

While you wait for the start of this test, you may fill in the FRONT AND BACK of the BUBBLE FORM and read this cover. BUT, keep these test papers face up and flat on your desk.

Instructor: Prof. Murdock

Duration: 110 minutes. You MUST STAY for at least 60 minutes

Allowed aids: A non-programmable calculator and the aid sheets provided with this test

Format: This test includes these question papers and a BUBBLE FORM. There are 37 multiple choice questions with point values that may vary from 1 to 3 points each for a total of 71 points. The point value for each question is shown by [1pt], [2pts], [3pts]. Most questions have choices (A) – (E). For questions with fewer choices, the correct answer is ALWAYS one of those offered (e.g. if the choices are (A) – (D), then (E) is NOT a possible correct answer.)

Once the start of the test is announced, you may detach this cover page: the aid sheet is on the reverse side. These question papers and the aid sheets will not be collected. Only the BUBBLE FORM is collected.

You must record your answers on the BUBBLE FORM. In ALL cases what is (or is not) indicated on the BUBBLE FORM is your FINAL ANSWER. Marks are based SOLELY on the BUBBLE FORM, which must be completed before the end of the test is announced.

- On the FRONT of the BUBBLE FORM:
 - Print your 9 (or 10) digit student number in the boxes AND darken each number in the corresponding circles.
 - Print your last name and initial in the boxes AND darken each letter in the corresponding circles.
 - Fill in the upper left region of the form.
 - *****Your FORM CODE is A. Darken the circle with the letter A.*****
 - Failing to indicate your FORM CODE means that your answers will be out of sync compared to the solution key used to mark your paper. It is entirely *your responsibility* to properly indicate your FORM CODE.
 - You earn 1 point (not a joke) for correctly marking your FORM CODE on the front of your bubble form *and* as the answer for Question 1.
- On the BACK of the BUBBLE FORM:
 - Write in your name, sign, and record your answers.
- Use a pencil and make dark solid marks that fill the bubble completely.
- Erase completely any marks you want to change; Crossing out a marked box is incorrect.
- Choose the best answer for each question. If more than one answer is selected that question earns 0 points.
- For questions with numeric answers that require rounding, round your final answer to be consistent with the choices offered. Use standard rounding rules.

Sample mean: $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$ **Sample variance:** $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2}{n-1} - \frac{(\sum_{i=1}^n x_i)^2}{n(n-1)}$ **Sample s.d.:** $s = \sqrt{S^2}$

Sample coefficient of variation: $CV = \frac{s}{\bar{X}}$ **Sample covariance:** $s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{n-1} = \frac{\sum_{i=1}^n x_i y_i}{n-1} - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n(n-1)}$

Sample interquartile range: $IQR = Q3 - Q1$ **Sample coefficient of correlation:** $r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n z_{x_i} z_{y_i}}{n-1}$

Addition rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ **Conditional probability:** $P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$

Complement rules: $P(A^c) = P(A') = 1 - P(A)$ $P(A^c|B) = P(A'|B) = 1 - P(A|B)$

Multiplication rule: $P(A \text{ and } B) = P(A|B)P(B) = P(B|A)P(A)$

SIMPLE REGRESSION:

OLS line: $\hat{y}_i = b_0 + b_1 x_i$ $b_1 = \frac{s_{xy}}{s_x^2} = r \frac{s_y}{s_x}$ $b_0 = \bar{Y} - b_1 \bar{X}$

$SST = \sum_{i=1}^n (y_i - \bar{Y})^2 = SSR + SSE$ $s_y^2 = \frac{SST}{n-1}$

$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{Y})^2$ $SSE = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$

Coefficient of determination: $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$ **and** $R^2 = (r)^2$

Residuals: $e_i = y_i - \hat{y}_i$

Standard deviation of residuals: $s_e = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^n (e_i - 0)^2}{n-2}}$

(1) [1pt] Bubble your FORM CODE (A). (To earn the point, mark it for both Q1 and on the front of your bubble form.)

