ECO220Y1Y, Test #2, Prof. Murdock

November 16, 2018, 9:10 – 11:00 am

U of T E-MAIL: _____@MAIL.UTORONTO.CA

SURNAME (LAST NAME):																		
GIVEN NAME (FIRST NAME):																		

UTORID: (e.g. LIHAO118)							

Instructions:

- You have 110 minutes. Keep these test papers and the *Supplement* closed and face up on your desk until the start of the test is announced. You must stay for a minimum of 60 minutes.
- You may use a non-programmable calculator.
- There are 6 questions (most with multiple parts) with varying point values worth a total of 100 points.
- This test includes these 8 pages plus the *Supplement*. The *Supplement* contains the aid sheets (formulas) and readings, figures, tables, and other materials required for some test questions. For each question referencing the *Supplement*, review *all* materials. ***The Supplement will NOT be graded:*** write your answers on these test papers. When we announce the end of the test, hand these test papers to us (you keep the *Supplement*).
- Write your answers clearly, completely and concisely in the designated space provided immediately after each question. An answer guide ends each question to let you know what is expected. For example, a quantitative analysis (which shows your work), a fully-labelled graph, and/or sentences.
    - Anything requested by the question and/or the answer guide is required.
    - Similarly, limit yourself to the answer guide. For example, if the answer guide does not request sentences, provide only what is requested (e.g. quantitative analysis). Leave yourself time to complete all questions rather than overdoing some questions and running out of time.
    - What is ***acceptable rounding***? Unlike online quizzes, written tests and exams do not specify high-precision rounding requirements. Marking TAs are instructed to accept all reasonable rounding.
    - For questions with multiple parts (e.g. (a) – (c)), ***attempt each part***.
- ***Your entire answer must fit in the designated space provided immediately after each question.*** No extra space/pages are possible. You *cannot* use blank space for other questions nor can you write answers on the *Supplement*. ***Write in PENCIL and use an ERASER as needed*** so that you can fit your final answer (including work and reasoning) in the appropriate space. Questions give more blank space than is needed for an answer (with typical handwriting) worth full marks. ***Follow the answer guides and avoid excessively long answers.***

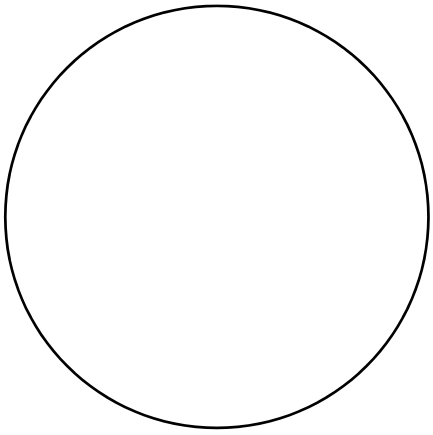
(1) See ***Supplement for Question (1)***: *Toronto's Segregated Immigrant Population.*

(a) [4 pts] What is the probability that a randomly selected Torontonian lives in a low income neighborhood? Use the *most directly relevant* information in the *Supplement* and the definitions of events. Write the probability in formal notation (shows if it is joint, marginal, or conditional) and compute its value. Show your work.

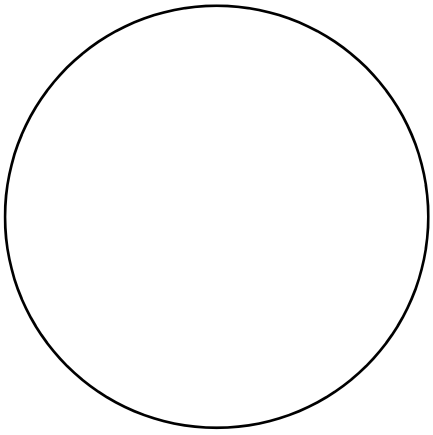
(b) [5 pts] What is the probability that a randomly selected Torontonian is a recent immigrant in a high income neighborhood? Use the *most directly relevant* information in the *Supplement* and the definitions of events. Write the probability in formal notation (shows if it is joint, marginal, or conditional) and compute its value. Show your work.

(c) [5 pts] Suppose Toronto were *EXACTLY* as in the *Supplement EXCEPT* immigrants were not segregated. How would the three pie charts look with no segregation of immigrants but everything else exactly the same? Answer by filling in the pie charts below. Put the numbers and labels *inside* the pie chart circles like the *Supplement*.

