Assignment 2: Bivariate and Multiple Regression

ECO 321

DUE: Thursday March 8 in class

Instructions: You need to write a Stata do-file to answer question (2). Your do-file must produce a log file that contains your Stata output. You must submit your log file along with your answers to the assignment. Assignments that do not contain log files will be marked down. You may work in groups (up to 4 people), but each student must submit his/her own assignment. If you work in a group and do not submit your own assignment, you will receive a zero for the assignment. Remember that late assignments will not be accepted.

1. Consider the OLS estimates of the multiple regression model:

$$\hat{Y}_i = \underset{SE(\hat{\beta}_0)}{\hat{\beta}_0} + \underset{\underset{2.2}{SE(\hat{\beta}_1)}}{\overset{3.5}{\hat{\beta}_1}} X_{i1} + \underset{\underset{1}{SE(\hat{\beta}_2)}}{\overset{2.1}{\hat{\beta}_2}} X_{i2}$$

where $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ are the estimates of the coefficients, and $SE(\hat{\beta}_0)$, $SE(\hat{\beta}_1)$, and $SE(\hat{\beta}_2)$ are the standard errors of the estimates.

a) What is the definition of consistency? Which assumptions do we need for $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ to be consistent?

b) We want to test the statistical significance of β_2. Write down the null hypothesis, alternative hypothesis, select α, and write down the t-statistic.

c) Write down the steps you would use to find the p-value for the statistical significance test for β_2. Now suppose $\hat{\beta}_2 = 2.1$ and $SE(\hat{\beta}_2) = 1$. Find the p-value.

d) Write down the general expression for a 99% confidence interval for β_1.

e) Suppose $\hat{\beta}_1 = 3.5$ and $SE(\hat{\beta}_1) = 2.2$. What is the 99% confidence interval for a one-unit change in X_{i1} ($\Delta X_{i1} = 1$)? What is the 99% confidence interval for a 8-unit change in X_{i1} ($\Delta X_{i1} = 8$)? Using only the confidence intervals, what can we conclude about the statistical significance of β_1?

(f) Based on your answers to parts (c) and (e), how might you rewrite the model? (Hint: Would you exclude any variable, and if so, why?)

2. Smoking and birth outcomes. The data set bwght.dta contains two variables: the birth weight (in ounces) of a newborn (*bwght*) and the number of cigarettes the mother smoked per day during pregnancy (*cigs*).

a) Load the data set in Stata (via the `use` command). Construct a dummy variable (using the `gen` command) named *anycig* that equals one if the mother smoked at least one cigarette per day and zero otherwise. Use the `tab` command (or whatever command you like) to determine what share (percent) of moms smoked during pregnancy. ✓ via tab

b) We want to test the null hypothesis that the average birth weight of babies born to smokers vs. non-smokers is the same. Use the `reg` command to estimate the model: $\boxed{bwght_i = \beta_0 + \beta_1 anycig_i + u_i.}$ How do you interpret $\hat{\beta}_1$ in this regression? How do you interpret $\hat{\beta}_0$? Can you reject the null hypothesis that the average birth weights of babies born to smokers vs. non-smokers is the same (use $\alpha = 0.05$)? ✓

c) Rather than looking at the effect of smoking versus not smoking, we want to determine how smoking intensity affects birth weights. Use the `reg` command to estimate: $bwght_i = \beta_0 + \beta_1 cigs_i + u_i$. Report your regression results in standard form. What is the marginal effect of smoking an additional cigarette on birthweight? Is the effect statistically significant?

*d) Using your regression results in part (c), compute the predicted birth weight of a child whose mother does not smoke, and of a child whose mother smokes a pack a day (assume 20 cigarettes per pack). Is the difference between these values statistically significant?

e) Does the simple regression in part (c) necessarily capture a causal relationship between the child's birth weight and the mother's smoking habits? Explain. (Hint: What is contained in u_i?)

f) What is homoskedasticity, and is it likely to be a valid assumption in this example?

g) Estimate the regression from part (c) without assuming homoskedasticity (i.e. use the `reg` command, but specify the option `robust` to account for heteroskedasticity in the error terms). Compare your standard errors to the standard errors from the regression that assumes homoskedasticity. What is the difference? Do you suspect heteroskedasticity is a problem in this data? Is there a benefit or cost to using robust standard errors?