► **Questions (2) – (4):** A 2015 report “Race, Class, and College Access: Achieving Diversity in a Shifting Legal Landscape” (<http://www.acenet.edu/news-room/Documents/Race-Class-and-College-Access-Achieving-Diversity-in-a-Shifting-Legal-Landscape.pdf>) surveyed U.S. universities. One question asked each institution to rank five different priority categories.

Figure 5. Ranking of Institutional Priorities

Priority	Rank				
	1	2	3	4	5
Enrolling students with outstanding admissions-test scores and high school grades to improve the institution’s academic profile	56%	14%	9%	14%	8%
Racial/ethnic diversity	12%	30%	28%	20%	8%
International and U.S. geographic diversity	7%	23%	30%	25%	14%
Socioeconomic diversity	8%	19%	22%	26%	24%
Improving national rankings (e.g., <i>U.S. News & World Report</i>)	6%	16%	11%	14%	54%

(2) [2pts] It could also report the mean ranking of each priority. What is the mean for “Socioeconomic diversity”?

- (A) 3.2
- (B) 3.4
- (C) 3.6
- (D) 3.8
- (E) 4.0

(3) [2pts] It could also report the median ranking of each priority. Which priorities have a median ranking of 2?

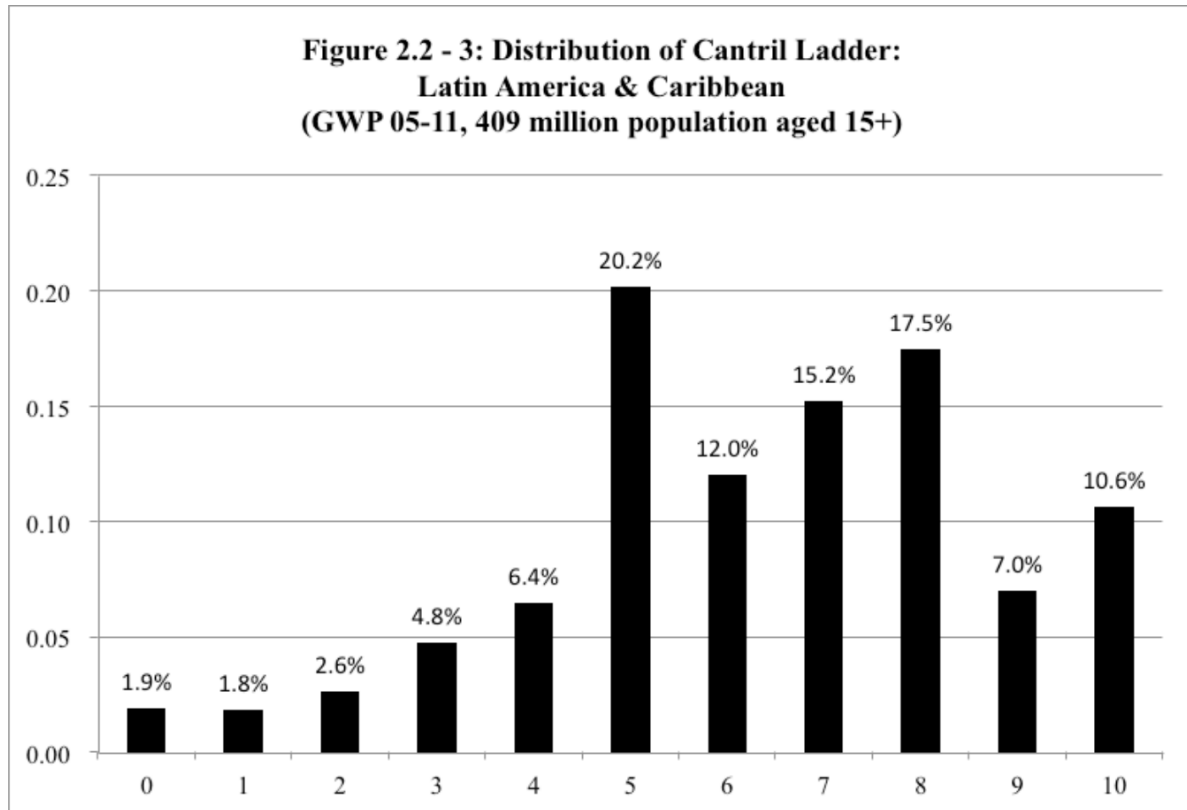
- (A) Only “Enrolling students with...”
- (B) Only “Racial/ethnic diversity”
- (C) Only “International and U.S. geographic diversity”
- (D) Both “Racial/ethnic diversity” and “International and U.S. geographic diversity”
- (E) None of the five institutional priorities have a median ranking of 2

(4) [2pts] It could also report the IQR of rankings of each priority. What is the IQR for “Enrolling students with...”?

