Univariate Analysis:
Comparing means, medians, and proportions

Dorry L. Segev, MD, PhD
Associate Professor of Surgery
Vice Chair for Research
Department of Surgery
Johns Hopkins University Medical School

American College of Surgeons
Outcomes Research Course, 2014
Univariate Analysis

Most statistical analyses in medicine involve comparisons between treatments, procedures, or patients

Univariate analysis compares two groups on a single dimension

- Baseline characteristics (e.g., mean age) in two groups
- Outcomes (e.g., mortality rates) in each group
- “Unadjusted” analysis
Hypothesis testing

- **Null hypothesis**: Patients given treatment A and treatment B have the same outcomes

- **Alternative hypothesis**: Patients given treatment A and treatment B have different outcomes

- **P-value**: Probability we could have obtained the observed data if the null hypothesis were true
Which test?

- Depends on the type of data
  - Continuous
    - Normal distribution
      - Comparing means
        - Students t-test
    - Skewed distribution
      - Comparing medians
        - Wilcoxon rank-sum test
  - Categorical
    - Chi square test
Who was “Student”?
(Maybe the only interesting story in statistics)

- W. S. Gossett, a statistician employed at the Guinness brewery
- Guinness did not allow employees to publish their research, Gossett's work on the t-test appears under the name "Student"
- The t-test was developed as a way of measuring how closely the yeast content of a particular batch of beer corresponded to the brewery's standard.
Hypothesis tests: The Test Statistic

- Behind the scenes math generates a “test statistic”
- Inversely related to the P-value – larger test statistics yield smaller p-values
- T-statistic, F-statistic, chi-square statistic

General Form = \frac{\text{Difference between groups}}{\text{Variability within groups}}
Hypothesis tests: The Test Statistic

- Variability small, large test statistic, small p-value, means different
- Variability large, small test statistic, large p-value, means not different
Comparing proportions in STATA

Modification of the tabulate command

P-value comparing the proportion in each category appears at the bottom of the table

Type the command:

```
tab died age80, chi
```

STATA output:

```
        Died during hospitalization |   Age greater than 80 years |
                          No |   Yes |       Total       |
-------------------------------|-----------------------------|
          Alive | 4,132 | 398 |      4,530       |
           Dead |   110 |  21 |       131       |
-------------------------------|-----------------------------|
            Total | 4,242 | 419 |      4,661       |

Pearson chi2(1) =  8.1677   Pr = 0.004
```
Comparing means in STATA

Compare mean ages for patients who died and those who lived

Mean age of 65.6 for those who lived and 71.2 for those who died

Middle p-value is the one of interest: probability that the difference between the means = 0
Comparing medians in STATA

Compare median length of stay for patients who died and those who lived

P-value: probability that the length of stay for those who lived and died is the same (difference = 0)

Type the command:

```
ranksum los, by(died)
```

STATA output:

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

<table>
<thead>
<tr>
<th></th>
<th>obs</th>
<th>rank sum</th>
<th>expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>died</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4530</td>
<td>10498664</td>
<td>10559430</td>
</tr>
<tr>
<td>1</td>
<td>131</td>
<td>366127</td>
<td>305361</td>
</tr>
<tr>
<td>combined</td>
<td>4661</td>
<td>10084791</td>
<td>10084791</td>
</tr>
</tbody>
</table>

unadjusted variance 2.305e+08
adjustment for ties -201239.7
adjusted variance 2.277e+08

H0: los(died==0) = los(died==1)

