Significance Tests

- **Example 1: Are college freshmen getting more stressed?**
- **A 2010 UCLA Higher Education Research Institute poll took a random sample of 201,818 college freshmen.**
- **29.1% of the sampled freshmen reported being frequently "overwhelmed."**
- **In a similar poll from 2009, the percentage overwhelmed was 27.1%.**
- **Has the true proportion of freshmen who are overwhelmed increased from 2009 to 2010?**
- **Or is the increase in overwhelmed percentage in 2010 simply attributable to random chance?**

Clicker Quiz 1

In the UCLA stress example, the sample size was very large (201,818 freshmen in 2010, and a similar number in 2009). Because of this, what can we conclude?

A. The true proportions of stressed freshmen should be similar to the respective sample proportions of stressed freshmen.

B. The sample proportions are untrustworthy since the sample sizes are so large.

C. The poll is actually a census of all college freshmen in America.

D. None of the above are true.

Another Significance Test Example

- **Example 2: A sample of coffee drinkers undergoes a taste test to determine whether fresh-brewed coffee or instant coffee is preferred.**
- **The experimenter suspects that more of the population prefers fresh-brewed coffee.**
- **In math notation, we could represent this belief as: $p > 0.5$.**
- **Here, p is the proportion of the *whole population* that prefers fresh-brewed.**
- **In our sample, suppose 36 out of 50 prefer fresh-brewed. That is, our $\hat{p}$ is 36/50 = 0.72.**
- **Is this sample proportion large enough that we would conclude that the population $p > 0.5$?**

Reasoning of Tests of Significance

- **Recall that the sample result in Example 2 was that 36 out of 50 sampled drinkers preferred fresh-brewed.**
- **The significance test in Example 2 asks the question: "If the true proportion preferring fresh-brewed is no more than 0.5 (say, if it equals 0.5), then is it *unlikely* that we'd see a sample result like we did?"**
- **We can answer this question based on our knowledge of the *sampling distribution* of $\hat{p}$.**
- **When we have a fairly large sample size, the *sampling distribution* of $\hat{p}$ is roughly Normal with a mean equal to the true p and a standard deviation of**

$$\sqrt{\frac{p(1-p)}{n}}.$$

Sampling Distribution of $\hat{p}$

- **Let's pretend that in fact $p = 0.5$.**
- **The sampling distribution gives us the probability of seeing as extreme or more extreme of a value of $\hat{p}$ than the one we actually did see.**
- **The term "extreme" here means "favorable toward what the experimenter is trying to show."**
- **In this example, he's trying to show that $p > 0.5$.**

Sampling Distribution of $\hat{p}$

- **If $p = 0.5$, then the probability of seeing a $\hat{p} \geq 0.72$ is 0.001 (really small!)**
- **Conclusion: "If the true proportion preferring fresh-brewed equals 0.5, then it is *VERY unlikely* that we'd see a sample result as extreme as we did!"**
- **Are you convinced that p, the true proportion preferring fresh-brewed, is actually bigger than 0.5?**
- ***What if* our sample proportion had been 28/50 = 0.56? Then if $p = 0.5$, the probability of seeing a $\hat{p} \geq 0.56$ would be 0.20. Is that unlikely enough to convince you that p is really bigger than 0.5?**

Two Types of Hypothesis

- **The *null hypothesis*, denoted H_0, is the claim we are testing.**
- **This is usually some statement of “no effect,” “no difference,” or “previous belief.”**
- **In our examples, the null hypothesis always has an equals sign!**
- **The *alternative hypothesis*, denoted H_a, is what we are seeking evidence for.**
- **This is usually some statement of “some new effect,” “some difference,” or “suspected belief.”**
- **The alternative hypothesis contains one of these signs: $< \quad > \quad \neq$.**
- **These hypotheses are ALWAYS statements about a parameter, not about a statistic!**

Clicker Quiz 2

In the coffee example, what is the null hypothesis?

