8. Qualitative Information

Binary or Dummy Variables

Read Wooldridge (2013), Chapter 8

Outline

I. Describing Qualitative Information
II. A Single Dummy Independent Variable
III. Using Dummy Variable for Multiple Categories
IV. Interactions Involving Dummy Variables
V. Chow Test

I. Describing Qualitative Information

- Quantitative vs. Qualitative Variables
  Quantitative - *wage, educ, sales, exper, tenure, profit*
  Qualitative - gender, race, or policy evaluation
- Examples: male/female; married/single; part time/full time; capital intensive/labor intensive
- Qualitative factors often come in the form of binary information – zero or one (0 or 1)
  female = 1 Otherwise = 0
  married = 1 Otherwise = 0

Table 7.1
A Partial Listing of the Data in WAGE1.RAW

person	*wage*	*educ*	*exper*	*female*	*married*
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
.	.	.	.	.	.
525	11.56	16	5	0	1
526	3.50	14	5	1	0

II. A Single Dummy Independent Variable

- Consider a model
  $wage = \beta_0 + \delta_0 female + u$
  binary variable
  $female = 1$, if female
  $female = 0$, otherwise (if male)

- Comparing mean wages of male and female
  $E(wage|female=1) = \beta_0 + \delta_0$ ($female$=1)
  $E(wage|female=0) = \beta_0$ ($female$=0)
  $\delta_0 = E(wage|female) - E(wage|male)$
  δ_0 : the difference in hourly wage between female and male (the base group)

Example: Comparison-of-means-test between two groups:
$wage = \beta_0 + \delta_0 female + u$

$\widehat{wage} = 7.099 - 2.51 female$
s.e. (.210) (.3034)
t-stat [33.81] [-8.28]
n=526 R^2=.115667 R^2-bar = .113979

1) What is $\hat{\delta}_0$?
2) What is the average hourly wage for male and female?

Example: add *educ* to the model:
$wage = \beta_0 + \delta_0 female + \beta_1 educ + u$

$\widehat{wage} = .6228 - 2.27 female + .50652 educ$
s.e (.672) (.2790) (.05391)
[t-stat] [0.527] [-3.15] [10.05]
n=526 R^2=.2588 R^2-bar=.2559

Interpretation:
1) What is $\hat{\beta}_0$?
2) What is $\hat{\delta}_0$?
3) What is $\hat{\beta}_1$?
4) Graph with different intercepts.

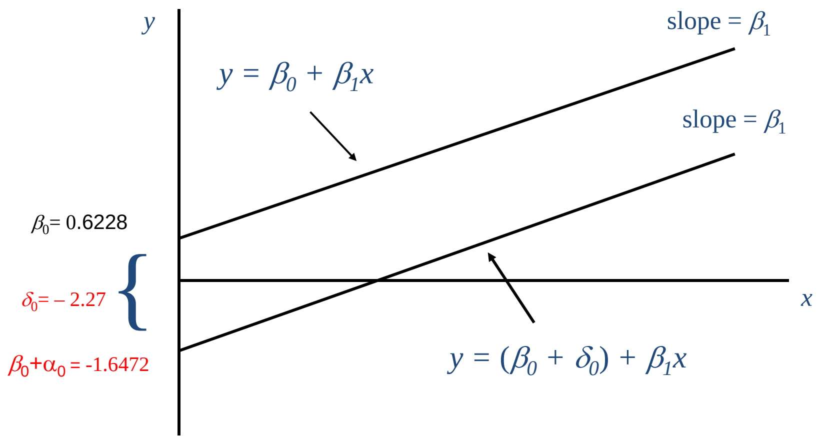
Different Intercept but same slope

$E(wage|female = 1, educ) = \beta_0 + \delta_0 + \beta_1 educ$
$= -1.6472 + .5065 educ$

$E(wage|female=0, educ) = \beta_0 + \beta_1 educ$
$= .6228 + .5065 educ$

Can you test whether the mean wage for female with high school diploma is equal to zero?

Example of $\delta_0 < 0$



Can we include both *male* and *female* in the model?