3. In this question we want to explore the importance (or lack thereof) of unionization on household earnings. In particular, some economic theories suggest that collective bargaining (such as unionization) matters more for traditionally marginalized groups—female and minority workers. Consider the models:

$$FamIncome_i = \beta_0 + \beta_1 HAge_i + \beta_2 WAge_i + u_i \qquad (1)$$

50K male wife

$$FamIncome_i = \beta_0 + \beta_1 HAge_i + \beta_2 WAge_i + \beta_3 HUnion_j + \beta_4 WUnion_i + u_i \qquad (2)$$

where $FamIncome_i$ is a family's annual income, $HAge_i$ is the husband's age, $WAge_i$ is the wife's age, $HUnion_i$ is a dummy variable indicating that the husband is in a union, and $WUnion_i$ is a dummy variable for the wife's union status. Suppose average family income in the sample is \$50,000. Estimating these models using OLS gives the following results:

Restricted Unrestricted

	(1)	(2)
HAge	-90.8	205
	(101)	(125)
WAge	495	322
	(112)	(134)
HUnion		-1805
		(1351)
WUnion		8972
		(1594)
Constant		24164
		(2380)
R^2	0.0173	0.0441
n	2574	2574

Brazil
Mex
Japan
Italy
USA

a) Consider regression (1). How do you interpret the coefficient on $WAge$? Is the coefficient estimate on $WAge$ in regression (1) economically significant?

b) According to regression (2), how much more (in dollars) can the average (married) family expect to earn in ten years (holding union status constant)?

c) Construct a t-statistic for the significance test of $HAge$ in regression (2). Is the relationship between husband's age and family earnings statistically significant?

d) How do you interpret the coefficient on $HUnion$?

e) What is the average effect on family income of a husband and wife both joining a labor union, if neither were previously in a union?

f) Extra Credit: Assuming homoskedasticity, are the coefficients on unionization jointly statistically significant?

```
 1 . do "/var/folders/pj/w6clq82j4lsdvzfglnglbbb40000gn/T//SD43000.000000"

 2 . /*
   > For Econometrics Assignment # 2: On Bivariate and Multiple Regression
   >
   > Last modified by: Monica Elgawly
   > Last modified on: Saturday, March 3, 2018
   >
   > Notes:
   >
   > */
 3 .
 4 . clear all

 5 . cd "/Users/monicaelgawly/Downloads"
   /Users/monicaelgawly/Downloads

 6 .
 7 . *loads the stata data file named for various h
 8 . use "bwght.dta"
   file bwght.dta not found
   r(601);

   end of do-file

   r(601);
```

