Lecture 13

Last time we went through the steps of hypothesis testing specific to testing proportions:

Step 1: Assumptions

- The data are obtained using randomization
- The sample size is large enough that the sampling distribution of $\hat{p}$ is approximately normal

Step 2: Hypotheses

One-sided alternative hypothesis:

$$H_o : p \leq p_o \; VS. \; H_a : p > p_o \tag{1}$$

$$H_o : p \geq p_o \; VS. \; H_a : p < p_o \tag{2}$$

Two-sided alternative hypothesis:

$$H_o : p = p_o \; VS. \; H_a : p \neq p_o \tag{3}$$

Step 3: Test statistic

$$Z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} = \frac{\textit{sample proportion} - \textit{null hypothesis proportion}}{\textit{standard deviation when null hypothesis is true}} \tag{4}$$

This Z-score measures the number of standard errors between the sample proportion $\hat{p}$ and the null hypothesis p_o.

Step 4: P-value

The **P-value** is the probability that the test statistic takes a value like the observed test statistic or even more extreme, if the null hypothesis is true.

Hypothesis	**P-value**
$H_o : p \leq p_o \; VS. \; H_a : p > p_o$	$P(Z > \textit{observed test statistic})$
$H_o : p \geq p_o \; VS. \; H_a : p < p_o$	$P(Z < \textit{observed test statistic})$
$H_o : p = p_o \; VS. \; H_a : p \neq p_o$	$P(\lvert Z \rvert > \textit{observed test statistic})$

Step 5: Conclusion

If the P-value is small, then it is unlikely (the probability is small) to observe our sample data taken randomly from a population with the null hypothesized parameter value. Therefore, it is unlikely that the null hypothesized parameter value is correct so "reject" H_o.
If the P-value is large, then it is likely (the probability is large) to observe our sample data taken randomly from a population with the null hypothesized parameter value. Therefore, it is likely that the null hypothesized parameter value is correct so we "fail to reject" H_o.

Interpreting the P-value

A significance test analyzes the strength of the evidence against the null hypothesis, H_o. We start by presuming that H_o is true, putting the *burden of proof* of H_a. The approach taken is the indirect on of *proof by contradiction*. To convince ourselves that H_a is true, we must show that the data contradict H_o, by showing they'd be unusual if H_o were true. We analyze whether the data would be unusual if H_o were true by finding the P-value. (see Step 5 on previous page)

Two-sided significance tests

Sometimes we're interested in investigating whether a proportion falls above or below some point. *Example*: Can we conclude whether the population proportion who voted for a particular candidate is above 1/2 or below 1/2?

$$H_o : p = \frac{1}{2} \; VS. \; H_a : p \neq \frac{1}{2}$$

For two-sided tests, the values that are more extreme than the observed test statistic value are ones that fall farter out in the tails in <u>either</u> direction. The P-value is a "two-tailed" probability under the standard normal curve, because these are the test statistic values that provide even stronger evidence in favor of $H_a : p \neq p_o$ than the observed value. We calculate this by finding the tail probability in a single tail and then doubling it, since the distribution is symmetric.

Example: A recent study investigated whether dogs can be trained to distinguish a patient with bladder cancer by smelling certain compounds released in the patient's urine (C.M. Willis, *British Medical Journal*, 329, 2004). Six dogs of varying breeds were trained to discriminate between urine from patients with bladder cancer and urine from control patients without it. The dogs were taught to indicate which among several specimens was from the bladder cancer patient by lying beside it.

Each of the six dogs was tested with nine trials. In each trial, one urine sample from a bladder cancer patient was randomly placed among six control urine samples. In the total of 54 trials with the six dogs, the dogs made the correct selection 22 times.

Let p denote the probability that a dog makes the correct selection on a given trial.

Step 1:

Step 2:

Step 3:

Step 4:

Step 5:

Comparing the P-value to the significance level

Sometimes we need to make a solid decision about whether the data provide sufficient evidence to reject H_o. BEFORE SEEING THE DATA, we decide how small the P-value would need to be to reject H_o. We might decide that we will reject H_o if the P-value ≤ 0.05. The cutoff point of 0.05 is called the **significance level**.

P-value	**Decision about** H_o
≤ 0.05	Reject H_o
> 0.05	Fail to reject H_o

We either "reject H_o" or "fail to reject H_o". If the P-value is larger than 0.05 then the data do not contradict H_o sufficiently for us to reject it. If P-value si less than or equal to 0.05 then the data provide enough evidence to "reject H_o". Recall that H_a had the *burden of proof* and in this case that the proof is sufficient. When we reject H_o we say the results are **statistically significant**.

Example:
Therapeutic touch (TT) practitioners claim to improve or hel many medical conditions by using their hands to manipulate a "human energy field" they perceive above the patient's skin. (The patient does not have to be touched). A test investigating this claim used the following experiment (Rosa, L. et al, *JAMA*, 279, 1998):
A TT practitioner was blind-folded. In each trial the researcher placed her hand over either the right or left hand of the TT practitioner, the choice being determined by flipping a coin. The TT practitioner was asked to identify whether his or her right or left hands was closer to the hand of the researcher. Let p denotethe probability of a correct prediction by a TT practitioner. With random guessing, $p = \frac{1}{2}$. However, the TT practitioners claimed that they could do better than random guessing. They claimed that $p > \frac{1}{2}$.
In a set of 150 trials with 15 TT practitioners (10 trials each), the TT practitioners were correct with 70 of their 150 predictions.
Step 1:

Step 2:

Step 3:

Step 4:

Step 5:

Why we cannot "accept" H_o

A small P-value means that the sample data would be unusual if H_o were true. If the P-value is not small, such as 0.79, then the null hypothesis is plausible. In the case, the conclusion is reported as "fail to reject H_o" since the data do not contradict H_o.

The population proportion has many plausible values besides the number in the null hypothesis. For instance, consider the TT example, with $p =$ the probability of a correct prediction by a TT practitioner. We did not reject $H_o : p = 0.50.$. Thus, p may equal 0.50, but other values are also believable. A 95% confidence interval for p is

$$\hat{p} \pm 1.96\sqrt{\tfrac{p(1-p)}{n}} = 0.467 \pm 1.96\sqrt{\tfrac{0.467(0.533)}{150}} = (0.39, 0.55)$$

Even though insufficient evidence exists to reject H_o, it is wrong to accept it since p could be plausibly be any value between 0.39 and 0.55.

We can say that we accept the alternative hypothesis H_a when, for a a sufficiently small P-value, the entire range of believable values falls within the range of numbers contained in H_a.

NOTE: A two-sided test coincides with the ordinary approach for confidence intervals, which are two-sided, obtained by adding and subtracting some quantity from the point estimate.