No! problem of perfect collinearity (dummy variable trap)

$\widehat{wage} = \beta_0(1) + \gamma_0 male + \delta_0 female + \beta_1 educ + u$

$\therefore male + female = 1$

- Different Base group

  $wage = \alpha_0 + \gamma_0 male + \beta_1 educ + u$

  male = 1 if male

  = 0 otherwise (if female)

  $\gamma_0 = E(wage|male=1, educ) - E(wage|male = 0, educ)$

- What is γ_0?

Dependent variable : log(*y*)

$\log(\widehat{wage}) = \hat{\beta}_0 + \hat{\delta}_0 female + \hat{\beta}_1 educ + \hat{\beta}_2 exper + \hat{\beta}_3 tenure$

$= .501 - .301 female + .087 educ + .0046 exper + .0174 tenure$

t-stat [4.92] [-8.09] [12.6] [2.85] [5.84]

- Interpret the coefficient of female ($\hat{\delta}_0$).

  Women earn about 30.1% less than men, for the same levels of *educ, exper,* and *tenure.*

- Approximate vs. Exact proportionate change

  $\log(wage_F) - \log(wage_M)$ = - 0.301

  $[wage_F/wage_M - 1] \approx [\exp(-0.301) - 1]$ = - 0.2599

Regress log(*wage*) on *female*, *educ*, *exper*, and *tenure*

Dependent Variable: LOG(WAGE)

Sample: 1 526

Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.501348	0.101902	4.919885	0
FEMALE	-0.30115	0.037246	-8.085414	0
EDUC	0.087462	0.006939	12.60463	0
EXPER	0.004629	0.001627	2.845087	0.0046
TENURE	0.017367	0.002976	5.835235	0

R-squared	0.39227	Mean dependent var	1.623268
Adjusted R-squared	0.387604	S.D. dependent var	0.531538
S.E. of regression	0.415959	Akaike info criterion	1.093001
Sum squared resid	90.14445	Schwarz criterion	1.133545
Log likelihood	-282.459	F-statistic	84.07213
Durbin-Watson stat	1.775217	Prob(F-statistic)	0

Other use: Project Evaluation

- Example: The goal is to study the effect of capital intensity on state enterprise employment.
  - Control group : labor intensive project
  - Experiment group: capital intensive project

$empnum = \beta_0 + \delta_0 K510 + \beta_1 invsize + \beta_2 years + u$

K510	=1	capital intensive project
	= 0	otherwise (labor intensive)
empnum		workers employed (persons)
invsize		amount of investment (million baht)
years		duration of the project (years)

$\widehat{empnum} = 760.2 - 203 K510 + 0.148 invsize - 123.7 years$

s.e. (182.6) (183.4) (.0168) (33.09)

t-stat [4.16] [-1.11] [8.80] [-3.74]

$n=106$ $R^2=.433282$ $R^2\text{-bar}=.416614$

- Interpretation:

1) Interpret the coefficient on *K510*?
2) Is K510 statistically significant?
3) Interpret the coefficients on *invsize* and *years*?

Regress *empnum* on K510, *invsize, years*

Dependent Variable: EMPNUM
Method: Least Squares
Sample: 1 154
Included observations: 106
Excluded observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	760.1944	182.6296	4.162492	0.0001
K510	-203.3838	183.3694	-1.109148	0.27
INVSIZE	0.147838	0.016804	8.798054	0
YEARS	-123.7395	33.08859	-3.739644	0.0003

R-squared	0.433282	Mean dependent var	622.8517
Adjusted R-squared	0.416614	S.D. dependent var	997.0641
S.E. of regression	761.5549	Akaike info criterion	16.14561
Sum squared resid	59156512	Schwarz criterion	16.24611
Log likelihood	-851.7172	F-statistic	25.99455
Durbin-Watson stat	1.240076	Prob(F-statistic)	0

III. Using Dummy Variables for Multiple Categories

- Consider the model:
  $\log(wage) = \beta_0 + \delta_0 female + \delta_1 married + u$

  Two dummy variables: *female* and *married*

- δ_1 is the proportional differential in wages between those who are married and those who are not (the base group).