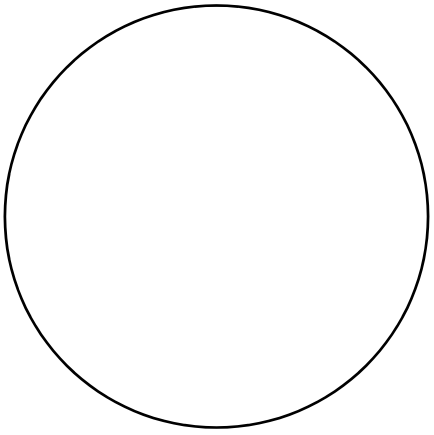
Low Income Neighbourhoods



Middle Income Neighbourhoods



High Income Neighbourhoods



(d) [6 pts] Are Events **M** and **H** *independent*? Include both a formal analysis and a context-specific explanation.
Answer with supporting analysis & 1 – 3 sentences.

(e) [8 pts] Are Events **N** and **L** *independent*? Include both a formal analysis and a context-specific explanation.
Answer with supporting analysis & 2 – 3 sentences.

(2) See ***Supplement for Question (2)***: *Electric Vehicles: Price versus Performance.*

(a) [2 pts] For 2013-2015, what can we say about the intercept, which is a in $\hat{y} = a + bx$? Answer by filling in the blank with: "about $20,000 ($\pm$ $2,000)," "well below $20,000," or "well above $20,000."

The intercept for the ***Line for 2013-2015*** is _____.

(b) [3 pts] Given the excerpt and Figure 3 in the *Supplement*, what do the authors say about how the relationship between base MSRP and battery range of electric vehicles is changing over time? Answer by filling in each blank with: "dramatically rising," "dramatically declining," or "staying about constant."

From 2011/12 through 2017/18, for the relationship between base MSRP and battery range of electric vehicles, the

intercept is _____ and the slope is _____.

(c) [4 pts] The ***Line for 2017-2018*** is approximately $\hat{y} = -34{,}580 + 330x$. Given this, what would the OLS line be if battery range were measured in kilometers? Note that 1 mile equals 1.60934 kilometers. Answer with the equation for the new OLS line.

(d) [4 pts] The ***Line for 2017-2018*** is approximately $\hat{y} = -34{,}580 + 330x$. Given this, what would the OLS line be if base MSRP were measured in Canadian dollars? Use the annual 2017 exchange rate posted by the Bank of Canada: $1USD equals $1.2986 Canadian. Answer with the equation for the new OLS line.

(3) See ***Supplement for Question (3)***: *Currie and Schwandt (2016): Per capita Income Versus Poverty Rank.*

(a) [8 pts] *Ignoring* the violations of the underlying conditions, *interpret* the OLS slope. What does -116.3 mean? In addition, address its sign and explain whether or not the sign is surprising. Answer with 2 – 3 sentences.

(b) [8 pts] The outlier condition is one of the violated underlying conditions. If we dropped the outlier, what would happen to the value of the s_e? *Why?* What would happen to the value of the R^2? *Why?* Answer with 2 – 3 sentences.

(4) See ***Supplement for Question (4)***: *Real GDP per capita in Australia.*

(a) [6 pts] In the first graph, what does the value 644 mean? Answer with 1 sentence that *interprets* it.

(b) [5 pts] For each of the two decades, it has been verified that a scatter plot of the *natural logarithm* of real GDP per capita in 2011US$ (abbreviated gdp) versus year is also straight for the Australian data. For the decade from 2000 through 2010, approximately what would b be in $\widehat{\ln(gdp)} = a + b * year$? Answer with a quantitative analysis.

(c) [8 pts] How do we compare 644 with 673? Include a good suggestion for a more informative comparison. Also, what can we learn from the comparison in this specific context? Answer with 2 – 4 sentences.

(5) Recall the 2016 Statistics Canada report "The evolution of income mobility in Canada" discussed in class (https://www150.statcan.gc.ca/n1/pub/75f0002m/75f0002m2016001-eng.htm). There is a 1.2% chance that a Canadian taxfiler in the 10^{th} income decile in 2007 drops down to the 1^{st} income decile in 2012.