2 = -4.934

Prob > |z| = 0.0001
```

You may also hear this called the Mann-Whitney test.
Why do we need a different test for medians?

Non-parametric methods compare ranks (e.g., Wilcoxon rank-sum test) and are not dependent on the underlying distribution (“distribution free” methods)
BIOSTATISTICS LABORATORY PART 2:

UNIVARIATE ANALYSIS: COMPARING MEANS, MEDIANs, AND PROPORTIONS

Learning objectives:
1) Understand that the type of variable (continuous vs. dichotomous) determines which statistical test should be used.
2) Learn the commands for comparing proportions using the Chi-square test.
3) Learn the commands to compare two means with a t-test.
4) Learn the commands to compare two medians using the Wilcoxon rank-sum test.
Laboratory Part 2: Extra Stuff

Comparing more than two means:
Analysis of Variance

Comparing more than two medians:
Kruskal-Wallis Test
Introduction to Multivariate Regression & Interpreting STATA Output

Justin B. Dimick, MD, MPH
H.K. Ransom Professor of Surgery
Director, Center for Healthcare Outcomes & Policy
University of Michigan

American College of Surgeons
Outcomes Research Course, 2014
Multivariate analysis

• Widely used in health services research
  – **Observational studies**: Adjust for confounding
  – **Quality assessment**: Account for patient severity

• Theory complex, but essentials are straightforward

• We will focus on essentials and reading STATA output
Which test?

Depends on the type of data

Continuous

Dichotomous

Normal distribution

Linear regression

Skewed distribution

Log-transformation

Logistic regression

Linear regression
Linear regression

Dependent variable: Continuous

- Simple linear regression: one dependent and one independent variable
- Multiple linear regression (multivariate): multiple independent variables
- Independent variables may be any combination of continuous, dichotomous, or categorical
- Output: Coefficients and intercept
Equation for a straight line:
\[ y = mx + b \]

Linear regression:
\[ y_i = \beta_0 + \beta_1 x_i \]
Linear regression

Plot of weight vs. length

Model as a straight line

\[ Y = mx + b \]

Weight = Coefficient*Length + Constant
Linear regression

Weight = 33*Length - 3186 lbs
Log-Transformation: Length of Stay

Length of stay has a right-skewed distribution and must be log transformed to achieve a normal distribution.
Model the relationship between log(los) and old age (>80 years)?

Coefficient of 0.291

What does that mean?
Log-transformation:
How do you interpret the coefficient?

- Log length of stay has no real meaning
- Take the anti-log of the coefficient
- The antilog of the coefficient should be interpreted as the percent change—rather than the absolute change—associated with a change in one unit of the independent variable

Regression of log(length of stay) on age>80 years
Coefficient for an age > 80 years = 0.291
Antilog of 0.291 = 1.34

- Patients greater than 80 years old have length of stay 34% greater (1.34 times) than those younger than 80 years old
Logistic regression

Dependent variable: Dichotomous

- Simple logistic regression: one dependent and one independent variable
- Multiple logistic regression (multivariate): multiple independent variables
- Independent variables may also be any combination of continuous, dichotomous, or categorical
- Output: Odds Ratios
Logistic regression explained

How can you make a line out of a variable that is either “0” or “1”?  
Example:  
- **Outcome** = Death after surgery  
- **Exposure** = Severity of illness
What is the relationship?

Died after surgery

Patient illness severity
What is the relationship?

Probability of death after surgery

Patient illness severity

Pr(died) vs 10 quantiles of $x_{2_3}$
Logistic regression explained

• But probability still goes from 0 to 1?
• Need to transform probability into something that goes from \(-\infty\) to \(\infty\)…
• How about odds? \(\text{odds} = \frac{p}{1-p}\)
• But there is another problem:
  Odds have asymmetry, i.e., the odds of being 0 are not the inverse of the odds of being 1.

Example. Consider a group with 9 females and 1 male.
  Odds of being female = \(\frac{0.90}{0.10} = 9\) to one
  Odds of being male = \(\frac{0.10}{0.90} = 0.11\) to one
Logistic regression explained

Deal with asymmetry using natural log

\[ \ln(9) \text{ is } 2.217 \]
\[ \ln(1/9) \text{ is } -2.217 \]

Log(odds) or \( \log \frac{p}{1-p} \) = logit

Logistic regression: \( \logit(y_i) = \beta_0 + \beta_1 x_i \)
\[ \log[p_{died}/(1-p_{died})] = \beta_0 + \beta_1 \text{severity} \]

where “p_{died}” is the probability of death and “severity” is the severity of illness variable
Logistic Regression in STATA

What is the relationship between old age (>80 years) and mortality?

Odds Ratio of 1.98 for those older than age 80

Patients older than 80 are almost twice as likely to die after heart surgery

STATA Command:
```
logistic died age80
```
Not that easy…

- Next steps in logistic regression:
  - Which variables do you include?
    - Those significant in univariate analysis
    - Stepwise regression (backwards & forwards)
  - Test for interactions (first and second order)
  - Test for multicollinearity (correlated variables)
  - Model diagnostics
    - Calibration (Hosmer-Lemeshow test)
    - Discrimination (C-index, area under ROC curve)
Introduction to multilevel modeling

Justin B. Dimick, MD, MPH
H.K. Ransom Professor of Surgery
Director, Center for Healthcare Outcomes & Policy
University of Michigan

American College of Surgeons
