# Lecture 16

# Case Study: Gender Discrimination

# Gender Discrimination

-
-
-
-
-
-

# Data

```
## Registered S3 method overwritten by 'rvest':
##   method            from
##   read_xml.response xml2
```

| | Promoted | | |
|---|---|---|---|
| | Yes | No | Total |
| Gender | | | |
| Total | | | |

**% of males promoted**

**% of females promoted**

# Practice

# Practice

# Two Competing Claims

**Null Hypothesis**

**independent**

---

**versus**

---

**Alternative Hypothesis**

**dependent**

# We Will Return!

# Variability in estimates

# Young, Underemployed and Optimistic

*Coming of Age, Slowly, in a Tough Economy*

**Young adults hit hard by the recession.** A plurality of the public (41%) believes young adults, rather than middle-aged or older adults, are having the toughest time in today's economy. An analysis of government economic data suggests that this perception is correct. The recent indicators on the nation's labor market show a decline in the

**Tough economic times altering young adults' daily lives, long-term plans.** While negative trends in the labor market have been felt most acutely by the youngest workers, many adults in their late 20s and early 30s have also felt the impact of the weak economy. Among all 18- to 34-year-olds, fully half (49%) say they have taken a job they didn't want just to pay the bills, with 24% saying they have taken an unpaid job to gain work experience. And more than one-third (35%) say that, as a result of the poor economy, they have gone back to school. Their personal lives have also been affected: 31% have postponed either getting married or having a baby (22% say they have postponed having a baby and 20% have put off getting married). One-in-four (24%) say they have moved back in with their parents after living on their own.

# Margin of error

**The general public survey** is based on telephone interviews conducted Dec. 6-19, 2011, with a nationally representative sample of 2,048 adults ages 18 and older living in the continental United States, including an oversample of 346 adults ages 18 to 34. A total of 769 interviews were completed with respondents contacted by landline telephone and 1,279 with those contacted on their cellular phone. Data are weighted to produce a final sample that is representative of the general population of adults in the continental United States. Survey interviews were conducted under the direction of Princeton Survey Research Associates International, in English and Spanish. Margin of sampling error is plus or minus 2.9 percentage points for results based on the total sample and 4.4 percentage points for adults ages 18-34 at the 95% confidence level.

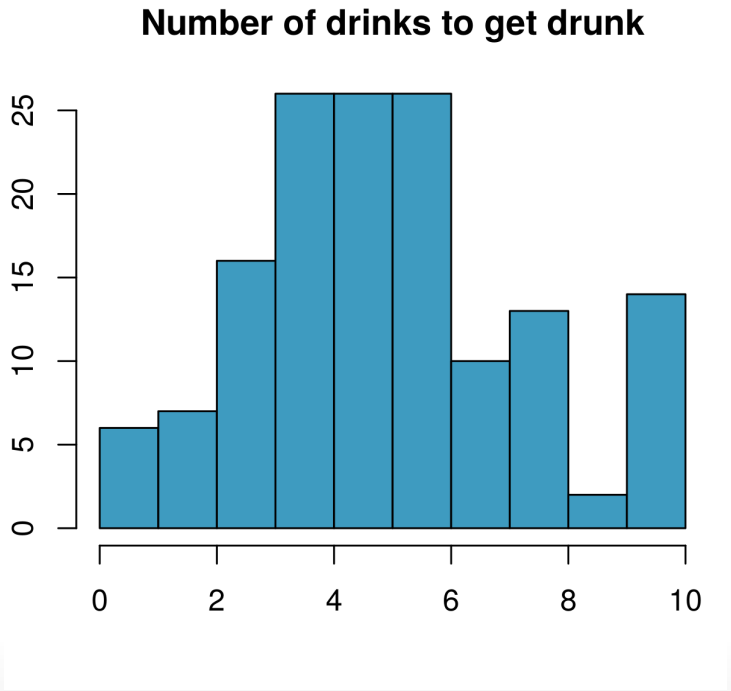
- ±
- ±

# Parameter estimation

- population parameters
- sample statistics point estimates
- 
- margin of error
-

# Parameter estimation

Not the same, but only somewhat different.