- (A) 0
- (B) 1
- (C) 2
- (D) 3
- (E) 4

► **Questions (5) – (6):** Recall the “2012 World Happiness Report” edited by John Helliwell, Richard Layard and Jeffrey Sachs (<http://www.earth.columbia.edu/sitefiles/file/Sachs%20Writing/2012/World%20Happiness%20Report.pdf>).

In the Gallup World Poll respondents are asked (using fresh annual samples of 1,000 respondents aged 15 or over in each of more than 150 countries) to evaluate the quality of their lives on an 11-point ladder scale running from 0 to 10, with the bottom rung of the ladder (0) being the worst possible life for them and 10 being the best possible.” [This is called the Cantril Ladder.]



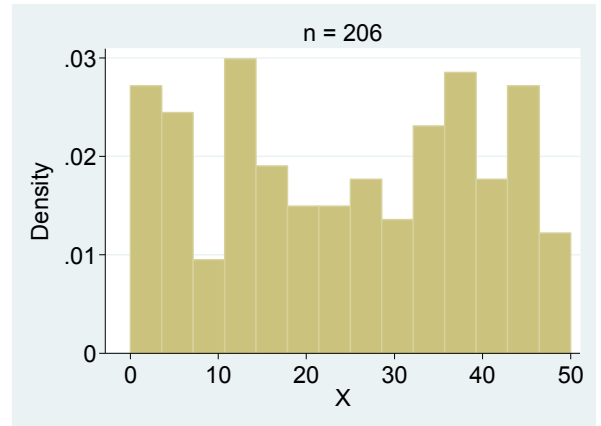
(5) [2pts] What is the *median* evaluation of the quality of life for the *first (lowest) decile*?

- (A) 0
- (B) 0.5
- (C) 1
- (D) 2
- (E) 3

(6) [2pts] What is the *mean* evaluation of the quality of life for the *fourth (highest) quartile*?

- (A) 9.006
- (B) 9.128
- (C) 9.214
- (D) 9.337
- (E) 9.410

► **Questions (7) – (8):** Consider this histogram of a variable named X with 206 observations.

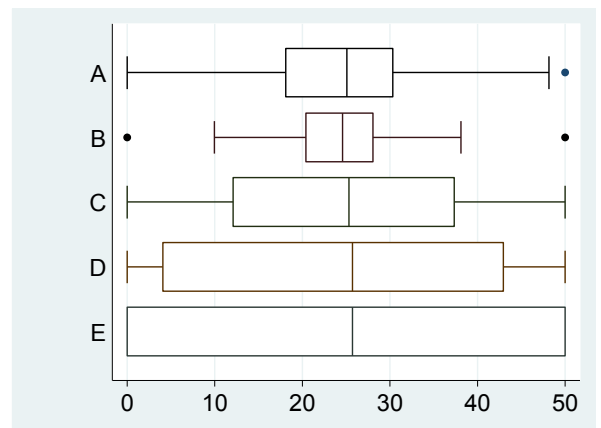


(7) [2pts] Approximately how many observations fall within the tallest bin?

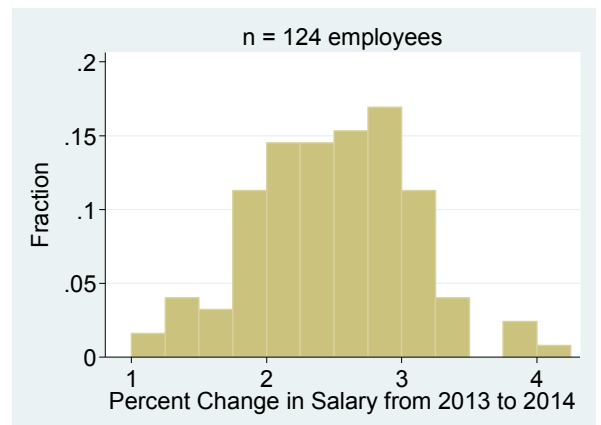
- (A) 3
- (B) 6
- (C) 10
- (D) 15
- (E) 22

(8) [2pts] Which of these box plots could be of the variable X?

