One goal of science: determine whether current ways of thinking about the world are adequate for predicting and understanding events.

Not all models have to predict all things...
...but a model probably has to predict something to be useful for anything.
At first, comparing predictions to observational data seems straightforward...

- Theory generates comparative static predictions in the form:
  \[
  \frac{\partial y}{\partial x} > 0 \quad \text{or} \quad \frac{\partial y}{\partial x} < 0 \quad \text{or} \quad \frac{\partial y}{\partial x} = C
  \]

- We estimate a linear model on data and determine:
  \[
  \frac{\partial y}{\partial x} = \frac{2}{2x} (\beta_0 + \beta_1 x + \hat{\beta}_2 z) = \hat{\beta}_1
  \]

- Maybe we can just compare the empirical derivative to the comparative static prediction and look for a match!

\[
\text{does } \hat{\beta}_1 \text{ look like } \frac{\partial y}{\partial x} ?
\]

Upon reflection, this is a bit harder than you might think. Suppose we find:

\[
\frac{\partial y}{\partial x} = \beta_1 = C + \varepsilon, \varepsilon \sim \mathcal{N}.
\]

Would we reject the hypothesis that \( \beta_1 = C \)? If not, two reasons:

- **Probably not**
  1. The difference between \( C \) and \( C + \varepsilon \) is not large & important enough to matter. This is an assessment of the substantive significance of the difference between \( C \) and \( C + \varepsilon \).
Our estimate of $\hat{\beta}_i$ is imperfect and intrinsically random. As a result, statistically there is no difference between $c$ and $c + \epsilon$. This is an assessment of the statistical significance of the difference between $c$ and $c + \epsilon$. 


Classical hypothesis testing focuses on the second reason, a result's *statistical significance*.

The majority of work in statistics (and our topic today) is about establishing statistical significance and for our purposes it’s only important to recognize that this is different from substantive significance.

The reasons why we need a test of statistical significance...

Our estimate of \( \hat{\beta} \) is random because the error term \( u \) is random, or possibly because \( \beta \) is itself, random (though this possibility is not a part of the classical exposition of hypothesis tests).

Randomness in \( u \) → that our precise estimates of \( \hat{\beta} \) will depend on precisely what sample (and consequently which values of \( u \)) that we are example.
Why is randomness of our estimate of $\tilde{\beta}$ problematic?
The average value of $\hat{\beta}$ is greater than zero, but there is some proportion of the time that we will observe a $\approx 0$ or even a negative value of $\hat{\beta}$ ... could give misleading results!
Presume that we have a theoretical prediction that $\beta > 0$. There are four possibilities:

<table>
<thead>
<tr>
<th>State of the world</th>
<th>Estimated Effect</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta &gt; 0^* \neq \hat{\beta}$</td>
<td>$\beta &gt; 0$</td>
<td>true positive</td>
</tr>
<tr>
<td>$\beta &gt; 0$</td>
<td>$\beta \leq 0^*$</td>
<td>false negative</td>
</tr>
<tr>
<td>$\beta \leq 0^*$</td>
<td>$\beta &gt; 0^*$</td>
<td>false positive</td>
</tr>
</tbody>
</table>
| $\beta < 0$     | $\hat{\beta} < 0$ | true negative

- Type I error
- Type II error
Statistical tests are generally designed to minimize false positives at the expense of false negatives... why?

Idea: better to ignore a correct hypothesis/theory and miss out on opportunities/knowledge/policy than to falsely accept a theory and have misguided conclusions that destructive policy and wasted research effort.

When might another set of tradeoffs apply?

pregnancy test

We can show the tradeoff between power (ability to avoid false negatives) and size (ability to avoid false positives) in an explicit analysis, and will do so later in the lecture.

next time.
Null: $\beta \leq 0$, Alternative: $\beta > 0$

Only accept the alternative if $Pr(\hat{\beta} \geq \hat{\beta}_0 | \beta = 0) \leq \alpha$, where $\alpha$ is some critical value that gives the chance of a Type I error / false positive

Conventionally, $\alpha = 0.05$

The probability of a false negative is never assessed—and it can be different from case to case even when $\alpha$ is constant!

- We'll return to that shortly.
The distribution of $\hat{\beta}$ under the null

One little problem: how did we draw $f(\hat{\beta}|\text{null})$ in the previous graph?

Answer: we don’t necessarily know the distribution (at least, I haven’t told you anything that about it yet), but we can write down a statistic for which we do know

Alternative: we can write down a statistic for which we know the distribution:

$$
Z = \frac{\hat{\beta} - \beta_{\text{null}}}{\sqrt{\text{var}(\hat{\beta})}} \sim \text{se}(\hat{\beta})
$$

Recall: $\text{var}(\hat{\beta}) = \sigma^2 (\mathbf{x}^\prime \mathbf{x})^{-1}$, VCV of OLS

$\hat{\beta}_i$, $\text{var}(\hat{\beta}_i)$ = its diagonal element VCV.

