# 1. Theory

Question 1

a) i)

Given by lecture:

TSS = ESS + RSS

TSS = $\Sigma_{i=1}^{n} y_i^2$

ESS = $\Sigma_{i=1}^{n} \hat{y}_i^2$

RSS = $\Sigma_{i=1}^{n} e_i^2$

Formal proof of TSS = ESS + RSS:

$$\text{TSS} = \Sigma_{i=1}^{n} y_i^2 = \Sigma_{i=1}^{n}(y_i - \bar{y})^2 = \Sigma_{i=1}^{n}(y_i - \bar{y} + \widehat{y_i} - \widehat{y_i})^2 = \Sigma_{i=1}^{n}\left((\widehat{y_i} - \bar{y}) + (y_i - \widehat{y_i})\right)^2 = \Sigma_{i=1}^{n}\left((\widehat{y_i} - \bar{y}) + \widehat{e_i}\right)^2 = \Sigma_{i=1}^{n}((\widehat{y_i} - \bar{y})^2 + 2\widehat{e_i}(\widehat{y_i} - \bar{y}) + \widehat{e_i}^2) = \Sigma_{i=1}^{n}(\widehat{y_i} - \bar{y})^2 + 2\Sigma_{i=1}^{n}\widehat{e_i}(\widehat{y_i} - \bar{y}) + \Sigma_{i=1}^{n}\widehat{e_i}^2 = \Sigma_{i=1}^{n}(\widehat{y_i} - \bar{y})^2 + \Sigma_{i=1}^{n}\widehat{e_i}^2 + 2\Sigma_{i=1}^{n}\widehat{e_i}(\widehat{\beta_0} + \widehat{\beta_1 x_{i1}} + (\dots) + \widehat{\beta_k x_{ik}} - \bar{y}) = \Sigma_{i=1}^{n}(\widehat{y_i} - \bar{y})^2 + \Sigma_{i=1}^{n}\widehat{e_i}^2 + 2(\widehat{\beta_0} - \bar{y})\Sigma_{i=1}^{n}\widehat{e_i}^2 + 2\widehat{\beta_1}\Sigma_{i=1}^{n}\widehat{e_i}x_{i1} + (\dots) + 2\widehat{\beta_k}\Sigma_{i=1}^{n}\widehat{e_i}x_{i1} = \Sigma_{i=1}^{n}(\widehat{y_i} - \bar{y})^2 + \Sigma_{i=1}^{n}\widehat{e_i}^2 = \Sigma_{i=1}^{n}\hat{y}_i^2 + \Sigma_{i=1}^{n}\widehat{e_i}^2 = ESS + RSS$$

ii)

Formal proof of $R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{e\prime e}{\tilde{y}\prime\tilde{y}}$

$$R^2 = \frac{ESS}{TSS} = \frac{\Sigma_{i=1}^{n}\hat{y}_i^2}{\Sigma_{i=1}^{n}y_i^2} = \frac{ESS}{ESS + RSS} = \frac{\Sigma_{i=1}^{n}\hat{y}_i^2}{\Sigma_{i=1}^{n}\hat{y}_i^2 + \Sigma_{i=1}^{n}e_i^2} = \frac{\Sigma_{i=1}^{n}\hat{y}_i^2 + \Sigma_{i=1}^{n}e_i^2 - \Sigma_{i=1}^{n}e_i^2}{\Sigma_{i=1}^{n}\hat{y}_i^2 + \Sigma_{i=1}^{n}e_i^2}$$

$$= \frac{\Sigma_{i=1}^{n}y_i^2 - \Sigma_{i=1}^{n}e_i^2}{\Sigma_{i=1}^{n}y_i^2} = 1 - \frac{\Sigma_{i=1}^{n}e_i^2}{\Sigma_{i=1}^{n}y_i^2} = 1 - \frac{RSS}{TSS} = 1 - \frac{e'e}{\tilde{y}'\tilde{y}}$$

b)

Formal proof of $R^2 = corr^2(\mathbf{y}, \hat{y}) = \rho_{y,\hat{y}}^2 = \frac{ESS}{TSS}$

➔ First take the square root of $\rho_{y,\hat{y}}^2$



Tarik Benli 15-719-818, Ramon Gmür 16-705-220

$$\rho_{y,\hat{y}} = \frac{\sigma_{y,\hat{y}}^2}{\sqrt{\sigma_y^2 * \sigma_{\hat{y}}^2}} = \frac{\Sigma_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\Sigma_{i=1}^n (y_i - \bar{y})^2 * \Sigma_{i=1}^n (\hat{y}_i - \bar{y})^2}} = \frac{\Sigma_{i=1}^n (y_i + \hat{y}_i - \hat{y}_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\Sigma_{i=1}^n (y_i - \bar{y})^2 * \Sigma_{i=1}^n (\hat{y}_i - \bar{y})^2}}$$

$$= \frac{\Sigma_{i=1}^n \left(y_i\hat{y}_i + \hat{y}_i^2 - \hat{y}_i^2 - \bar{y}\hat{y}_i - \bar{y}y_i + \hat{y}_i\bar{y} - \bar{y}\,\hat{y}_i + \bar{y}^2\right)}{\sqrt{\Sigma_{i=1}^n (y_i - \bar{y})^2 * \Sigma_{i=1}^n (\hat{y}_i - \bar{y})^2}}$$

$$= \frac{\Sigma_{i=1}^n ((y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + (\hat{y}_i - \bar{y})^2)}{\sqrt{\Sigma_{i=1}^n (y_i - \bar{y})^2 * \Sigma_{i=1}^n (\hat{y}_i - \bar{y})^2}}$$

$$= \frac{\Sigma_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sqrt{\Sigma_{i=1}^n (y_i - \bar{y})^2 * \Sigma_{i=1}^n (\hat{y}_i - \bar{y})^2}} = \sqrt{\frac{\Sigma_{i=1}^n (\hat{y}_i - \bar{y})^2}{\Sigma_{i=1}^n (y_i - \bar{y})^2}} = \sqrt{\frac{ESS}{TSS}}$$