- (A) A
- (B) B
- (C) C
- (D) D
- (E) E



► **Questions (9) – (10):** For 124 employees this histogram shows the percent change in salary between 2013 and 2014.



(9) [2pts] Approximately what is the sample s.d. of the percent change in salary?

- (A) 0.25 percentage points
- (B) 0.6 percentage points
- (C) 1 percentage point
- (D) 1.25 percentage points
- (E) 1.5 percentage points

(10) [2pts] Given the histogram, which of these could be the 95th percentile of the percent change in salary?

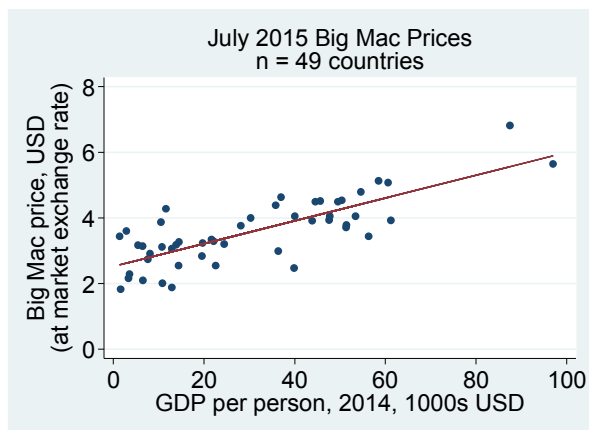
- (A) 3.1
- (B) 3.3
- (C) 3.8
- (D) 4.0
- (E) 4.1

► **Questions (11) – (20):** “The Big Mac Index” is computed and reported every six months by *The Economist* <http://www.economist.com/content/big-mac-index>. It reports both the Raw Index and the Adjusted Index.

The average price of a Big Mac in America in July 2015 was \$4.79; in China it was only \$2.74 at market exchange rates. So the “raw” Big Mac index says that the yuan was undervalued by 43% [(100*(2.74 – 4.79)/4.79)].

The adjusted index uses the “line of best fit” between Big Mac prices and GDP per person for 48 countries (plus the euro area). The difference between the price predicted by the [solid] line for each country, given its income per person, and its actual price gives a supersized measure of currency under- and over-valuation.

The Economist posts the raw data that I downloaded (July 2015). Here is the relevant scatter plot and OLS results.



OLS Results: $BIGMAC_P\text{-hat} = 2.5225 + 0.0349 * GDP_PC_1000USD$, $n=49$, $R\text{-squared} = 0.6226$

(11) [2pts] What kind of data are illustrated in the scatter diagram?

- (A) experimental, panel data
- (B) experimental, time series data
- (C) observational, panel data
- (D) observational, time series data
- (E) observational, cross-sectional data

(12) [2pts] In Canada a Big Mac is \$4.54 USD and GDP per capita is \$50,398 USD. According to the OLS line what is the predicted price of a Big Mac in Canada?

- (A) \$4.11
- (B) \$4.15
- (C) \$4.17
- (D) \$4.22
- (E) \$4.28

(13) [2pts] In Pakistan a Big Mac is \$3.44 USD and GDP per capita is \$1,343 USD. What is the residual for Pakistan?

- (A) \$0.87
- (B) \$1.09
- (C) \$1.12
- (D) \$1.16
- (E) \$1.28

(14) [1pt] In Japan a Big Mac is \$2.99 USD. What is the *Raw Index* for Japan?

- (A) The Japanese Yen is undervalued by 24%
- (B) The Japanese Yen is undervalued by 30%
- (C) The Japanese Yen is undervalued by 38%
- (D) The Japanese Yen is undervalued by 42%

(15) [2pts] In Japan a Big Mac is \$2.99 USD and GDP per capita is \$36,332 USD. Given its GDP per capita, does it seem that the Japanese Yen is undervalued?