2014 Outcomes Research Research Course
Overview

• Provide conceptual overview of multilevel modeling
• Review the applications of this technique that are useful in health services research
• Introduce the STATA exercise
Multilevel models

- Many synonyms: Hierarchical, empirical Bayes, random effects models

- Regression technique designed to fit models with multiple levels (hierarchical data structure)
  - Patient, surgeon, hospital, region

- Explicitly model the variation at each level
When should you use multilevel models?

When data is clustered, i.e., patients in hospitals
Why should you use multilevel models?

• So NEJM and JAMA to take your paper
• Ignoring clustering will bias both your coefficients (just wrong) and confidence intervals (too narrow)
• The statistical assumption that outcomes are independent across observations is violated
Two reasons to use multilevel models

1. Adjust for clustering
   - Coefficients are right
   - Standard errors (wider)

2. Several useful applications
   - Adjusting for reliability
     - Eliminate “noise” variation
   - Understanding variation
     - How much is explained?
     - How much is at each level?
Conceptual overview

Hospital costs

Hospital 1
Mean $15,000

Hospital 2
Mean $25,000
Average $20,000

Hospital effect
Random effects

- What is a random effect?
  - Hospital level residual
  - or, the difference between the overall mean and the hospitals mean

- You can think of the random effect as the variation at the hospital level not accounted for by observable variables
  - Patient- or hospital-level variables
**Equations:**

Linear model:

\[ y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \]

Multilevel linear model:

\[ y_{ij} = \beta_0 + \beta_1 x_{ij} + \mu_i + \varepsilon_{ij} \]

\( \mu_i \) is the random intercept and \( \sim N(0, \tau^2) \)

Multilevel logistic model:

\[ \logit(y_{ij}) = \beta_0 + \beta_1 x_{ij} + \mu_i + \varepsilon_{ij} \]
Understanding the math:

\[ y_{ij} = \beta_0 + \beta_1 x_{ij} + \mu_i + \varepsilon_{ij} \]
You can estimate random effects

- Random effects (hospital specific) are estimated using an empirical Bayes method
- New information considered in the context of prior expectations – e.g., the overall mean mortality rate
- Random effects, or deviations from the overall average, are “shrunk” back
- Degree of shrinkage is proportional to the reliability, a measure of precision
Empirical Bayes: Shrinkage factor

Reliability = \frac{\text{Between hospital variation}}{\text{Between hospital variation} + \text{within hospital variation}}

\text{Reliability} = \frac{\text{Variance (µi)}}{\text{Variance (µi) + Variance (εij)/n}}
Adjusting for reliability:
Profiling hospital performance

• Hospital “random effects” are ideally suited for profiling hospital quality
• The calculus of quality:
  \[ \text{Outcome} = \text{severity} + \text{quality} + \text{error} \]
• Patient covariates adjust for severity and the “shrinkage” adjusts for error
Ranking Hospitals on Surgical Mortality: The Importance of Reliability Adjustment

Justin B. Dimick, Douglas O. Staiger, and John D. Birkmeyer

Objective. We examined the implications of reliability adjustment on hospital mortality with surgery.

Data Source. We used national Medicare data (2003–2006) for three surgical procedures: coronary artery bypass grafting (CABG), abdominal aortic aneurysm (AAA) repair, and pancreatic resection.

Study Design. We conducted an observational study to evaluate the impact of reliability adjustment on hospital mortality rankings. Using hierarchical modeling, we adjusted hospital mortality for reliability using empirical Bayes techniques. We assessed the implication of this adjustment on the apparent variation across hospitals and the ability of historical hospital mortality rates (2003–2004) to forecast future mortality (2005–2006).

Principal Findings. The net effect of reliability adjustment was to greatly diminish apparent variation for all three operations. Reliability adjustment was also particularly important for identifying hospitals with the lowest future mortality. Without reliability adjustment, hospitals in the “best” quintile (2003–2004) with pancreatic resection had a mortality of 7.6 percent in 2005–2006; with reliability adjustment, the “best” hospital quintile had a mortality of 2.7 percent in 2005–2006. For AAA repair, reliability adjustment also improved the ability to identify hospitals with lower future mortality. For CABG, the benefits of reliability adjustment were limited to the lowest volume hospitals.

Conclusion. Reliability adjustment results in more stable estimates of mortality that better forecast future performance. This statistical technique is crucial for helping patients select the best hospitals for specific procedures, particularly uncommon ones, and should be used for public reporting of hospital mortality.

Key Words. Quality, surgery, hospital, mortality, hierarchical
Impact of reliability adjustment on hospital mortality rates

Risk-adjusted mortality rates (%)

Pancreatic Cancer Resection

After adjusting for reliability

Mortality rates after filtering out the statistical noise

Observed mortality rates

Dimick et al. Health Services Research 2010
Partitioning variation

• Understanding the proportion of variation at each level of nesting
  – Patients, surgeons, hospital, regions
• Explicitly calculated in hierarchical models
• Can help you understand the appropriate level to target interventions
Understanding Variation in the Adoption of a New Technology in Surgery