➔ And now square it: $\frac{ESS}{TSS} = \rho_{y,\hat{y}}^2 = corr^2(\mathbf{y}, \hat{y}) = R^2$

One can interpret $R^2$ as the squared correlation coefficient between the true value $y_i$ and the estimated value $\hat{y}_i$. In a regression model $R^2$ measures how good the estimated value explains the true value. In other words, it explains the variation in the estimated $\hat{y}_i$ and its true value $y_i$. Thus, this can be translated to the correlation between these two variables as shown above. One can generally say: The higher $R^2$ or $\rho_{y,\hat{y}}^2$ the better the model can predict the true values where the $R^2$ is in a range between 0 and 1 and the correlation between -1 and 1.

c) By transforming our X variables linearly, one cannot gain more information out of our variables than before. For each datapoint the true value of $y_i$ and the estimated value of $\hat{y}_i$ do not change at all by linear transformation. It neither change the relation nor the underlying data. Since we have showed above that $R^2 = \rho_{y,\hat{y}}^2$, one can see that the correlation between $y_i$ and $\hat{y}_i$ is not affected either and the calculated $R^2$ remains the same.

d) Intuitively the residual sum squared always decreases if one adds another regressor into the model (see formal proof 1.e) and so $R^2$ will increase. That is because with a new regressor one might gather more information from the data, since the estimated values may be predicted more precisely. Only in a special case, if the added regressor is perfectly correlated with an already existing one in the model, the RSS would stay the same, since one cannot gather more information by adding this new but perfectly correlated variable.

e)

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\Sigma_{i=1}^{n} e_i^2}{\Sigma_{i=1}^{n} y_i^2} = 1 - \frac{\Sigma_{i=1}^{n}\left(y_i - x_i\hat{\beta} - \bar{e}\right)^2}{\Sigma_{i=1}^{n} y_i^2}$$

If one adds an additional regressor $x_i$, the $\hat{\beta}$ will increase and therefore the RSS is getting smaller as a direct consequence of this. Furthermore, $R^2$ will increase. In other words, the precision of the prediction is increasing until the error term is hypothetically equal to zero.

f) Working with $R^2$ could lead to multiple types of misinterpretation:

- The **type** of data implies different values of $R^2$. While time-series data often have a high $R^2$, the exact opposite is often true for cross-section data.
- Different functional forms for $y_i$ can **change** $R^2$ (for example: using $\log(y_i)$ often increases $R^2$)
- Adding explanatory variables to a model always **increase** $R^2$
    - This may lead one towards overfitting the model with too many variables. So, one can increase the $R^2$ even by just adding a large set of totally random predictors.

Furthermore, it is only an in-sample measurement: That means $R^2$ only measures the precision within the sample. A better way to measure the precision between samples is for example the measurement techniques of Cross-Validation.

## 2. Empirical Question

a) Big school dummy

i) The coefficient on *classize* is 0.134 and statistically significant. In this case, this means that the average marks in a grammar tests rises by 0.134 per additional student in class.

ii) The new coefficient on *classize* shrinks to 0.102, the one of the new dummy *big school* is 1.246 (both are statistically significant). This means that, while an additional student only brings a rise in the average grammar test marks of 0.102 by controlling for school size, the bigger schools have higher averages.

b) Natural log

i) The coefficient is -0.0603 on *classsize* (-0.0007 in the log-model) and -0.335 on *pct_dis* (-0.0048 in the log model). The effect of *classize* is now even smaller than before and now slightly negative. The effect of the percentage of disadvantaged kids is also negative. With a rise of the amount of disadvantaged kids of 1%, the average mark decreases by -0.335 points. The model with the logged grammar scores shows the coefficients as an (approximated) percentage change with respect to the constant. The coefficients in the log model show percentage changes, meaning with one unit change in *classize* or *pct_dis*, the grammar marks change by 0.07% and 0.5%.

ii) In the regression of grammar scores on *classize* and *pct_dis*, the coefficient of the latter variable can be interpreted as follows: If pct_dis rises by one unit, the average grammar mark decreases by 0.335 points controlling for class size.

c) Small size dummy

i) The coefficient of small size tells us that small classes score 2.56 points higher on grammar tests than big classes, controlling for the percentage of disadvantaged kids. In terms of economical significance, this coefficient seems rather large, comparing it with previous coefficients and with respect to the constant. In Stata, the two-sided test $\beta_2 = 0$ yields a p-value of 0.2016. Assuming that a small sizes class has a positive effect on grammar scores, we also use a one-sided test $\beta_2 \leq 0$, yielding a p-value of 0.1008 (which in this scenario is exactly 50% of the two-sided p-value). This means we cannot reject our hypotheses.

We recommend using the two-sided hypothesis test, because with *small_size* we are looking at an extreme attribute of class sizes and should not assume that the effect of *small_size* is positive (only 8 classes fall into the category small sized). This lack of observations for small class sizes explains why the results are far from significant. Calculation of the hypothesis test that $\beta_2 = 0$ by hand:

1: $H_0: \beta_2 = 0, H_A: \beta_2 \neq 0$

2: Degrees of freedom: $N - K = 1967 - 2 = 1965$

$s^2 = \frac{1}{N-K} \sum_{i=1}^{n} e_i^2$

t-value: $\frac{2.560}{2.004} = 1.277$

3: At the 5% level: $P(t(_{1965}) \leq \overline{t_{0.975}}) = 0.975 -> \overline{t_{0.975}} = 1.96$

$|1.277| > 1.96 -> H_0$ cannot be rejected

ii) First, we need to regress the grammar scores only on *pct_dis* and then take the residuals from this model (see table below: 1). Then we regress *small_size* on *pct_dis* too and also take the residuals (see table below: 2). Now we can regress the residuals from the first model one the residuals on the second model. The resulting coefficient is the coefficient of grammar score on small sized classes (and is exactly the same as in i).

