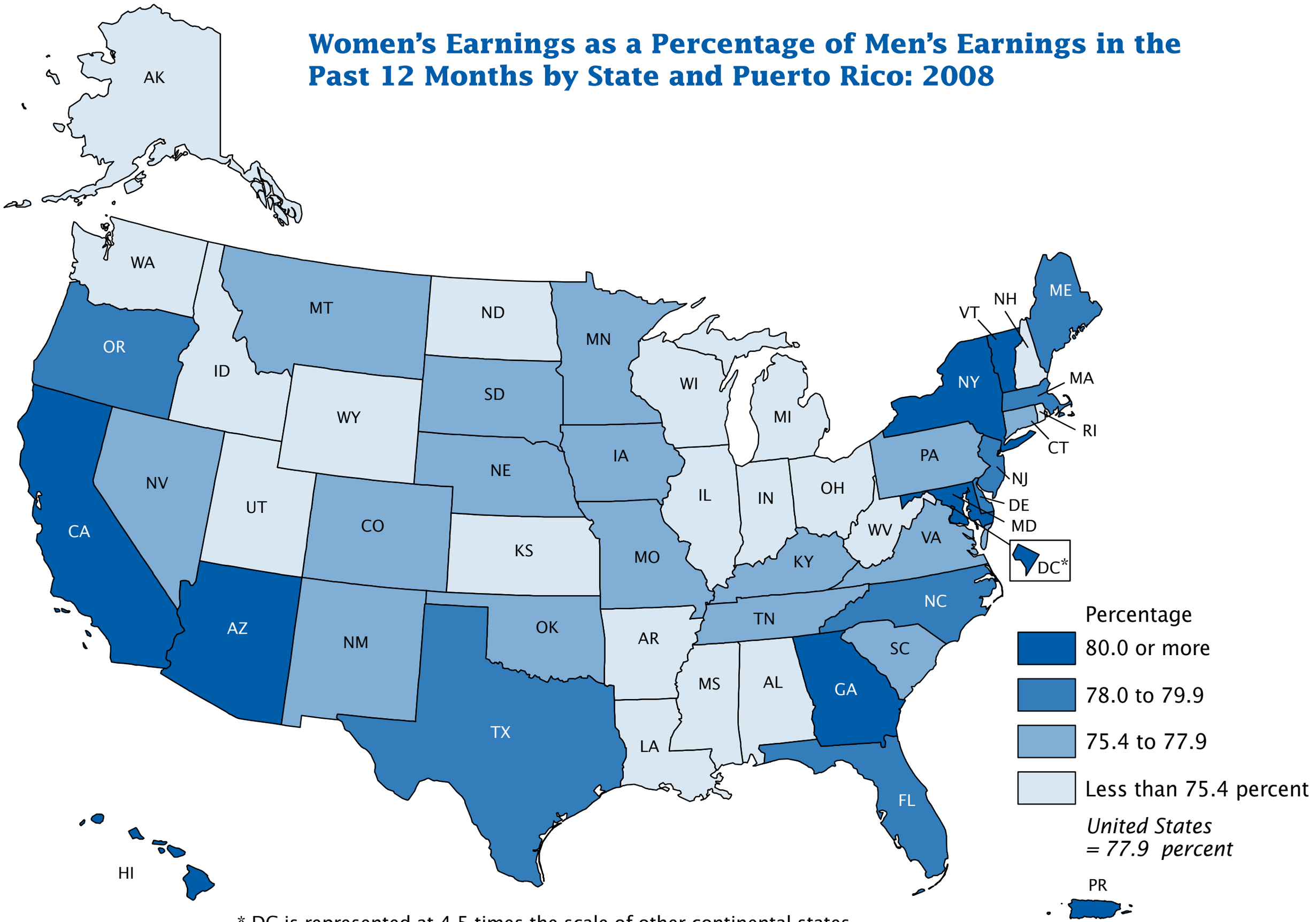
Using Randomization Tests and IPUMS-USA* to Investigate the Gender Wage Gap

Laura Schultz, Rowan University

Historically, women have often earned less than men for performing the same job, a phenomenon known as the gender wage gap. Given that salary data are notoriously skewed, investigating the gender wage gap provides an engaging context for introducing students to modern randomization tests as an alternative to more traditional nonparametric tests. The first randomization test included in this classroom activity addresses whether the mean salary of male accountants and auditors is significantly greater than that of females with similar qualifications, and the second task investigates whether the ratio of the median salary of female accountants/auditors to the median salary of their male counterparts is significantly less than 1. (A ratio of 1 would indicate that the median female and male salaries are the same.) Both randomization tests utilize custom JMP scripts that I have written. The data set, a subset of a sample collected as part of the 2008 American Communities Survey that I downloaded from IPUMS-USA, consists of the total personal earnings for samples of full-time female and male accountants/auditors ages 25-34 with Bachelor's degrees in the Philadelphia metropolitan area. I chose this particular data set due to its appeal to the accounting and finance majors I teach at a university in the greater Philadelphia region. Given the flexibility of the IPUMS-USA interface, this classroom activity could be adapted to investigate the gender wage gap using samples from other professions, age groups, or geographical regions.

