Two simple examples

- Lady tasting tea
- Human energy fields

These examples provide the intuition behind statistical inference
Fisher’s exact test

- A simple approach to inference
- Only applicable when outcome probabilities known
- Lady tasting tea example
  - Claims she can tell whether the milk was poured first
  - In a test, 4/8 teacups had milk poured first
  - The lady correctly detects all four
  - Should we believe that she has milk-first detection ability?
- To answer this question, we ask, “What is the probability she did this by chance?”
  - If likely to happen by chance, then we shouldn’t be convinced
  - If very unlikely, then we should maybe believe her
  - This is the basic question behind statistical inference
    - Null hypothesis
    - People seem poorly equipped to make these inferences, in part because they forget about failures, but notice success: e.g. Dog ESP, miracles
    - Other examples: fingerprints, DNA, HIV tests, regression coefficients, mean differences, etc.
  - Answer?
    - 70 ways of choosing four cups out of eight
    - How many ways can she do so correctly?
By chance, she would only guess all four correctly with probability \((1/70 = 0.014)\). So, we can be quite confident in her milk-first detection ability.
Second simple example
Healing touch: human energy field detection

“A Close Look at Therapeutic Touch”
Linda Rosa; Emily Rosa; Larry Sarner; Stephen Barrett.
1998.

JAMA
(279: 1005 – 1010)
Human energy field: Prob. of success by chance

\[ f(k; n, p) = \binom{n}{k} p^k (1 - p)^{n-k} \]

n (the number of trials) = 10
k (number of successes)
p (prob. of success for the null model) = .5

Binomial distribution

# of successful detections
Human energy field detection:
Confidence in ability

# of successful detections out of 150 trials
**Table 2.——Statistical Analysis**

<table>
<thead>
<tr>
<th>Statistical Function</th>
<th>Initial Test (n = 15)</th>
<th>Follow-up Test (n = 13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (95% confidence interval)</td>
<td>4.67 (3.67-5.67)</td>
<td>4.08 (3.17-4.99)</td>
</tr>
<tr>
<td>SD</td>
<td>1.74</td>
<td>1.44</td>
</tr>
<tr>
<td>α (1-tailed test)</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>t statistic</td>
<td>-0.7174</td>
<td>-2.222</td>
</tr>
<tr>
<td>Upper critical limit of Student t distribution</td>
<td>1.761</td>
<td>1.782</td>
</tr>
<tr>
<td>Alternative hypothesis, μ = 6.67</td>
<td>0.9559</td>
<td>0.9801</td>
</tr>
<tr>
<td>Alternative hypothesis, μ = 7.50</td>
<td>0.999644</td>
<td>0.999953</td>
</tr>
</tbody>
</table>

**Figure 2.——Distribution of test results.**

Null hypothesis

- In both cases, we calculated the probability of making the correct choice by chance and compared it to the observed results.
- Thus, our null hypothesis was that the lady and the therapists lacked any of their claimed ability.
- What’s the null hypothesis that Stata uses by default for calculating p values?
- Always consider whether null hypotheses other than 0 might be more substantively meaningful.
  - E.g., testing whether the benefits from government programs outweigh the costs.
Assessing uncertainty

- With more complicated statistical processes, larger samples, continuous variables, Fisher’s exact test becomes difficult or impossible.

- Instead, we use other approaches, such as calculating standard errors and using them to calculate confidence intervals.

- The intuition from these simple examples, however, extends to the more complicated one.
Standard error: Baseball example

- In 2006, Manny Ramírez hit .321
- How certain are we that, in 2006, he was a .321 hitter? Confidence interval?
- To answer this question, we need to know how precisely we have estimated his batting average
- The standard error gives us this information, which in general is (where $s$ is the sample standard deviation)

**Equation?**

$$ \text{std. err.} = \frac{s}{\sqrt{n}} $$
Baseball example

The standard error (s.e.) for proportions (percentages/100) is?

\[ \sqrt{\frac{p(1-p)}{n}} \]

For \( n = 400 \), \( p = .321 \), s.e. = .023

Which means, on average, the .321 estimate will be off by .023
Baseball example: postseason

- 20 at-bats
  - N = 20, p = .400, s.e. = .109
  - Which means, on average, the .400 estimate will be off by .109