Multiple Categories

- Example: Hourly wage equation with two categories

$$\log(wage) = 1.653 - .359 female + .233 married$$

s.e.	(.0417)	(.0426)	(.0436)
t-stat	[39.66]	[-8.44]	[5.36]

$n = 526$; $R^2 = .18438$; R^2-bar = .18126

Regress log(*wage*) on *female* and *married*

Dependent Variable: LOG(WAGE)
Method: Least Squares
Sample: 1 526
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.653447	0.041688	39.66281	0
FEMALE	-0.35934	0.04257	-8.441102	0
MARRIED	0.233371	0.043568	5.356471	0

R-squared	0.18438	Mean dependent var	1.623268
Adjusted R-squared	0.181261	S.D. dependent var	0.531538
S.E. of regression	0.480958	Akaike info criterion	1.379614
Sum squared resid	120.9807	Schwarz criterion	1.403941
Log likelihood	-359.839	F-statistic	59.1149
Durbin-Watson stat	1.792627	Prob(F-statistic)	0

Example: Effect of Capital Intensity

- Suppose we want to divide investment projects into three categories
  1) 0-30 percent of total project fund invested in capital
  2) 30-60 percent invested in capital
  3) 60-100 percent invested in capital

- Dummy variables: 0-30% capital investment as the base group

K36	=1	30-60 percent capital investment
	=0	Otherwise
K610	=1	60-100 percent capital investment
	= 0	Otherwise

$$empnum = \beta_0 + \delta_0 K36 + \delta_1 K610 + \beta_1 invsize + \beta_2 years + u$$

$$\widehat{empnum} = 967.9 - 359.3K36 - 397.3K610 + .148invsize - 126.0years$$

t-stat	[3.17]	[-1.10]	[-1.31]	[8.79]	[-3.78]

$n=106$, $R^2=0.436065$, R^2-bar=.413731

Interpretation:

1) What is the economic significance of $\hat{\delta}_0$? Interpret .
2) What is the economic significance of $\hat{\delta}_1$? Interpret .
3) Should we include K36 and K610 in the model?

View/Coefficient Tests/Redundant Variables..

Redundant Variables: K36 K610			
F-statistic	0.861298	Probability	0.425694
Log likelihood ratio	1.792628	Probability	0.408071

View/Coefficient Tests/Wald - Coefficient Restrictions

Wald Test:			
Equation: Untitled			
Null Hypothesis:	C(2)=0		
	C(3)=0		
F-statistic	0.861298	Probability	0.425694
Chi-square	1.722595	Probability	0.422613

Regress *empnum* on *K36, K610, invsize* and *years*

Dependent Variable: EMPNUM
Sample: 1 154
Included observations: 106
Excluded observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	967.899	305.2325	3.171022	0.002
K36	-359.283	327.6044	-1.096698	0.2754
K610	-397.298	302.7533	-1.312282	0.1924
INVSIZE	0.14796	0.016836	8.788151	0
YEARS	-125.994	33.28816	-3.784936	0.0003

R-squared	0.436065	Mean dependent var	622.8517
Adjusted R-squared	0.413731	S.D. dependent var	997.0641
S.E. of regression	763.4342	Akaike info criterion	16.15955
Sum squared resid	58866008	Schwarz criterion	16.28519
Log likelihood	-851.456	F-statistic	19.52466
Durbin-Watson stat	1.256433	Prob(F-statistic)	0

Wage differences among four our groups

marrmale = 1, =0 otherwise
marrfem = 1, =0 otherwise
singfem = 1, =0 otherwise
singmale - base group

$$\log(wage) = \delta_0 + \delta_1 marrmale + \delta_2 marrfem + \delta_3 singfem + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 tenure + \beta_5 tenure^2 + u$$

obs	WAGE	FEMALE	MARRIED	MARRMALE	MARRFEM	SINGFEM
1	3.1	1	0	0	0	1
2	3.24	1	1	0	1	0
3	3	0	0	0	0	0
4	6	0	1	1	0	0
5	5.3	0	1	1	0	0
..	..	..	..	..	..	..
..	..	..	..	..	..	..

