L&S 39G
Health, Human Behavior, and Data
Prof. Ryan Edwards

Class 2
Health metrics, social indicators, inequality, and policy

February 1, 2016
Today’s agenda

• Announcements, term paper topics

• In-class data exercises
  - Class 1: Smoking, life expectancy, Hispanics
  - Class 2: Jumping the gun on scatter plots

• Readings
  - Bhattacharya Chaps 1-2
  - Cutler, Deaton, and Lleras-Muney (2006) on historical mortality
Special Office Hours

• Because of a conference and Presidents’ Day, we next meet **in 3 weeks** on Monday, Feb 22

• Let’s schedule times for you to come visit me individually & discuss your term papers

• Links to WeJoinIn are on bCourses, and they’re:

  - Feb 8-12: [https://wejoinin.com/sheets/rmzxv](https://wejoinin.com/sheets/rmzxv)

  - Feb 15-19: [https://wejoinin.com/sheets/bhoip](https://wejoinin.com/sheets/bhoip)
Term papers

• Good writing may be more valuable than good coding

• It’s certainly more fungible across industries

• Students in previous classes have found the term paper helpful in setting them up for cool stuff at Cal

  - Economics majors take econometrics and write a paper; they can also write a thesis for honors
Term paper topics are due this Friday, Feb 5

• Guidelines:
  
  — Must be data-focused, not a book report nor a philosophical or theoretical piece

  — “Topics” consist of three sentences:
    
    1. This is the question I am asking & why

    2. These are the data I will analyze

    3. This is what I think I find (and what it will mean)

• Let’s discuss the end product and look at some examples
Don’t procrastinate

• Starting early helps you in two ways

• You’ll get done sooner with less stress

• You’ll have more time to identify a project that you can actually pull off in the time provided

• Just get it done. It doesn’t have to be a work of art
# Deadlines are on Fridays

<table>
<thead>
<tr>
<th>Feb 5</th>
<th><strong>Topic statement with 3 sentences</strong></th>
<th>Mar 25</th>
<th>(Spring Break)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feb 12</td>
<td><em>meet with me individually this week</em></td>
<td>Apr 1</td>
<td></td>
</tr>
<tr>
<td>Feb 19</td>
<td><em>or this week</em></td>
<td>Apr 8</td>
<td></td>
</tr>
<tr>
<td>Feb 26</td>
<td><strong>Produce one table or figure using your data</strong></td>
<td>Apr 15</td>
<td>Mostly done draft</td>
</tr>
<tr>
<td>Mar 4</td>
<td></td>
<td>Apr 22</td>
<td></td>
</tr>
<tr>
<td>Mar 7</td>
<td></td>
<td>Apr 29</td>
<td></td>
</tr>
<tr>
<td>Mar 11</td>
<td></td>
<td>May 6</td>
<td>Final draft</td>
</tr>
<tr>
<td>Mar 18</td>
<td><strong>Rough draft: One table or figure with written discussion</strong></td>
<td>May 13</td>
<td>Emailed response to final draft</td>
</tr>
</tbody>
</table>

Deadlines are on Fridays.
Methods

• Use any method you want

• Simple histograms, which will be discussed in DS 8, are fine

• Use any program you want

• MS Excel is fine, Python is fine, literally anything

• The tools may vary, but your desired output is the same: figures or charts and/or tables
Wikipedia

- Use it to jumpstart your knowledge
- Never
- Ever
- Ever
- Ever cite it as a source. Never.
- It doesn’t cite itself. It’s not a source, it’s practically anonymous cloud opinions
Basic structure of a scientific paper

I. Introduction

II. Data and methods

III. Results

IV. Conclusion

V. References

VI. Appendix (charts, figures)
I. Introduction

• State the question you are trying to answer

• Motivate it
  - Why do we care about this question?
  - Have other people looked at this? What did they find?

• State your hypothesized answer to the question

• You can also preview results if you like, or wait for dramatic effect
II. Data and methods

• What data are you using?
  