Number of drinks to get drunk
0
5
10
15
20
25
0
2
4
6
8
10

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7 | 16 | 3 | 31 | 5 | 46 | 4 | 61 | 10 | 76 | 6 | 91 | 4 | 106 | 6 | 121 | 6 | 136 | 6 |
| 2 | 5 | 17 | 10 | 32 | 9 | 47 | 3 | 62 | 7 | 77 | 6 | 92 | 0.5 | 107 | 2 | 122 | 5 | 137 | 7 |
| 3 | 4 | 18 | 8 | 33 | 7 | 48 | 3 | 63 | 4 | 78 | 5 | 93 | 3 | 108 | 5 | 123 | 3 | 138 | 3 |
| 4 | 4 | 19 | 5 | 34 | 5 | 49 | 6 | 64 | 5 | 79 | 4 | 94 | 3 | 109 | 1 | 124 | 2 | 139 | 10 |
| 5 | 6 | 20 | 10 | 35 | 5 | 50 | 8 | 65 | 6 | 80 | 5 | 95 | 5 | 110 | 5 | 125 | 2 | 140 | 4 |
| 6 | 2 | 21 | 6 | 36 | 7 | 51 | 8 | 66 | 6 | 81 | 6 | 96 | 6 | 111 | 5 | 126 | 5 | 141 | 4 |
| 7 | 3 | 22 | 2 | 37 | 4 | 52 | 8 | 67 | 6 | 82 | 5 | 97 | 4 | 112 | 4 | 127 | 10 | 142 | 6 |
| 8 | 5 | 23 | 6 | 38 | 0 | 53 | 2 | 68 | 7 | 83 | 6 | 98 | 4 | 113 | 4 | 128 | 4 | 143 | 6 |
| 9 | 5 | 24 | 7 | 39 | 4 | 54 | 4 | 69 | 7 | 84 | 8 | 99 | 2 | 114 | 9 | 129 | 1 | 144 | 4 |
| 10 | 6 | 25 | 3 | 40 | 3 | 55 | 8 | 70 | 5 | 85 | 4 | 100 | 5 | 115 | 4 | 130 | 4 | 145 | 5 |
| 11 | 1 | 26 | 6 | 41 | 6 | 56 | 3 | 71 | 10 | 86 | 10 | 101 | 4 | 116 | 3 | 131 | 10 | 146 | 5 |
| 12 | 10 | 27 | 5 | 42 | 10 | 57 | 5 | 72 | 3 | 87 | 5 | 102 | 7 | 117 | 3 | 132 | 8 | | |
| 13 | 4 | 28 | 8 | 43 | 3 | 58 | 5 | 73 | 5.5 | 88 | 10 | 103 | 6 | 118 | 4 | 133 | 10 | | |
| 14 | 4 | 29 | 0 | 44 | 6 | 59 | 8 | 74 | 7 | 89 | 8 | 104 | 8 | 119 | 4 | 134 | 6 | | |
| 15 | 6 | 30 | 8 | 45 | 10 | 60 | 4 | 75 | 10 | 90 | 5 | 105 | 3 | 120 | 8 | 135 | 6 | | |