John M. Hollingsworth, MD, MS,* Sarah L. Krein, PhD, RN,†‡ Rodney L. Dunn, MS,*
J. Stuart Wolf, Jr., MD,* and Brent K. Hollenbeck, MD, MS*§

**Background:** Despite data supporting the superiority of laparoscopy over conventional surgery for donor nephrectomy, prior work reveals that this technology’s adoption has been slow. However, the underlying cause for its gradual adoption remains unclear. For this reason, a multilevel analysis was conducted using a population-based cohort to evaluate the amount of variation in laparoscopic donor nephrectomy use attributable to the individual, provider, and hospital levels.

**Methods:** Using the Nationwide Inpatient Sample (2000–2003), discharges for donor nephrectomies were identified, distinguishing between those surgeries performed by laparoscopic versus conventional approaches. To examine variation in laparoscopy use, we fitted generalized linear mixed models with random intercepts. The principal model had no explanatory variables included and allowed for an understanding of the basic partitioning of the data’s variability. Patient-, provider-, and hospital-level factors were then added to the model to determine how much variability at each level might be explained by these covariates. For some indications, laparoscopy affords considerable improvements in short-term convalescence and offers similar efficacy when compared with its conventional (ie, through large incisions) counterpart. Well into the third decade of the laparoscopic revolution, many of the structural barriers that may have initially precluded the adoption of laparoscopy have abated, as evidenced by its extensive use for cholecystectomy and fundoplication.

However, the diffusion of laparoscopy for some surgical approaches has been comparatively gradual, including for donor nephrectomy. Despite several observational and experimental studies demonstrating similar patient safety and renal graft function and superior quality of life and patient convalescence among those treated laparoscopically, national trends indicate that the majority of kidney donors underwent a conventional procedure in 2003 more than a decade after the laparoscopic nephrec-
TABLE 3. Percent of Residual Variance Attributable to the Individual, Provider, and Hospital Levels of Care

<table>
<thead>
<tr>
<th>Levels of Care (%)</th>
<th>Individual</th>
<th>Provider</th>
<th>Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empty model</td>
<td>37.2</td>
<td>27.6</td>
<td>35.2</td>
</tr>
<tr>
<td>After including operative year</td>
<td>39.9</td>
<td>26.6</td>
<td>33.5</td>
</tr>
<tr>
<td>After including operative year and patient characteristics</td>
<td>39.5</td>
<td>27.1</td>
<td>33.4</td>
</tr>
<tr>
<td>After including operative year and patient and surgeon characteristics</td>
<td>43.5</td>
<td>25.3</td>
<td>31.2</td>
</tr>
<tr>
<td>After including operative year and patient, surgeon, and hospital characteristics</td>
<td>45.9</td>
<td>27.3</td>
<td>26.8</td>
</tr>
</tbody>
</table>

CONCLUSIONS

These data reveal that the provider and hospital, combined, contribute to the majority of variation in laparoscopy use among kidney donors. For this reason, interventions focused at these levels would most likely hasten laparoscopy’s uptake in the community. Further, this type of multilevel analysis may prove useful for understanding those barriers and facilitators to the adoption of other surgical technologies.
Understanding variation: Evaluating hospital quality indicators

- Random effects models explicitly estimate the (non-random) variation among hospitals
- The change in the hospital-level variance after entering a quality measure can be used to compare quality indicators, i.e., the decline in variance of the random effect after adding the quality measure

\[
\text{Proportion of variance explained} = \frac{\text{Variance (without) } - \text{Variance (with)}}{\text{Variance (without)}}
\]
Hospital Complication Rates With Bariatric Surgery in Michigan

Nancy J. O. Birkmeyer, PhD
Justin B. Dimick, MD, MPH
David Share, MD, MPH
Abdelkader Hawasli, MD
Wayne J. English, MD
Jeffrey Genaw, MD
Jonathan F. Finks, MD
Arthur M. Carlin, MD
John D. Birkmeyer, MD
for the Michigan Bariatric Surgery Collaborative

Context Despite the growing popularity of bariatric surgery, there remain concerns about perioperative safety and variation in outcomes across hospitals.

Objective To assess complication rates of different bariatric procedures and variability in rates of serious complications across hospitals and according to procedure volume and center of excellence (COE) status.

Design, Setting, and Patients Involving 25 hospitals and 62 surgeons statewide, the Michigan Bariatric Surgery Collaborative (MBSC) administers an externally audited, prospective clinical registry. We evaluated short-term morbidity in 15,275 Michigan patients undergoing 1 of 3 common bariatric procedures between 2006 and 2009. We used multilevel regression models to assess variation in risk-adjusted complication rates across hospitals and the effects of procedure volume and COE designation (by the American College of Surgeons or American Society for Metabolic and Bariatric Surgery) status.

Main Outcome Measure Complications occurring within 30 days of surgery.