  - State the source, like:
    
    • The American Community Survey extracted from IPUMS for 1980-2000
    
    • The Integrated Health Interview Series
  
  - Which variables did you look at? What time periods? Was the unit of analysis individuals? Households? Something else?
  
  - You could use a table to summarize the data set with averages of variables, number of observations
III. Results

- Show graphs or tables, in-line or at the end of the paper

- Number them like “Figure 1” and refer to “Figure 1” in the main body of the text

- For each graph or table, state what you see
  - Refer to the figure numbers in order
  - What is the observed relationship? Why are we looking at this, what does it tell us?
IV. Conclusion

• Summarize your findings (briefly!)
• Take a stand, and be clear
• But also keep a healthy skepticism of your results
• What are the limitations of your analysis?
  - (Don’t say you ran out of time)
  - Selection bias
  - Reverse causality
  - Measurement problems
  - Omitted variables
• If you had more time, what other analysis would you do?
V. References

• Also known as a bibliography

• Cite everything

• Cite your data sources

• Include URLs if relevant
Tips

• GET STARTED thinking about this. Don’t procrastinate!

• Have a friend proofread it. It’s only 5 pages

• Too long? Cut. Or move material into endnotes

• Label your sections clearly, make sure graphs are readable on the screen and printed

• Number your pages!

• Save a tree and print double-sided
Two examples of great papers

The first uses very straightforward graphing of data

The second uses complicated regression analysis!

You should use methods that:

• you understand
• are appropriate to the task, no more & no less
Accessibility of Mental Health Services for Whites and Non-Whites: How Much Do Income and Language Differences Explain?
Introduction:

Many college students including myself lead an irregular and unhealthy lifestyle as a consequence of the heavy workload from school. We have unhealthy diets, lack of exercise, and very limited sleep. Furthermore, we are constantly exposed to high volume of stress from school and relationships. Such irregular lifestyles can severely harm our health, specifically our mental health. Regardless of the severity of an individual’s mental health problem, he or she will experience an inability to perform any tasks productively. Personal relationships with families and friends are also jeopardized as a result of the individual’s anxiety and mood disorders. Therefore, it is essential to seek professional help when one suffers or potentially suffers from mental illness.
Although the importance of seeking professional mental healthcare applies to everyone in any society, evidence from public mental healthcare services in the United States shows that racial and ethnic minorities receive less mental health services than White Americans. As an Asian American, I am thus motivated to find possible linguistic barriers. In comparison to the income of Whites, does the lower income of the minorities serve as a financial barrier to the costly mental health services? Do differences in the languages spoken generate lack of communication and understanding between the racial minority patients and their mental health providers, resulting in lower rates of mental help services attained by minorities? Using data from the California Health Interview Survey and the National Institute of Mental Health, I discovered that differences in the income of Whites and Non-Whites have very limited effects on the lower rates of mental health services for Non-Whites. On the other hand, linguistic barriers greatly reduce the minorities' demand and accessibility of mental healthcare.
and care between the two groups. To carry out this analysis, I took the last available set of data from the National Institute of Mental Health containing the number of individuals seeking psychiatric services, which was surveyed in 1985. The data was already divided between White and Non-White; however, it was given as the number of people seeking for psychiatric services. Since I am more interested in the proportion of people within each of the two groups seeking mental services, I divide the numbers by the total population of White or Non-White to obtain the percentages of White and the percentage of Non-White obtaining mental services. As Figure 1 illustrates, although the percentage of individuals seeking psychiatric services is small for both White and Non-White, ranging from 1% to 7%, White Americans receive more services relative to the racial minorities. This difference in proportion is prevalent at every age category.
Conclusion:

After analyzing how differences in income and language between Whites and minorities contribute to the lower amount of mental healthcare received by the minorities in the United States, I am surprised by part of the results. Contrary to popular belief that the lower income of minorities greatly reduces their ability to afford expensive mental healthcare, the demand for mental health services is similar across all income levels. Furthermore, when holding income constant, there remains a lower demand by minorities.