A. $\boldsymbol{p} = 0.5$

B. $\boldsymbol{p} > 0.5$

C. $\hat{\boldsymbol{p}} = 0.5$

D. $\hat{\boldsymbol{p}} > 0.5$

Clicker Quiz 3

In the coffee example, what is the alternative hypothesis?

A. $\boldsymbol{p} < 0.5$

B. $\boldsymbol{p} > 0.5$

C. $\boldsymbol{p} \neq 0.5$

D. $\hat{\boldsymbol{p}} > 0.5$

P-value

- **Recall that our decision about whether H_0 was true was based on the probability of observing a sample outcome *as extreme or more extreme* as the one we did observe, if H_0 were in fact true.**
- **"Extreme" here means "favorable toward the alternative hypothesis".**
- **This probability is called the *P-value* of the significance test.**
- **The smaller the P-value is, the stronger the evidence against H_0.**
- ***Coffee example*: If $\hat{p} = 0.72$, then the P-value is 0.001 (really small!). If $\hat{p} = 0.56$, then the P-value is 0.20 (not so small).**

Statistical Significance

- **We will let software or calculators calculate the P-value for us, so we won't worry about how to calculate it.**
- **We do need to interpret the P-value and decide whether it is small enough to make us reject the null hypothesis in favor of the alternative.**
- **Our *significance level* (denoted α) is the cutoff value for deciding whether the P-value is "small enough."**
- **If our P-value is less than or equal to α, we will *reject* H_0 and say the data are "statistically significant."**
- **If our P-value is greater than α, we will *not reject* H_0 and we say the data are "not statistically significant."**

Clicker Quiz 4

We test $H_0 : \boldsymbol{p} = 0.5$ against $H_a : \boldsymbol{p} < 0.5$, using a significance level of α = 0.05. Our p-value turns out to be 0.036. What is our conclusion?

A. The data are statistically significant; reject H_0 and conclude $\boldsymbol{p} = 0.5$.

B. The data are statistically significant; reject H_0 and conclude $\boldsymbol{p} < 0.5$.

C. The data are NOT statistically significant; do not reject H_0 and we conclude $\boldsymbol{p} = 0.5$.

D. The data are NOT statistically significant; do not reject H_0 and we conclude $\boldsymbol{p} < 0.5$.

Clicker Quiz 5

We test $H_0 : \boldsymbol{p} = 0.5$ against $H_a : \boldsymbol{p} < 0.5$, using a significance level of α = 0.05. Our p-value turns out to be 0.087. What is our conclusion?

A. **The data are statistically significant; reject H_0 and conclude $\boldsymbol{p} = 0.5$ is possible.**

B. **The data are statistically significant; reject H_0 and conclude $\boldsymbol{p} < 0.5$.**

C. **The data are NOT statistically significant; do not reject H_0 and we conclude $\boldsymbol{p} = 0.5$ is possible.**

D. **The data are NOT statistically significant; do not reject H_0 and we conclude $\boldsymbol{p} < 0.5$.**

Meaning of Statistical Significance

- Recall that "statistically significant" means that such a result is "unlikely to have occurred by chance" (if H_0 were true).
- "Statistically significant" does not mean "practically important." (more about this in the next chapter)

Meaning of Statistical Significance (Continued)

- **Usually scientists choose a significance level such as 0.05 or 0.01.**
- **In a report, it's best to report the actual P-value; then any reader can decide whether it is "small enough" for her purposes.**
- **Sometimes as a rule of thumb, people say a P-value $<$ 0.10 indicates "some evidence against H_0."**
- **A P-value $<$ 0.05 indicates "moderate evidence against H_0."**
- **A P-value $<$ 0.01 indicates "strong evidence against H_0."**
- **These are just loose guidelines, however.**

Different Types of Alternative Hypothesis

- **In the coffee example, our alternative hypothesis was $H_a : \boldsymbol{p} > 0.5$.**
- **This is a *one-sided* alternative because we only seek evidence that $\boldsymbol{p}$ is on one side of 0.5.**
- **An alternative hypothesis like $H_a : \boldsymbol{p} < 0.5$ is also a *one-sided* alternative.**

