# Lecture 17

## Types of Errors in Hypothesis Tests Continued

### Significance tests are less useful than confidence intervals

- A significance test merely indicates whether the particular parameter value in $H_o$ (such as $\mu = 0$ is plausible.

When a P-value is small, the significance test indicates that the hypothesized value is not plausible, but it tells us little about which potential parameter values *are* plausible.

- A confidence interval is more informative because it displays the entire set of believable values.

A confidence interval shows how badly $H_o$ may be false by showing whether the values in the interval are far from the $H_o$ value. It helps us to determine whether the difference between the true value and the $H_o$ value has practical importance.

### How likely is a Type II error?

We now know that the probability of a Type I error is the significance level $\alpha$ of the test. Given that $H_o$ is really true, when $\alpha = 0.05$, the probability of reject it equals 0.05.

Given that $H_o$ is really false, a Type II error results from *not* rejecting $H_o$. This probability has more than one value, because $H_a$ contains a range of possible values for the parameter. Each value in $H_a$ has its own probability of a Type II error. We will learn how to find the probability of a Type II error at a particular value.

*Example*:
Recall the astrologer example. For each person's horoscope, an astrologer must predict which of three personality charts is the actual one. Let $p$ denote the probability of a correct prediction by an astrologer. Consider the test of $H_o p = 1/3$ against $H_a : p > 1/3$, using the 0.05 significance level. Suppose an experiment plans to use $n = 116$ people, as this experiment did.

For what value of the sample proportion can we reject $H_o$?

For testing $H_o : p = 1/3$ with $n = 116$, the standard deviation for the test statistic, Z, is

$$\sigma_{p_o} = \sqrt{\frac{p_o(1-p_o)}{n}} = \sqrt{\frac{(1/3)(2/3)}{116}} = 0.0438$$

For $H_a : p > 1/3$, a test statistic of $Z = 1.645$ has a P-value (right-tail probability) of 0.05. If $Z \geq 1.645$ then P-value $\leq 0.05$ and we can reject $H_o$.

Therefore, we reject $H_o$ when $\hat{p}$ falls at least 1.645 standard deviations above p=1/3:

$$\hat{p} \geq 1/3 \ + \ 1.645(\sigma_{p_o}) \ = \ 1/3 \ + \ 1.645(0.0438) \ = \ 0.405$$

We reject $H_o$ when $\hat{p} \geq 0.405$

The National Council for Geocosmic Research claimed that $p$ would be 0.50 or higher. If truly $p = 0.50$, for what values of the sample proportion would we make a Type I error, failing to reject $H_o$ even though it's false?

In this case $H_o$ is really false so a Type II error can occur when we fail to reject $H_o$. From the previous question we know that we do not reject $H_o$ if $\hat{p} < 0.405$ causing a Type II Error.

If truly $p = 0.50$, what is the probability that a significance test based on this experiment will make a Type II error?

If the true value of $p$ is 0.50, then the true sampling distribution of $\hat{p}$ is centered around 0.50 ($\mu_{\hat{p}} = 0.50$).

From the previous question we know that the probability of a Type II error is the probability that $\hat{p} < 0.405$ when $p = 0.50$. When $p = 0.50$, the standard deviation of $\hat{p}$ for a sample size of 116 is $\sqrt{\frac{0.5(0.5)}{116}} = 0.0464$.

For a normal sample distribution of $\hat{p}$ with mean 0.50 and standard deviation 0.0464, the $\hat{p}$ value of 0.405 has a Z-score of

$$Z = \frac{0.405 - 0.50}{0.0464} = -2.04$$

The left-tail probability below -2.04 for the standard normal distribution equals 0.0207. Therefore, when $p = 0.50$ the probability of making a Type II error and failing to reject $H_o : p = 1/3$ is only 0.02.

The probability of a Type II error increases when the true parameter value moves close to $H_o$.
*Example:* Find the probability of a Type II error when the true $p = 0.40$ instead of 0.50.

For a fixed significance level $\alpha$, P(Type II error) decreases:
- as the parameter value moves farther into the $H_a$ values and away from the $H_o$ values.
- as the sample size increases.

Also, recall that P(Type II error) increases as $\alpha$ decreases.

## The Power of a Test

When $H_o$ is false, you want the probability of rejecting it to be high. The probability of reject $H_o$ is called the **power** of the test. For a particular value of the parameter from the range of the alternative hypothesis values,

**Power = 1 - P(Type II error)**

In the astrology example where the true $p = 0.50$, the P(Type II error)=0.02 so the power of the test at $p = 0.50$ is 1-.02=0.98. The higher the power the better.

*Example:* Recall the Therapeutic Touch (TT) example. The data did not support the Tt practitioners' claim to be able to detect a human energy field. The P-value was not small for testing $H_o : p = 0.50$ against $H_a : p > 0.50$, where $p$ is the probability of a correct prediction about which hand was near the researcher's hand. The medical journal article about the study stated, "The statistical power of this experiment was sufficient to conclude that if TT practitioners could reliably detect a human energy field, the study would have demonstrated this." For the test of $H_o : p = 0.50$ with one of the sets of trials, the power was reported as 0.96 if actually P=2/3.

Interpret the probability of 0.96.

The power of 0.96 is the probability of correctly rejecting $H_o$ when it is false. If the actual probability of correct predictions by TT practitioners was 2/3, there was a 96% chance of data such that the significance test performed would reject $H_o$.

What is a Type II error in the context of this experiment?

A Type II error occurs if we do not reject $H_o : p = 0.50$, when actually TT practitioners *can* predict correctly more than half the time. The consequence would be to question the truthfulness of what they have been practicing for over 30 years, when they actually do have some ability.

If truly $p = 2/3$, what is the probability of a Type II error?

If $p = 2/3$, the value of P(Type II error)=1-(Power at $p = 2/3$)=1-0.96=0.04. A Type II error is unlikely if truly $p = 2/3$.