*S. Ruggles, J.T. Alexander, K. Genadek, R. Goeken, M.B. Schroeder, and M. Sobek. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

Women's Earnings as a Percentage of Men's Earnings in the Past 12 Months by State and Puerto Rico: 2008



* DC is represented at 4.5 times the scale of other continental states.

Sources: U.S. Census Bureau, American Community Survey, 2008; and Puerto Rico Community Survey, 2008.

IPUMS USA

http://usa.ipums.org/usa/index.shtml

Google

MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA

IPUMS USA

Home | Select Data | FAQ | Contact Us | Login | Data Cart

DATA

Browse and Select Data
Download or Revise Extracts
Analyze Data Online
IPUMS Registration

DOCUMENTATION

FAQ
User's Guide
Variables
Samples

RESOURCES

Enumeration Forms
Published Census Volumes
Errata and Revisions

RESEARCH

Citation and Use
Bibliography
Related Sites

CONTACT US

Technical Support
IPUMS Staff
How to Help

Integrated Public Use Microdata Series

census microdata for social and economic research

IPUMS-USA is a project dedicated to collecting and distributing United States census data. Its goals are to:

- Collect and preserve data and documentation
- Harmonize data
- Disseminate the data absolutely free!

Use it for GOOD -- never for EVIL

IPUMS-USA NEWS

New! Cluster and Strata Variables available 1960-2010
Revised 1980-2010 data
NYT Interactive Map
2006-2010 ACS/PRCS data
Migration Revisions
Data release schedule

OTHER MPC PROJECTS

Linked Representative Samples
ATUS-X
IPUMS-International
IPUMS-CPS
NAPP
NHGIS
IHIS

SDA [Use classic interface] Selected Study: 2008 ACS sample

Analysis | Create Variables | Codebook | Getting Started

Frequencies or crosstabulation | Comparison of means | Correlation matrix | Comparison of correlations | Multiple regression | Logit/Probit regression | List values of individual cases

Variable Selection: Help

Selected: pwmetro View

Copy to: Vars to List Filter

- United States 2008
  - Household variables
  - Person variables
    - Technical
    - Family Interrelationship
    - Demographic
    - Race, Ethnicity, and Nativity
    - Health Insurance
    - Education
    - Work
    - Income
    - Occupational Standing
    - Migration
    - Disability
    - Veteran Status
    - Place of Work and Travel Time
      - pwstate2 - Place of work: state, 1980 onward
      - pwmetro - Place of work: metropolitan area
      - pwcity - Place of work: city

SDA Listcase Program

Help: General / Percents

Variables to List: (You can tab from one input box to the next)

1:	age	2:	sex	3:	race	4:	educd
5:	occ	6:	incearn	7:		8:	
9:		10:		11:		12:	
13:		14:		15:		16:	

Selection Filter(s): pwmetro(6160), educd(101), occ(80 *Example: age(18-50)*

Summaries of each variable listed (valid codes only)

☐ Sum ☐ Mean ☐ Minimum ☐ Maximum

How to display variables:

○ Code value ◉ Category label (if any) ○ Both

☑ Color coding ☐ Question text

Title: Total Personal Income for Full-Time Accountants/Auditors Age 25-3

Start Listing Clear Fields

Total Personal Income for Full-Time Accountants/Auditors Age 25-34 with Bachelor's Degrees Working in the Philadelphia Metropolitan Area, 2008 ACS

SDA 3.5: Listcase

United States 2008

Mar 22, 2012 (Thu 08:05 AM CDT)