| Y-VARIABLE: | (mrkgrm) | (small_size) | (residuals1) |
|---|---|---|---|
| VARIABLES | 1 | 2 | 3 |
| pct_dis | -0.327*** | 9.57e-06 | |
| | (0.00977) | (0.000110) | |
| residuals2 | | | 2.560 |
| | | | (2.003) |
| Constant | 77.11*** | 0.00394* | -6.64e-09 |
| | (0.183) | (0.00206) | (0.128) |
| Observations | 1,967 | 1,967 | 1,967 |
| R-squared | 0.363 | 0.000 | 0.001 |

iii) To show that $\hat{\beta}_1 = \bar{y} - \bar{X}_{-1}\hat{\beta}_{-1}$, we need the means of *mrkgrm*, *small_size* and *pct_dis*. Now by deducting the means of *small_size* and *pct_dis* multiplied with their respective coefficient from the mean of *mrkgrm*, we get $\hat{\beta}_1$:

$$\overline{mrkgrm} - \overline{small_size} \times 2.5597 - \overline{pct_dis} \times (-0.32677) = 77.099$$

iv) The correct interpretation is that small classes have an average grammar score that is 3.65% higher than in bigger classes.

d) Many disadvantaged dummy

i) The joint hypothesis that $\beta_3 = 0$ and $\beta_4 = 0$ has an F-value of 401.85 (p-value = 0). Calculating this by hand shows the same results (small difference due to rounding):

$$\frac{(R_U^2 - R_R^2)/q}{(1 - R_U^2)/(N - K)} = \frac{(0.3005 - 0.0142)/2}{(1 - 0.3005)/(1966\text{-}3)} = 401.72$$

In conclusion, we reject the joint hypothesis that $\beta_3 = 0$ and $\beta_4 = 0$. Many disadvantaged kids (alone and combined with class size) affects grammar scores.

ii) The effect of having 10 additional students in a class with less than 10% disadvantaged kids is –1.1 (since *many_dis* is a dummy with value 0, both $\beta_3$ and $\beta_4$ are not relevant in this specific case here)

e) Separated regressions

The table below shows the results separating classes with high and low percentages of disadvantaged kids as well as the results from d). The coefficient on *classize* in (2) is exactly the same as in (3), which makes sense because both only consider class size for classes with less than 10% disadvantaged kids. For classes with more than 10% disadvantaged kids, one can calculate that the separate regressions (1) and (2) and the combined regression (3) also yield the same results:

$$63.81 + classize \times 0.159$$
$$= 79.52 + classize \times (-0.110) + (-15.71) + (classize \times many\_dis) \times 0.269$$

In conclusion, model (3) is a combination of models (1) and (2)

| VARIABLES | (1)<br>high dis | (2)<br>low dis | (3)<br>d) |
|---|---|---|---|
| classize | 0.159*** | -0.110*** | -0.110*** |
| | (0.0373) | (0.0255) | (0.0291) |
| many_dis | | | -15.71*** |
| | | | (1.346) |
| classize × many_dis | | | 0.269*** |
| | | | (0.0438) |
| Constant | 63.81*** | 79.52*** | 79.52*** |
| | (1.103) | (0.821) | (0.937) |
| Observations | 858 | 1,109 | 1,967 |
| R-squared | 0.021 | 0.017 | 0.301 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

f) Region dummies

The region dummies cannot all be included in the same model because of multicollinearity (dummy variable trap). If one wants to include region dummies, one needs to be omitted.

g) The table below shows the results for regression separated by region. While the coefficients of *pct_dis* are all negative and in the same range, the coefficients on *classize* are not.

| VARIABLES | (1)<br>Reg1 | (2)<br>Reg2 | (3)<br>Reg3 | (4)<br>Reg4 | (5)<br>Reg5 | (6)<br>Reg6 |
|---|---|---|---|---|---|---|
| classize | -0.0901** | -0.0550 | 0.168** | 0.00741 | 0.0212 | -0.0758 |
| | (0.0451) | (0.0823) | (0.0669) | (0.0459) | (0.0429) | (0.0541) |
| pct_dis | -0.249*** | -0.252*** | -0.213*** | -0.490*** | -0.319*** | -0.404*** |
| | (0.0218) | (0.0309) | (0.0275) | (0.0389) | (0.0195) | (0.0232) |
| Constant | 81.01*** | 77.17*** | 69.65*** | 80.18*** | 75.24*** | 79.62*** |
| | (1.311) | (2.511) | (2.226) | (1.479) | (1.557) | (1.883) |
| Observations | 255 | 195 | 267 | 276 | 574 | 400 |
| R-squared | 0.344 | 0.266 | 0.257 | 0.382 | 0.373 | 0.460 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

We propose a model that includes all the regions. To do so, we use the region dummies and omit region 5 (the region with most observations):

| VARIABLES | mrkgrm |
|---|---|
| classize | 0.00654 |
| | (0.0218) |
| pct_dis | -0.309*** |
| | (0.0103) |
| Reg1 | 3.501*** |
| | (0.431) |
| Reg2 | 0.969** |
| | (0.467) |
| Reg3 | 0.399 |
| | (0.410) |
| Reg4 | 3.258*** |
| | (0.419) |
| Reg6 | 0.249 |
| | (0.360) |
| Constant | 75.55*** |
| | (0.801) |
| Observations | 1,967 |
| R-squared | 0.400 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This model shows positive coefficients for all regions, suggesting a positive effect of class size on grammar scores. The effect for region 5 is shown in the coefficient of *classize*, for the other regions one needs to add the effect of the respective region dummy.

h) Subsample with only one class

i) The coefficient of *sc_boys* is -0.302 and statistically significant, meaning that with one additional boy in the school, the average grammar scores decrease by 0.302 points. The coefficient of *classize* is 0.0961, suggesting that the average grammar scores increase by this much with one additional student in class. However, this effect is not statistically significant anymore.

ii) The coefficient of *sc_boys* is -0.206. This means, controlling for the number of girls per school, an additional boy in the school decreases the average grammar scores by 0.206 points, which is slightly less than before.

iii) From the estimation in h-ii) one cannot say anything about the exact effect of one pupil in general, because the effects differ with respect to gender. However, we can expect the variance in the number of girls and boys to be roughly the same:

$$Cov(sc\_boys, sc\_girls) \approx Var(sc\_boys) = Var(sc\_girls)$$

The standard deviations in model h-ii) suggest this expectation to be true. Also, the correlation between *sc_boys* and *sc_girls* is with 0.7918 close to 1.