(a) [10 pts] For a random sample of 30 Canadian taxfilers in the 10^{th} decile in 2007, consider a random variable counting how many taxfilers in the sample drop down to the 1^{st} decile in 2012. Graph the probability distribution of this random variable. You may skip graphing events with less than a 0.1% chance of occurring. Answer with a quantitative analysis & a fully-labelled graph.

(b) [7 pts] For a random sample of 100 Canadian taxfilers in the 10th decile in 2007, consider a random variable counting how many taxfilers in the sample are ***above*** the 1st decile (i.e. in the 2nd through 10th deciles) in 2012. What is the mean and standard deviation of this random variable? What would be the shape of its distribution: symmetric, positively skewed, negatively skewed, or approximately Bell shaped (Normal)? Answer with a quantitative analysis & the correct phrase to describe the shape.

(6) [7 pts] See ***Supplement for Question (6)***: *Asiaphoria and PWT 9.0.* What is the coefficient of correlation between the annual growth rate during the 1970-1990 period and the annual growth rate during the 1990-2010 period? Answer with a quantitative analysis.

This *Supplement* contains the aid sheets (formulas) and readings, figures, tables, and other materials required for some test questions. For each question referencing this *Supplement*, carefully review *all* materials.

Sample mean: $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ **Sample variance:** $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2}{n-1} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$ **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: $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$ **Conditional probability:** $P(A|B) = \frac{P(A\ and\ B)}{P(B)}$

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

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

Expected value: $E[X] = \mu = \sum_{all\ x} xp(x)$ **Variance:** $V[X] = E[(X-\mu)^2] = \sigma^2 = \sum_{all\ x}(x-\mu)^2 p(x)$

Covariance: $COV[X,Y] = E[(X-\mu_X)(Y-\mu_Y)] = \sigma_{XY} = \sum_{all\ x}\sum_{all\ y}(x-\mu_X)(y-\mu_Y)p(x,y)$

Laws of expected value:

$E[c] = c$

$E[X + c] = E[X] + c$

$E[cX] = cE[X]$

$E[a + bX + cY] = a + bE[X] + cE[Y]$

Laws of variance:

$V[c] = 0$

$V[X + c] = V[X]$

$V[cX] = c^2 V[X]$

$V[a + bX + cY] = b^2 V[X] + c^2 V[Y] + 2bc * COV[X,Y]$

$V[a + bX + cY] = b^2 V[X] + c^2 V[Y] + 2bc * SD(X) * SD(Y) * \rho$

where $\rho = CORRELATION[X,Y]$

Laws of covariance:

$COV[X,c] = 0$

$COV[a + bX, c + dY] = bd * COV[X,Y]$

Combinatorial formula: $C_x^n = \frac{n!}{x!(n-x)!}$ **Binomial probability:** $p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$ for $x = 0,1,2,\ldots,n$

If *X* is Binomial $(X \sim B(n,p))$ **then** $E[X] = np$ **and** $V[X] = np(1-p)$

SIMPLE REGRESSION:

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

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

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

$s_y^2 = \frac{SST}{n-1}$

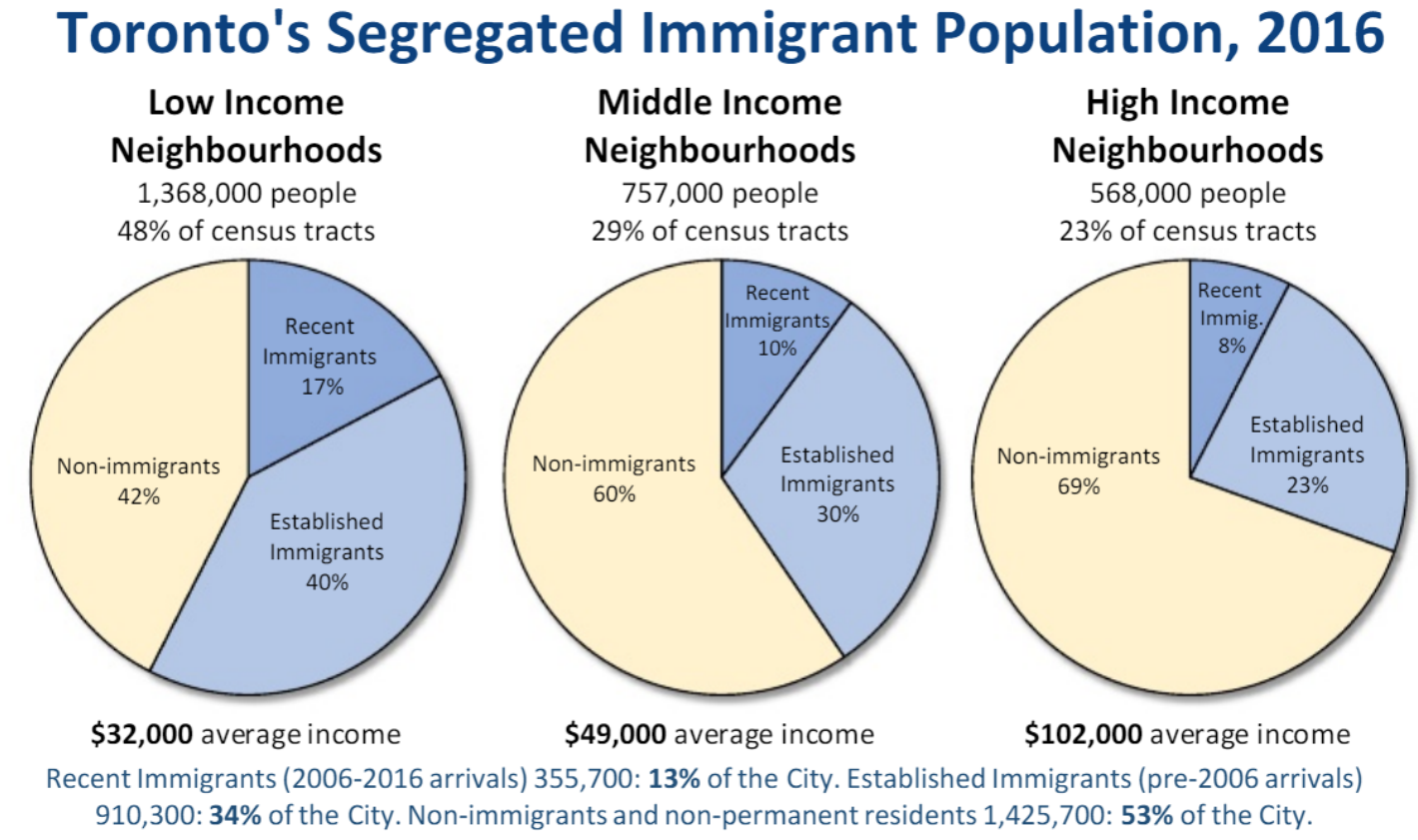
Coefficient of determination: $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = (r)^2$

Supplement for Question (1): Consider *one* of the figures appearing in a September 30th, 2018 article in *The Star* titled "Toronto is segregated by race and income. And the numbers are ugly." This figure is titled "Toronto's Segregated Immigrant Population, 2016" and a copy of it is below.

You must use these **definitions of events** in your answers:

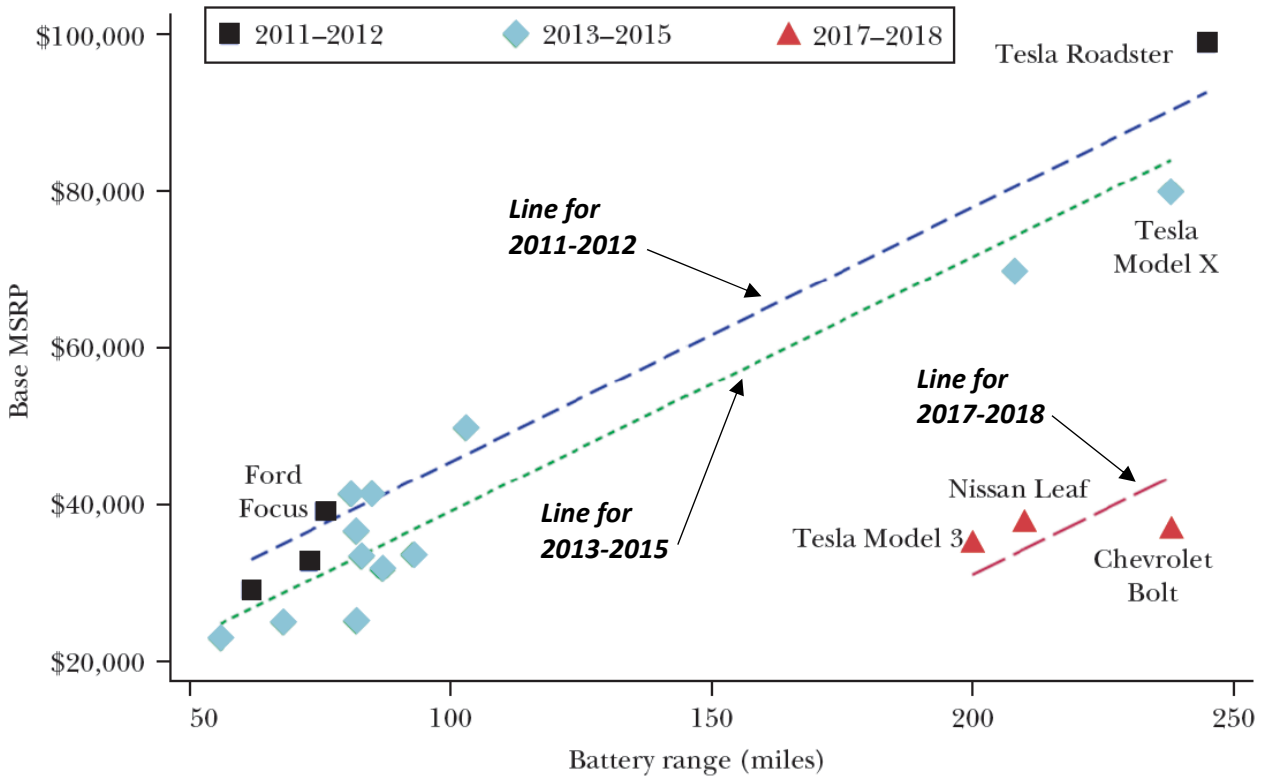
- **N** : The event that a randomly selected Torontonian is a **non-immigrant**
- **R** : The event that a randomly selected Torontonian is a **recent immigrant**
- **E** : The event that a randomly selected Torontonian is an **established immigrant**
- **L** : The event that a randomly selected Torontonian lives in a **low income** neighborhood
- **M** : The event that a randomly selected Torontonian lives in a **middle income** neighborhood
- **H** : The event that a randomly selected Torontonian lives in a **high income** neighborhood

Note: In the figure below, some population numbers are rounded to the nearest 1,000 whereas others are rounded to the nearest 100. This is the reason why they do not sum to exactly the same total population. This tiny discrepancy – caused by rounding – does not affect the meaning of the figure.



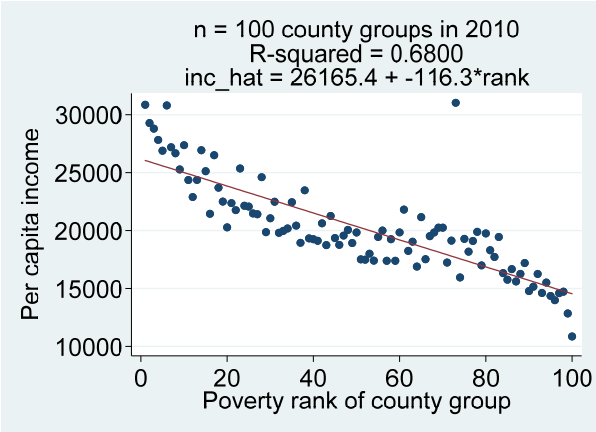
Supplement for Question (2): Consider Gillingham and Stock (2018) "The Cost of Reducing Greenhouse Gas Emissions" (https://doi.org/10.1257/jep.32.4.53). The title of Figure 3 [below] is "Electric Vehicle Manufacturers' Suggested Retail Price (MSRP) Plotted against the Battery Range Shows Impressive Technology Improvements within a Short Time."