weight = $-2390 + 143G - 245S$

nonsmoking = 0

weight = $-2390 + 143G - \cancel{245(0)}$

$\boxed{\text{weight} = -2390 + 143\,\text{Gest}}$

smoking = 1

weight = $-2390 + 143G - 245(1)$

$\boxed{\text{weight} = -2635 + 143\,\text{Gest}}$

weight = $120.5 + (-8.9)(\text{anycig})$

```
 9 . do "/var/folders/pj/w6clq82j4lsdvzfglnglbbb40000gn/T//SD43000.000000"

10 . /*
   > For Econometrics Assignment # 2: On Bivariate and Multiple Regression
   >
   > Last modified by: Monica Elgawly
   > Last modified on: Saturday, March 3, 2018
   >
   > Notes:
   >
   > */
11 .
12 . clear all

13 . cd "/Users/monicaelgawly/Downloads/metrics"
   /Users/monicaelgawly/Downloads/metrics

14 .
15 . *loads the stata data file named for various heights
16 . use "bwght.dta"
```

```
17 .
18 . * gen command creates a dummy variable equating 1 to a mother who smokes at leas
   > t a cig/day 0 otherwise
19 .
20 . gen anycig==1 if cigs>0 & anycig==0 if cigs==0
   == invalid name
   r(198);

   end of do-file

   r(198);

21 . br

22 . do "/var/folders/pj/w6clq82j4lsdvzfglnglbbb40000gn/T//SD43000.000000"

23 . /*
   > For Econometrics Assignment # 2: On Bivariate and Multiple Regression
   >
   > Last modified by: Monica Elgawly
   > Last modified on: Saturday, March 3, 2018
   >
   > Notes:
   >
   > */
24 .
25 . clear all

26 . cd "/Users/monicaelgawly/Downloads/metrics"
   /Users/monicaelgawly/Downloads/metrics

27 .
28 . *loads the stata data file named for various heights
29 . use "bwght.dta"

30 .
31 . *gen command creates a dummy variable equating 1 to a mother who smokes at least
   >  a cig/day 0 otherwise
32 .
33 . gen anycig=1

34 . replace anycig=0 if cigs==0
   (993 real changes made)

35 .
36 .
37 .
38 .
   end of do-file
```

```
     anycig |   -8.92383    1.58908    -5.62   0.000    -12.04157   -5.806087
      _cons |   120.4693   .6325948   190.44   0.000     119.2281    121.7104
-----------------------------------------------------------------------------

58 .
59 . *regression used to view total number of sample points
60 .
61 .
62 .
63 .
   end of do-file

64 . do "/var/folders/pj/w6clq82j4lsdvzfglnglbbb40000gn/T//SD43000.000000"

65 . /*
   > For Econometrics Assignment # 2: On Bivariate and Multiple Regression
   >
   > Last modified by: Monica Elgawly
   > Last modified on: Saturday, March 3, 2018
   >
   > Notes:
   >
   > */
66 .
67 . clear all

68 . cd "/Users/monicaelgawly/Downloads/metrics"
   /Users/monicaelgawly/Downloads/metrics

69 .
70 . *loads the stata data file named for various heights
71 . use "bwght.dta"

72 .
73 . *gen command creates a dummy variable equating 1 to a mother who smokes at least
   >  a cig/day 0 otherwise
74 .
75 . gen anycig=1

76 . replace anycig=0 if cigs==0
   (993 real changes made)

77 .
78 . *tab command to determine what percentage of moms smoked during pregnancy
79 . *in English, where (anycig=1)/(total number of moms sampled)
80 .
81 . reg bwght anycig

      Source |       SS           df       MS      Number of obs   =     1,180
```

```
39 . br

40 . do "/var/folders/pj/w6clq82j41sdvzfglnglbbb40000gn/T//SD43000.000000"

41 . /*
   > For Econometrics Assignment # 2: On Bivariate and Multiple Regression
   >
   > Last modified by: Monica Elgawly
   > Last modified on: Saturday, March 3, 2018
   >
   > Notes:
   >
   > */
42 .
43 . clear all

44 . cd "/Users/monicaelgawly/Downloads/metrics"
   /Users/monicaelgawly/Downloads/metrics

45 .
46 . *loads the stata data file named for various heights
47 . use "bwght.dta"

48 .
49 . *gen command creates a dummy variable equating 1 to a mother who smokes at least
   >  a cig/day 0 otherwise
50 .
51 . gen anycig=1

52 . replace anycig=0 if cigs==0
   (993 real changes made)

53 .
54 . *tab command to determine what percentage of moms smoked during pregnancy
55 . *in English, where (anycig=1)/(total number of moms sampled)
56 .
57 . reg bwght anycig

      Source |       SS           df       MS      Number of obs   =     1,180
-------------+----------------------------------   F(1, 1178)      =     31.54
       Model |  12531.7427         1  12531.7427   Prob > F        =    0.0000
    Residual |  468107.677     1,178  397.374938   R-squared       =    0.0261
-------------+----------------------------------   Adj R-squared   =    0.0252
       Total |  480639.419     1,179  407.667022   Root MSE        =    19.934

------------------------------------------------------------------------------
       bwght |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

	SS	df	MS			
				F(1, 1178)	=	31.54
Model	12531.7427	1	12531.7427	Prob > F	=	0.0000
Residual	468107.677	1,178	397.374938	R-squared	=	0.0261
				Adj R-squared	=	0.0252
Total	480639.419	1,179	407.667022	Root MSE	=	19.934

bwght	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
anycig	-8.92383	1.58908	-5.62	0.000	-12.04157	-5.806087
_cons	120.4693	.6325948	190.44	0.000	119.2281	121.7104

```
82 .
83 . *regression used to view total number of sample points
84 .
85 . reg bwght anycig if anycig==1
   note: anycig omitted because of collinearity
```

Source	SS	df	MS	Number of obs	=	187
				F(0, 186)	=	0.00
Model	0	0	.	Prob > F	=	.
Residual	70664.3636	186	379.915934	R-squared	=	0.0000
				Adj R-squared	=	0.0000
Total	70664.3636	186	379.915934	Root MSE	=	19.491

bwght	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
anycig	0	(omitted)				
_cons	111.5455	1.425355	78.26	0.000	108.7335	114.3574

```
86 .
87 .
   end of do-file

88 .
```

```
/*
For Econometrics Assignment # 2: On Bivariate and Multiple Regression

Last modified by: Monica Elgawly
Last modified on: Saturday, March 3, 2018

Notes:

*/

clear all
cd "/Users/monicaelgawly/Downloads/metrics"