Results Overall, 7.3% of patients experienced perioperative complications, most of which were wound problems and other minor complications. Serious complications were most common after gastric bypass (3.6%; 95% confidence interval [CI], 3.2%-4.0%), followed by sleeve gastrectomy (2.2%; 95% CI, 1.2%-3.2%), and laparoscopic adjustable gastric band (0.9%; 95% CI, 0.6%-1.1%) procedures (P<.001). Mortality occurred in 0.04% (95% CI, 0.001%-0.13%) of laparoscopic adjustable gastric band, 0 sleeve gastrectomy, and 0.14% (95% CI, 0.08%-0.25%) of the gastric bypass patients. After adjustment for
# Results of Multilevel Models

## Table 4. Results of Model Fitting

<table>
<thead>
<tr>
<th>Level</th>
<th>Variable</th>
<th>Coefficient (P Value)</th>
<th>Empty Model</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient level</td>
<td>Predicted risk&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.02 (&lt;.001)</td>
<td>1.04 (&lt;.001)</td>
<td>1.03 (&lt;.001)</td>
<td>1.02 (&lt;.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgeon level</td>
<td>Average annual surgeon volume</td>
<td>−0.299 (.003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Center level</td>
<td>Average annual hospital volume</td>
<td>0.260 (.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COE status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.166 (.50)</td>
</tr>
<tr>
<td>Model information criterion&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Akaike information criterion&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3616</td>
<td>3497</td>
<td>3492</td>
<td>3495</td>
<td>3499</td>
<td></td>
</tr>
<tr>
<td>Covariance parameter</td>
<td>Standard deviation of intercepts</td>
<td>0.3895</td>
<td>0.3438</td>
<td>0.1942</td>
<td>0.2501</td>
<td>0.3371</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reduction in between center</td>
<td>22</td>
<td>75</td>
<td>59</td>
<td>25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Based on a multivariate logistic regression model including all significant patient risk factors for serious complications (age, body mass index [calculated as weight in kilograms divided by height in meters squared], male sex, mobility limitations, prior history of venous thromboembolism, and total number of comorbid conditions) and procedure type (laparoscopic adjustable gastric band, sleeve gastrectomy, or gastric bypass).

<sup>b</sup>Akaike information criterion is defined as minus twice log likelihood plus 2 degrees of freedom.

Understanding variation: Serious morbidity with bariatric surgery

<table>
<thead>
<tr>
<th>Variables</th>
<th>Proportion of hospital-level variation explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient risk factors</td>
<td>22%</td>
</tr>
<tr>
<td>Hospital volume</td>
<td>75%</td>
</tr>
<tr>
<td>Surgeon volume</td>
<td>59%</td>
</tr>
<tr>
<td>Center of excellence status</td>
<td>25%</td>
</tr>
</tbody>
</table>
Learning objectives:
To create multilevel models in STATA and then evaluate the usefulness of a random effects model to determine how much hospital-level variation in outcomes after cardiac surgery is explained by patient risk factors.
STATA lab

• Work your way through the handout, which includes commands and output

• You will need a new Maryland CABG dataset that has a few additional variables (hospital level)

  Maryland.CABG.2001_hospital.dta
Advanced Statistical Modeling: Propensity Scores

Dorry L. Segev, MD, PhD
Associate Professor of Surgery
Vice Chair of Research
Department of Surgery
Johns Hopkins University Medical School

American College of Surgeons
Outcomes Research Course, 2014
Introduction

• RCT’s the gold standard for assessing treatment effects
  • Control of both known and unknown confounding variables
  • Not always feasible (ethics and cost)

• Traditional methods for controlling for confounding in observational studies
  • Multivariate analysis
  • Matching
Propensity Score: Definition

• A measure of likelihood that a person would have been treated based on their covariates

• Uses the probability that subject would have been treated to adjust the estimate of treatment effect to create a “quasi-experiment”

• Reduces the entire collection of observed covariates to a single composite variable
Propensity Score Calculation and Analytic Approaches

1. Fit logistic regression model where the dependent variable is the treatment

2. Predict each patient’s probability (propensity) for treatment

3. Analytic approaches
   - Multivariate modeling
   - Matching
   - Stratification
Why use propensity scores?

• Most studies have found that multivariate adjustment and propensity scores are equivalent.

• The one exception is the case of rare outcome variables where a multivariate model with many covariates cannot be estimated.
Preoperative Placement of Inferior Vena Cava Filters and Outcomes After Gastric Bypass Surgery
Example: IVC Filters and Outcomes after Gastric Bypass Surgery

- **Study population**: 6,376 gastric bypass patients 2006-8
- **Comparison groups**: IVC filter (8.5%) vs no IVC filter (91.5%)
- **Outcomes**: Post-op VTE, serious complications, death/permanent disability
- **Analytic method**: Propensity score stratification
Step 1. Fit a logistic model with IVC filter as the outcome variable and all covariates associated with treatment with an IVC filter as independent variables.
Example Figure. Mean propensity score within strata

Step 2a. Estimate propensity score
Step 2b. Stratify on propensity score
Step 2c. Test balancing property to ensure mean propensity score is not different for treated and control patients in each block
Example Step 3. Use fixed effects regression to calculate a summary measure of the treatment effect within blocks