- (A) No, because a Big Mac costs only \$2.99 USD in Japan versus \$4.79 in the US
- (B) Yes, because it has a negative residual (is below the OLS line in the scatter plot)
- (C) No, because Japan has a fairly low GDP per capita, which explains why prices are lower there
- (D) Yes, because the Raw and Adjusted Indices always have the same sign (i.e. both say over- or undervalued)
- (E) No, because Japan lies close to the OLS line: \$2.99 is roughly what it would predict given that GDP per capita

(16) [2pts] How should you interpret the R-squared (0.6226) of the OLS regression?

- (A) Across countries, when GDP per capita (USD) is 1% higher, Big Mac prices (USD) are 0.62% higher
- (B) Across countries, when the Big Mac price (USD) is one s.d. higher, GDP per capita (USD) is 0.62 s.d.'s higher
- (C) Across countries, when GDP per capita (USD) is one s.d. higher, the Big Mac price (USD) is 0.62 s.d.'s higher
- (D) Across countries, 62% of variation in Big Mac prices (USD) is explained by variation in GDP per capita (USD)

(17) [2pts] How should you interpret the intercept (2.5225) of the OLS regression?

- (A) In countries with zero GDP, Big Macs cost \$2.52 USD
- (B) In no country does a Big Mac cost less than \$2.52 USD
- (C) In countries with zero GDP, on average Big Macs cost \$2.52 USD
- (D) In no country does the OLS line predict the Big Mac cost less than \$2.52: it is a base or starting value
- (E) In countries with GDP per capita of \$2,523 USD, the OLS line (unreasonably) predicts Big Macs are free

(18) [2pts] How should you interpret the slope (0.0349) of the OLS regression? On average ____.

- (A) in countries with GDP per capita (USD) that is 1% higher, Big Mac prices (USD) are 3.5% higher
- (B) in countries with GDP per capita (USD) that is \$1 higher, Big Mac prices (USD) are 0.035% higher
- (C) in countries with GDP per capita (USD) that is 1% higher, Big Mac prices (USD) are \$0.035 higher
- (D) in countries with GDP per capita (USD) that is \$1 higher, Big Mac prices (USD) are \$0.035 higher
- (E) in countries with GDP per capita (USD) that is \$1,000 higher, Big Mac prices (USD) are \$0.035 higher

(19) [2pts] On average countries with GDP per capita (USD) that is 1 s.d. higher have Big Mac prices (USD) that ____.

- (A) are 0.035 s.d.'s higher
- (B) are 0.388 s.d.'s higher
- (C) are 0.584 s.d.'s higher
- (D) are 0.623 s.d.'s higher
- (E) are 0.789 s.d.'s higher

(20) [3pts] Also, here is a variance-covariance matrix for the Raw and Adjusted indices for the 49 countries.

	raw_index adj_index
-----+-----	
raw_index	449.276
adj_index	210.848 297.749

What is the coefficient of correlation between the Raw Index and the Adjusted Index?

- (A)** 0.58
- (B)** 0.60
- (C)** 0.63
- (D)** 0.66
- (E)** 0.68

► **Questions (21) – (27):** In the 2014 NBER paper “Learning Millennial-Style” participants watch a video and then choose among four credit card offers on a computer. Participants are randomly assigned the baseline video or that video plus a recap, which reviews how to choose a credit card. Next, participants were randomly shown one of two screens with the four credit card offers: one screen had a link to the terms for each card and the other screen had that plus “misleading ads” (e.g. “low fees” even if the card has fairly high fees). Suppose a researcher tries this with U of T students in ECO100Y. Of the 2,900 students invited, 1,800 complete it.

Contingency Table: Credit Card Choices			
	<i>Choice of Card</i>		
<i>Treatment</i>	Optimal	Suboptimal	Total
Baseline Video, No Ads	233	199	432
Baseline Video, Misleading Ads	119	340	459
Video Plus Recap, No Ads	300	168	468
Video Plus Recap, Misleading Ads	212	229	441
Total	864	936	1800

(21) [1pt] Which kind of data are these credit card choice data?