- 10 at-bats
  - N = 10, p = .400, s.e. = .159
  - Which means, on average, the .400 estimate will be off by .159
Using Standard Errors, we can construct “confidence intervals”

- **Confidence interval (ci):** an interval between two numbers, where there is a certain specified level of confidence that a population parameter lies

- $ci = \text{sample parameter} \pm \text{multiple} \times \text{sample standard error}$
N = 20; avg. = .400; s = .489; s.e. = .109

Confidence interval

\[
\begin{align*}
\text{Lower Limit} &= 0.400 - 2 \times 0.109 = 0.185 \\
\text{Upper Limit} &= 0.400 + 2 \times 0.109 = 0.615
\end{align*}
\]

s.e. is estimate of \( \sigma \)
Much of the time, we fail to appreciate the uncertainty in averages and other statistical estimates

- Postseason statistics
- Boardgames
- Life
Two types of inference

- Testing underlying traits
  - E.g., can lady detect milk-poured first?
  - E.g., does democracy improve human lives?

- Testing inferences about a population from a sample
  - What percentage of the population approves of President Bush?
  - What’s average household income in the United States?
Example of second type of inference:

Testing inferences about a population from a sample

Family income in 2006
Certainty about mean of a population based on a sample: Family income in 2006

$\overline{X} = 65.8$, $n = 31,401$, $s = 41.7$

Source: 2006 CCES
Calculating the Standard Error on the mean family income of $65.8 thousand dollars

Equation?

$$std.\ err. = \frac{S}{\sqrt{n}}$$

For the income example,

std. err. = $41.6/177.2 = $0.23$ thousands of dollars

$X = 65.8$, $n = 31401$, $s = 41.7$
N = 31,401; avg. = 65.8; s = 41.6; s.e. = s/\sqrt{n} = .2

The Picture
Where does the bell-shaped curve come from?

That is, how do we know that two \( \pm \) standard errors covers 95% of the distribution?
Could this possibly be right? Why?

- Central limit theorem
Central Limit Theorem

As the sample size $n$ increases, the distribution of the mean $\overline{X}$ of a random sample taken from practically any population approaches a normal distribution, with mean $\mu$ and standard deviation $\sigma/\sqrt{n}$.
Illustration of Central Limit Theorem: Exponential Distribution

Mean = 250,000
Median = 125,000
σ = 283,474
Min = 0
Max = 1,000,000
Consider 10,000 samples of $n = 100$

N = 10,000  
Mean = 249,993  
s = 28,559

What will the distribution of these means look like?
Consider 1,000 samples of various sizes

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>100</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>250,105</td>
<td>250,498</td>
<td>249,938</td>
</tr>
<tr>
<td>s</td>
<td>90,891</td>
<td>28,297</td>
<td>9,376</td>
</tr>
</tbody>
</table>
Convince yourself by playing with simulations

In small samples (n <30), these statistics are not normally distributed. Instead, they follow the t-distribution.

We’ll discuss that complication next class.
Another example

- Let’s say we draw a sample of tuitions from 15 private universities. Can we estimate what the average of all private university tuitions is?
  - $N = 15$
  - Average = $29,735$
  - $s = 2,196$
  - $s.e. = \frac{s}{\sqrt{n}} = \frac{2,196}{\sqrt{15}} = 567$
N = 15; avg. = 29,735; s = 2,196; s.e. = \frac{s}{\sqrt{n}} = 567

The Picture

\begin{align*}
29,735 - 2 \times 567 &= 29,168 \\
29,735 - 567 &= 29,168 \\
29,735 + 567 &= 30,302 \\
29,735 + 2 \times 567 &= 30,869
\end{align*}
Confidence Intervals for Tuition Example

- 68% confidence interval
  - $29,735 + 567 = [$29,168 to $30,302]

- 95% confidence interval
  - $29,735 + 2*567 = [$28,601 to $30,869]

- 99% confidence interval
  - $29,735 + 3*567 = [$28,034 to $31,436]
Using z-scores
The z-score or the “standardized score”

Equation?

\[ Z = \frac{x - \bar{x}}{\sigma_x} \]

Using z-scores to assess how far values are from the mean
What if someone (ahead of time) had said, “I think the average tuition of major research universities is $25k”?

