Announcements:

- Extra office hours for final:
  - Fri, Dec 3: Jason will have 1-3 (usual is 1-2).
  - Mon, Dec 6: 10-noon (my office or note on door)
- Pick up old hw and exams in office hrs
- Bring #2 pencil, 4 sheets of notes, calculator.
- Final exam will consist of:
  - 30 to 35 multiple choice on all material, 2 points each
  - 30 to 40 points of free response on new material only

Homework (not to hand in, solutions posted):
12.73, 12.80, 13.50, 13.51, 17.2, 17.14

Cautions, Interpretations, and Other Useful Information

Sections 12.5, 13.5, 13.6,
A little bit from 13.7 (on power),
13.8 and Chapter 17

12.5 Statistical Significance vs Practical Importance

Cautions about Sample Size and Statistical Significance:

- A *small to moderate effect* in the population is hard to detect. With a small sample, the result has little chance of being statistically significant.
- With a **large sample**, ***even a small and unimportant effect*** in the population ***may be statistically significant***.

Type 2 Errors and Power (review)

When the alternative hypothesis is true, the probability of making the correct decision is called the **power** of a test.

Factors that affect probability of a type 2 error and power:

1. **Sample size**; larger n increases power without affecting the probability of a type 1 error.
2. **Level of significance**; larger α increases power, by increasing the probability of a type 1 error. So there is a trade-off.
3. **Actual value of the population parameter**; This is not in the researcher's control. The farther the truth falls from the null value (in H_a direction), the lower the probability of a type 2 error, and the higher the power.

Example

- In a test for ESP with 4 choices: p, the proportion correct by chance alone (in the long run), should be .25. So we would test:
  H_0: p = .25 (no ESP) vs H_a: p > .25 (ESP)
- Suppose the *truth* is p = .33. What is the *power* of the test, i.e. the probability that we *correctly* reject the null hypothesis? ($\alpha = .05$)

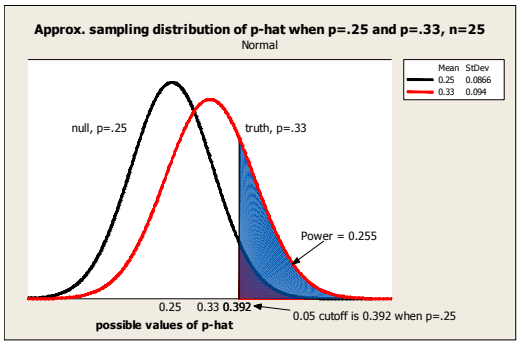
Sample Size	25	50	100	1000
Power	.25	.38	.57	.9999

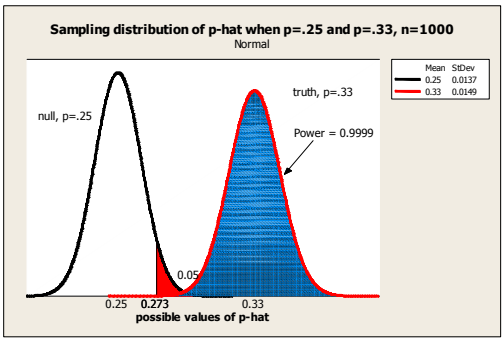
Why does this happen?

- Power depends on:
  - Sample size
  - Level of significance used
  - Actual difference between null value and truth
- How it depends on sample size:
  In ESP example, reject H_0 if $z > 1.645$ where

$$z = \frac{\hat{p} - .25}{\sqrt{\frac{(.25)(.75)}{n}}}$$

n	25	1000
s.e.	.09	.014





Real Importance versus Statistical Significance

- The *p*-value does ***not*** provide information about the ***magnitude*** of the effect.
- The **magnitude** of a statistically significant effect can be **so small** that the **practical effect is not important**.
- If sample size **large** enough, almost **any null hypothesis can be rejected**.
- It's best to find a confidence interval too – it provides the estimated magnitude.