- (A)** panel, experimental data
- (B)** panel, observational data
- (C)** cross-sectional, experimental data
- (D)** cross-sectional, observational data

(22) [1pt] What is the probability that a participant made an optimal choice?

- (A)** 0.48
- (B)** 0.49
- (C)** 0.50
- (D)** 0.53
- (E)** 0.54

(23) [2pts] What is the probability that a participant shown the baseline video with no ads made an optimal choice?

- (A) 0.12
- (B) 0.27
- (C) 0.47
- (D) 0.54
- (E) 0.62

(24) [2pts] What is the probability that a participant who made an optimal choice saw misleading ads?

- (A) 0.18
- (B) 0.37
- (C) 0.38
- (D) 0.41
- (E) 0.42

(25) [2pts] Which of these statements are TRUE?

- (A) Seeing the misleading ads is disjoint from choosing the optimal card
- (B) Seeing the misleading ads is independent of choosing the optimal card
- (C) Choosing the optimal card is disjoint from choosing the suboptimal card
- (D) Choosing the optimal card is independent of choosing the suboptimal card
- (E) All of the above

(26) [2pts] If the treatments had *absolutely no influence* on the participants' choices and 936 chose the suboptimal card and 459 saw the baseline video with misleading ads, then what number should replace 340 in the contingency table?

- (A) 239
- (B) 254
- (C) 267
- (D) 291
- (E) It would still be 340

(27) [2pts] Which of these describe valid concerns about the credit card choice study?

- (A) These results may not generalize to a broader target population such Canadian Millennials
- (B) The stratified random sampling approach used means survey weights must be used when analyzing the data
- (C) Given that the Millennial generation in Canada is millions of young people, a sample size of only 1,800 is too small and would result in substantial non-response bias
- (D) Because there would be a lot of variability among ECO100 students in terms of pre-existing understanding of financial decisions and current debt: the results would likely suffer from an endogeneity bias
- (E) All of the above

(28) [2pts] A manufacturing process has two independent stages. The first stage has a mean of 4.25 hours and a s.d. of 1.05 hours. The second has a mean of 0.75 hours with a s.d. of 0.45 hours. What is the s.d. of *total* production time?

- (A) 1.14 hours
- (B) 1.22 hours
- (C) 1.31 hours
- (D) 1.50 hours
- (E) 1.70 hours

(29) [3pts] 8% of accounts have a problem an auditor should report. However, the auditor sometimes makes mistakes: s/he reports a problem when there is no problem 7% of the time; s/he reports there is no problem when there is a problem 6% of the time. If the auditor reports a problem, what is the probability that there really is a problem with the account?

- (A) 0.40
- (B) 0.42
- (C) 0.46
- (D) 0.50
- (E) 0.54

► **Questions (30) – (31):** The OECD maintains the *PISA International Database*. PISA is a test given across countries to measure students’ skills and knowledge after completing all compulsory education. It reports results for reading, math, and science. The 2012 PISA data for 65 countries record each country’s average score in each category. Suppose we round each country’s average score to the nearest 50 (“grouped data”). For just the 34 OECD member nations here is a cross-tabulation of the science and reading scores.

reading	science				Total
	400	450	500	550	
400	1	0	0	0	1
450	0	2	0	0	2
500	0	3	22	4	29
550	0	0	0	2	2
Total	1	5	22	6	34