- Note that $25,000 is well out of the 99% confidence interval, [28,034 to 31,436]
- Q: How far away is the $25k estimate from the sample mean?
  - A: Do it in z-scores: $(29,735 - 25,000)/567 = 8.35$
More confidence interval calculations

- Proportions
- Difference in means
- Difference in proportions
Constructing confidence intervals of proportions

Let us say we drew a sample of 1,000 adults and asked them if they approved of the way George Bush was handling his job as president. (March 13-16, 2006 Gallup Poll) Can we estimate the % of all American adults who approve?

- N = 1000
- p = .37
- s.e. = \( \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.37(1-.37)}{1000}} = 0.02 \)
N = 1,000; p. = .37; s.e. = \sqrt{p(1-p)/n} = .02

The Picture
Confidence Intervals for Bush approval example

- 68% confidence interval = .37±.02 = [.35 to .39]
- 95% confidence interval = .37±2*.02 = [.33 to .41]
- 99% confidence interval = .37±3*.02 = [ .31 to .43]
What if someone (ahead of time) had said, “I think Americans are equally divided in how they think about Bush.”

- Note that 50% is well out of the 99% confidence interval, [31% to 43%]

- Q: How far away is the 50% estimate from the sample proportion?
  - A: Do it in z-scores: \((.37-.5)/.02 = -6.5\)
Constructing confidence intervals of differences of means

- Let’s say we draw a sample of tuitions from 15 private and public universities. Can we estimate what the difference in average tuitions is between the two types of universities?
- \( N = 15 \) in both cases
- Average = 29,735 (private); 5,498 (public); diff = 24,238
- \( s = 2,196 \) (private); 1,894 (public)
- \( s.e. = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{4,822,416}{15} + \frac{3,587,236}{15}} = 749 \)
N = 15 twice; diff = 24,238; s.e. = 749

The Picture
Confidence Intervals for difference of tuition means example

68% confidence interval = $24,238 \pm 749 = [23,489 \text{ to } 24,987]$

95% confidence interval = $24,238 \pm 2 \times 749 = [22,740 \text{ to } 25,736]$

99% confidence interval = $24,238 \pm 3 \times 749 = [21,991 \text{ to } 26,485]$
What if someone (ahead of time) had said, “Private universities are no more expensive than public universities”

- Note that $0 is well out of the 99% confidence interval, [$21,991 to $26,485]
- Q: How far away is the $0 estimate from the sample proportion?
  - A: Do it in z-scores: (24,238-0)/749 = 32.4
Constructing confidence intervals of difference of proportions

- Let us say we drew a sample of 1,000 adults and asked them if they approved of the way George Bush was handling his job as president. (March 13-16, 2006 Gallup Poll). We focus on the 600 who are either independents or Democrats. Can we estimate whether independents and Democrats view Bush differently?
- \( N = 300 \) ind; 300 Dem.
- \( p = .29 \) (ind.); .10 (Dem.); \( \text{diff} = .19 \)
- \( \text{s.e.} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{.29(1-.29)}{300} + \frac{.10(1-.10)}{300}} = .03 \)
diff. p. = .19; s.e. = .03

The Picture

- Mean
- \(\sigma\)
- \(\sigma\)
- \(\sigma\)
- \(\sigma\)
- \(\sigma\)

\[y = .398942\]

- .19 - .03 = .16
- .19 + .03 = .22
- .19 - 2* .03 = .13
- .19 + 2* .03 = .25

- 68% Mean
- 95%
- 99%
Confidence Intervals for Bush Ind/Dem approval example

- 68% confidence interval = $0.19 \pm 0.03 = [0.16 \text{ to } 0.22]$
- 95% confidence interval = $0.19 \pm 2 \times 0.03 = [0.13 \text{ to } 0.25]$
- 99% confidence interval = $0.19 \pm 3 \times 0.03 = [0.10 \text{ to } 0.28]$
What if someone (ahead of time) had said, “I think Democrats and Independents are equally unsupportive of Bush”?

- Note that 0% is well out of the 99% confidence interval, [10% to 28%]
- Q: How far away is the 0% estimate from the sample proportion?
  - A: Do it in z-scores: \((0.19 - 0) / 0.03 = 6.33\)
Constructing confidence intervals for regression coefficients

Let’s look at the relationship between the mid-term seat loss by the President’s party at midterm and the President’s Gallup poll rating.