*loads the stata data file named for various heights
use "bwght.dta"

*gen command creates a dummy variable equating 1 to a mother who smokes at least a cig/day 0 otherwise

gen anycig=1
replace anycig=0 if cigs==0

*tab command to determine what percentage of moms smoked during pregnancy
*in English, where (anycig=1)/(total number of moms sampled)

reg bwght anycig

*regression used to view total number of sample points

reg bwght anycig if anycig==1
```

For the linear regression there are 3 assumptions where if they hold, the OLS estimators are (1) unbiased
(2) consistent
& (3) normally distributed when the sample is large.

If, in addition, the regression errors are homoskedastic & if the regression errors are normally distributed, then the OLS t-statistic computed using homoskedasticity-only standard errors has a student t-distribution when the null hypothesis is true.

Econometrics
Professor Van Parys
Monica Elgawly
3_ March 2018

Project

Title: Assignment 2: Bivariate and Multiple Regression

(1.) Description: a.) Consistency = let's say we repeat the experiment multiple times. As we increase sample size in these experiments, we can say the following: $\lim_{n\to\infty} \hat{\beta}_1 = \beta_1$ which translates to mean that with increasing sample size the $\hat{\beta}_1$, sample slope coefficient can be more confidently as stated in approxiamation to the true value of β_1. With consistency, smaller variance is created for efficiency.

Date

*As in Lecture #4, When assumptions #1, 2 and 3 are true, then $\hat{\beta}_1$ and $\hat{\beta}_0$ are normally distributed in large samples (as well as $\hat{\beta}_2$), so $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ will get closer to the true population parameters as n goes to infinity. As n increases, the variances of $\hat{\beta}_1$ and $\hat{\beta}_0$ go toward zero. *Note that unbiasedness & consistency are similar but consistency describes what happens to $\hat{\beta}_1$ and $\hat{\beta}_0$ as n gets large.

Assumption ① $E(u_i|X_i) = 0$ where the conditional distribution of u_i given X_i has a mean of zero.
② (X_i, Y_i), $i = 1, \ldots, n$, are independently & identically distributed
③ large outliers are unlikely.

Concept:

b) $H_0: \hat{\beta}_2 = 0$
$H_a: \hat{\beta}_2 \neq 0$
$\alpha = 0.05$

So if the p-value is sufficiently small, we can show $\hat{\beta}_2$ is not statistically significant.

$$t\text{-statistic} = \frac{\text{estimator} - \text{hypothesized value}}{\text{standard error of the estimator}} = \frac{2.1 - 0}{1.0} = 2.1$$

c) To find the p-value for the statistical significance test for β_2
① State your null and alternate hypotheses. The null hypothesis is being where we state what we think what would happen normally. It is the assumption we're hoping is true. The alternate hypothesis is what we conclude to be true if the p-value is found to be too small. If the p-value is too small, it represents the probability that whatever this data represents, the null hypothesis is too strange to be accepted as true.

② Organize the data to locate the z-score that corresponds to a standardized test statistic. This test statistic is calculated via various formulas depending on the scenario of the data. This test statistic is also known as the critical value.

Action Steps: ③ Calculate/find the p-value z or t-score via the table and/or further calculation.

Date

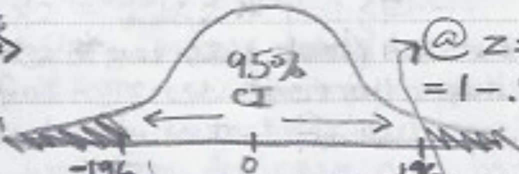
④ Compare p-value to critical value to decide if you support or reject the null. utilizing the model of the distribution.

with the numerical data given, we calculate our $\hat{t}$ to be = 2.1.

$\hat{t} = |2.1| > 1.96 = t_{crit_{.05}}$

$p = \text{prob}(z \text{ at } 2.1) \Rightarrow$

Our hypothesis is $\beta = 0$ so we have to test both tails to check if β is in either region.



this is where the null would be true.

@ z = 2.1 the area above is $= 1 - .9821 = .0179$

Multiply by 2 for two tails: $(2)(.0179) = .0358$

Compare to $\alpha = .05$:

$.0358 < .05$ $\boxed{\text{We reject the null.}}$