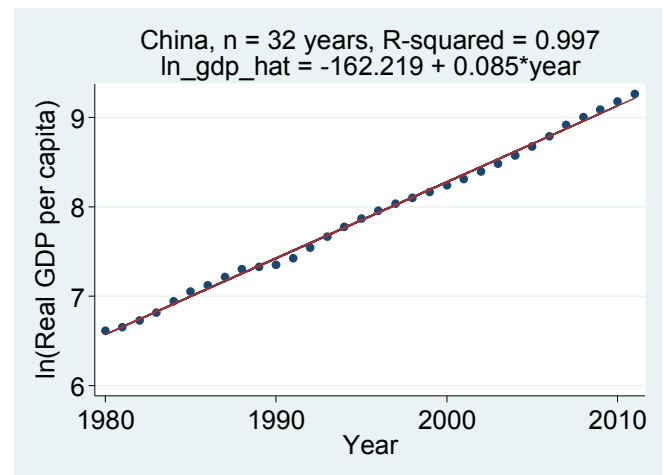
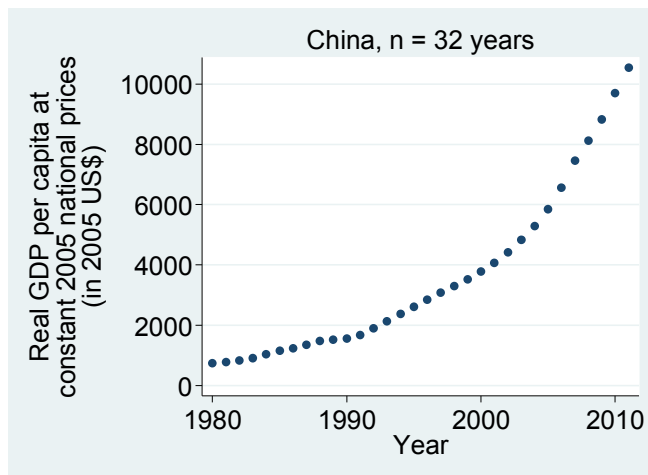
(30) [2pts] What can we conclude from the cross-tabulation for OECD countries?

- (A) There is no relationship between reading and science scores
- (B) There is a positive relationship between reading and science scores
- (C) There is a weak negative relationship between reading and science scores
- (D) There is a strong negative relationship between reading and science scores
- (E) After rounding to the nearest 50, the reading and science scores are perfectly correlated

(31) [1pt] For each of these 34 countries we have three variables: math, reading, and science scores in 2012. Which kind of data are these?

- (A) panel
- (B) time series
- (C) cross-sectional

► **Questions (32) – (35):** Recall the readings and study materials assigned prior to this test for “Asiaphoria Meets Regression to the Mean,” *NBER Working Paper 20573*, Oct. 2014, by Lant Pritchett and Larry Summers. All results in this test use the most recent PWT 8.1 data.¹ Consider these graphs and OLS results given in the title of the second graph.



(32) [2pts] How should the number 0.085 be *interpreted*?

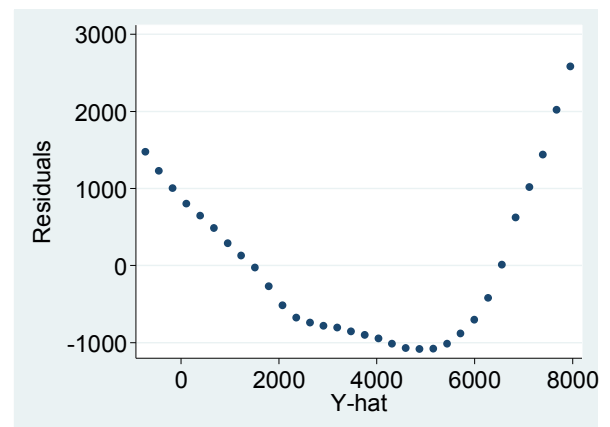
- (A) It should not be interpreted because the underlying relationship is exponential
- (B) Between 1980 and 2011 China’s GDP per capita has grown on average by 8.5% annually
- (C) Between 1980 and 2011 China’s GDP per capita has grown on average by 0.085% annually
- (D) Between 1980 and 2011 China’s GDP per capita has grown on average by 850 US\$ annually
- (E) Each year between 1980 and 2011 the natural log of GDP per capita has gone up by 0.085 units

(33) [2pts] What would the authors (Pritchett and Summers) say about the graphs above?

- (A) Given the excellent fit of the OLS line for China, it is reasonable to extrapolate out to the next decade
- (B) Recently (in 2015), China’s growth has slowed down just like the graphs and OLS results above predict
- (C) Regression to the mean implies that in the long run annual growth rates will vary little across countries
- (D) Be careful predicting China’s future growth: the OLS line explains little of the variation from 1980 - 2011
- (E) In the future China’s growth is expected to be more like an average country than like its past growth rate

(34) [2pts] Here is a plot of the residuals versus y-hat for regressing Real GDP per capita on year (i.e. fitting an OLS line to the *first* graph above). Which statements about this diagnostic plot are TRUE?