Slope = 1.97
N = 14
s.e.r. = 13.8
s_x = 8.14
s.e. slope =

\[
\frac{s.e.r.}{\sqrt{n-1}} \times \frac{1}{s_x} = \frac{13.8}{\sqrt{13}} \times \frac{1}{8.14} = 0.47
\]
The Stata output

```
. reg loss Gallup if year>1948

Source |       SS       df       MS              Number of obs =      14
-------------+----------------------------------------------------------------
Model |  3332.58872     1  3332.58872           Prob > F      =  0.0013
Residual |  2280.83985    12  190.069988           R-squared     =  0.5937
          |                    -------------+------------------------------         Adj R-squared =  0.5598
          |                    -------------+------------------------------         Root MSE      =  13.787
          |  5613.42857    13  431.802198
------------------------------------------------------------------------------
        |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
loss |                  Gallup |    1.96812   .4700211     4.19   0.001     .9440315    2.992208
       | _cons | -127.4281   25.54753    -4.99   0.000    -183.0914   -71.76486
------------------------------------------------------------------------------
```

Number of obs = 14
F(  1,    12) = 17.53
Prob > F = 0.0013
R-squared = 0.5937
Adj R-squared = 0.5598
Root MSE = 13.787
The Picture

N = 14; slope = 1.97; s.e. = 0.47

\[ \pm 1.97 \]

-2\sigma \quad -\sigma \quad \sigma \quad 2\sigma

68% Mean

95%

99%
Confidence Intervals for regression example

- 68% confidence interval = 1.97 ± 0.47 = [1.50 to 2.44]
- 95% confidence interval = 1.97 ± 2*0.47 = [1.03 to 2.91]
- 99% confidence interval = 1.97 ± 3*0.47 = [0.62 to 3.32]
What if someone (ahead of time) had said, “There is no relationship between the president’s popularity and how his party’s House members do at midterm”?

- Note that 0 is well out of the 99% confidence interval, [0.62 to 3.32]
- Q: How far away is the 0 estimate from the sample proportion?
  - A: Do it in z-scores: \((1.97-0)/0.47 = 4.19\)
z vs. \( t \)
If $n$ is sufficiently large, we know the distribution of sample means/coeffs. will obey the normal curve.
When the sample size is large (i.e., > 150), convert the difference into $z$ units and consult a $z$ table

$$Z = \frac{(H_1 - H_0)}{s.e.}$$
**Regression example**

\[ Z = \frac{(H_1 - H_{null})}{s.e.} \]

Large sample (n = 1000)
Slope (b) = 2.1
s.e. = 0.9

Calculate p-value for one-tailed test \( H_{null} = 0 \)

\[ Z = \frac{(2.1 - 0)}{0.9} = 2.3 \]

p-value (using handout)

\[ Pr(Z > 2.3) < 0.011 \]

Interpretation: probability that we would observe a coefficient of 2.1 by chance is less than 0.011.

For two-tailed test: \( Pr(|Z| > 2.3) < 1 - 2 \times 0.4893 \) (calculations differ by table)
$t$

(when the sample is small)
When the sample size is small (i.e., <150), convert the difference into $t$ units and consult a $t$ table