Example 12.9 *Birth Month and Height*

Headline: ***Spring Birthday Confers Height Advantage***

Austrian study of heights of 507,125 military recruits.
Results were highly statistically significant (tiny p-value), test of difference in means, independent samples

Men born in spring were, on average, about 0.6 cm taller than men born in fall, i.e. about 1/4 inch (Weber et al., *Nature,* 1998, 391:754–755).

Sample size so large that **even a very small difference** was ***statistically significant***.

13.5 Relationship Between Tests and Confidence Intervals

For two-sided tests (for one or two means):
H_0: parameter = null value and H_a: parameter ≠ null value

- If the null value is ***covered*** by a $(1 - \alpha)100\%$ confidence interval, the null hypothesis is **not rejected** and the test is **not statistically significant** at level α.
- If the null value is ***not covered*** by a $(1 - \alpha)100\%$ confidence interval, the null hypothesis is **rejected** and the test is **statistically significant** at level α.

Note: 95% confidence interval ⇔ 5% significance level
99% confidence interval ⇔ 1% significance level

Example 13.7 *Mean TV hours (M vs F)*

Question: Does the population mean daily TV hours differ for male and female college students?

Sex	N	Mean	StDev	SE Mean
Male	59	2.37	1.87	0.24
Female	116	1.95	1.51	0.14

95% CI for mu (Male) − mu (Female): (−0.14, 0.98)
T-Test of mu (Male) − mu (Female) (vs not =): T = 1.49 P = 0.140 DF = 97

95% CI for difference in population means: **(-0.14, +0.98)**
Test H_0: $\mu_1 - \mu_2 = 0$ versus H_a: $\mu_1 - \mu_2 \neq 0$ using $\alpha = 0.05$
The **null value of 0 hours is in this interval**.
Thus the difference in the sample means of 0.42 hours is not significantly different from 0.

Note this does *not* mean we conclude difference = 0!

Confidence Intervals and One-Sided Tests

When testing the hypotheses:
H_0: parameter = null value versus a ***one-sided*** alternative, compare the null value to a $(1 - 2\alpha)100\%$ confidence interval:

- If the **null value is covered by the interval**, the test is **not statistically significant** at level α.
- For the alternative **H_a: parameter > null value**, the test is **statistically significant** at level α if the **entire interval falls *above* the null value**.
- For the alternative **H_a: parameter < null value**, the test is **statistically significant** at level α if the **entire interval falls *below* the null value**.

Example 13.8 *Ear Infections and Xylitol*

Comparing proportions of kids with ear infections taking placebo and xylitol:

95% CI for $p_1 - p_2$ is 0.020 to 0.226

Reject H_0: $p_1 - p_2 = 0$ and accept H_a: $p_1 - p_2 > 0$ with $\alpha = 0.025$, because the entire confidence interval falls *above* the null value of 0.

Note that the *p*-value for the test was 0.01, which is less than 0.025.

13.6 Choosing an Appropriate Inference Procedure

- **Confidence Interval or Hypothesis Test?**
  Is the main purpose to *estimate* the numerical value of a parameter **or** to *make a "maybe not/maybe yes" conclusion* about a specific hypothesized value for a parameter?
- If there is no obvious/natural null value, use C.I.
- Often it makes sense to do both. They give different information.

Determining the Appropriate Parameter

(See examples, pgs 579-80)

- Is the response variable *for each unit* categorical (yes, no; agree, don't agree; etc.) or quantitative (height, IQ, weight gain, etc.)?
- Is there one sample or two?
- If two, independent or paired?

Variable type (parameter type)	One sample (No pairing)	Paired Data	Two independent samples
Categorical (Proportions)	p	*none*	$p_1 - p_2$
Quantitative (Means)	μ	μ_d	$\mu_1 - \mu_2$

13.8 Evaluating Significance in Research Reports

Tips for reading about studies in the news.