Regress log(*wage*) on *marrmale, marrfem, singfem, educ, exper, exper²*, *tenure*, and *tenure²*

Dependent Variable: LOG(WAGE)
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.321378	0.100009	3.213492	0.0014
MARRMALE	0.212676	0.055357	3.841881	0.0001
MARRFEM	-0.198268	0.057835	-3.428132	0.0007
SINGFEM	-0.11035	0.055742	-1.979658	0.0483
EDUC	0.07891	0.006694	11.78733	0
EXPER	0.026801	0.005243	5.111835	0
EXPER^2	-0.000535	0.00011	-4.847105	0
TENURE	0.029088	0.006762	4.301614	0
TENURE^2	-0.000533	0.000231	-2.305553	0.0215

R-squared	0.460877	Mean dependent var	1.623268
Adjusted R-squared	0.452535	S.D. dependent var	0.531538

Example: Estimated wage differential between groups

$\log(\widehat{wage}) = \hat{\beta}_0 + \hat{\delta}_1 marrmale + \hat{\delta}_2 marrfem + \hat{\delta}_3 singfem + \ldots$

$\log(\widehat{wage}) = .321 + .213 marrmale - .198 marrfem - .110 singfem + \ldots$

s.e.	(.100)	(.055)	(.578)	(.056)	
t-stat	[3.21]	[3.84]	[-3.42]	[-1.98]	

n=526, R^2=.460877 R^2-bar=.452535

Interpretation

1) What is the base category?
2) Interpret $\hat{\delta}_1$, $\hat{\delta}_2$, and $\hat{\delta}_3$
3) What is the proportionate difference between single and married women (the new base group)?
4) Test whether the proportionate difference between single women and married women (the base group) is statistically significant.

$\log(\widehat{wage}) = \hat{\beta}_0 + \hat{\delta}_1 marrmale + \hat{\delta}_2 marrfem + \hat{\delta}_3 singfem + \hat{\beta}_1 educ + \hat{\beta}_2 exper + \hat{\beta}_3 exper^2 + \hat{\beta}_4 tenure + \hat{\beta}_5 tenure^2$

Summary: four groups

- Intercept for the base group – the overall intercept in the model.
- Dummy variable coefficient – the estimated proportional differential in wages between the group of interest and the base group.

IV. Interactions Involving Dummy Variables

Model: interaction terms between two dummies

$\log(wage) = \beta_0 + \delta_1 female + \delta_2 married + \delta_3 female*married + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 tenure + \beta_5 tenure^2 + u$

- Interaction term : female*married

  $\Delta\log(wage)/\Delta married = \delta_2 + \delta_3 female$

  $\Delta\log(wage)/\Delta female = \delta_1 + \delta_3 married$

Interactions

Example:

$\log(\widehat{wage}) = .321 - .110 female + .213 married - .301 female*married + \ldots$

s.e.	(0.100)	(0.056)	(0.055)	(0.072)
p-value	{.0014}	{.0483}	{.0001}	{0}

Does married premium depend on gender?

Regress log(*wage*) on *female, married, female*married educ, exper, exper^2, tenure* and *tenure^2*

Dependent Variable: LOG(WAGE)
Sample: 1 526
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.321378	0.100009	3.213492	0.0014
FEMALE	-0.11035	0.055742	-1.979658	0.0483
MARRIED	0.212676	0.055357	3.841881	0.0001
FEMALE*MARRIED	-0.30059	0.071767	-4.188461	0
EDUC	0.07891	0.006694	11.78733	0
EXPER	0.026801	0.005243	5.111835	0
EXPER^2	-0.00054	0.00011	-4.847105	0
TENURE	0.029088	0.006762	4.301614	0
TENURE^2	-0.00053	0.000231	-2.305553	0.0215

R-squared	0.460877	Mean dependent var	1.623268
Adjusted R-squared	0.452535	S.D. dependent var	0.531538

Comparing Models
(*)log($\widehat{wage}$)= .321 -.110*female* +.213*married* -.301*female*married* + ...
(**)log($\widehat{wage}$) = .321 +.213*marrmale* –.198*marrfem* -.110*singfem* + ...