$$t = \frac{H_1 - H_{null}}{s.e.}$$

**Mid-term seat loss example**

Slope = 1.97  
$s.e._{slope} = 0.47$

What’s $H_1$?  
What’s $H_{null}$?  
$t = \frac{(H_1 - H_{null})}{s.e.}$

$t = \frac{(1.97 - 0)}{0.47}$

$t = 4.19$
Reading a *t* table

<table>
<thead>
<tr>
<th>df</th>
<th>.05</th>
<th>.025</th>
<th>.01</th>
<th>.005</th>
<th>.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.07</td>
<td>3.29</td>
<td>3.37</td>
<td>3.78</td>
<td>6.32</td>
</tr>
<tr>
<td>2</td>
<td>2.88</td>
<td>2.99</td>
<td>4.02</td>
<td>5.84</td>
<td>6.96</td>
</tr>
<tr>
<td>3</td>
<td>2.77</td>
<td>2.92</td>
<td>4.31</td>
<td>6.61</td>
<td>7.38</td>
</tr>
<tr>
<td>4</td>
<td>2.66</td>
<td>2.86</td>
<td>4.54</td>
<td>6.94</td>
<td>7.74</td>
</tr>
<tr>
<td>5</td>
<td>2.56</td>
<td>2.79</td>
<td>4.74</td>
<td>7.08</td>
<td>7.94</td>
</tr>
<tr>
<td>6</td>
<td>2.47</td>
<td>2.70</td>
<td>4.90</td>
<td>7.21</td>
<td>8.08</td>
</tr>
<tr>
<td>7</td>
<td>2.39</td>
<td>2.62</td>
<td>5.04</td>
<td>7.33</td>
<td>8.19</td>
</tr>
<tr>
<td>8</td>
<td>2.32</td>
<td>2.56</td>
<td>5.16</td>
<td>7.44</td>
<td>8.29</td>
</tr>
<tr>
<td>9</td>
<td>2.26</td>
<td>2.50</td>
<td>5.27</td>
<td>7.53</td>
<td>8.38</td>
</tr>
<tr>
<td>10</td>
<td>2.20</td>
<td>2.44</td>
<td>5.36</td>
<td>7.59</td>
<td>8.44</td>
</tr>
<tr>
<td>11</td>
<td>2.15</td>
<td>2.39</td>
<td>5.45</td>
<td>7.65</td>
<td>8.50</td>
</tr>
<tr>
<td>12</td>
<td>2.10</td>
<td>2.34</td>
<td>5.52</td>
<td>7.70</td>
<td>8.55</td>
</tr>
<tr>
<td>13</td>
<td>2.06</td>
<td>2.29</td>
<td>5.59</td>
<td>7.74</td>
<td>8.59</td>
</tr>
<tr>
<td>14</td>
<td>2.02</td>
<td>2.25</td>
<td>5.65</td>
<td>7.77</td>
<td>8.63</td>
</tr>
<tr>
<td>15</td>
<td>1.98</td>
<td>2.21</td>
<td>5.71</td>
<td>7.80</td>
<td>8.66</td>
</tr>
<tr>
<td>16</td>
<td>1.94</td>
<td>2.17</td>
<td>5.76</td>
<td>7.83</td>
<td>8.68</td>
</tr>
<tr>
<td>17</td>
<td>1.90</td>
<td>2.13</td>
<td>5.81</td>
<td>7.85</td>
<td>8.70</td>
</tr>
<tr>
<td>18</td>
<td>1.86</td>
<td>2.09</td>
<td>5.86</td>
<td>7.87</td>
<td>8.72</td>
</tr>
<tr>
<td>19</td>
<td>1.83</td>
<td>2.06</td>
<td>5.90</td>
<td>7.88</td>
<td>8.74</td>
</tr>
<tr>
<td>20</td>
<td>1.80</td>
<td>2.02</td>
<td>5.94</td>
<td>7.89</td>
<td>8.75</td>
</tr>
<tr>
<td>25</td>
<td>1.70</td>
<td>1.93</td>
<td>6.21</td>
<td>8.12</td>
<td>8.87</td>
</tr>
<tr>
<td>50</td>
<td>1.60</td>
<td>1.83</td>
<td>6.76</td>
<td>8.56</td>
<td>9.03</td>
</tr>
<tr>
<td>100</td>
<td>1.50</td>
<td>1.75</td>
<td>7.27</td>
<td>8.91</td>
<td>9.20</td>
</tr>
</tbody>
</table>

Note that the *t*-distribution with infinite df is the standard normal distribution.

![t-distribution graph](image-url)
Testing hypotheses in Stata with \texttt{ttest}

What if someone (ahead of time) said, “Private university tuitions did not grow from 2003 to 2004”

- Mean growth = $1,632
- Standard deviation on growth = 229
- Note that $0 is well out of the 95\% confidence interval, [$1,141 to $2,122]
- Q: How far away is the $0 estimate from the sample proportion?
  - A: Do it in z-scores: \((1,632-0)/229 = 7.13\)
The Stata output

. gen difftuition=tuition2004-tuition2003
. ttest diff = 0

One-sample t test

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Err.</th>
<th>Std. Dev.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>difftuition</td>
<td>15</td>
<td>1631.6</td>
<td>228.6886</td>
<td>885.707</td>
<td>1141.112 - 2122.088</td>
</tr>
</tbody>
</table>

mean = mean(difftuition)                                      t = 7.1346
Ho: mean = 0                                     degrees of freedom = 14
Ha: mean < 0                          Ha: mean != 0                          Ha: mean > 0
Pr(T < t) = 1.0000                      Pr(|T| > |t|) = 0.0000                      Pr(T > t) = 0.0000