- If a study is important to you, try to find the original journal article or contact the researcher for more information.
- Use the *p*-value to make your *own* decision, based on severity of type 1 and type 2 errors.

Tips for reading about studies in the news, continued...

- If a study reports "no difference" or "no relationship" find out the sample size and see if it has low power.
- Be careful about interpreting "significant" effects based on large samples.
- If possible, get a confidence interval to go with a test.
- Determine whether **multiple testing** was a problem – if enough tests are done, some will be significant *by chance alone* (5% of all *true* null hypotheses).

Chapter 17: Turning Information into Wisdom

What Has Statistics Done for Us?

Nearly every area of knowledge has been advanced by statistical studies. But you must use them wisely!

Information developed through the use of statistics has …

- enhanced our understanding of how life works,
- helped us learn about each other,
- allowed control over some societal issues, and
- helped individuals make informed decisions

How/When Can We Go Beyond the Data in the Sample?

There are two important questions:

- *Can we **extend** the results from a sample **to a population**?*
- *Can we make a **cause and effect conclusion**?*

Random, Representative, or Restrictive Sample?

Quote from a statistics book: "Inferences to ***populations*** can be drawn from random sampling studies, but not otherwise."

But true random samples almost ***impossible*** to obtain.

Fundamental Rule for Using Data for Inference is that available data can be used to make inferences about a much larger group *if the data can be considered to be representative with regard to the question(s) of interest.*

Randomized Experiments, Observational Studies, and Causal Conclusions

The most common error by media:
Conclude a ***causal relationship*** has been established, when it is *not* warranted by the way the study was conducted.

Rule for Concluding Cause and Effect is that cause-and-effect relationships can be inferred from randomized experiments, but *not* from observational studies.

Using Non-statistical Considerations to Assess Cause and Effect

Here are some hints that *may suggest* **cause and effect from observational studies** (but it would need to be verified by non-statistical methods, or randomized experiment):

- There is a ***reasonable explanation*** of cause and effect. (Ex: High-fat diet raises heart attack risk)
- Connection happens under ***varying conditions*** in a number of studies. (Ex: smoking & lung cancer)
- Potential ***confounding variables*** are ***ruled out*** by measuring and analyzing them. (Ex: Mom smokers and lower infant IQ – looked at education, etc.)

17.2 Some reasons statistical methods are useful:

- As individuals, we need to ***make personal decisions***.
- As a society, we want to ***have some control*** over things.
- As intelligent and curious beings, we want to ***understand*** things.
- As social and curious beings, we want to ***know about other people***.

17.3 Making Personal Decisions

Think about decisions in framework of hypothesis testing, and consider consequences of errors.

H_0: I will be better off if I ***take no action***.

H_a: I will be better off if I ***do take action***.

- **Type 1 error**: taking action when you would have been better off not doing so.
- **Type 2 error**: taking no action when you would have been better off taking action.

Ex: *Should you take daily aspirin?*

As we have seen, statistical studies show that taking an aspirin a day may lower the risk of heart disease.

H_0: Aspirin will not help you
H_a: Aspirin will lower your risk of heart disease

	Likelihood	H_0 chosen	H_a chosen
H_0 true	Probably depends on genetics, diet, etc.	No gain or loss	Side effects of aspirin
H_a true		Could die of heart disease	Could save your life

You must assess the likelihood of each hypothesis and weigh the possible consequences of each choice before you decide.

17.4 Control of Societal Risks

•Statistical studies are used to guide policy decisions, make legal decisions, etc.

•For example, randomized experiments are used by the FDA to decide what drugs to approve.

•Lawmakers, government regulatory agencies, and other decision makers must weigh decisions and their consequences.

Example 17.4 *Older Drivers and Vision*

Efforts to improve driving safety often are based on results of statistical studies. Examples include requiring the wearing of seat belts, banning drinking and driving, and banning cell phone use while driving.

Laws passed to help protect drivers and passengers based on findings. But there is a trade-off between protection and personal freedom.