## Example:

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 7 | 16 | 3 | 31 | 5 | 46 | 4 | 61 | 10 | 76 | 6 | 91 | 4 | 106 | 6 | 121 | 6 | 136 | 6 |
| 2 | 5 | 17 | 10 | 32 | 9 | 47 | 3 | 62 | 7 | 77 | 6 | 92 | 0.5 | 107 | 2 | 122 | 5 | 137 | 7 |
| 3 | 4 | 18 | 8 | 33 | 7 | 48 | 3 | 63 | 4 | 78 | 5 | 93 | 3 | 108 | 5 | 123 | 3 | 138 | 3 |
| 4 | 4 | 19 | 5 | 34 | 5 | 49 | 6 | 64 | 5 | 79 | 4 | 94 | 3 | 109 | 1 | 124 | 2 | 139 | 10 |
| 5 | 6 | 20 | 10 | 35 | 5 | 50 | 8 | 65 | 6 | 80 | 5 | 95 | 5 | 110 | 5 | 125 | 2 | 140 | 4 |
| 6 | 2 | 21 | 6 | 36 | 7 | 51 | 8 | 66 | 6 | 81 | 6 | 96 | 6 | 111 | 5 | 126 | 5 | 141 | 4 |
| 7 | 3 | 22 | 2 | 37 | 4 | 52 | 8 | 67 | 6 | 82 | 5 | 97 | 4 | 112 | 4 | 127 | 10 | 142 | 6 |
| 8 | 5 | 23 | 6 | 38 | 0 | 53 | 2 | 68 | 7 | 83 | 6 | 98 | 4 | 113 | 4 | 128 | 4 | 143 | 6 |
| 9 | 5 | 24 | 7 | 39 | 4 | 54 | 4 | 69 | 7 | 84 | 8 | 99 | 2 | 114 | 9 | 129 | 1 | 144 | 4 |
| 10 | 6 | 25 | 3 | 40 | 3 | 55 | 8 | 70 | 5 | 85 | 4 | 100 | 5 | 115 | 4 | 130 | 4 | 145 | 5 |
| 11 | 1 | 26 | 6 | 41 | 6 | 56 | 3 | 71 | 10 | 86 | 10 | 101 | 4 | 116 | 3 | 131 | 10 | 146 | 5 |
| 12 | 10 | 27 | 5 | 42 | 10 | 57 | 5 | 72 | 3 | 87 | 5 | 102 | 7 | 117 | 3 | 132 | 8 | | |
| 13 | 4 | 28 | 8 | 43 | 3 | 58 | 5 | 73 | 5.5 | 88 | 10 | 103 | 6 | 118 | 4 | 133 | 10 | | |
| 14 | 4 | 29 | 0 | 44 | 6 | 59 | 8 | 74 | 7 | 89 | 8 | 104 | 8 | 119 | 4 | 134 | 6 | | |
| 15 | 6 | 30 | 8 | 45 | 10 | 60 | 4 | 75 | 10 | 90 | 5 | 105 | 3 | 120 | 8 | 135 | 6 | | |

**Sample mean** $\frac{8+6+10+4+5+3+5+6+6+6}{10} = 5.9$

# Sampling distribution

# Sampling distribution

Approximately 5.39, the true population mean.

# Sampling distributions - via CLT

# Central limit theorem

## Central limit theorem

$$\bar{x} \sim \mathcal{N}\left(\text{mean} = \mu, \text{SE} = \frac{\sigma}{\sqrt{n}}\right),$$

**standard error**

$\sigma$ $s$

# Central limit theorem

.

. $SE = \frac{\sigma}{\sqrt{n}}$ $n$

$SE$

-

# CLT - conditions

- **Independence:**
  - 
  - $n <$

# CLT - conditions

- **Independence:**
  - 
  - $n <$
- **Sample size/skew:**
  - 
  - $n > 30$

# Confidence intervals

# Why do we report confidence intervals?

- confidence interval
- fishing with a spear
  fishing with a net
-

# Average number of exclusive relationships

$$\bar{x} = 3.2 \qquad s = 1.74$$

$$\text{point estimate} \pm 2 \times \text{SE}$$

$$SE = \frac{s}{\sqrt{n}} = \frac{1.74}{\sqrt{50}} \approx 0.25$$

# Average number of exclusive relationships

$$\bar{x} = 3.2 \qquad s = 1.74$$

$$\begin{aligned} \bar{x} \pm 2 \times SE &= 3.2 \pm 2 \times 0.25 \\ &= (3.2 - 0.5, 3.2 + 0.5) \\ &= (2.7, 3.7) \end{aligned}$$

