9.1: Hypotheses and Test Procedures

Taylor

University of Virginia

Motivation

So far we've been talking about estimating a parameter θ. We talked about different point estimates, and different types of intervals.

"Frequently, however, the objective of an investigationis not to estimate a parameter but to decide which of two contradictory claims about the parameter is correct."

This sort of thing is called *hypothesis testing*. And a lot of this section will be definitions.

Definition

A **statistical hypothesis** is a claim or assertion either about the value of a single parameter, about the values of several parameters, or about the form of an entire probability distribution.

Definition

The **null hypothesis**, denoted H_0, is the claim that is initially assumed to be true (the "prior" belief claim). The **alternative hypothesis**, denoted by H_a is the assertion that is contradictory to H_0.

We reject H_0 in favor of H_a if our data tells us to. If it doesn't, we continue to believe in H_0. So our decisions are either reject H_0 or fail to reject H_0.

Definition

A **test of hypothesis** is a method for using sample data to decide whether the null hypothesis should be rejected.

Generally our null hypothesis will involve an $=$

1. $H_0 : \theta = 4$

and our alternative hypothesis will take one of the three forms:

1. $H_a : \theta > 4$
2. $H_a : \theta < 4$
3. $H_a : \theta \neq 4$

Sometimes we will be more abstract and instead of using a specific number, like 4, we'll use θ_0. This value is called the **null value**.

Definitions

A test procedure is specified by two things:

1. a **test statistic**, a function of the sample data on which the decision is to be based
2. a **rejection region**, a set of all the test statistic values for which H_0 will be rejected

We reject the null hypothesis H_0 if and only if our test statistic falls inside the rejection region.

What could go wrong?

A **type 1 error** consists of rejecting H_0 when it's true...the probablity of which is usually denoted by α.

A **type 2 error** consists of not rejecting H_0 when it's false. The probability of this is usually denoted by β.

Example 9.1 on page 429

"An automobile model is known to sustain no visible damage 25% of the time in 10-mph crash tests. A modified bumper design has been proposed in an effort to increase this percentage. Let p denote the proportion of all 10-mph crashes with this new bumper that result in no visible damage. The hypotheses to be tested are $H_0 : p = .25$ versus $H_a : p > .25$. The test will be based on an experiment involving $n = 20$ independent crashes with prototypes of the new design. Intuitively, H_0 should be rejected if a substantial number of the crashes show no damage."

Example

Let's have our test statistic be X, the number of crashes with no visible damage. If H_0 is true, $p = .25$, and $X \sim \text{Binomial}(20, .25)$.

They give us a rejection region of $R_8 = \{8, 9, 10, \ldots\}$.

$$\begin{aligned}\alpha = P(\text{ Type 1 Error }) &= P(\text{ Reject when } H_0 \text{ is true })\\ &= P(X \geq 8 \text{ when } p = .25)\\ &= 1 - P(X \leq 7 | p = .25)\\ &= .102\end{aligned}$$

Example continued

Recall that type 2 error is when H_0 is false. In this case $H_a : p > .25$. But this is not specific enough; we need a fixed value of p to be able to calculate cdfs.

Let's pick .3.

$$\begin{aligned}\beta(.3) = P(\text{ type 2 error when } p \text{ is } .3) &= P(X < 8 | p = .3) \\ &= P(X \leq 7 | p = .3) \\ &= .772\end{aligned}$$

This says that when p is only slightly above .25 (.3 in this case), that we're going to have a hard time rejecting H_0. 77 % of the time we will make a mistake here.

Example 9.2 on page 430

"The drying time of a type of paint under specified test conditions is known to be normally distributed with mean value 75 min and standard deviation 9 min. Chemists have proposed a new additive designed to decrease average drying time. It is believed that drying times with this additive will remain normally distributed with $\sigma = 9$. Because of the expense associated with the additive, evidence should strongly suggest an improvement in average drying time before such a conclusion is adopted. Let μ denote the true average drying time when the additive is used. The appropriate hypotheses are $H_0 : \mu = 75$ versus $H_a : \mu < 75$. Only if H_0 can be rejected will the additive be declared successfull and used."

Example 9.2 on page 430

For this problem we have a random sample of size $n = 25$. Again, our data are all normally distributed, independently, with mean μ and $\sigma = 9$ and our hypotheses are $H_0 : \mu = 75$ versus $H_a : \mu < 75$.

Consider the rejection region $R = (-\infty, 70.8]$. We know that $\bar{X} \sim \mathcal{N}(\mu, 1.8^2)$

Example 9.2 on page 430

$$
\begin{aligned}
\alpha &= P(\bar{X} \leq 70.8 \text{ when } H_0 \text{ is true })\\
&= P(\bar{X} \leq 70.8 | \mu = 75)\\
&= P\left(\frac{\bar{X} - 75}{1.8} \leq \frac{70.8 - 75}{1.8}\right)\\
&= \Phi\left(\frac{70.8 - 75}{1.8}\right)\\
&= .01
\end{aligned}
$$

this is good.

Motivation

Let's consider when $\mu = 72$, which is part of the alternative hypothesis

$$\begin{aligned}
\beta &= P(\bar{X} > 70.8 | \mu = 72) \\
&= P\left(\frac{\bar{X} - 72}{1.8} > \frac{70.8 - 72}{1.8}\right) \\
&= 1 - \Phi\left(\frac{70.8 - 72}{1.8}\right) \\
&= .7486
\end{aligned}$$

this is bad.

Motivation

For any given sample size and experiment, there is a tradeoff between α and β whenever you move around the rejection region. Too far away, and you'll never reject H_0 (high type 2), but too close and you'll always accidentally reject (high type 1).

Motivation

Traditionally, one will pick a maximum level for α. Then, out of all the possible test choices that fulfill this criterion, one will select the test with the highest power (or lowest type 2 error). So we control type 1 error, and hope for the best on type 2 error.

The choice for your type 1 error is called the **significance level**. The corresponding test procedure is called the **level α test**.