hat\beta_3 \cdot 100\%$

$\hat\beta_1$ is $\widehat{\log y}$ for $\log x_1 = x_2 = 0$

for ② if x_1 increases on 1 unit y increases on $\hat\beta_2$ units

if x_2 increases on 1 unit y increases on $\hat\beta_3$ units

δ – difference in intercept ~~for~~ between the referent category and the other one compared with it

λ – difference ~~im~~ in marginal effect of x_1 on y between the referent category and the other one compared with it (holding other variables constant)

№3

$n = 300$

$GPA_i = \beta_0 + \beta_1 \cdot CLASS_i + \varepsilon_i$

if there is an important omitted variable (talent), then the estimates ~~with the~~ of class will be biased (if class is correlated with talent)

~~E~~ if $GPA_i = \beta_0 + \beta_1 CLASS_i + \beta_2 TALENT_i + \varepsilon_i$

$$\hat{\beta}_1 = \beta_1 + \beta_2 \frac{\sum(CLASS_i - \overline{CLASS})(TALENT_i - \overline{TALENT})}{\sum(CLASS_i - \overline{CLASS})} + \frac{\sum(CLASS_i - \overline{CLASS})(\varepsilon_i - \bar{\varepsilon})}{\sum(CLASS_i - \overline{CLASS})}$$

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 \cdot \frac{\sum(CLASS_i - \overline{CLASS})(TALENT_i - \overline{TALENT})}{\sum(CLASS_i - \overline{CLASS})}$$

if corr (CLASS, TALENT) $= 0$ there will be no bias

if corr > 0 and $\beta_2 > 0 \Rightarrow \beta_1$ will be biased upwards

~~if corr~~ (which is a sensible : for this model)

$\Rightarrow$ s.e. of coefs will be invalid

if ~~talent~~ there is a problem of dependence of the regressor and disturbance term, instrumental variable might help to make and OLS estimator consistent. (otherwise $\text{plim } \hat{\beta}_i^{OLS} \neq \beta_i$)

Probably distance from a university is correlated with class attendance (but not perfectly) is not dependent on $\varepsilon \Rightarrow$ it can be used to partly replace CLASS variable

$$\hat{\beta}_1^{IV} = \frac{\sum(DIST - \overline{DIST})(GPA - \overline{GPA})}{\sum(DIST - \overline{DIST})(CLASS - \overline{CLASS})} = \frac{\sum(DIST - \overline{DIST}) \cdot}{\sum(DIST - \overline{DIST}) \cdot}$$

$$\frac{\cdot\, (\beta_0 + \beta_1 CLASS + \varepsilon_i - \beta_0 - \beta_1 \overline{CLASS} - \bar{\varepsilon})}{\cdot\, (class - \overline{CLASS})} = \beta_1 + \frac{\sum(DIST - \overline{DIST})(\varepsilon_i - \bar{\varepsilon})}{\sum(DIST - \overline{DIST})(class - \overline{class})}$$

$$\Rightarrow \text{plim } \hat{\beta}_1^{IV} = \beta_1 + \frac{cov(dist, \varepsilon)}{cov(DIST, class)} = \beta_1 \Rightarrow \text{consistent}$$

№ 4

$E = \beta_0 + \beta_1 X + \beta_2 Z + u$

$\sigma_u = f(Z)$

1) Goldfeld - Quandt test

$$F\text{-stat} = \frac{250/(50-3)}{200/(50-3)} = 1{,}25$$

Fcrit for $F(40,40)$ and $\alpha = 0{,}05$

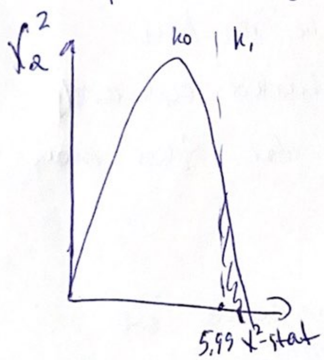
1,69 (for $F(60,60)$ 1,5)

H0 is not rejected =>

=> homosceadasticity is not rejected

2) White test

χ^2-stat = 100 · 0,3 = 30



H0 is rejected => there is some association between the variance of u and the regressors.

~~White test may~~

Goldfeld - Quandt test is used to understand wheth if ~~...~~ σ_u^2 is proportional to X, but this might not be the case

3) Two methods to deal with heteroskedasticity:

1) devide everything by the variable, with ~~on~~ which u can be proportional.

2) use nonlinear model (if the original model is logarithmic ~~a~~ with homosk. u, then it will be heterosk. with natural unit model)