```
xtdlogit sercomp ivcfilter, fe or nolog
note: multiple positive outcomes within groups encountered.
note: 1 group (17 obs) dropped because of all positive or all negative outcomes.

Conditional fixed-effects logistic regression
Group variable: mbid

Number of obs  =  6359
Number of groups =  12

Obs per group: min =   49
               avg =  529.9
               max =  2090

LR chi2(1)     =  2.27
Prob > chi2    =  0.1319

Log likelihood  =  -1012.1127

|             | OR  | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|-------------|-----|-----------|-----|------|----------------------|
| ivcfilter   | 1.402053 | .3091726   | 1.53 | 0.125 | .9100442  2.160063  |
```
Example Figure. Propensity adjusted rates of complications in patients with and without IVC filters.
Example Figure. Propensity adjusted rates of complications in hospitals with high and low rates of use of IVC filters.
Propensity Score Matching Methods

- **Nearest Neighbor Matching**: randomly select a treated subject and find a control with closest PS.

- **Caliper Matching**: randomly select a treated subject and randomly find a control within a predefined common support region (ex: $\frac{1}{4}$ of SE of estimated logit PS)

- **Mahalanobis Distance Matching**: Distance is determined by PS and selected individual covariates. Find the closest match.
Example

Table 1. Patient characteristics in the propensity-matched and unmatched groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>IVC Filter</th>
<th>Matched Controls</th>
<th>p-value</th>
<th>Unmatched Controls</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>1,077</td>
<td>1,077</td>
<td></td>
<td>33,323</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>48</td>
<td>49</td>
<td>0.295</td>
<td>46</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Body mass index</td>
<td>58</td>
<td>57</td>
<td>0.061</td>
<td>47</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Male gender</td>
<td>32</td>
<td>31</td>
<td>0.546</td>
<td>21</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Black race</td>
<td>27</td>
<td>25</td>
<td>0.667</td>
<td>15</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Private Insurance</td>
<td>62</td>
<td>64</td>
<td>0.305</td>
<td>74</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Smoking in past year</td>
<td>2</td>
<td>2</td>
<td>0.883</td>
<td>2</td>
<td>0.440</td>
</tr>
<tr>
<td>Mobility Limitations</td>
<td>18</td>
<td>18</td>
<td>0.780</td>
<td>5</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Lung Disease</td>
<td>43</td>
<td>43</td>
<td>1.000</td>
<td>25</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Cardiovascular Disease</td>
<td>21</td>
<td>21</td>
<td>0.874</td>
<td>10</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Hypertension</td>
<td>72</td>
<td>72</td>
<td>0.737</td>
<td>53</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>59</td>
<td>59</td>
<td>0.930</td>
<td>50</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>GERD</td>
<td>50</td>
<td>52</td>
<td>0.490</td>
<td>49</td>
<td>0.417</td>
</tr>
<tr>
<td>Peptic Ulcer Disease</td>
<td>5</td>
<td>4</td>
<td>0.228</td>
<td>3</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Cholelithiasis</td>
<td>30</td>
<td>30</td>
<td>0.963</td>
<td>27</td>
<td>0.018</td>
</tr>
<tr>
<td>Urinary Incontinence</td>
<td>25</td>
<td>25</td>
<td>0.960</td>
<td>22</td>
<td>0.029</td>
</tr>
<tr>
<td>Renal Failure</td>
<td>0.4</td>
<td>0.6</td>
<td>0.526</td>
<td>0.2</td>
<td>0.298</td>
</tr>
<tr>
<td>Diabetes</td>
<td>46</td>
<td>48</td>
<td>0.546</td>
<td>33</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Liver Disorder</td>
<td>4</td>
<td>4</td>
<td>0.584</td>
<td>5</td>
<td>0.184</td>
</tr>
<tr>
<td>Prior history of VTE</td>
<td>39</td>
<td>39</td>
<td>0.965</td>
<td>2</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Sleep Apnea</td>
<td>70</td>
<td>68</td>
<td>0.209</td>
<td>43</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Musculoskeletal Disorder</td>
<td>78</td>
<td>80</td>
<td>0.221</td>
<td>77</td>
<td>0.189</td>
</tr>
<tr>
<td>History of Hernia Repair</td>
<td>5</td>
<td>6</td>
<td>0.924</td>
<td>3</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Psychological Disorder</td>
<td>49</td>
<td>49</td>
<td>0.796</td>
<td>47</td>
<td>0.267</td>
</tr>
<tr>
<td>Total Comorbidities</td>
<td>6</td>
<td>6</td>
<td>0.922</td>
<td>4</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Step 1. Fit a logistic model with IVC filter as the outcome variable and all covariates associated with treatment with an IVC filter as independent variables.