It’s always OK to be surprised by results!
Figure 1: % of White vs % of Non-White Receiving Psychiatric Services: United States 1985

Source: National Institute of Mental Health, 1985. Table Estimated Admissions to Mental Centers, US; Table Population By Race, US.
Figure 2: Percentage Receiving Professional Mental Services within Each Income Level, Categorized by Race:
United States 2003

Length that fits this paper very well

6.25 pages of text

Four figures

One table

One page of just 3 references
“Cigarette Smoking and Work-Related Stress: A Multiple Regression Data Analysis”
I. Introduction

The human race has made enormous strides in improving living standards and increasing longevity – even in the treatment and prevention of cancer. Incidences of all forms of cancer have been declining in the past few years, except those related to behavior. This would include lung and throat cancer, both of which have been strongly linked to cigarette smoking. Tobacco use, particularly cigarette smoking, is the leading preventable cause of death in the United States today. Hence, it is important to understand the major causes of cigarette smoking to prevent addiction and the probable onset of cancer.

One of the most-commonly cited reasons for why people smoke is to relieve stress.

In this paper, I try to establish a positive correlation between smoking cigarettes and work-related stress. Statistics on the state-specific prevalence of adult cigarette use and unemployment rates are used to estimate the responsiveness of adult smoking to changes in the level of work-related stress among states in the US for the year 2000. Using multiple regression analysis, I found that for every 1% increase in the unemployment rate, the percent of adult cigarette use increases by approximately 0.284%.

Additional variables, such as gender, percentage of Whites, and level of education, are also included to control for omitted variable bias.
II. Data and Hypothesis

Statistics on adult cigarette use of the 50 states and the District of Columbia are taken from a recent study by the Center for Disease Control and Prevention (CDC) of data from the 2000 Behavioral Risk Factor Surveillance System (BRFSS). BRFSS is a state-based, random-digit-dialed telephone survey of the noninstitutionalized US population over the 18 years of age. Current smokers were defined as those who reported having smoked more than 100 cigarettes during their lifetime and who currently smoked every day or some days. Table 1 shows the mean of adult cigarette use for the United States in 2000 is 22.8%. Although standard deviation is only 0.03%, the range of percentages among the 50 states reach as high as 30.5% and as low as 12.9%. This is more than a two-fold difference across observations.
(1) $cig = fn(u)$

Equation (1) states that adult cigarette use ($cig$) is a function of unemployment rates ($u$). Since differences in gender ($g$), population of Whites ($r$), and education ($e$) may also be individually correlated with both unemployment and smoking, these variables are added into the model one-by-one to determine its effect on the base association. The final model (2) regresses adult cigarette use on unemployment rates in addition to the other three variables:

(2) $cig = fn(u, g, r, e)$

Each of the regressions approximates a linear relationship between adult cigarette use and the relevant explanatory variables.

Because stress is often stated as a reason for smoking, I anticipate a positive correlation between adult cigarette use and unemployment. The higher the level of unemployment the more work-related stress in the population thus leading to an increased prevalence of smoking within the state. However, I
III. Empirical Implementation

The fundamental multiple regression model is

\[ Y_i = B_1 X_{1i} + B_2 X_{2i} + B_3 X_{3i} + B_4 X_{4i} + u_i \]

where the subscript \( i \) indicates the \( i \)th of the 51 observations (states) in the sample and \( u_i \) is the error term. \( Y_i \) is the \( i \)th observation of the dependent variable, adult cigarette use, and \( X_i \) is the \( i \)th observation on each of the regressors (gender, population of Whites, and education). A multiple regression model permits estimating the effect on adult cigarette use of changing one variable while holding the other regressors constant. In other words, it provides a way to isolate the effect on the dependent variable of one explanatory variable while holding the others constant.\(^6\) This model also postulates a linear relationship between the \( X \)'s and \( Y \) and applies the method of ordinary-least squares to approximate the best-fit line.