Now we can compare the two models:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \times sc\_girls + \hat{\beta}_2 \times sc\_boys + \epsilon$$

$$\hat{y} = \hat{\gamma}_0 + \hat{\gamma}_1 \times (sc\_girls + sc\_boys) + \mu$$

Solving the second model for $\hat{\gamma}_1$, we get:

$$\hat{\gamma}_1 = \frac{\mathrm{Cov}(\hat{y},\ \mathrm{sc\_girls})}{\mathrm{Var}(\mathrm{sc\_girls})} \frac{\mathrm{Var}(\mathrm{sc\_girls})}{\mathrm{Var}(\mathrm{sc\_girls} + \mathrm{sc\_boys})} + \frac{\mathrm{Cov}(\hat{y},\ \mathrm{sc\_boys})}{\mathrm{Var}(\mathrm{sc\_boys})} \frac{\mathrm{Var}(\mathrm{sc\_boys})}{\mathrm{Var}(\mathrm{sc\_girls} + \mathrm{sc\_boys})}$$

Combining this equation with auxiliary models only including *sc_boys* or *sc_girls* solved for their respective coefficients, we get the following formula:

$$\hat{\gamma}_1 = \frac{\mathrm{Var}(\mathrm{sc\_girls}) + \mathrm{Cov}(\mathrm{sc\_grils},\ \mathrm{sc\_boys})}{\mathrm{Var}(\mathrm{sc\_girls}\ + \mathrm{sc\_boys})} \hat{\beta}_1 + \frac{\mathrm{Var}(\mathrm{sc\_boys}) + \mathrm{Cov}(\mathrm{sc\_grils},\ \mathrm{sc\_boys})}{\mathrm{Var}(\mathrm{sc\_girls}\ + \mathrm{sc\_boys})} \hat{\beta}_2$$

Using the expectation with the similar variances and the correlations stated at the beginning of h-iii), we get the following approximation:

$$\hat{\gamma}_1 \approx \frac{1}{2}\left(\hat{\beta}_1 + \hat{\beta}_2\right) = \frac{1}{2}(0.096 - 0.206) = -0.055$$

Compared to the coefficient-value of the regression of *mrkgrm* on *classize* (-.0475), it seems that one can say something about the effect of increasing the class size by one pupil, even though the effect is different for boys and girls.

i) It is very unlikely that assumption 2 holds. There are definitely omitted variables that influence the grammar scores. Possible examples are cultural background, family income, school types (public or private) or the degree of preparation the schools provide for this standardized grammar tests.

# 3. Log-file

See attachment

```
  ___  ____  ____  ____  ____(R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/
  Statistics/Data Analysis
```

```
(863 real changes made)
      name:  <unnamed>
       log:  C:\Users\ramon\Desktop\UZH\Empirical Methods\Problem Sets\Problem Set 2\Stata\log_gm
  log type:  smcl
 opened on:  11 Nov 2019, 17:42:25

. 
. use "C:\Users\ramon\Desktop\UZH\Empirical Methods\Problem Sets\Problem Set 2\Stata\class_size_p

. 
. *Empirical Question
. 
. **a)
. 
. gen big_school = 0

. replace big_school = 1 if n_classes > 2
(863 real changes made)

. 
. ***a-i)
. 
. reg mrkgrm classize

      Source |       SS           df       MS      Number of obs   =     1,967
-------------+----------------------------------   F(1, 1965)      =     28.22
       Model |  1396.59756         1  1396.59756   Prob > F        =    0.0000
    Residual |  97259.1452     1,965  49.4957482   R-squared       =    0.0142
-------------+----------------------------------   Adj R-squared   =    0.0137
       Total |  98655.7428     1,966  50.1809475   Root MSE        =    7.0353

------------------------------------------------------------------------------
      mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    classize |   .1341112   .0252472     5.31   0.000     .0845971    .1836254
       _cons |    68.6283   .7839904    87.54   0.000     67.09076    70.16584
------------------------------------------------------------------------------

. 
. outreg2 using "PS2_regressiona.doc", replace ctitle(a-i)
PS2_regressiona.doc
dir : seeout

. 
. ***a-ii)
. 
. reg mrkgrm classize big_school

      Source |       SS           df       MS      Number of obs   =     1,967
-------------+----------------------------------   F(2, 1964)      =     21.03
       Model |  2068.58052         2  1034.29026   Prob > F        =    0.0000
    Residual |  96587.1622     1,964  49.1787995   R-squared       =    0.0210
-------------+----------------------------------   Adj R-squared   =    0.0200
       Total |  98655.7428     1,966  50.1809475   Root MSE        =    7.0128

------------------------------------------------------------------------------
      mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    classize |   .1019125   .0266311     3.83   0.000     .0496844    .1541407
  big_school |   1.246412   .3371876     3.70   0.000     .5851293    1.907695
       _cons |   69.06062   .7901793    87.40   0.000     67.51094     70.6103
------------------------------------------------------------------------------
```

```
20 .
21 . outreg2 using "PS2_regressiona.doc", append ctitle(a-ii)
   PS2_regressiona.doc
   dir : seeout

22 .
23 . **b)
24 .
25 . drop big_school

26 .
27 . ***b-i)
28 .
29 . gen ln_mrkgrm = log(mrkgrm)

30 . reg mrkgrm classize pct_dis
```

| Source | SS | df | MS | | | |
|---|---|---|---|---|---|---|
| Model | 36025.757 | 2 | 18012.8785 | Number of obs | = | 1,967 |
| Residual | 62629.9858 | 1,964 | 31.8889948 | F(2, 1964) | = | 564.86 |
| | | | | Prob > F | = | 0.0000 |
| | | | | R-squared | = | 0.3652 |
| | | | | Adj R-squared | = | 0.3645 |
| Total | 98655.7428 | 1,966 | 50.1809475 | Root MSE | = | 5.647 |