<table>
<thead>
<tr>
<th>Variable</th>
<th>IVC Filter</th>
<th>Matched Controls</th>
<th>p-value</th>
<th>Unmatched Controls</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>1,077</td>
<td>1,077</td>
<td></td>
<td>33,323</td>
<td></td>
</tr>
<tr>
<td>Gastric bypass</td>
<td>73</td>
<td>69</td>
<td></td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Duodenal switch</td>
<td>0.7</td>
<td>0.8</td>
<td></td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Procedure Length (minutes)</td>
<td>114</td>
<td>116</td>
<td>0.427</td>
<td>95</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Medical VTE prophylaxis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-operative heparin:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfractionated</td>
<td>36</td>
<td>38</td>
<td>0.246</td>
<td>34</td>
<td>0.306</td>
</tr>
<tr>
<td>Low molecular weight</td>
<td>60</td>
<td>54</td>
<td>0.017</td>
<td>53</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Post-operative heparin:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfractionated</td>
<td>7</td>
<td>10</td>
<td>0.023</td>
<td>19</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Low molecular weight</td>
<td>70</td>
<td>68</td>
<td>0.326</td>
<td>64</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Post-discharge heparin:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low molecular weight</td>
<td>72</td>
<td>66</td>
<td>0.003</td>
<td>16</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
# Race and Outcomes of Bariatric Surgery

## Table 1. Baseline demographic and clinical characteristics in the matched study cohorts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unmatched</th>
<th></th>
<th>Matched</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Black</td>
<td>Black</td>
<td>p-value</td>
<td>Non-Black</td>
</tr>
<tr>
<td>n</td>
<td>18,271</td>
<td>3,407</td>
<td></td>
<td>2,095</td>
</tr>
<tr>
<td>Age</td>
<td>47</td>
<td>43</td>
<td>&lt;0.0001</td>
<td>44</td>
</tr>
<tr>
<td>Body mass index (start)</td>
<td>48</td>
<td>50</td>
<td>&lt;0.0001</td>
<td>49</td>
</tr>
<tr>
<td>Body mass index (pre-op)</td>
<td>47</td>
<td>49</td>
<td>&lt;0.0001</td>
<td>48</td>
</tr>
<tr>
<td>Male</td>
<td>23</td>
<td>15</td>
<td>&lt;0.0001</td>
<td>17</td>
</tr>
<tr>
<td>Private insurance</td>
<td>73</td>
<td>77</td>
<td>&lt;0.0001</td>
<td>79</td>
</tr>
<tr>
<td>Smoking</td>
<td>42</td>
<td>27</td>
<td>&lt;0.0001</td>
<td>33</td>
</tr>
<tr>
<td>Mobility Limitations</td>
<td>6</td>
<td>6</td>
<td>0.379</td>
<td>6</td>
</tr>
<tr>
<td>Lung Disease</td>
<td>26</td>
<td>29</td>
<td>&lt;0.0001</td>
<td>28</td>
</tr>
<tr>
<td>Cardiovascular Disease</td>
<td>55</td>
<td>59</td>
<td>&lt;0.0001</td>
<td>58</td>
</tr>
<tr>
<td>Hypertension</td>
<td>53</td>
<td>58</td>
<td>&lt;0.0001</td>
<td>56</td>
</tr>
<tr>
<td># meds</td>
<td>1</td>
<td>2</td>
<td>&lt;0.0001</td>
<td>2</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>53</td>
<td>42</td>
<td>&lt;0.0001</td>
<td>45</td>
</tr>
<tr>
<td>Gastroesophageal reflux</td>
<td>50</td>
<td>44</td>
<td>&lt;0.0001</td>
<td>46</td>
</tr>
<tr>
<td>Peptic ulcer disease</td>
<td>3</td>
<td>4</td>
<td>0.021</td>
<td>3</td>
</tr>
<tr>
<td>Cholelithiasis</td>
<td>28</td>
<td>17</td>
<td>&lt;0.0001</td>
<td>21</td>
</tr>
<tr>
<td>Urinary incontinence</td>
<td>26</td>
<td>18</td>
<td>&lt;0.0001</td>
<td>21</td>
</tr>
</tbody>
</table>
## Race and Outcomes of Bariatric Surgery

### Table 1. Baseline demographic and clinical characteristics in the matched study cohorts, continued.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unmatched</th>
<th>Matched</th>
<th>p-value</th>
<th>Unmatched</th>
<th>Matched</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Non-Black</td>
<td>Black</td>
<td></td>
<td>Non-Black</td>
<td>Black</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18,271</td>
<td>3,407</td>
<td></td>
<td>2,095</td>
<td>2,095</td>
<td></td>
</tr>
<tr>
<td>Renal failure</td>
<td>0.2</td>
<td>0.7</td>
<td>&lt;0.0001</td>
<td>0.1</td>
<td>0.4</td>
<td>0.035</td>
</tr>
<tr>
<td>Non-insulin dependent diabetes</td>
<td>30</td>
<td>27</td>
<td>0.003</td>
<td>29</td>
<td>28</td>
<td>0.391</td>
</tr>
<tr>
<td>Insulin-dependent diabetes</td>
<td>11</td>
<td>10</td>
<td>0.017</td>
<td>10</td>
<td>9</td>
<td>0.753</td>
</tr>
<tr>
<td>Liver disorder</td>
<td>5</td>
<td>2</td>
<td>&lt;0.0001</td>
<td>3</td>
<td>2</td>
<td>0.302</td>
</tr>
<tr>
<td>Venous thromboembolism</td>
<td>4</td>
<td>5</td>
<td>0.031</td>
<td>5</td>
<td>4</td>
<td>0.286</td>
</tr>
<tr>