IV. Results

Table 2 to 5 show coefficients for each of the relevant explanatory variables along with their corresponding t-statistics and p-values at the 5% significance level. A joint hypothesis test is also included at the bottom of each table. The null hypothesis for the joint test is that unemployment rates have no effect on adult cigarette use. If this is true, the regression coefficient for state-specific unemployment rates will be zero.
For the base model (Equation (1)) in Table 2, the coefficient for state unemployment rates \( (B_{ii}) \) is 0.284 with a standard error of 0.337. This finding supports the hypothesis that there is a positive correlation between unemployment rates, hence work-related stress, and adult cigarette use. However, the coefficient is insignificant at the 5% level. This means that there is a 95% probability that the coefficient for state unemployment rates could be equal to zero. Because this is a relatively high probability, the positive association is unconvincing. The joint hypothesis test is the same as the t-statistic in this particular case since there is only one regressor.

The author is using a standard statistical approach to assess whether the estimated relationship between unemployment and smoking is abnormal enough not to be just noise.
V. Conclusion

The findings in this paper suggest that there is little to no relationship between unemployment levels and adult cigarette use. All four regressions show a relatively low, insignificant but positive coefficient for the state-specific unemployment rates variable. However, work-related stress may be measured more accurately by a different variable rather than unemployment rates. It may be that state unemployment rates are a poor indicator for the level of work-related stress. The data in this paper is insufficient to completely reject the original assumption that there is a positive correlation between work-related stress and cigarette smoking. Another factor to consider is the recent implementation of laws prohibiting smoking in certain locations, such as offices.

Very nice exposition of how results contradict the hypothesis, but in a way that doesn’t rule out future investigations
### Table 1

**Descriptive Statistics**

(Standard deviation in parenthesis)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult Cigarette Use</td>
<td>22.8%</td>
<td>30.5%</td>
<td>12.9%</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State Unemployment Rates</td>
<td>3.9%</td>
<td>6.7%</td>
<td>2.2%</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of Females</td>
<td>50.8%</td>
<td>52.9%</td>
<td>48.3%</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of Whites</td>
<td>78.5%</td>
<td>96.9%</td>
<td>24.3%</td>
</tr>
<tr>
<td></td>
<td>(0.145)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-School Graduates</td>
<td>81.9%</td>
<td>88.3%</td>
<td>72.9%</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2

Regression Coefficients and Joint Hypothesis Test
For Adult Cigarette Use and State Unemployment Rates

|                                | Coefficient | Standard Error | t stat | p > |t| | Lower 95% | Upper 95% |
|--------------------------------|-------------|----------------|--------|-----|---|-----------|-----------|
| State Unemployment Rates       | 0.284       | 0.337          | 0.843  | 0.398 |   | -0.376    | 0.946     |

Ho = Unemployment rates have no effect on adult cigarette use (coefficient = 0)

F-stat 0.712
p value 0.398


This is sparsely explained here, but it is somewhat clear from the title that:

smoking = coefficient \times \text{state unemployment rate}
Readings
Readings this week

  - Overview of historical human development since Enlightenment, health among developing countries today
  - A look at health and mortality within countries
  - Focus is on trying to understand why health improves
  - Is it just income? Rich people, rich countries are healthier?

• Bhattacharya Chapters 1 and 2
  - Motivation for studying health economics, first look at randomized controlled trials in the RAND Health Insurance Experiment & Oregon
Figure 1

The Preston Curve: Life Expectancy versus GDP Per Capita

Source: Reproduced from Deaton (2003, Figure 1).
Note: Circles are proportional to population.
Historical mortality declines

- Stronger declines in mortality rates during infancy and youth, because of great progress against infectious disease. Why?