Article: "Visual Field Loss Ups Elderly Car Crashes"

"A 40% loss in range of vision among older drivers more than doubles their risk for a car accident"

What should lawmakers do? This is an observational study. Multiple testing might be a problem. Etc….

17.5 Understanding Our World

- Statistical studies are done to help us understand ourselves and our world, without involving any decisions.
- Scan any major news source and find reports of many interesting studies done to help us understand the world.
- Many studies are exploratory in nature and results are controversial. That's part of why they make interesting news. You should now be able to interpret and understand them better!

Example 17.7 *Gender and Memory*

Memory is a fascinating ability, and most of us wish we had more of that ability.

Studies about memory may eventually help us understand ways to improve it.

Article: "Women Remember Item Location Better"

"When it comes to memory, women have more skill than confidence, and men have more confidence than skill."

Based on experiment using computer-based tests of 300 healthy men and women, done at University of Florida.

Reported in Yahoo Health News.

17.6 Getting to Know You

- We are curious about how others think and behave.
- What do they do with their time?
- Are we in the majority with our opinions on controversial issues?
- Are people basically honest?
- Many questions are answered by surveying random or representative samples.
- Most national governments have agencies that collect samples to answer some of these questions on a routine basis.

Lifestyle Statistics

U.S.Census Bureau collects data on many aspects of American life. Ongoing "***Current Population Survey***" polls a random sample of U.S. households on wide variety of topics. **Trends in lifestyle decisions can be tracked over the years.**

Median age at first marriage in 2009 was 28.1 for men and 25.9 for women, *highest* since reporting began in 1890.

Lowest ages were in *1956*: 22.5 for men, 20.1 for women.

Biggest difference occurred in *1890*: median of 26.1 for men, 4.1 years older than the median age of 22.0 for women.

Since 1948 the gap in ages has always been under 3 years, but median has always been higher for men than women.

17.7 Ten Guiding Principles

1. A **representative sample** can be used to make inferences about a larger population, but descriptive statistics are *the only useful results* for an unrepresentative sample.
2. **Cause and effect** can be inferred from randomized experiments, but not from observational studies, where confounding variables are likely to cloud the interpretation.
3. A **conservative estimate of *sampling* error** in a survey is the margin of error $1/\sqrt{n}$. Provides a bound on the difference between true proportion and sample proportion that holds for at least 95% of properly conducted surveys.
4. The **margin of error does *not* include nonsampling error**, such as errors due to biased wording, nonresponse, etc.
5. When the **individuals measured make up the whole population**, there is no need for statistical inference because the truth is known.
6. A **significance test based on a very large sample** is likely to produce a ***statistically significant result*** even if the true value is close to the null value. Wise to examine the magnitude of the parameter with a confidence interval to determine if result has ***practical importance***.

7. A **significance test based on a small sample** may not produce a statistically significant result even if true value differs substantially from null. That's why it's important not to *accept* a null hypothesis.
8. When deciding how readily to reject the null hypothesis (**what significance level to use**), important to consider consequences of type 1 and type 2 errors. If a type 1 error has serious consequences, the level of significance should be small. If a type 2 error is more serious, a higher level of significance should be used.

9. ***Examining many hypotheses could find one or more statistically significant results just by chance***, so find out how many tests were conducted when you read about a significant result. It's common in large studies to find that one test attracts media attention, so it's important to know if that test was the only one out of many conducted that achieved statistical significance.

10. Sometimes you will read that researchers were ***surprised to find "no effect"*** and ***study "failed to replicate" an earlier finding of statistical significance***.

    Possible explanations: Sample size too small and the test had low power.

    Or, result in first study was a type 1 error -- likely if the effect was moderate and was part of larger study that covered multiple hypotheses.

And the final word….

If you learn and understand only what was covered in this lecture today

you will be way ahead of most of the population,

and you will be able to make much more informed decisions throughout your life!