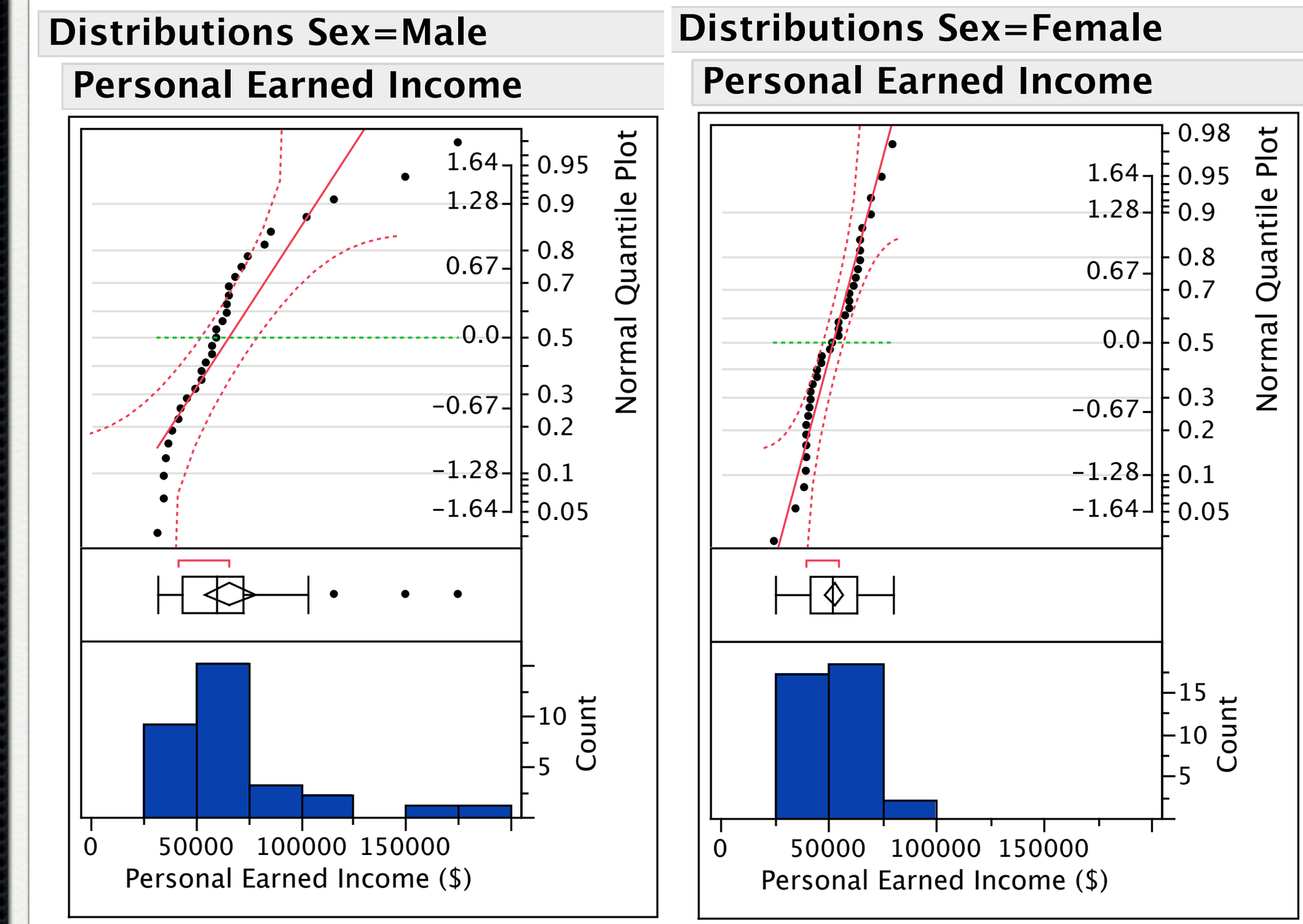
Variables

Role	Name	Label	Range	MD	Dataset
Variable	**age**	Age	0-95		1
Variable	**sex**	Sex	1-2		1
Variable	**race**	Race	1-9		1
Variable	**educd**	Educational attainment	1-116		1
Variable	**occ**	Occupation	0-9920		1
Variable	**incearn**	Total personal earned income	-9999-1251000		1
Filter	**pwmetro(6160)**	Place of work: metropolitan area(=Philadelphia, PA/NJ)	0-9360		1
Filter	**educd(101)**	Educational attainment(=Bachelor's degree)	1-116		1
Filter	**occ(800)**	Occupation	0-9920		1
Filter	**empstat(1)**	Employment status(=Employed)	0-3		1
Filter	**age(25-34)**	Age	0-95		1
Filter	**uhrswork(40-**)**	Usual hours worked per week	0-99		1