Different Types of Alternative Hypothesis (Continued)

- ***Example 2*: We toss a coin numerous times to determine whether it is balanced. If p is the probability of “heads”, we test $H_0 : \boldsymbol{p} = 0.5$ against $H_a : \boldsymbol{p} \neq 0.5$.**
- **This is a *two-sided* alternative because we seek evidence that p is *either* side of 0.5.**
- **We still make our conclusion by comparing the P-value to α in the usual way.**

Clicker Quiz 6

Which is a two-sided alternative hypothesis?

A. $\boldsymbol{p} = 0.5$

B. $\boldsymbol{p} < 0.5$

C. $\boldsymbol{p} > 0.5$

D. $\boldsymbol{p} \neq 0.5$

Clicker Quiz 7

In the coin-tossing experiment, we test $H_0 : \boldsymbol{p} = 0.5$ against $H_a : \boldsymbol{p} \neq 0.5$, using $\boldsymbol{\alpha}$ = 0.05. We toss a coin 50 times and get 21 heads. Our p-value turns out to be 0.52. What is our conclusion?

A. The data are statistically significant; reject H_0 and conclude the coin may be balanced.

B. The data are statistically significant; reject H_0 and conclude the coin is unbalanced.

C. The data are NOT statistically significant; do not reject H_0 and we conclude the coin may be balanced.

D. The data are NOT statistically significant; do not reject H_0 and we conclude the coin is unbalanced.

Significance Tests about a Population Mean

- **We can similarly perform a significance test about a population mean $\boldsymbol{\mu}$.**
- **Our null hypothesis is $H_0 : \boldsymbol{\mu} =$ [some number].**
- **The alternative hypothesis could be one-sided like $H_a : \boldsymbol{\mu} <$ [some number], or $H_a : \boldsymbol{\mu} >$ [some number].**
- **Or it could be two-sided like $H_a : \boldsymbol{\mu} \neq$ [some number].**
- **We take a random sample and calculate the sample mean $\bar{x}$ (recall $\bar{x}$ has a normal sampling distribution).**
- **If the value of $\bar{x}$ is so "extreme" that it is unlikely to have occurred when H_0 is true, then we reject H_0 in favor of H_a.**
- **Again, we reject H_0 when the P-value is less than or equal to α.**

Clicker Quiz 8

We test whether the mean SAT math score for Iowa students is greater than 500, so we test $H_0 : \mu = 500$ against $H_a : \mu > 500$, using α = 0.05. Based on a sample of 75 students yielding $\bar{x} = 523.4$, our p-value is 0.048. What is our conclusion?

A. The data are statistically significant; reject H_0 and conclude Iowa's mean math SAT score exceeds 500.

B. The data are statistically significant; reject H_0 and conclude Iowa's mean math SAT score may not exceed 500.

C. The data are NOT statistically significant; do not reject H_0 and we conclude Iowa's mean math SAT score exceeds 500.

D. The data are NOT statistically significant; do not reject H_0 and we conclude Iowa's mean math SAT score may not exceed 500.

Clicker Quiz 9

The mean systolic blood pressure of males aged 35-44 is 128. Is this mean SBP different for executives? We test $H_0 : \mu = 128$ against $H_a : \mu \neq 128$, using α = 0.05. Based on a sample of 72 executives yielding $\bar{x} = 126.1$, our p-value is 0.29. Conclusion?

A. The data are NOT statistically significant; do not reject H_0 and we conclude the mean SBP for executives may be 128.

B. The data are statistically significant; reject H_0 and conclude the mean SBP for executives is not 128.

C. The data are statistically significant; reject H_0 and conclude the mean SBP for executives is less than 128.

D. The data are statistically significant; reject H_0 and conclude the mean SBP for executives is greater than 128.