<td>Sleep apnea</td>
<td>46</td>
<td>43</td>
<td>0.001</td>
<td>43</td>
<td>42</td>
<td>0.333</td>
</tr>
<tr>
<td>Musculoskeletal disorder</td>
<td>77</td>
<td>76</td>
<td>0.283</td>
<td>77</td>
<td>76</td>
<td>0.283</td>
</tr>
<tr>
<td>History of hernia repair</td>
<td>3</td>
<td>2</td>
<td>0.093</td>
<td>2</td>
<td>2</td>
<td>0.670</td>
</tr>
<tr>
<td>Psychological disorder</td>
<td>50</td>
<td>33</td>
<td>&lt;0.0001</td>
<td>40</td>
<td>39</td>
<td>0.613</td>
</tr>
<tr>
<td>Total # comorbidities</td>
<td>5</td>
<td>4</td>
<td>&lt;0.0001</td>
<td>4</td>
<td>4</td>
<td>0.578</td>
</tr>
<tr>
<td>Procedure type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAGB</td>
<td>24</td>
<td>23</td>
<td>0.004</td>
<td>27</td>
<td>25</td>
<td>0.018</td>
</tr>
<tr>
<td>SG</td>
<td>22</td>
<td>24</td>
<td></td>
<td>22</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>LRYGB</td>
<td>52</td>
<td>51</td>
<td></td>
<td>49</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>ORYGB</td>
<td>1.5</td>
<td>1.7</td>
<td></td>
<td>1.3</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>BPD</td>
<td>1.2</td>
<td>0.6</td>
<td></td>
<td>1.0</td>
<td>0.6</td>
<td></td>
</tr>
</tbody>
</table>
### Table 2. Baseline SES in the matched study cohorts

<table>
<thead>
<tr>
<th>SES Variable</th>
<th>Unmatched</th>
<th>Matched</th>
<th>p-value</th>
<th>Unmatched</th>
<th>Matched</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marital status (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>69</td>
<td>42</td>
<td>&lt;0.0001</td>
<td>54</td>
<td>52</td>
<td>0.078</td>
</tr>
<tr>
<td>Widowed/Divorced</td>
<td>16</td>
<td>23</td>
<td>&lt;0.0001</td>
<td>18</td>
<td>24</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Never married</td>
<td>15</td>
<td>35</td>
<td>&lt;0.0001</td>
<td>28</td>
<td>24</td>
<td>0.024</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;High School</td>
<td>4</td>
<td>5</td>
<td>0.262</td>
<td>3</td>
<td>5</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>High School</td>
<td>21</td>
<td>14</td>
<td>&lt;0.0001</td>
<td>19</td>
<td>14</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>College</td>
<td>64</td>
<td>69</td>
<td>&lt;0.0001</td>
<td>68</td>
<td>68</td>
<td>0.766</td>
</tr>
<tr>
<td>Graduate School</td>
<td>10</td>
<td>12</td>
<td>&lt;0.0001</td>
<td>11</td>
<td>13</td>
<td>0.199</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working</td>
<td>65</td>
<td>65</td>
<td>0.966</td>
<td>66</td>
<td>66</td>
<td>0.770</td>
</tr>
<tr>
<td>Unemployed</td>
<td>5</td>
<td>5</td>
<td>0.116</td>
<td>5</td>
<td>5</td>
<td>0.424</td>
</tr>
<tr>
<td>Homemaker/Student</td>
<td>8</td>
<td>6</td>
<td>0.001</td>
<td>7</td>
<td>6</td>
<td>0.226</td>
</tr>
<tr>
<td>Retired</td>
<td>10</td>
<td>6</td>
<td>&lt;0.0001</td>
<td>10</td>
<td>7</td>
<td>0.004</td>
</tr>
<tr>
<td>Disabled</td>
<td>11</td>
<td>16</td>
<td>&lt;0.0001</td>
<td>12</td>
<td>15</td>
<td>0.015</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>2</td>
<td>0.761</td>
<td>1</td>
<td>1</td>
<td>0.689</td>
</tr>
</tbody>
</table>
### Race and Outcomes of Bariatric Surgery

#### Table 2. Baseline SES in the matched study cohorts, continued

<table>
<thead>
<tr>
<th>SES Variable</th>
<th>Unmatched</th>
<th>Matched</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Income</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;$10,000</td>
<td>6</td>
<td>12</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$10,000-$24,999</td>
<td>14</td>
<td>18</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$25,000-$44,999</td>
<td>24</td>
<td>30</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$45,000-$75,000</td>
<td>29</td>
<td>26</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>$75,000</td>
<td>27</td>
<td>15</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Zip Code Level SES</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>30</td>
<td>50</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Medium</td>
<td>34</td>
<td>23</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>High</td>
<td>35</td>
<td>26</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Propensity Scores Lab

• Learning Objectives
  • To learn commands for creating and analyzing propensity scores in STATA and then to use them to evaluate the outcomes of laparoscopic and open appendectomy.

• Instructions
  • Work your way through the handout, which includes commands and output
  • Need to download the pscore and psmatch2 commands from the STATA website
  • You will need a the NSQIP dataset: ACS.NSQIP.Appy.dta