You could test difference in means with

ttest tuition2004 = tuition2003
A word about standard errors and collinearity

- The problem: if $X_1$ and $X_2$ are highly correlated, then it will be difficult to precisely estimate the effect of either one of these variables on $Y$
How does having another collinear independent variable affect standard errors?

\[
s.e.(\hat{\beta}_1) = \sqrt{\frac{1}{N - n - 1} \frac{S_Y^2}{S_{X_1}^2} \frac{1 - R_Y^2}{1 - R_{X_1}^2}}
\]

R\(^2\) of the “auxiliary regression” of X\(_1\) on all the other independent variables
Example: Effect of party, ideology, and religiosity on feelings toward Bush

<table>
<thead>
<tr>
<th></th>
<th>Bush Feelings</th>
<th>Conserv.</th>
<th>Repub.</th>
<th>Religious</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bush Feelings</td>
<td>1.0</td>
<td>.39</td>
<td>.57</td>
<td>.16</td>
</tr>
<tr>
<td>Conserv.</td>
<td></td>
<td>1.0</td>
<td>.46</td>
<td>.18</td>
</tr>
<tr>
<td>Repub.</td>
<td></td>
<td></td>
<td>1.0</td>
<td>.06</td>
</tr>
<tr>
<td>Relig.</td>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
</tbody>
</table>
### Regression Table

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>32.7 (0.85)</td>
<td>32.9 (1.08)</td>
<td>32.6 (1.20)</td>
<td>29.3 (1.31)</td>
</tr>
<tr>
<td>Repub.</td>
<td>6.73 (0.244)</td>
<td>5.86 (0.27)</td>
<td>6.64 (0.241)</td>
<td>5.88 (0.27)</td>
</tr>
<tr>
<td>Conserv.</td>
<td>---</td>
<td>2.11 (0.30)</td>
<td>---</td>
<td>1.87 (0.30)</td>
</tr>
<tr>
<td>Relig.</td>
<td>---</td>
<td>---</td>
<td>7.92 (1.18)</td>
<td>5.78 (1.19)</td>
</tr>
<tr>
<td>N</td>
<td>1575</td>
<td>1575</td>
<td>1575</td>
<td>1575</td>
</tr>
<tr>
<td>R²</td>
<td>.32</td>
<td>.35</td>
<td>.35</td>
<td>.36</td>
</tr>
</tbody>
</table>
Pathologies of statistical significance
Understanding and using “significance”
Substantive versus statistical significance

- Which variable is more statistically significant?
  - $X_1$
- Which variable is more important?
  - $X_2$
- Importance (size) is often more relevant

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>$X_1$</td>
<td>0.500*</td>
<td>0.055**</td>
</tr>
<tr>
<td></td>
<td>(0.244)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>$X_2$</td>
<td>0.600</td>
<td>0.600</td>
</tr>
<tr>
<td></td>
<td>(0.305)</td>
<td>(0.305)</td>
</tr>
<tr>
<td>N</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.32</td>
<td>.20</td>
</tr>
</tbody>
</table>

*p<.05, **p <.01
Substantive versus statistical significance (again)

- Think about point estimates, such as means or regression coefficients, as the center of distributions
- Let \( B^* \) be of value of a regression coefficient that is large enough for substantive significance
- Which is substantively significant?
- (a)
Substantive versus statistical significance (again)

Which is more substantively significant? That is, which is larger?
- Depends, but probably (d)

Don’t confuse lack of statistical significance with no effect
- Lack of statistical significance usually implies uncertainty, not no effect
Degree of significance

- We often use 95% confidence intervals, which correspond with $p < .05$
- Is an effect statistically significant if it is $p < .06$? (that is, 95% CI encompasses zero)
  - Yes!
  - For many data sets, anything less than $p < .20$ is informative
  - Treat significance as a continuous variable
    - E.g., if $p < .20$, we should be roughly 80% sure that the coefficient is different from zero. If $p < .10$, we should be roughly 90% sure that the coefficient is different from zero. Etc.
Don’t make this mistake
Understanding and using “significance”