Interpretation:

1) In Model (*), set female=1 and married=1. Is there any relationship to Model (**)?
2) What is the intercept, $\hat{\beta}_0$?
3) What is the estimated proportional wage differential between married female and single male?
4) Test whether the proportional wage differential between married female and single men is statistically significant.

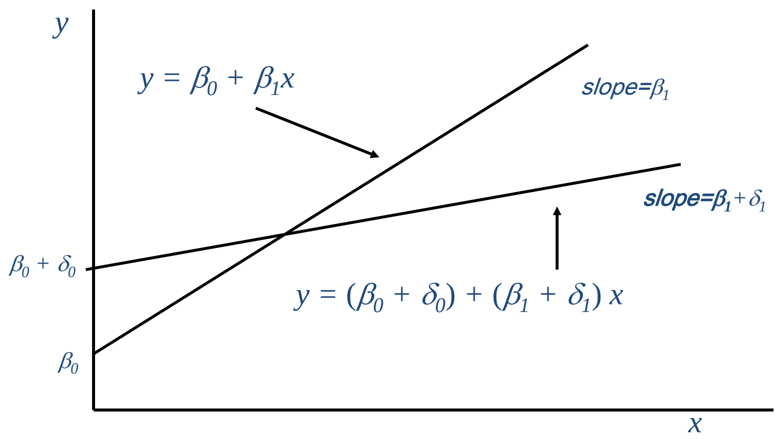
log ($\widehat{wage}$) = $\hat{\beta}_0 + \hat{\delta}_1 female + \hat{\delta}_2 married + \hat{\delta}_3 female*married + \hat{\beta}_1 educ + \hat{\beta}_2 exper + \hat{\beta}_3 exper^2 + \hat{\beta}_4 tenure + \hat{\beta}_5 tenure^2$

Interaction between binary and quantitative variables: allowing for different slope and Intercept

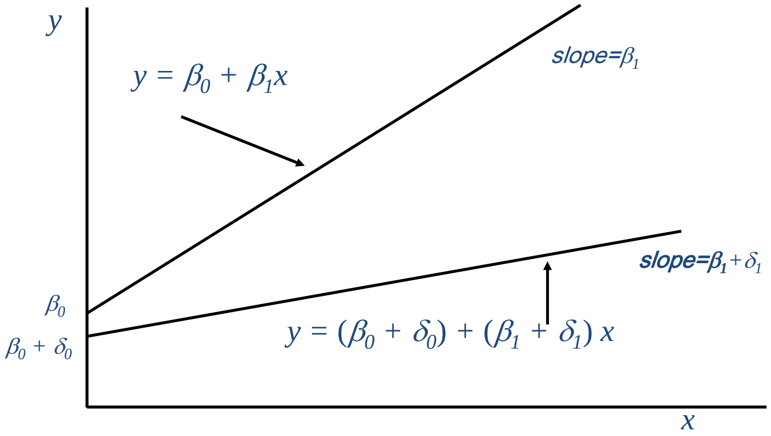
- Consider the model
  $wage = \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 female*educ + u$
  For *female*: intercept = $\beta_0 + \delta_0$
  slope = $\beta_1 + \delta_1$
  For *male:* Intercept = β_0
  Slope = β_1
- Interpretation
  δ_0 : the difference in intercepts between women and men.
  δ_1 : the difference in slopes between women and men

  Graphs: Case 1: $\delta_0 > 0$, $\delta_1 < 0$
  Case 2: $\delta_0 < 0$, $\delta_1 < 0$

Example of $\delta_0 > 0$ and $\delta_1 < 0$



Example of $\delta_0 < 0$ and $\delta_1 < 0$



Example: $wage = \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 female*educ + u$

$\widehat{wage}$ =	.2004	-1.19*female*	+.539*educ*	-.086female**educ*
s.e.	(.844)	(1.33)	(.064)	(.1036)
t-stat	[.238]	[-.90]	[8.40]	[-.83]
p-value	{.812}	{.366}	{.000}	{.407}

Interpretation:

1) What is the estimated return to education for men?
2) Interpret $\hat{\delta}_0$ and Test.
3) Interpret $\hat{\delta}_1$ and Test.
4) Test whether there are differences in regression functions between male and female.

