STAT 513 - Theory of Statistical Inference

Chapter 10: Hypothesis Testing

Recall: Statistical inference involves using sample data to make conclusions about a larger ____________.

- In point estimation and interval estimation, the goal is to estimate a parameter (a number that summarizes some aspect of a population).
- In hypothesis testing, the goal is to:

- Usually, this statement specifies a particular value of some parameter(s).

Example 1: We gather experimental data to test whether drug 1 yields the same blood pressure reduction as drug 2, on average.

Example 2: We gather sample data to test whether the proportion of defective parts at a factory is no more than 3%.

10.2 Elements of a Hypothesis Test

- Statistical hypotheses are statements about one or more parameters.
- We have two kinds of hypothesis:

① The ____________ hypothesis (or ____________ hypothesis, denoted H_a) represents a theory that the researcher suspects to be true, or seeks evidence to "prove")

② The ________ hypothesis (denoted H_0) is the ____________ of H_a (the opposite).

- H_0 often represents some "previously held belief", "status quo", or "lack of effect".

<u>Example 1 again:</u> The parameter of interest is:

<u>Example 2 again:</u> The parameter of interest is:

- If we gather a set of sample data and it would be <u>highly</u> <u>unlikely</u> to observe such data if H_0 were true, then we have evidence against ___ and in favor of _____.

- The usual approach to hypothesis testing is based on a <u>test statistic</u> (some relevant function of the data $Y_1, \ldots, Y_n$).

<u>Example 2</u>: Suppose we sample 100 parts and count the number that are defective. Let $Y =$ the number defective out of 100. Suppose we find $Y = 14$ defective parts in our sample.

- If $H_0: p = 0.03$ is true, is it <u>impossible</u> to observe 14 defective parts out of 100?
- If $H_0: p = 0.03$ is true, is it <u>unlikely</u> to observe 14 defective parts out of 100?
- If our observed test statistic is <u>extreme enough</u>, it provides evidence that in fact H_0 is _____ and H_a is ______.

- The set of values of the test statistic that lead us to reject H_0 (and conclude H_a) is called the rejection region (RR).

Example 2 again: If Y is our test statistic, our rejection region should consist of values of Y that are excessively ______:

- We will need a rigorous method for choosing this cutoff k.

Errors in Hypothesis Testing

- "Proving" the research hypothesis H_a does not imply removing all doubt (as in a mathematical proof).
- If H_0 is true, it is still possible to observe a "highly unlikely" test statistic value that falls in the RR (which would cause the researcher to reject H_0).
- If H_0 is rejected when H_0 is in fact true, this is called a ____________.

- We set up the rejection region so that the probability of a Type I error is some small number α, like:

- This value α is chosen by the researcher and is called the ____________________.

- A ____________________ occurs when H_0 is not rejected, yet H_0 is in fact false.

- We denote the probability of a Type II error as

Example 2 again: If we let our $RR = \{y \geq 7\}$, then what is α?

- Suppose the true proportion of defectives is actually 8% (implying H_0 is false and we should reject H_0). Using the same RR, what is β?

- Using this RR, the P[Type I error] is ________, but the P[Type II error] is somewhat __________.
- If we change the RR to, say, $\{y \geq 5\}$, then β will __________ but α will __________.

<u>Exercise:</u> Show that if $RR = \{y \geq 5\}$ here, $\alpha =$ and, when the true $p = 0.08$, $\beta =$.

- <u>Tradeoff</u>: α and β are __________ related.

<u>Note</u>: The value of β depends on the true parameter value: The farther the parameter is away from the null region, the lower β is (the ______ likely that we won't reject H_0).

- Conundrum: How to decrease β while still keeping α small?

  Answer:

10.3 Common Large-Sample Z-tests

- There are several common hypothesis tests in which the test statistic involves an estimator that (for large samples) has an approximately normal distribution.
- Consider a test about some parameter θ, in which an estimator $\hat{\theta}$ has an approximately normal sampling distribution, with mean θ and standard error $\sigma_{\hat{\theta}}$.
- If θ_0 is some hypothesized value of θ, our null hypothesis may be written:
- Depending on the research question, we could have one of three possible alternatives:

- We need to choose a test statistic whose ________________ (distribution if H_0 is true) is known, so that we can determine the RR corresponding to some specific α.
- Recall that if $H_0: \theta = \theta_0$ is <u>true</u>, then

has

- The RR will differ depending on which alternative we use:

<u>Example</u> (Upper-Tail Test):

If H_a is $\theta > \theta_0$, then ________ values of $\hat{\theta}$ (and thus of Z) will lead us to reject H_0 in favor of H_a.

- So RR =
- What choice of k will give us $P[\text{Type I error}] = \alpha$?

- Let P[Type I error] =
so that

- Clearly $k =$

- So RR =

Example (Lower-Tail Test): If H_a is $\theta < \theta_0$, then ________ values of Z lead to rejecting H_0 in favor of H_a.

- To preserve P[Type I error] $= \alpha$, the RR =

Picture:

Example (Two-Tailed Test): If H_a is $\theta \neq \theta_0$, then ________________ values of Z lead to rejecting H_0 in favor of H_a.

- Our RR should be of the form

- Note that choosing preserves P[Type I error] $= \alpha$:

Picture:

Common Examples of large-sample z-tests

- Recall: With large sample(s), the following estimators have approximately normal sampling distributions:

Estimator ($\hat{\theta}$)	Parameter (θ)	Standard Error ($\sigma_{\hat{\theta}}$)

- When the standard error contains unknown quantities, given the large sample(s) it is reasonable to plug in consistent estimators for these.

- So we have the following summary of available large-sample z-tests:

Test about	H_0	Possible H_a	Test Statistic	RR

- The parameter of interest and which alternative to choose are implied by the problem and the research question.