| mrkgrm | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| classize | -.0602863 | .0211063 | -2.86 | 0.004 | -.1016794 | -.0188931 |
| pct_dis | -.3348571 | .0101615 | -32.95 | 0.000 | -.3547856 | -.3149286 |
| _cons | 79.05196 | .7043112 | 112.24 | 0.000 | 77.67068 | 80.43323 |

```
31 . outreg2 using "PS2_regressionbi.doc", replace ctitle(normal)
   PS2_regressionbi.doc
   dir : seeout

32 . reg ln_mrkgrm classize pct_dis
```

| Source | SS | df | MS | | | |
|---|---|---|---|---|---|---|
| Model | 7.53443573 | 2 | 3.76721786 | Number of obs | = | 1,967 |
| Residual | 12.9082879 | 1,964 | .006572448 | F(2, 1964) | = | 573.18 |
| | | | | Prob > F | = | 0.0000 |
| | | | | R-squared | = | 0.3686 |
| | | | | Adj R-squared | = | 0.3679 |
| Total | 20.4427236 | 1,966 | .01039813 | Root MSE | = | .08107 |

| ln_mrkgrm | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| classize | -.0007001 | .000303 | -2.31 | 0.021 | -.0012944 | -.0001059 |
| pct_dis | -.0048256 | .0001459 | -33.08 | 0.000 | -.0051117 | -.0045395 |
| _cons | 4.367718 | .0101113 | 431.96 | 0.000 | 4.347888 | 4.387548 |

```
33 . outreg2 using "PS2_regressionbi.doc", append ctitle(log)
   PS2_regressionbi.doc
   dir : seeout

34 .
35 . ***b-ii)
36 .
37 . **c)
38 .
39 . gen small_size = 0

40 . replace small_size = 1 if classize <= 10
   (8 real changes made)

41 .
42 . reg mrkgrm small_size pct_dis

         Source |       SS           df       MS      Number of obs   =     1,967
   -------------+----------------------------------   F(2, 1964)      =    559.74
          Model |  35817.7919         2  17908.8959   Prob > F        =    0.0000
       Residual |  62837.9509     1,964  31.9948833   R-squared       =    0.3631
   -------------+----------------------------------   Adj R-squared   =    0.3624
          Total |  98655.7428     1,966  50.1809475   Root MSE        =    5.6564

   ------------------------------------------------------------------------------
         mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
   -------------+----------------------------------------------------------------
     small_size |   2.559678   2.003923     1.28   0.202    -1.370361    6.489718
        pct_dis |  -.3267693   .0097728   -33.44   0.000    -.3459354   -.3076032
          _cons |   77.09925   .1834885   420.19   0.000     76.73939     77.4591
   ------------------------------------------------------------------------------

43 . outreg2 using "PS2_regressionc.doc", replace ctitle(c)
   PS2_regressionc.doc
   dir : seeout

44 . test _b[small_size]=0

    ( 1)  small_size = 0

          F(  1,  1964) =    1.63
               Prob > F =    0.2016

45 . local sign_ss = sign(_b[small_size])

46 . display "Ho: coef <= 0  p-value = " ttail(r(df_r),`sign_ss'*sqrt(r(F)))
   Ho: coef <= 0  p-value = .10081773

47 .
48 . ***c-i)
49 .
```

```
50 . ****Hand- and Stata-Testing!!!
51 .
52 . ***c-ii)
53 .
54 . reg mrkgrm pct_dis

      Source |       SS           df       MS      Number of obs   =     1,967
-------------+----------------------------------   F(1, 1965)      =   1117.49
       Model |  35765.5896         1  35765.5896   Prob > F        =    0.0000
    Residual |  62890.1531     1,965   32.005167   R-squared       =    0.3625
-------------+----------------------------------   Adj R-squared   =    0.3622
       Total |  98655.7428     1,966  50.1809475   Root MSE        =    5.6573

------------------------------------------------------------------------------
      mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     pct_dis |  -.3267448   .0097743   -33.43   0.000    -.3459139   -.3075757
       _cons |   77.10933   .1833481   420.56   0.000     76.74975     77.4689
------------------------------------------------------------------------------

55 . outreg2 using "PS2_regressioncii.doc", replace ctitle(1)
   PS2_regressioncii.doc
   dir : seeout

56 . predict residuals1, residuals

57 .
58 . reg small_size pct_dis

      Source |       SS           df       MS      Number of obs   =     1,967
-------------+----------------------------------   F(1, 1965)      =      0.01
       Model |  .00003067         1   .00003067   Prob > F        =    0.9307
    Residual |  7.96743247     1,965  .004054673   R-squared       =    0.0000
-------------+----------------------------------   Adj R-squared   =   -0.0005
       Total |  7.96746314     1,966  .004052626   Root MSE        =    .06368

------------------------------------------------------------------------------
  small_size |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     pct_dis |   9.57e-06     .00011     0.09   0.931    -.0002062    .0002253
       _cons |   .0039382   .0020637     1.91   0.056    -.0001091    .0079854
------------------------------------------------------------------------------

59 . outreg2 using "PS2_regressioncii.doc", append ctitle(2)
   PS2_regressioncii.doc
   dir : seeout

60 . predict residuals2, residuals

61 .
62 . reg residuals1 residuals2

      Source |       SS           df       MS      Number of obs   =     1,967
-------------+----------------------------------   F(1, 1965)      =      1.63
       Model |  52.2022502         1  52.2022502   Prob > F        =    0.2015
    Residual |  62837.9512     1,965  31.9786011   R-squared       =    0.0008
-------------+----------------------------------   Adj R-squared   =    0.0003
       Total |  62890.1534     1,966  31.9888878   Root MSE        =     5.655
```

```
------------------------------------------------------------------------------
  residuals1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  residuals2 |   2.559679   2.003413     1.28   0.202    -1.369359    6.488716
       _cons |  -6.64e-09   .1275051    -0.00   1.000    -.2500594    .2500594
------------------------------------------------------------------------------

63 . outreg2 using "PS2_regressioncii.doc", append ctitle(3)
   PS2_regressioncii.doc
   dir : seeout

64 .
65 . ***c-iii)
66 .
67 . egen mean_mrkgrm = mean(mrkgrm)

68 . egen mean_small_size = mean(small_size)

69 . egen mean_pct_dis = mean(pct_dis)

70 .
71 . display mean_mrkgrm - mean_small_size*2.559768 - mean_pct_dis*-.3267693
   77.099243

72 .
73 . ***c-iv)
74 .
75 . reg ln_mrkgrm small_size pct_dis

      Source |       SS           df       MS      Number of obs   =     1,967
-------------+----------------------------------   F(2, 1964)      =    570.24
       Model |  7.5099808         2  3.7549904   Prob > F        =    0.0000
    Residual |  12.9327428     1,964  .0065849   R-squared       =    0.3674
-------------+----------------------------------   Adj R-squared   =    0.3667
       Total |  20.4427236     1,966  .01039813   Root MSE        =    .08115

------------------------------------------------------------------------------
   ln_mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  small_size |   .0365346   .0287485     1.27   0.204    -.0198462    .0929154
     pct_dis |  -.0047317   .0001402   -33.75   0.000    -.0050067   -.0044568
       _cons |   4.345013   .0026323  1650.62   0.000     4.339851    4.350176
------------------------------------------------------------------------------

76 . outreg2 using "PS2_regressionciv.doc", replace ctitle(c-iv)
   PS2_regressionciv.doc
   dir : seeout

77 .
78 .
79 . **d)
```

```
80 .
81 . gen many_dis = 0

82 . replace many_dis = 1 if pct_dis > 10
   (858 real changes made)

83 . gen many_disXclassize = many_dis*classize

84 . reg mrkgrm classize many_dis many_disXclassize
```