Cases listed: 68

#	age	sex	race	educd	occ	incearn
1	34	Male	White	Bachelor's degree	800	116,000

The Salary Data



The Randomization Test Statistics

- Difference between two sample means

$$\left(\overline{x}_{Male} - \overline{x}_{Female} = \$13482\right)$$

- Ratio of two sample medians

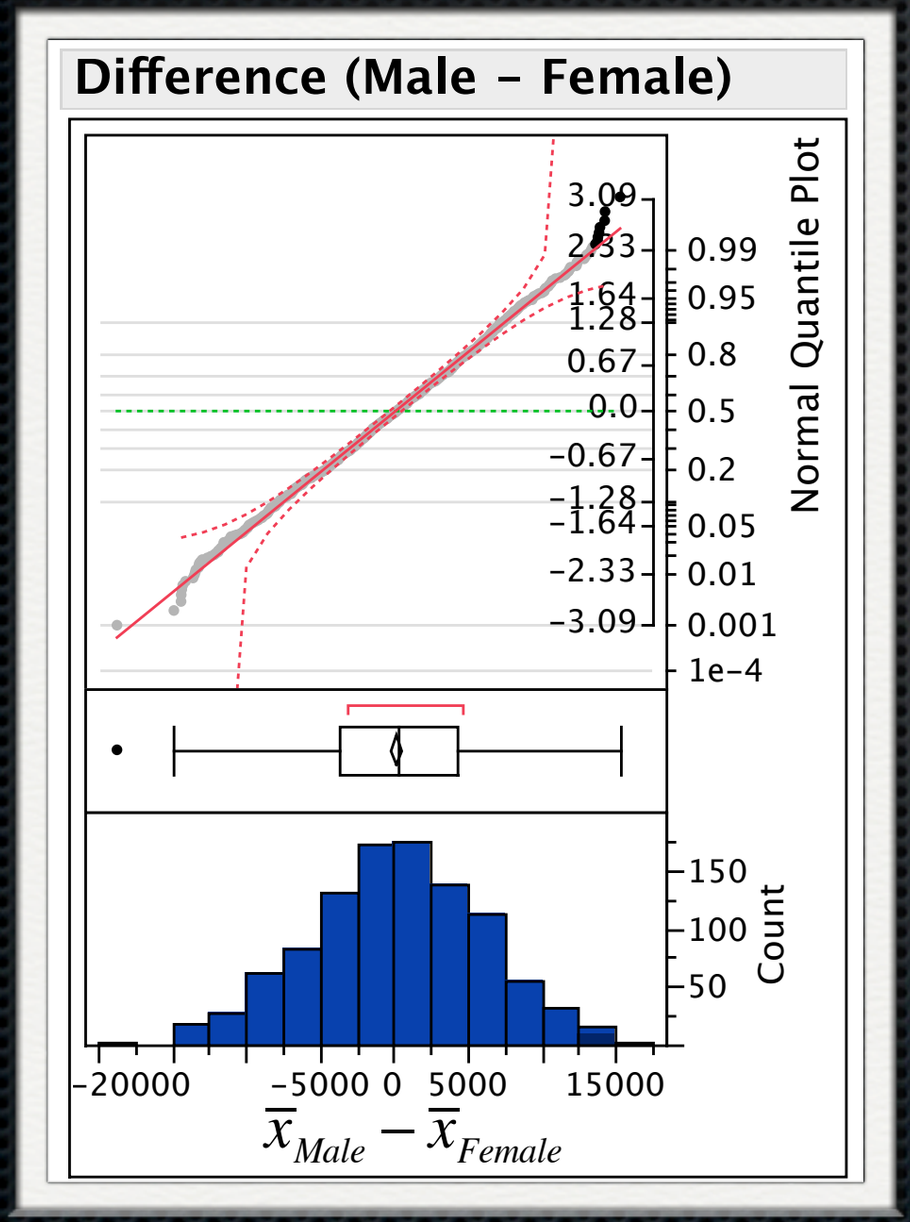
$$\left(\frac{\tilde{x}_{Female}}{\tilde{x}_{Male}} = 0.867\right)$$

The Randomization Procedure

- Compute the test statistic for the original data.
- Scramble the group labels and recalculate the test statistic for the permutation resample.
- Repeat many times; I generated 999 resamples.
- Construct the Monte Carlo permutation distribution and use it to estimate the *P*-value by locating the original test statistic.