Regress *wage* on *female educ* and *female*educ*

Dependent Variable: WAGE
Sample: 1 526
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.200496	0.843562	0.237678	0.8122
FEMALE	-1.19852	1.32504	-0.904518	0.3661
EDUC	0.539476	0.064223	8.400054	0
FEMALE*EDUC	-0.086	0.103639	-0.829795	0.407

R-squared	0.259796	Mean dependent var	5.896103
Adjusted R-squared	0.255542	S.D. dependent var	3.693086
S.E. of regression	3.186469	Akaike info criterion	5.163279
Sum squared resid	5300.17	Schwarz criterion	5.195715
Log likelihood	-1353.94	F-statistic	61.07022
Durbin-Watson stat	1.825925	Prob(F-statistic)	0

View/Coefficient Tests/Redundant Variables

Redundant Variables: FEMALE FEMALE*EDUC			
F-statistic	33.51095	Probability	0
Log likelihood ratio	63.53858	Probability	0

View/Coefficient Tests/Wald – Coefficient Restrictions

Wald Test:			
Equation: Untitled			
Null Hypothesis:	C(2)=0		
	C(4)=0		
F-statistic	33.51095	Probability	0
Chi-square	67.0219	Probability	0

V. Chow test: a test for differences in Regression Function

(0) Male : $wage = \beta_0 + \beta_1 educ + u_M$ $n_M = 274$

(1) Female : $wage = \alpha_0 + \alpha_1 educ + u_F$ $n_F = 252$

(P) All : $wage = b_0 + b_1 educ + u$ $n = 526$

Male: $\widehat{wage} = 0.200 + 0.539 educ$
t-stat [.197] [6.97] $SSR_M = 4009.931$

Female: $\widehat{wage} = .998 + .453 educ$
t-stat [-1.37] [7.82] $SSR_F = 1290.239$

Pooled: $\widehat{wage} = -.905 + 0.541 educ$
t-stat [-1.32] [10.17] $SSR = 5980.62$

Data for Chow Test

Obser	Female	educ	wage
1	0	11	3
2	0	8	6
3	0	12	5.3
..	..	..	..
273	0	15	4.67
274	0	16	11.56
275	1	11	3.1
276	1	12	3.24
..	..	..	..
525	1	10	2.27
526	1	14	3.5

Chow Breakpoint is the observation 275

Pooled Data: Regress *wage* on *educ*

Dependent Variable: WAGE
Method: Least Squares
Sample: 1 526
Included observations: 526

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.90485	0.684968	-1.321013	0.1871
EDUC	0.541359	0.053248	10.16675	0

R-squared	0.164758	Mean dependent var	5.896103
Adjusted R-squared	0.163164	S.D. dependent var	3.693086
S.E. of regression	3.37839	Akaike info criterion	5.27647
Sum squared resid	5980.682	Schwarz criterion	5.292688
Log likelihood	-1385.71	F-statistic	103.3627
Durbin-Watson stat	1.724873	Prob(F-statistic)	0

Male: Regress *wage* on *educ*

Dependent Variable: WAGE
Method: Least Squares
Sample: 1 274
Included observations: 274

Choose Estimate.
In the "Equation Specification“ box, replace with observations between 1-274.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.200496	1.016462	0.197249	0.8438
EDUC	0.539476	0.077386	6.971203	0

R-squared	0.151585	Mean dependent var	7.099489
Adjusted R-squared	0.148465	S.D. dependent var	4.160858
S.E. of regression	3.839582	Akaike info criterion	5.535877
Sum squared resid	4009.931	Schwarz criterion	5.56225
Log likelihood	-756.415	F-statistic	48.59767
Durbin-Watson stat	1.918787	Prob(F-statistic)	0

Female: Regress *wage* on *educ*

Dependent Variable: WAGE
Method: Least Squares
Sample: 275 526
Included observations: 252