  - Improved nutrition?
    - More income buys more food and improves health
    - Also technologies improve crop yields, more food, improves health
    - McKeown: tuberculosis ("consumption")
    - Fogel: big increases in caloric intake & height after mid-1800s
    - But was the timing right? What about the lack of a clear health gradient between well-fed aristocrats and poorly-fed commoners? And did diseases cause poor nutritional absorption?
- **Public health initiatives?**

  - There’s something besides income that matters for health, because country X, with the same incomes today as country Y in the past, has better health

  - Macro and micro initiatives
    - Macro: quarantines, public works like piping/plumbing, chlorination, pasteurizing
    - Micro: boiling, washing hands, ventilating rooms

  - A lot of this required the germ theory of disease, not accepted until 1880s

  - Deaths from water and food-borne diseases like typhoid, cholera, dysentery, fell a lot because of water purification
- Urbanization?
  - At first, living in cities was *bad for health* because of the easier spread of infectious disease

- Vaccination?
  - Big vaccination didn’t take off until the 20th century, but there were cruder technologies before that, like “variolation” — rubbing people with smallpox
  - In 1775, Washington ordered the entire Continental Army variolated
  - But except for TB, most diseases degraded quality rather than quantity of life … so this can’t have been a huge source of improvement
- Medical treatments?

  • Huge for understanding gains against cardiovascular disease (heart attacks) after 1960

    - They say the rest of the gains are due to smoking cessation

  • Gains against infant mortality stemming from improved neonatal care especially for low birthweight

  • Cutler clearly thinks this is important, Fogel not so much: “The main thing that physicians do is to make life more bearable: reduce morbidity and tell people how to take care of themselves.” (Escape from Hunger and Premature Death, 2004)
- Early life conditions

  • “Fetal origins” or “womb with a view” hypothesis of Barker (1990):
    - Nutritionally deprived fetuses get biologically reprogrammed to sustain themselves through their reproductive years
  
  • Supported by robust independent correlation between birthweight and adult height
  
  • Mixed evidence of poor health in adulthood for surviving fetuses during famines and trauma
  
  • Doblhammer and Vaupel (2001) find month-of-birth effects that show improved longevity after age 50 for kids born during the autuminal harvest
  
  • Cutler et al. think these effects in any event are likely to be small compared to those of cardiovascular risk factors in adulthood, but they also effectively “shrug” and suggest wait-and-see
Summary of historical mortality declines

1. From mid-1700s to mid-1800s
   Improved incomes may have improved nutrition and thus health

2. From late 1800s to 1930s
   Public health era, when urbanization raised disease prevalence at first, then cleaner water and healthy practices improved health

3. From the 1930s
   The “Era of Big Medicine” starting with vaccination and antibiotics, continuing with intensive personal interventions
Table 1

The Worldwide Structure of Mortality in 2002

<table>
<thead>
<tr>
<th></th>
<th>Treatments/Prevention</th>
<th>World</th>
<th>Low-income countries</th>
<th>High-income countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deaths per 100,000</td>
<td></td>
<td>916</td>
<td>1,113</td>
<td>846</td>
</tr>
<tr>
<td>Percent of total deaths by age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children (0–4)</td>
<td></td>
<td>18.4%</td>
<td>30.2%</td>
<td>0.9%</td>
</tr>
<tr>
<td>Elderly (60+)</td>
<td></td>
<td>50.8</td>
<td>34.2</td>
<td>75.7</td>
</tr>
<tr>
<td>Percentage of deaths from chronic diseases</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer</td>
<td>Partially preventable and treatable</td>
<td>12.4</td>
<td>6.3</td>
<td>26.2</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>Partially preventable and treatable</td>
<td>29.3</td>
<td>21.5</td>
<td>38.1</td>
</tr>
<tr>
<td>Numbers of deaths, millions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory infections*</td>
<td>Antibiotics</td>
<td>3.96</td>
<td>2.90</td>
<td>0.34</td>
</tr>
<tr>
<td>HIV/AIDS</td>
<td>Anti-retroviral therapy</td>
<td>2.78</td>
<td>2.14</td>
<td>0.02</td>
</tr>
<tr>
<td>Perinatal deaths*</td>
<td>Pre- and post-natal care</td>
<td>2.46</td>
<td>1.83</td>
<td>0.03</td>
</tr>
<tr>
<td>Diarrheal diseases*</td>
<td>Oral rehydration therapy</td>
<td>1.80</td>
<td>1.54</td>
<td>—</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>Preventable with public health; usually treatable</td>
<td>1.57</td>
<td>1.09</td>
<td>0.01</td>
</tr>
<tr>
<td>Malaria*</td>
<td>Partially preventable; treatable</td>
<td>1.27</td>
<td>1.24</td>
<td>—</td>
</tr>
<tr>
<td>DPT/Polio/Measles*</td>
<td>Vaccinations</td>
<td>1.12</td>
<td>1.07</td>
<td>—</td>
</tr>
</tbody>
</table>