When we get a z greater than or less than 1.96, it forces us to conclude that the $\beta \neq 0$ because its z-score is not within the accepted range of values.

d) A 99% confidence interval for β_1 is (1) the set of values that cannot be rejected using a two-sided hypothesis test w/ a 1% significance level. And (2) it is an interval that has a 99% probability of containing the true value of β_1; that is, in 99% of the possible samples that might be drawn, the confidence interval will contain the true value of β_1.

Title: 99% confidence interval for $\beta_1 = \left[\hat{\beta}_1 - 2.576SE(\hat{\beta}_1),\ \hat{\beta}_1 + 2.576SE(\hat{\beta}_1)\right]$

Description: Date

e) The predicted change in Y associated with the change in X is $\beta_1 \Delta X$. Because we can construct a confidence interval for β_1, we can construct a confidence interval for the predicted effect $\beta_1 \Delta X$.

$$\hat{\beta}_1 - 2.576\, SE(\hat{\beta}_1) \Longrightarrow \left(\hat{\beta} - 2.576\, SE(\hat{\beta}_1)\right) * \Delta X$$

∴ for $\Delta X_{i1} = 1$, we get $[3.5 - (2.576)(2.2),\ 3.5 + (2.576)(2.2)] * 1$

$= [-2.1672, 9.1672]$ ∴ The effect of a one unit change in X_{i1} could be as little as -2.1672 or as great as 9.1672, creating a change in Y_i anywhere between approx. -2 to 9 points with a 99% confidence level. (p. 153 txtbook)

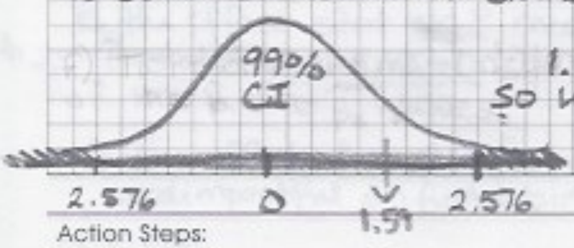
Concept:

∴ With a $\Delta X_{i1} = 8$, the impact is greater on the range of Y_{i1}.

$[3.5 - (2.576)(2.2),\ 3.5 + (2.576)(2.2)] * 8 \Longrightarrow [-17.3376, 73.3376]$

∴ Here, the change in X_{i1} could be as little as -17 to 73 approximately, creating a change in Y_{i1} to this degree of points with a 99% confidence level.

To conclude about the statistical significance of β_1: $\hat{t} = \frac{3.5 - 0}{2.2} = 1.59$



1.59 is within the range of expected values so we say it's not statistically significant.

p-value @ 1.59 = $2 * (1 - .9441) = \boxed{.1118}$

p-value @ 2.576 = $2 * (1 - .9950) = \boxed{.01}$

Action Steps: Date

Since .1118 > .01 again we can conclude β_1 is not statistically significant.

Mental Trick for P-values!
$p \le .05$ significant
$p \le .01$ highly signific
$p > .05$ not significant

f) The hint on excluding a variable leads to commenting on the theory of omitted variable bias. If we omit $\beta_1 X_1$ due to β_1's lack of statistical significance we have to be sure it is uncorrelated with X_2. This is the case because if we were to leave $\beta_1 X_1$ in the equation it could skew regression results if $\beta_1 X_1$ is found to be a linear map of another variable. To be found as a linear function of another regressor is the same as saying the variable is a determinant of Y_i and if it's correlated with another regressor then the first least squares assumption is violated.

Project

2. Smoking & Birth Outcomes.

bwght.dta: birthweight(oz) = bwght

\# of cigarettes the mom smoked/day during pregnancy (cigs)

a) Title: use "bwght.dta"

gen anycig = 1 if cigs > 1

Description: replace anycig = 0 if cig = 0 ** ← KEY LINE! Date

to tabulate % of moms who smoked

y = dependent $\dfrac{(\Sigma \text{ of where anycig} = 1)}{(\Sigma \text{ of \# of data pts in sample})} = \dfrac{\text{anycig} = 1}{n}$

b) H_0: $bwght_s = bwght_{ns}$ or anycig of 1 = anycig of 0

reg to estimate $bwght_i = \beta_0 + \beta_1 anycig_i + u_i$ x = independent

What does $\hat{\beta}_1$ mean here? It represents the extent of the significance of the relationship between % of moms who smoked and their impact on their newborn's weight, that is the extent of the correlation between smoking & baby weight. $\hat{\beta}_0$ here represents the base level weight of the children without error accounted for.

reg bwght anycig (LC#8 slide #9)

Concept:

Table of % of moms who smoked: 187/1180 = .158 or 15.8%

Option 1: because t-stat is given in the regression

From the regression, the regression of |-5.46| > 1.95 @ 5% significance level so we can reject the null hypothesis that the birthweights between the children of smokers vs. nonsmokers is the same.

