

## ECO 274

### Applications of Statistics in Economics

#### Data Set for the final project/presentation

#### Group: 1

Data set: CARD (<http://fmwww.bc.edu/ec-p/data/wooldridge/card.des>)

(<https://faculty.utrgv.edu/diego.escobari/teaching/Datasets.html>)

Card (1995) used wage and education data for a sample of men in 1976 to estimate the return to education. He used a dummy variable for whether someone grew up near a four-year college (nearc4) as an instrumental variable for education. Use the data in CARD to answer this question. (i) How many children are in the sample live with mom, single mom, step parent and how many report married and black race? (ii) What is the average number of father's and mother's education? Is the average a good measure of the "typical" schooling year in this case? Explain. (iii) Among black whose schooling year is greater than the average, what is the mean hourly wage? (iv) Report the average hourly wage and its standard deviation in dollars/cents. (v) Choose any set of three variables (as the climate-risk data we discussed in the class) from CARD data set and write the hypothesis for a one-way and two-way ANOVA, and examine the hypothesis, and draw a meaningful conclusion.

#### Group: 2

Data set: GPA2 (<http://fmwww.bc.edu/ec-p/data/wooldridge/gpa2.des>)

(<https://faculty.utrgv.edu/diego.escobari/teaching/Datasets.html>)

Use the data in GPA2 to answer this question. (i) How many male are in the sample, and how many report their race as black? (ii) What is the average of SAT score for the black and white? Is the average a good measure of the "typical SAT score" for a race in this case? Explain. (iii) Among the athlete who is black, what is the average of verbal/math SAT score? (iv) Find the average SAT score of female-athlete in the sample. (v) Report the average college gpa after fall semester and its standard deviation. (v) Choose any set of three variables (as the climate-risk data we discussed in the class) from GPA2 data set and write the hypothesis for a one-way and two-way ANOVA, and examine the hypothesis, and draw a meaningful conclusion.

### Group: 3

Data set: BWGHT (<http://fmwww.bc.edu/ec-p/data/wooldridge/bwgght.des>)

(<https://faculty.utrgv.edu/diego.escobari/teaching/Datasets.html>)

Use the data in BWGHT to answer this question. (i) How many women are in the sample, and how many report smoking during pregnancy? (ii) What is the average number of cigarettes smoked per day? Is the average a good measure of the “typical” woman in this case? Explain. (iii) Among women who smoked during pregnancy, what is the average number of cigarettes smoked per day? How does this compare with your answer from part (ii), and why? (iv) Find the average of fatheduc in the sample. Why are only 1,192 observations used to compute this average? (v) Report the average family income and its standard deviation in dollars. (v) Choose any set of three variables (as the climate-risk data we discussed in the class) from BWGHT data set and write the hypothesis for a one-way and two-way ANOVA, and examine the hypothesis, and draw a meaningful conclusion.

### Group: 4

Data set: FERTIL2 (<http://fmwww.bc.edu/ec-p/data/wooldridge/fertil2.des>)

(<https://faculty.utrgv.edu/diego.escobari/teaching/Datasets.html>)

The data in FERTIL2 were collected on women living in the Republic of Botswana in 1988. The variable children refer to the number of living children. The variable electric is a binary indicator equal to one if the woman’s home has electricity, and zero if not.

(i) Find the smallest and largest values of children in the sample. What is the average of children? (ii) What percentage of women have electricity in the home? (iii) Compute the average number of children for those without electricity and do the same for those with electricity. Comment on what you find. (iv) From part (iii), can you infer that having electricity “causes” women to have fewer children? Explain. (v) Choose any set of three variables (as the climate-risk data we discussed in the class) from FERTIL2 data set and write the hypothesis for a one-way and two-way ANOVA, and examine the hypothesis, and draw a meaningful conclusion.