# Practice

- 
- 
- 
- 

# Practice

-
-
-
-

# A more accurate interval

## Confidence interval, a general formula

$$\text{point estimate} \pm z^\star \cdot SE$$

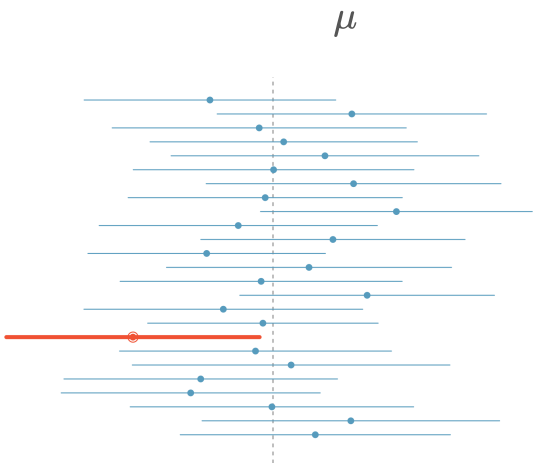
$$\bar{x}$$

- **Independence:**
  - 
  - $n <$
- **Sample size / skew:** $n \geq 30$

**Note:** $n < 30$

# Capturing the population parameter

- 
- $$\text{point estimate} \pm 2 \cdot SE$$
- 



# Width of an interval

# Width of an interval

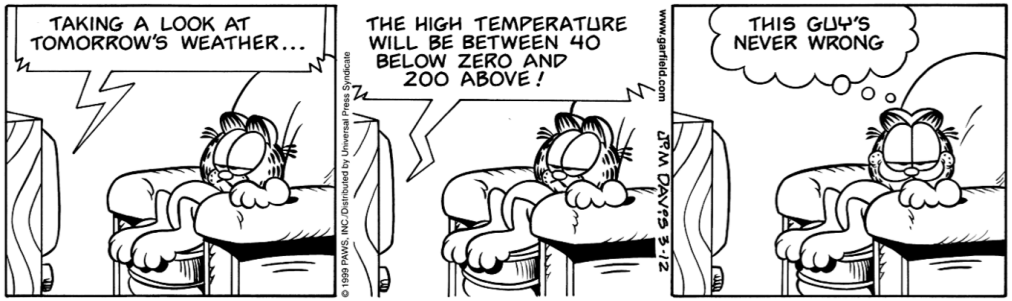
A wider interval.

# Width of an interval

A wider interval.

# Width of an interval

A wider interval.




If the interval is too wide it may not be very informative.