Source: Based on WHO data and subject to large margins of error, particularly for adult mortality in low-income countries, most of which lack complete vital registration systems.

Notes: DPT stands for diphtheria, pertussis (whooping cough) and tetanus. An asterisk (*) indicates that the disease is most commonly fatal in children, except respiratory disease in high-income countries. — indicates less than 10,000 deaths. “Low-income” and “high-income” are World Bank designations of countries; these can be thought of as corresponding to below $5,000 PPP and above $10,000 PPP in Figure 1. Perinatal deaths are deaths in the first seven days of life and are primarily associated with low birth weight.
Challenges in the developing world

• Much debate over how to improve health

• “Wealthier is Healthier,” Pritchett and Summers (1996)

• Cutler et al. say “income can only buy so much if the disease burden is overwhelming” (p. 111)

• Health delivery systems are not in good shape, with doctor absenteeism, incentives to overtreat, ineffective care

• Women’s education (and empowerment) continues to lag but is arguably critical, especially for children’s health

• Even if income raises health, reduced mortality may increase population growth and thus reduce income unless things adjust
Health inequalities within countries

- Whitehall studies of British civil servants: strong effects of rank on cardiovascular mortality and all-cause, but less for cancer

- In the U.S. and in other advanced economies, a strong gradient in mortality at all ages across race, education, income, wealth, almost always favoring the more well off

- Why?
Causes of the gradient

• Health care?
  - Least well off have worse care, but they’re also sicker to begin with
  - Universal health care doesn’t seem to change thing much, but it should if health care were an important cause of the gradient
    • Introduction of Medicare had unclear impacts on U.S. mortality, no clear improvement vs. the UK
    • Britain’s class gradient in health survived the introduction of the National Health Service in WWII

• Resources? (Income to buy food)
  - Access to cheap food now appears to be a real problem (obesity)
• Health behaviors?

- There are big gradients in smoking, drinking, exercise, preventive behaviors like mammograms, adherence to doctors’ suggestions.

- But gradients in mortality persist even after controlling for these observed behaviors.

- Worse, behavior by different groups differs even though information and prices seem to be the same.

(We’ll see the Grossman (1972) model next time)
• **Income and income inequality?**

  - Cutler et al. refer to a literature that says *psychosocial stress*, from low status and control, is bad for health

  • In humans, this would look like civil servants; in apes, this would look like anyone other than the alpha male

  • There are biological channels of this badness: stress, cortisol, inflammation, wear-and-tear

  - But health and *income* probably cause each other

  - Economists like *education* causing health because it represents knowledge and protective “human capital” but it too can be reverse-caused by health
Conclusions of Cutler et al.

- There’s no consensus on many issues!
- Knowledge and technology probably have played key roles, at the dawn of the transition and still now
- Economists tend to downplay the role of income
- Markets for developing health technology seem like they’re functioning well
- But this may imply the rich will benefit the most at first