The Results: Comparing Means

Monte Carlo Permutation Distribution of $\bar{x}_{Male} - \bar{x}_{Female}$

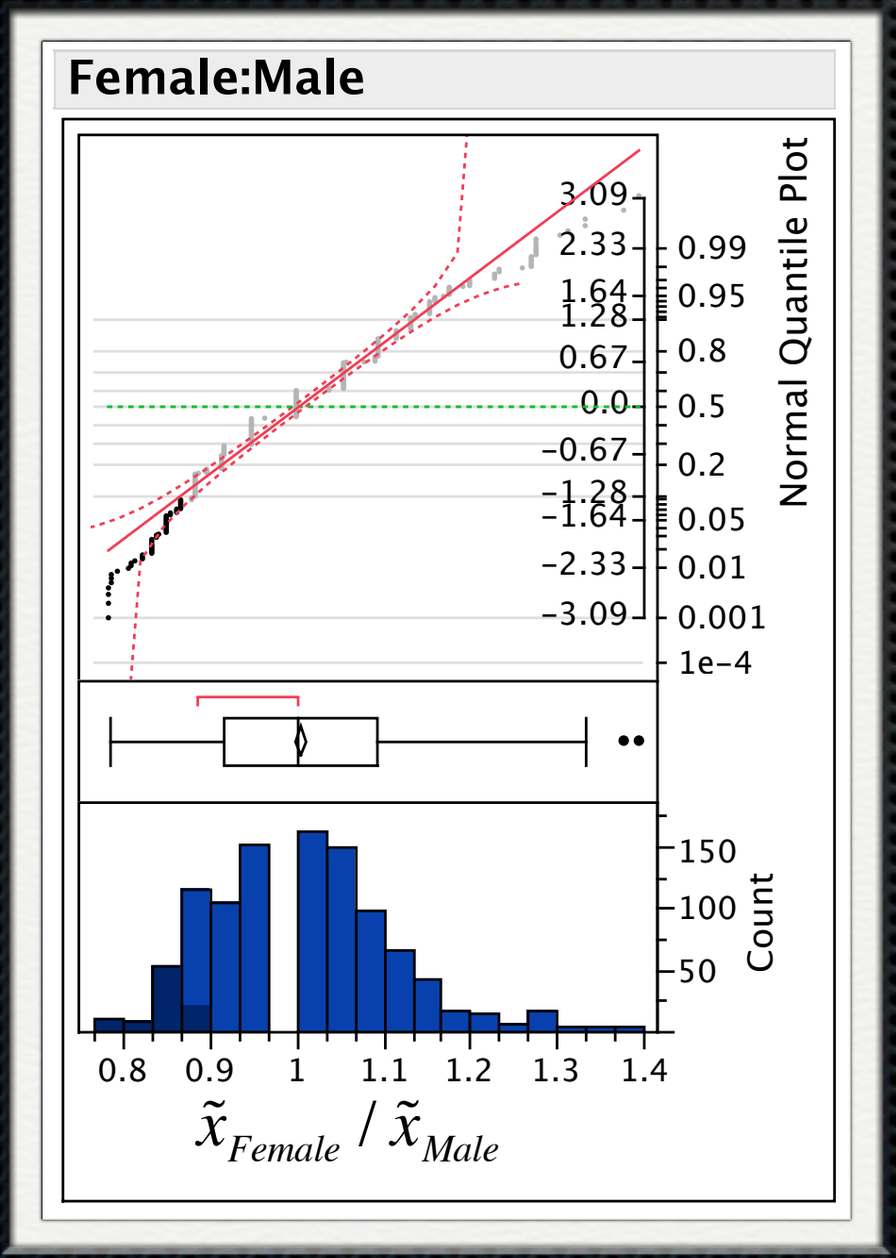


- Eight resamples resulted in test statistics as large or larger than the original test statistic $\left(\bar{x}_{Male} - \bar{x}_{Female} = \$13482\right)$.
- Thus, the approximate right-sided *P*-value is

$$P \approx \frac{8+1}{999+1} = .009$$

The Results: Comparing Medians

Monte Carlo Permutation Distribution of $\tilde{x}_{Female} / \tilde{x}_{Male}$



- 86 resamples resulted in test statistics as small or smaller than the original test statistic $\left(\frac{\tilde{x}_{Female}}{\tilde{x}_{Male}} = 0.867 \right)$.

- Thus, the approximate left-sided *P*-value is

$$P \approx \frac{86+1}{999+1} = .087$$

Conclusions

- Randomization tests provide a modern nonparametric approach for analyzing notoriously skewed salary data.
- When comparing means, we find evidence of a gender wage gap between the salaries earned by male and female accountants/ auditors with Bachelor's degrees who work in the Philadelphia metropolitan area ($\bar{x}_{Male} - \bar{x}_{Female} = \$13482$, $P = .009$, right-sided).
- However, there is not evidence of a significant wage gap when we consider the ratio of the median salaries of females and males in this same demographic ($\tilde{x}_{Female} / \tilde{x}_{Male} = 0.867$, $P = .087$, left-sided).