# Changing the confidence level

$$\text{point estimate} \pm z^\star \cdot SE$$

- $z^\star \cdot SE$ margin of error
- $z^\star$
- 
- $z^\star = 1.96$
- $z$

$z^\star$

# Practice

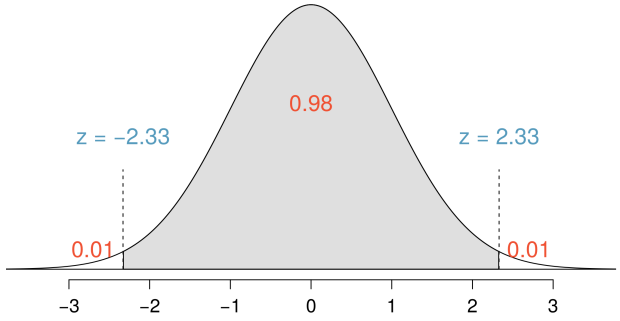
$z^\star$

- $Z = 2.05$
- $Z = 1.96$
- $Z = 2.33$
- $Z = -2.33$
- $Z = -1.65$

# Practice

- $Z = 2.05$
- $Z = 1.96$
- $Z = 2.33$
- $Z = -2.33$
- $Z = -1.65$

$z^\star$



# Hypothesis Testing

# Hypothesis Tests as a Trial

- $H_0$

  $H_A$
- —
-

  -
-

unlikely

# A Hypothesis Test as a Trial (continued)

- 
  - 
  - 
- fail to reject the null hypothesis
  - 
  - 

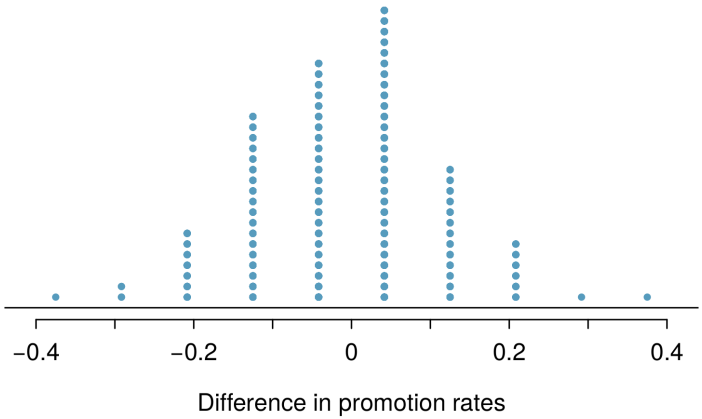
# A Hypothesis Test as a Trial (continued)

- 
- 
- 

# Recap: Concept of Hypothesis Testing

- null hypothesis $H_0$
- alternative hypothesis $H_A$

# Back to the example of gender discrimination ...



# Testing Hypotheses: CIs

# Testing hypotheses using confidence intervals

.

- $H_0 \;\; \mu = 3$
- $H_A \;\; \mu > 3$

# Testing hypotheses using confidence intervals

.

- $H_0 \;\; \mu = 3$
- $H_A \;\; \mu > 3$

.

.

# Hypothesis Testing

# Concept of Hypothesis Testing

- null hypothesis $H_0$
- alternative hypothesis $H_A$

# Testing hypotheses using confidence intervals

.

- $H_0 \;\; \mu = 3$
- $H_A \;\; \mu > 3$

# Testing hypotheses using confidence intervals

.

- $H_0 \;\; \mu = 3$
- $H_A \;\; \mu > 3$

.

.

# Recall

$$\mu$$

$$\bar{x} \pm \text{ME} = \bar{x} \pm z^\star \cdot \text{SE}$$

$$\text{SE}$$

$$\text{SE} = \frac{\sigma}{\sqrt{n}}.$$

# Examples

# Example 1

# Example 1

- $\mu = 3.7$
- $\bar{x} = 3.6$
- $s = 0.61$
- $n = 36$

# Example 1: the CI

$$\bar{x} \pm z^\star \cdot \frac{s}{\sqrt{n}} = 3.6 \pm z^\star \cdot \frac{0.61}{\sqrt{36}}$$

$z^\star$

$\alpha/2 = 0.005$

```
qnorm(1 - 0.01/2)
```

```
## [1] 2.575829
```

# Example 1: finishing

$$\bar{x} \pm z^\star \cdot \frac{s}{\sqrt{n}} = 3.6 \pm 2.576 \cdot \frac{0.61}{\sqrt{36}} = (3.3381, 3.8619)$$

includes

fail to reject the null hypothesis

# Number of university applications

# Setting the hypotheses

- parameter of interest all

-
-

# Setting the hypotheses (ctd.)

.

$$H_0 : \ \mu = 8$$

.

$$H_A : \ \mu > 8$$

# Number of university applications - conditions

-
-
-
-
-

# Using a CI

- $n = 206$
- $\bar{x} = 9.7$
- $s = 7$
- $\mu = 8$

$$\bar{x} \pm z^\star \cdot \frac{s}{\sqrt{n}} = 9.7 \pm 1.96 \cdot \frac{7}{\sqrt{206}} = (8.744, 10.656)$$

8 not do have evidence to reject the null

# Next Idea