## Group: 5

Data set: JTRAIN2 (<http://fmwww.bc.edu/ec-p/data/wooldridge2k/JTRAIN2.DES>)

(<https://faculty.utrgv.edu/diego.escobari/teaching/Datasets.html>)

The data in JTRAIN2 come from a job training experiment conducted for low-income men during 1976–1977; see Lalonde (1986).

(i) Use the indicator variable `train` to determine the fraction of men receiving job training. (ii) The variable `re78` is earnings from 1978, measured in thousands of 1982 dollars. Find the averages of `re78` for the sample of men receiving job training and the sample not receiving job training. Is the difference economically large? (iii) The variable `unem78` is an indicator of whether a man is unemployed or not in 1978. What fraction of the men who received job training are unemployed? What about for men who did not receive job training? Comment on the difference. (iv) Choose any set of three variables (as the climate-risk data we discussed in the class) from JTRAIN2 data set and write the hypothesis for a one-way and two-way ANOVA, and examine the hypothesis, and draw a meaningful conclusion.

## Group: 6

Data set: LOANAPP (<http://fmwww.bc.edu/ec-p/data/wooldridge/loanapp.des>)

(<https://faculty.utrgv.edu/diego.escobari/teaching/Datasets.html>)

Use the data in LOANAPP for this exercise. The binary variable `approve` is equal to one if a mortgage loan to an individual was approved. The explanatory variable `white` is a dummy variable equal to one if the applicant was white. The other applicants in the data set are black and Hispanic.

Use the data in LOANAPP to answer this question. (i) How many minorities are in the sample, and how many report bankruptcy? (ii) What is the average net worth? Is the average a good measure of the “typical” net worth in this case? Explain. (iii) Among white applicants who have more than 2 late payments, what is the average unemployment rate? (iv) Find the proportion of applicants with more than 12 years schooling by race in the sample. (v) Report the average years employed in line of work for married applicants and its standard deviation. (vi) Choose any set of three variables (as the climate-risk data we discussed in the class) from LOANAPP data set and write the hypothesis for a one-way and two-way ANOVA, and examine the hypothesis, and draw a meaningful conclusion.

## Group: 7

Data set: SMOKE (<http://fmwww.bc.edu/ec-p/data/wooldridge/smoke.des>)

(<https://faculty.utrgv.edu/diego.escobari/teaching/Datasets.html>)

Use the data in SMOKE for this exercise

(i) Find the smallest and largest values of years of schooling in the sample. What is the average of years of schooling? (ii) The variable *cigs* is the number of cigarettes smoked per day. How many people in the sample do not smoke at all? What fraction of people claim to smoke 20 cigarettes a day? Why do you think there is a pileup of people at 20 cigarettes? (iii) Given your answers to part (ii), does *cigs* seem a good candidate for having a normal distribution? (iv) What is the average annual income? Is the average a good measure of the “typical” annual income in this case? Explain. (v) Among states where state restaurant smoking restrictions apply, what is the average number of *cigs*. smoked per day? (vi) Choose any set of three variables (as the climate-risk data we discussed in the class) from SMOKE data set and write the hypothesis for a one-way and two-way ANOVA, and examine the hypothesis, and draw a meaningful conclusion.

## Group: 8

Data set: WAGE1 (<http://fmwww.bc.edu/ec-p/data/wooldridge2k/JTRAIN2.DES>)

(<https://faculty.utrgv.edu/diego.escobari/teaching/Datasets.html>)

Use the data in WAGE1 for this exercise. (i) Find the average education level of mother and father in the sample. What are the lowest and highest years of education? (ii) Find the average weekly hours in the sample. Does it seem high or low? What is the average monthly earnings? Is the average a good measure of the “typical” monthly income in this case? Explain. (iii) Find the proportion black and other races in the sample. What fraction of the married and black, whose IQ score is above the median score. (iv) How many participants in the sample are not from the urban area? How many are from the South region? (v) Choose any set of three variables (as the climate-risk data we discussed in the class) from WAGE1 data set and write the hypothesis for a one-way and two-way ANOVA, and examine the hypothesis, and draw a meaningful conclusion.