The z-statistic is a standardization of the distance between the observed and null prediction of $\beta$

$$
X_S = \frac{X - \mu_X}{\sigma_X}
$$

$X_S \sim \mu = 0, \sigma = 1$

We can prove that we know the distribution of the z-statistic, for any sample size.
\[ z = \frac{\hat{\beta} - \beta_0}{\sqrt{\text{var } \hat{\beta}}} \]

\[ z = \left( \sigma^2 (x'x)^{-1} \right)^{-1/2} (x'x)^{-1} x' \left[ y - \beta_0 \right] \]

\[ z = \sigma^{-1} (x'x)^{-1/2} (x'x)^{-1} x' \left( \beta_0 + u - \beta_0 \right) \]

\[ = \sigma^{-1} (x'x)^{-1/2} (x'x)^{-1} x' u \]

\[ = \sigma^{-1} (x'x)^{-1/2} \text{ constant.} \]

\[ \downarrow \text{ nonstochastic} \]

\[ \Rightarrow \text{ random} \]

Make an assumption about \( f(u) \): \( u \sim \Phi(0, \sigma^2) \).

If true, \( z \sim \Phi(0, 1) \).
When we assume that \( u \sim \Phi(0, \sigma^2) \), we now have the *Classical Normal Linear Regression Model*

- Important note: up to this point, we have referred to the CLRM -- the CLRM = the CLNRM without the normal error assumption

- The \( z \)-statistic and all tests that flow from similar ideas depend on CLNRM assumptions (at least, in small samples); hence hypothesis tests can be misleading if (all) of these assumptions do not hold
Problem with the z statistic: it assumes a known $\sigma$ and we don't actually know $\sigma$

Recall the formula we derived for estimating $\hat{\sigma}$

$$z = \left( x'x \right)^{-1/2} x' u$$

$$\hat{\sigma}^2 = \frac{1}{n-k} \; \hat{u}' \hat{u}$$

$$t = \left( x' x \right)^{-1/2} (x' u) \left[ \frac{1}{n-k} \; \hat{u}' \hat{u} \right]^{-1/2}$$

It too depends on the CLNRM

It turns out that any quantity $\frac{x}{y}$, where $x \sim \Phi(0,1)$ and $y \sim \chi^2(m)$, and where $x$ and $y$ are statistically independent, takes the $t(m)$ distribution

So we have:
\[ t = \left( X'X \right)^{-\frac{1}{2}} \left( X'u \right) \left[ \frac{1}{n-k} \hat{u}'\hat{u} \right]^{-\frac{1}{2}} \]

- The red quantity is \( \Phi(0,1) \), as we saw for the z formula
- The blue quantity is \( \chi^2 \) distributed \( \chi^2 \) because any quantity of the form \( \nu' \nu \) (where \( \nu \) is \( m \times 1 \)) is \( \chi^2 \) (\( m \)) if \( \nu \sim \Phi(0,1) \).

As \( n \to \infty \), the \( t \) distribution approaches \( \Phi(0,1) \)
What this all boils down to is: to conduct a test for statistical significance for a particular $\beta_i$ coefficient, if one is willing to accept the CLNRM assumptions in a small sample, one can calculate:

$$t = \frac{\hat{\beta}_i - \beta_0}{\sqrt{\text{var}(\hat{\beta}_i)}}$$

- We can then use the $t$ distribution to calculate
  $$\Pr(t(\hat{\beta}_i) \text{ or larger}|\beta_0) = \text{p-value}$$
- This means we can effectively compare $t(\hat{\beta}_i)$ to a critical value $t_c$ such that $\Pr(t_c|\text{null}) = \alpha$
- If $\alpha = 0.05$ then $t_c = 1.645$
- If $\alpha = 0.025$ then $t_c = 1.96$ which puts a total of 0.05 probability in both tails (0.025 in each tail)
- If $t(\hat{\beta}_i) > t_c$, then $\hat{\beta}_i$ is statistically significant
- Remember: $\text{var}(\hat{\beta}_i)$ comes out of the $i^{th}$ element of the VCV matrix,
  $$\hat{\sigma}^2 (X'X)^{-1} = \frac{1}{n-k} \hat{u}'\hat{u}$$
- As $n \to \infty$, t-testing is equivalent to z-testing, so Stata and R always report $t$ statistics
- R and Stata do most of this for you
  - R: summary command
  - Stata: regress command prints the relevant statistics
- Examples!
What if multiple $\beta$ values (or all the $\beta$ coefficients simultaneously) are tested for statistical significance?

For a model $y = X_1\beta_1 + X_2\beta_2 + u$, where $X_1$ and $X_2$ are $n \times k_1$ and $n \times k_2$ blocks of variables:

- Null hypothesis: $\beta_2 = 0$
- Alternative hypothesis: at least one element of $\beta_2 \neq 0$

The F-test compares the residuals from two regressions:

1. $\hat{y}_1 = X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 + v$  unrestricted
2. $\hat{y}_1 = X_1 \hat{\beta}_1 + u$  restricted

The F-statistic is:

$$F = \frac{\hat{u}'\hat{u} / (n - k_2)}{\hat{v}'\hat{v} / (n - k_1)} = \frac{\hat{u}'\hat{u} - \hat{v}'\hat{v} / r}{\hat{v}'\hat{v} / (n - k)}$$

$r = \# \text{ of restrictions}$

...this depends on the CLNRM assumptions because we need $u$ and $v$ to be normally distributed in order for their sum of squares to be distributed $\chi^2$. 
F distribution PDF, n-k=30

Examples!