Choose Estimate.
In the "Equation Specification" box,
replace with observations between 275-526.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.99803	0.728507	-1.369962	0.1719
EDUC	0.453477	0.057992	7.819667	0

R-squared	0.196522	Mean dependent var	4.587659
Adjusted R-squared	0.193308	S.D. dependent var	2.529363
S.E. of regression	2.271774	Akaike info criterion	4.486904
Sum squared resid	1290.239	Schwarz criterion	4.514915
Log likelihood	-563.35	F-statistic	61.14719
Durbin-Watson stat	1.983793	Prob(F-statistic)	0

Assumptions: $u_M \sim N(0, \sigma^2)$
$u_F \sim N(0, \sigma^2)$
u_M and u_F are independently distributed

Chow Test is basically F-test, a relative increase in SSR when moving from unrestricted to restricted model.

Step 1: Restricted model: The pooled data restrict the slopes and intercepts are equal ($\beta_0 = \alpha_0$ and $\beta_1 = \alpha_1$)

$SSR_R = 5980.682$ $DF_R = n\text{-}k\text{-}1 = 524$

Step 2: Unrestricted model. SSR is found by the sum of SSRs from two models

$SSRu = SSR_M + SSR_F = 4009.9 + 1290.2 = 5300$

$DF_U = (n_M\text{-}k\text{-}1) + (n_F\text{-}k\text{-}1) = n_M + n_F - 2(k+1) = n - 2(k+1) = 522$

F-statistic:
F =(5981-5300)/2 divided by 5300/522 = 33.5

Step 3:

$F = (SSR_R - SSR_U)/(DF_R - DF_U)$ dividend by SSR_U/DF_U

$$F = \frac{(SSR_R - SSR_U)/(k+1)}{SSR_U/(n-2(k+1))}$$

- Rejection Rule F > C
  C = $F_{2,522}$ (table G3) C = 3.00 (α=5%)
  = 4.61 (α=1%)
  Conclusion: Male and female regressions are different.

- Chow Test may tell us that the regression functions of male and female are different, but it does not tell us whether it is due to the intercept or slope.

Steps for Chow Test

Step 1: Obtain the original equation for both groups (Male: 1-274) Female (275-526)

Pooled: $\widehat{wage} = -.905 + 0.541 educ$

Step 2: In the "equation" window,
- Choose View/Stability Tests/Chow Breakpoint Test
- Pick the breakpoint observation: 275

Step 1: Pooled Data: Regress *wage* on *educ*

Dependent Variable: WAGE				
Method: Least Squares				
Sample: 1 526				
Included observations: 526				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.90485	0.684968	-1.321013	0.1871
EDUC	0.541359	0.053248	10.16675	0
R-squared	0.164758	Mean dependent var		5.896103
Adjusted R-squared	0.163164	S.D. dependent var		3.693086
S.E. of regression	3.37839	Akaike info criterion		5.27647
Sum squared resid	5980.682	Schwarz criterion		5.292688
Log likelihood	-1385.71	F-statistic		103.3627
Durbin-Watson stat	1.724873	Prob(F-statistic)		0

Step 2: Eviews In the "equation" window, choose View/Stability Tests/Chow Breakpoint Test

Chow Breakpoint Test: 275			
F-statistic	33.51095	Probability	0
Log likelihood ratio	63.53861	Probability	0

Pick the breakpoint observation: 275

Chow Test versus Dummy Variables

- Dummy Variables:

  $wage = \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 female*educ + u$
  - t-Test: $H_0 : \delta_0 = 0$
  - t-Test: $H_0 : \delta_1 = 0$
  - F-Test: $H_0 : \delta_0 = 0 , \delta_1 = 0$

- Better features of using dummies than Chow:

  (1) We can test certain slope or intercept whether the interested coefficient is significant.

  (2) There is only one model equation with dummies.

Recap of Binary Variable

- Describing Qualitative Information
- A Single Dummy Independent Variable
- Using Dummy Variable for Multiple Categories
- Interactions Involving Dummy Variables
- Chow Test