Summary

- Focus on substantive significance (effect size), not statistical significance
- Focus on degree of uncertainty, not on the arbitrary cutoff of $p = .05$
  - Confidence intervals are preferable to $p$-values
  - Treat $p$-values as a continuous variable
- Don’t confuse lack of statistical significance with no effect (that is, $p > .05$ does not mean $b = 0$)
  - Lack of statistical significance usually implies uncertainty, not no effect!
What to present

- Standard error
- CI
- t-value
- p-value
- Stars
- Combinations?
- Different disciplines have different norms, I prefer
  - Graphically presenting CIs
  - Coefficients with standard errors
  - No stars
  - (Showing data through scatter plots more important!)
<table>
<thead>
<tr>
<th>Variable</th>
<th>Main Effects</th>
<th></th>
<th>Southern Interactions</th>
<th></th>
<th>Border Interactions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>t-value</td>
<td>Coefficient</td>
<td>t-value</td>
<td>Coefficient</td>
<td>t-value</td>
</tr>
<tr>
<td>South</td>
<td>242.120</td>
<td>2.14</td>
<td>—</td>
<td></td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Border</td>
<td>262.120</td>
<td>1.32</td>
<td>—</td>
<td></td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>.032</td>
<td>.91</td>
<td>-.151</td>
<td>-2.28**</td>
<td>-.142</td>
<td>-1.31</td>
</tr>
<tr>
<td>Democrats (t - 1)</td>
<td>.508</td>
<td>20.10**</td>
<td>.435</td>
<td>7.70**</td>
<td>-.067</td>
<td>-.52</td>
</tr>
<tr>
<td>Compensation</td>
<td>.007</td>
<td>2.08*</td>
<td>-.005</td>
<td>-1.03</td>
<td>-.007</td>
<td>-.59</td>
</tr>
<tr>
<td>Presidential year</td>
<td>-27.079</td>
<td>-11.10**</td>
<td>22.778</td>
<td>6.19**</td>
<td>11.520</td>
<td>1.35</td>
</tr>
<tr>
<td>Presidential vote</td>
<td>.498</td>
<td>9.40**</td>
<td>-.452</td>
<td>5.85**</td>
<td>-.234</td>
<td>-1.39</td>
</tr>
<tr>
<td>Gubernatorial year</td>
<td>-.18.540</td>
<td>-7.86**</td>
<td>15.111</td>
<td>4.08**</td>
<td>-.440</td>
<td>-.04</td>
</tr>
<tr>
<td>Gubernatorial vote</td>
<td>.364</td>
<td>8.10**</td>
<td>-.328</td>
<td>-5.52**</td>
<td>-.044</td>
<td>-.20</td>
</tr>
<tr>
<td>Off year</td>
<td>-10.042</td>
<td>-10.62**</td>
<td>6.800</td>
<td>3.80**</td>
<td>2.818</td>
<td>.82</td>
</tr>
<tr>
<td>GNP growth</td>
<td>.149</td>
<td>2.74**</td>
<td>-.157</td>
<td>-1.49</td>
<td>-.148</td>
<td>-.79</td>
</tr>
</tbody>
</table>

Note: Thirty-one state intercepts not shown. Coefficients are unstandardized.

n = 1,035, adjusted $R^2 = .89$, SEE = 7.94
*p < .05; **p < .01
Statistical monkey business
(tricks to get $p < .05$)

- Bonferroni problem: using $p < .05$, one will get significant results about 5% (1/20) of the time by chance alone
- Reporting one of many dependent variables or dependent variable scales
  - Healing-with-prayer studies
  - Psychology lab studies
- Repeating an experiment until, by chance, the result is significant
  - Drug trials
  - Called file-drawer problem
Statistical monkey business
(tricks to get $p < .05$)

- Specification searches
  - Adding and removing control variables until, by chance, the result is significant
  - Exceedingly common
Statistical monkey business
Solutions

- With many dependent variables, test hypotheses on a simple unweighted average
- Bonferroni correction
  - If testing $n$ independent hypotheses, adjusts the significance level by $1/n$ times what it would be if only one hypothesis were tested
  - E.g., testing 5 hypotheses at $p < .05$ level, adjust significance level to $p/5 < .05/5 < .01$

- Show bivariate results
- Show many specifications
- Model averaging
- Always be suspicious of statistical monkey business!