| Source | SS | df | MS | | | |
|---|---|---|---|---|---|---|
| | | | | Number of obs | = | **1,967** |
| | | | | F(3, 1963) | = | **281.14** |
| Model | **29649.3319** | **3** | **9883.11065** | Prob > F | = | **0.0000** |
| Residual | **69006.4108** | **1,963** | **35.153546** | R-squared | = | **0.3005** |
| | | | | Adj R-squared | = | **0.2995** |
| Total | **98655.7428** | **1,966** | **50.1809475** | Root MSE | = | **5.929** |

| mrkgrm | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| classize | **-.1101571** | **.0291463** | **-3.78** | **0.000** | **-.1673179** | **-.0529962** |
| many_dis | **-15.71298** | **1.346291** | **-11.67** | **0.000** | **-18.35329** | **-13.07267** |
| many_disXclassize | **.2686753** | **.043792** | **6.14** | **0.000** | **.1827916** | **.354559** |
| _cons | **79.52074** | **.9369263** | **84.87** | **0.000** | **77.68326** | **81.35821** |

```
85 . outreg2 using "PS2_regressiond.doc", replace ctitle(d)
   PS2_regressiond.doc
   dir : seeout

86 .
87 . ***d-i)
88 .
89 . test _b[many_dis]=0

    ( 1)  many_dis = 0

          F(  1,  1963) =  136.22
               Prob > F =    0.0000

90 . test _b[many_disXclassize]=0, accumulate

    ( 1)  many_dis = 0
    ( 2)  many_disXclassize = 0

          F(  2,  1963) =  401.85
               Prob > F =    0.0000

91 .
92 .
93 . ****Ru^2=0.3005
```

```
 94 . reg mrkgrm classize

      Source |       SS           df       MS      Number of obs   =     1,967
-------------+----------------------------------   F(1, 1965)      =     28.22
       Model |  1396.59756         1  1396.59756   Prob > F        =    0.0000
    Residual |  97259.1452     1,965  49.4957482   R-squared       =    0.0142
-------------+----------------------------------   Adj R-squared   =    0.0137
       Total |  98655.7428     1,966  50.1809475   Root MSE        =    7.0353

------------------------------------------------------------------------------
      mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    classize |   .1341112   .0252472     5.31   0.000     .0845971    .1836254
       _cons |    68.6283   .7839904    87.54   0.000     67.09076    70.16584
------------------------------------------------------------------------------

 95 . ****Rr^2=0.0142
 96 . ****q=2, N-K=df=1,963
 97 . display ((0.3005-0.0142)/2)/((1-0.3005)/1963)
401.72044

 98 .
 99 .
100 . ***d-ii)
101 .
102 . **e)
103 .
104 . reg mrkgrm classize if many_dis == 1

      Source |       SS           df       MS      Number of obs   =       858
-------------+----------------------------------   F(1, 856)       =     18.08
       Model |  826.915211         1  826.915211   Prob > F        =    0.0000
    Residual |    39159.72       856  45.7473364   R-squared       =    0.0207
-------------+----------------------------------   Adj R-squared   =    0.0195
       Total |  39986.6352       857  46.6588509   Root MSE        =    6.7637

------------------------------------------------------------------------------
      mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    classize |   .1585182   .0372848     4.25   0.000     .0853379    .2316986
       _cons |   63.80775   1.102878    57.86   0.000     61.64309    65.97241
------------------------------------------------------------------------------

105 . outreg2 using "PS2_regressione.doc", replace ctitle(high dis)
PS2_regressione.doc
dir : seeout

106 . reg mrkgrm classize if many_dis == 0

      Source |       SS           df       MS      Number of obs   =     1,109
-------------+----------------------------------   F(1, 1107)      =     18.62
       Model |  502.144157         1  502.144157   Prob > F        =    0.0000
    Residual |  29846.6908     1,107  26.9617803   R-squared       =    0.0165
-------------+----------------------------------   Adj R-squared   =    0.0157
       Total |   30348.835     1,108  27.3906453   Root MSE        =    5.1925

------------------------------------------------------------------------------
      mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    classize |  -.1101571   .0255254    -4.32   0.000    -.1602407   -.0600735
       _cons |   79.52074   .8205313    96.91   0.000     77.91076    81.13071
------------------------------------------------------------------------------
```

```
107 . outreg2 using "PS2_regressione.doc", append ctitle(low dis)
    PS2_regressione.doc
    dir : seeout

108 . reg mrkgrm classize many_dis many_disXclassize

          Source |       SS           df       MS      Number of obs   =     1,967
    -------------+----------------------------------   F(3, 1963)      =    281.14
           Model |  29649.3319         3  9883.11065   Prob > F        =    0.0000
        Residual |  69006.4108     1,963   35.153546   R-squared       =    0.3005
    -------------+----------------------------------   Adj R-squared   =    0.2995
           Total |  98655.7428     1,966  50.1809475   Root MSE        =     5.929

    ------------------------------------------------------------------------------------
               mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
    ------------------+-----------------------------------------------------------------
             classize |  -.1101571   .0291463    -3.78   0.000    -.1673179   -.0529962
             many_dis |  -15.71298   1.346291   -11.67   0.000    -18.35329   -13.07267
    many_disXclassize |   .2686753    .043792     6.14   0.000     .1827916     .354559
                _cons |   79.52074   .9369263    84.87   0.000     77.68326    81.35821
    ------------------------------------------------------------------------------------

109 . outreg2 using "PS2_regressione.doc", append ctitle(d)
    PS2_regressione.doc
    dir : seeout

110 .
111 . **f)
112 .
113 . foreach regioncode in Reg1 Reg2 Reg3 Reg4 Reg5 Reg6{
      2.         gen `regioncode' = 0
      3.         replace `regioncode' = 1 if regioncode == "`regioncode'"