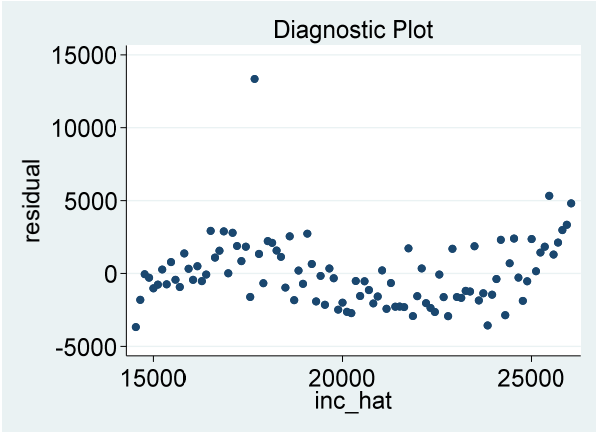
> **Excerpt, p. 66:** Figure 3 [below] plots electric vehicles that entered the market from model years 2011 to 2018 based on their suggested retail price (y-axis) and battery range (x-axis). The price-range frontier has strikingly shifted out: more recent market entrants have greater battery range at lower cost, underscoring this rapid improvement in technology.



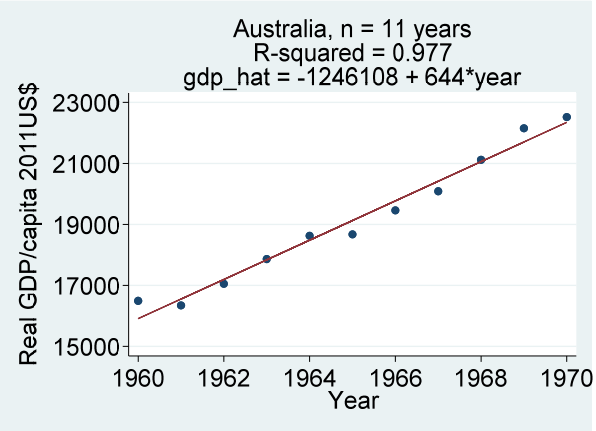
Note: Dates indicate year the model is introduced. Regression lines are fit with a common slope and different intercept for each group of model years. *Source:* J. Li (2017) and authors' calculations.

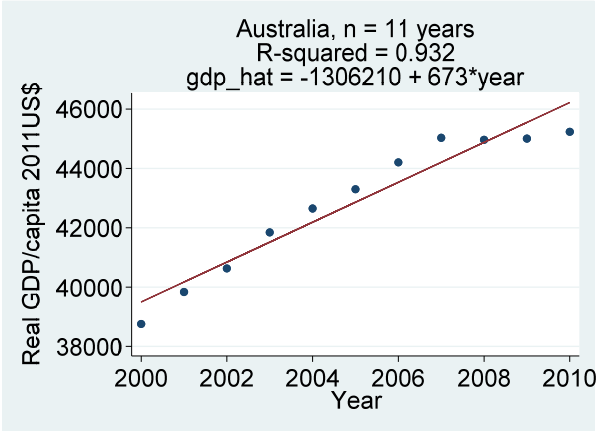
Supplement for Question (3): Recall Currie and Schwandt (2016) "Mortality Inequality: The Good News from a County-Level Approach." For the 2010 data, consider part of Figure A2 in the Appendix. In the left scatter plot below, on the x-axis is the poverty rank of each county group, where counties are grouped into 100 bins (each is 1 percent of the U.S. population). Page 40 explains "Counties [are] ranked by the share of their population below the poverty line [where a higher share means a higher rank]." On the y-axis is the per capita income of each county group (in constant 1999 USD).





Supplement for Question (4): The Penn World Tables (PWT) are a major database with some key economic indicators for many countries over many years. The scatter plots below use PWT 9.0 (released June 9, 2016, DOI: 10.15141/S5J01T). They show real GDP per capita (at constant 2011 national prices in 2011 US dollars) for two different decades for Australia. The title of each also gives the OLS regression results corresponding to the illustrated regression line.





Supplement for Question (6): Recall the NBER Working Paper 20573 "Asiaphoria Meets Regression to the Mean" by Pritchett and Summers (2014) and the PWT 9.0 data. Following the methods in Pritchett and Summers (2014) and using the PWT 9.0 data, the real GDP/capita annual growth rates are estimated for two two-decade long periods – from 1970 to 1990 and from 1990 to 2010 – for 142 countries. The last histogram shows how the annual growth rate changed from 1970-1990 to 1990-2010 across these 142 countries. For example, consider Canada: from 1970 to 1990 the growth rate was 0.020 whereas from 1990 to 2010 it was 0.018, which means a change of -0.002.