Since the coefficient estimate of smoking determines the birthweight difference & the coefficient estimate corresponding to the regressor "anycig" is -8.92383, it indicates that smoking any amount of cigarettes decrease birthweight by **9 oz** on average approximately.

Is the effect stat sig? means we're testing whether or not it's impact is zero.

Option 2: $\hat{t}$-statistic $= \dfrac{-8.92 - 0}{1.589} = |-5.61359\ldots| > 1.96$ We can reject the null hypothesis @ 5% level. of the effect being negligible.

∴ The effect of smoking any number of cigarettes can be seen as detrimental and impactful on future birth weight of a mother's future child during pregnancy.

Action Steps: Date

c) This is answering a different Question

Testing whether a one unit increase in the Δanycig would decrease bwght by an oz

$H_0: \beta_0 = -1$ vs $H_1: \beta_1 \neq -1$

$\hat{t} = \dfrac{-8.92 - (-1)}{1.589} = |-4.98427\ldots| > 1.96$ We again reject the null hypothesis @ the 5% level so the effect of smoking one more mariginal unit does not decrease weight by an ounce.

c) Write in Stata: margins, dydx(*)

So the marginal addition of one cigarette a day during pregnancy decreased baby birthweight by little more than half an ounce.

The dy/dx value given is -.5096199. With a null of H_0: sig = 0, we have the t = |-5.46| > 1.96 so we reject the null of significance of zero meaning the impact of cigarettes on a marginal level is statistically significant

) $FamIncome_i = \beta_0 + \beta_1 HAge_i + \beta_2 WAge_i + u_i$ (1)

$\underbrace{FamIncome_i}_{\text{family's annual income}} = \beta_0 + \beta_1 \underbrace{HAge_i}_{\text{husband's age}} + \beta_2 \underbrace{WAge_i}_{\text{wife's age}} + \beta_3 \underbrace{HUnion_I}_{\text{dummy var husband in a union}} + \beta_4 \underbrace{WUnion_i}_{\text{dummy var wife in a union}} + u_i$ (2)

50,000

a) The coefficient on WAge represents the marginal effect the age of the wife has on the dependent variable, the annual family income. To be economically significant, we have to relate the β_2 as the number where with increasing age, we see observed change in the annual income. In other words, it's to conclude that an additional year in age will lead to $+\beta_2$ change in annual income.

$50000 = \beta_0 - 90.8 HAge_i + 495 WAge_i$

Here in order to increase income by \$1 in family income, the wife needs to wait or advance in age $\frac{1}{495} = .002$ years or $(.002 \times 365 \approx 74)$ about 74 days. This is somewhat a reasonable portrayal of reality but not reasonable if she needs to make an additional \$30,000 or whatever. This is about \$5 per year so this is then not economically significant.

b) To hold union status constant, we give a coefficient of 1 for $HUnion_I$ and $WUnion_i$ and fill in the coefficients as such: where Union=1 × 10 years

ΔIncome in 10 year $= 24164 + 205 HAge + 322 WAge_i - 1805 HUnion_I + 8972 WUnion_i$ EQN (2)

~~ΔIncome in 10 years $= -90.8 HAge_i + 495 WAge_i$~~ EQN (1)

Why zero? Because when constant, the variable of Unions are not to impact the rest of the regression

For equation (2), substituting 10 to account for the change in 10 years gives $24164 + 205(10) + 322(10) - \cancel{1805(10)} + \cancel{8972(10)} = \$101,104$ more

c) Construct a t-statistic for the significance test of HAge in regression (2).

$t_{HA} = \frac{205-0}{125} = 1.64 < 1.96$ $\therefore$ HAge is not statistically significant in relation to annual family earnings.

d) The coefficient on HUnion represents the estimated slope, predicting the change in value of family earnings due to one unit change in the husband's union status.

e) As seeing the average effect is akin to observing a change in value as in part (d), the effect of both husband and wife joining a labor union is the equivalent of $HUnion_I = WUnion_i = 1$ and so the effect increases by the sum of the coefficients of the two variables: $(-1805) + (8972)$ according to the transition from regression 1 to regression 2 $\Rightarrow$ \$7167 is added to yearly family income.

f) "Restricted regression is one in which coefficients of some independent variable are ass

$F = \frac{(R_u^2 - R_r^2)/q}{(1-R_u^2)/(n-k_u-1)} =$