      4. }
    (255 real changes made)
    (195 real changes made)
    (267 real changes made)
    (276 real changes made)
    (574 real changes made)
    (400 real changes made)

114 .
115 . **g)
116 .
117 . reg mrkgrm classize pct_dis if regioncode == "Reg1"

          Source |       SS           df       MS      Number of obs   =       255
    -------------+----------------------------------   F(2, 252)       =     66.06
           Model |  2865.43406         2  1432.71703   Prob > F        =    0.0000
        Residual |  5465.35025       252  21.6878978   R-squared       =    0.3440
    -------------+----------------------------------   Adj R-squared   =    0.3388
           Total |  8330.78431       254  32.7983634   Root MSE        =     4.657

    ------------------------------------------------------------------------------
          mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
    -------------+----------------------------------------------------------------
        classize |  -.0900784   .0450526    -2.00   0.047    -.1788059   -.0013508
         pct_dis |  -.2490626   .0218086   -11.42   0.000    -.2920129   -.2061122
           _cons |   81.00851   1.311482    61.77   0.000     78.42564    83.59137
    ------------------------------------------------------------------------------
```

```
118 . outreg2 using "PS2_regressiong.doc", replace ctitle(Reg1)
    PS2_regressiong.doc
    dir : seeout

119 . foreach regioncode in Reg2 Reg3 Reg4 Reg5 Reg6{
      2.         reg mrkgrm classize pct_dis if regioncode =="`regioncode'"
      3.         outreg2 using "PS2_regressiong.doc", append ctitle(`regioncode')
      4. }

          Source |       SS           df       MS      Number of obs   =       195
    -------------+----------------------------------   F(2, 192)       =     34.84
           Model |  3017.96063         2  1508.98031   Prob > F        =    0.0000
        Residual |  8314.72655       192  43.3058675   R-squared       =    0.2663
    -------------+----------------------------------   Adj R-squared   =    0.2587
           Total |  11332.6872       194  58.4159133   Root MSE        =    6.5807

    ------------------------------------------------------------------------------
          mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
    -------------+----------------------------------------------------------------
        classize |   -.054951   .0822927    -0.67   0.505    -.2172648    .1073629
         pct_dis |  -.2521636   .0309288    -8.15   0.000    -.3131675   -.1911597
           _cons |   77.17122   2.510945    30.73   0.000     72.21864     82.1238
    ------------------------------------------------------------------------------
    PS2_regressiong.doc
    dir : seeout

          Source |       SS           df       MS      Number of obs   =       267
    -------------+----------------------------------   F(2, 264)       =     45.65
           Model |  2924.30927         2  1462.15463   Prob > F        =    0.0000
        Residual |  8456.43979       264  32.0319689   R-squared       =    0.2570
    -------------+----------------------------------   Adj R-squared   =    0.2513
           Total |  11380.7491       266  42.7847709   Root MSE        =    5.6597

    ------------------------------------------------------------------------------
          mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
    -------------+----------------------------------------------------------------
        classize |   .1678957   .0669434     2.51   0.013     .0360848    .2997067
         pct_dis |  -.2127071    .027496    -7.74   0.000    -.2668463   -.1585678
           _cons |   69.64526   2.225889    31.29   0.000      65.2625    74.02801
    ------------------------------------------------------------------------------
    PS2_regressiong.doc
    dir : seeout

          Source |       SS           df       MS      Number of obs   =       276
    -------------+----------------------------------   F(2, 273)       =     84.22
           Model |  4600.18127         2  2300.09063   Prob > F        =    0.0000
        Residual |  7455.58685       273  27.3098419   R-squared       =    0.3816
    -------------+----------------------------------   Adj R-squared   =    0.3770
           Total |  12055.7681       275  43.8391568   Root MSE        =    5.2259

    ------------------------------------------------------------------------------
          mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
    -------------+----------------------------------------------------------------
        classize |   .0074092   .0458776     0.16   0.872    -.0829096    .0977281
         pct_dis |  -.4898067   .0388518   -12.61   0.000    -.5662939   -.4133195
           _cons |   80.17505   1.479157    54.20   0.000     77.26304    83.08705
    ------------------------------------------------------------------------------
    PS2_regressiong.doc
    dir : seeout
```

```
      Source |       SS           df       MS      Number of obs   =       574
-------------+----------------------------------   F(2, 571)       =    169.65
       Model |  10101.2705         2  5050.63523   Prob > F        =    0.0000
    Residual |  16999.5884       571  29.7716085   R-squared       =    0.3727
-------------+----------------------------------   Adj R-squared   =    0.3705
       Total |  27100.8589       573  47.2964378   Root MSE        =    5.4563

------------------------------------------------------------------------------
      mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   classize |   .0212184   .0429338     0.49   0.621    -.0631092    .1055459
    pct_dis |  -.3191605   .0195322   -16.34   0.000    -.3575243   -.2807967
      _cons |   75.24485   1.556768    48.33   0.000     72.18715    78.30254
------------------------------------------------------------------------------
```