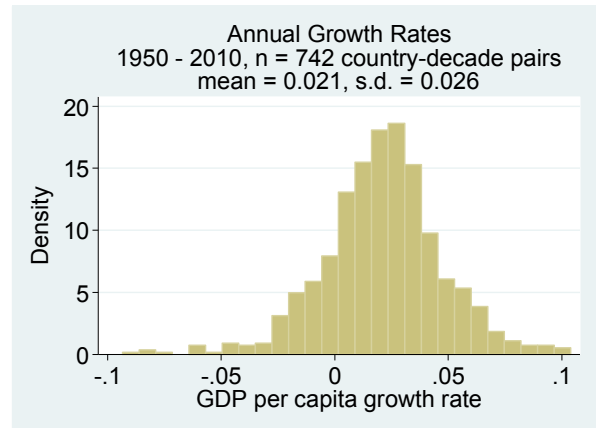
- (A) The mean of the residuals is zero
- (B) The residuals are measured in 2005 US\$
- (C) It shows a violation of the linearity condition
- (D) The OLS line systematically under-predicts recent GDP (e.g. the most recent five years of available data)
- (E) All of the above



¹ Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), “The Next Generation of the Penn World Table” forthcoming *American Economic Review*, available for download at www.ggd.net/pwt. PWT 8.1 is an updated version of PWT 8.0, covering the same countries and period. Released on: April 13, 2015. (DOI: 10.15141/S5NP4S, Retrieved June 8, 2015.)

(35) [2pts] Consider this histogram constructed using the PWT 8.1 data. It graphs the average annual growth rate for each decade for each country. Suppose the *standardized* growth rate in a particular country in a particular decade is -1. What does that mean? That country in that decade had an annual GDP growth rate ____.

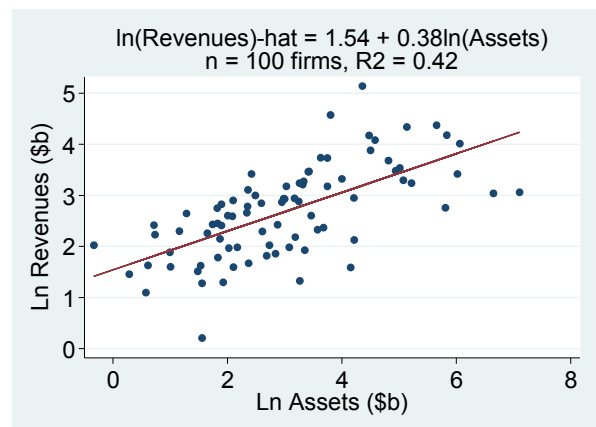
- (A) of negative 2.6 percent
- (B) of negative 1 percent
- (C) of negative 0.5 percent
- (D) of negative 0.026 percent
- (E) very slightly above zero growth



► **Questions (36) – (37):** For a random sample of 100 large and publically traded companies a researcher investigates the relationship between a firm’s annual revenue measured in billions of dollars and its assets also measured in billions of dollars. Here is a scatter plot, which includes the OLS results in the title.

(36) [2pts] How should you interpret 0.38? In these data, firms that have assets that are ____ higher typically have revenues that are ____ higher.

- (A) \$1 billion; \$0.38 billion
- (B) \$0.38 billion; \$1 billion
- (C) 1 percent; \$0.38 billion
- (D) 1 percent; 0.38 percent
- (E) 0.38 natural logs; 1 natural log



(37) [2pts] What would happen to 0.38 if the x variable were the natural log of assets measured in *millions* of dollars (and the y variable continued to be the natural log of revenues measured in billions of dollars)?

- (A) It would still be 0.38
- (B) It would become 38
- (C) It would become 380
- (D) It would become 0.0038
- (E) It would become 0.00038

REMEMBER, you must record your answers to the 37 multiple-choice questions on your BUBBLE FORM.