PS2 regressiong.doc
dir : seeout

```
      Source |       SS           df       MS      Number of obs   =       400
-------------+----------------------------------   F(2, 397)       =    168.84
       Model |  8889.58149         2  4444.79075   Prob > F        =    0.0000
    Residual |   10451.356       397  26.3258338   R-squared       =    0.4596
-------------+----------------------------------   Adj R-squared   =    0.4569
       Total |  19340.9375       399  48.4735276   Root MSE        =    5.1309

------------------------------------------------------------------------------
      mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   classize |  -.0758319   .0540577    -1.40   0.161    -.1821071    .0304432
    pct_dis |  -.4038452   .0231615   -17.44   0.000    -.4493797   -.3583107
      _cons |   79.62101   1.882575    42.29   0.000     75.91994    83.32207
------------------------------------------------------------------------------
```

PS2 regressiong.doc
dir : seeout

```
. 
. ****Model alternative: Dummy with ommitting Reg5:
. reg mrkgrm classize pct_dis Reg1 Reg2 Reg3 Reg4 Reg6

      Source |       SS           df       MS      Number of obs   =     1,967
-------------+----------------------------------   F(7, 1959)      =    186.70
       Model |  39479.1959         7  5639.88513   Prob > F        =    0.0000
    Residual |  59176.5468     1,959  30.2075277   R-squared       =    0.4002
-------------+----------------------------------   Adj R-squared   =    0.3980
       Total |  98655.7428     1,966  50.1809475   Root MSE        =    5.4961

------------------------------------------------------------------------------
      mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   classize |    .006539   .0218313     0.30   0.765     -.036276     .049354
    pct_dis |  -.3085222   .0102683   -30.05   0.000    -.3286603   -.2883842
       Reg1 |   3.501301   .4311002     8.12   0.000     2.655837    4.346764
       Reg2 |   .9691248   .4671522     2.07   0.038     .0529574    1.885292
       Reg3 |   .3992164   .4098365     0.97   0.330    -.4045451    1.202978
       Reg4 |   3.257748   .4188072     7.78   0.000     2.436394    4.079103
       Reg6 |   .2489347   .3595311     0.69   0.489     -.456169    .9540384
      _cons |   75.55303   .8013656    94.28   0.000     73.98141    77.12465
------------------------------------------------------------------------------
```

```
123 . outreg2 using "PS2_regressiongalt.doc", replace ctitle(alt)
    PS2_regressiongalt.doc
    dir : seeout

124 .
125 . **h)
126 .
127 . ***h-i)
128 .
129 . reg mrkgrm classize sc_boys if n_classes == 1

      Source |       SS           df       MS      Number of obs   =       240
-------------+----------------------------------   F(2, 237)       =      2.12
       Model |  378.269997         2  189.134999   Prob > F        =    0.1222
    Residual |  21139.3133       237  89.1954149   R-squared       =    0.0176
-------------+----------------------------------   Adj R-squared   =    0.0093
       Total |  21517.5833       239  90.0317294   Root MSE        =    9.4443

------------------------------------------------------------------------------
      mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    classize |   .0964143   .1099979     0.88   0.382    -.1202843    .3131129
     sc_boys |  -.3024193   .1529586    -1.98   0.049    -.6037514   -.0010871
       _cons |   72.92232   2.177034    33.50   0.000     68.63352    77.21113
------------------------------------------------------------------------------

130 . outreg2 using "PS2_regressionhi.doc", replace ctitle(1)
    PS2_regressionhi.doc
    dir : seeout

131 .
132 . ***h-ii)
133 .
134 . reg mrkgrm sc_girls sc_boys if n_classes == 1

      Source |       SS           df       MS      Number of obs   =       240
-------------+----------------------------------   F(2, 237)       =      2.12
       Model |  378.269997         2  189.134999   Prob > F        =    0.1222
    Residual |  21139.3133       237  89.1954149   R-squared       =    0.0176
-------------+----------------------------------   Adj R-squared   =    0.0093
       Total |  21517.5833       239  90.0317294   Root MSE        =    9.4443

------------------------------------------------------------------------------
      mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    sc_girls |   .0964143   .1099979     0.88   0.382    -.1202843    .3131129
     sc_boys |   -.206005   .1150097    -1.79   0.075     -.432577     .020567
       _cons |   72.92232   2.177034    33.50   0.000     68.63352    77.21113
------------------------------------------------------------------------------

135 . outreg2 using "PS2_regressionhii.doc", replace ctitle(1)
    PS2_regressionhii.doc
    dir : seeout
```

```
136 .
137 . correlate sc_boys sc_girls
    (obs=1,967)

                 | sc_boys sc_girls
    -------------+------------------
         sc_boys |  1.0000
        sc_girls |  0.7918   1.0000


138 . reg mrkgrm classize if n_classes == 1

          Source |       SS           df       MS      Number of obs   =       240
    -------------+----------------------------------   F(1, 238)       =      0.33
           Model |  29.6008824         1  29.6008824   Prob > F        =    0.5675
        Residual |  21487.9825       238  90.2856406   R-squared       =    0.0014
    -------------+----------------------------------   Adj R-squared   =   -0.0028
           Total |  21517.5833       239  90.0317294   Root MSE        =    9.5019

    ------------------------------------------------------------------------------
          mrkgrm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
    -------------+----------------------------------------------------------------
        classize |  -.0475071   .0829689    -0.57   0.567    -.2109544    .1159402
           _cons |   72.65987   2.186222    33.24   0.000     68.35305    76.96668
    ------------------------------------------------------------------------------

139 .
140 .
    end of do-file
```