Math 3070, Applied Statistics

Section 1

November 8, 2019

Lecture Outline, 11/8

Section 8.1

- Hypothesis Testing

Statistical Hypotheses

A *hypothesis* is an assertion about a distribution or its parameters. For example,

- Given a coin, one hypothesis is that each toss has probability $p = .5$ of coming up heads. Another hypothesis would be that $p \neq .5$.
- Given a certain type of candy bar, labeled as having a mass of 60 grams, one hypothesis is that the mean mass is as labeled, $\mu = 60$. Another hypothesis would be that the mean mass is smaller than labeled: $\mu < 60$.

Hypothesis Testing

In a hypothesis-testing problem, we consider two contradictory hypotheses H_0 and H_a.

- H_0, the *null hypothesis*, is the hypothesis which we initially presume to be true.
- The other hypothesis H_a is called the *alternative hypothesis*.
- The hypothesis H_0 is rejected only if the sample evidence strongly contradicts it. Otherwise we continue to believe that H_0 is plausible.
- The two possible outcomes of the analysis are that we *reject* the null hypothesis H_0, or we *do not reject* the null hypothesis.
- H_0 and H_a should be disjoint.

Example

Suppose someone claims that a coin is unfair, that it gives heads more than half the time. The null hypothesis would be $H_0 : p = .5$, that the coin is fair. The alternative hypothesis is $H_a : p > .5$.

To test the claim, we could toss the coin for several trials and reject H_0 if the number of heads we obtain is larger than a certain amount.

For example, one procedure would be to toss the coin 10 times and reject the null hypothesis if we obtain 8 heads or more.

Test Procedures

- In general, to test a statistical hypothesis, we select a *test statistic* that can be calculated from a random sample.
- Out of the possible values of the test statistic, we select a subset of values which are unlikely to occur if the null hypothesis H_0 is true; this subset is called the *rejection region*.
- We then obtain data from a random sample, calculate the test statistic for the data, and reject H_0 if the test statistic is in the rejection region.

For the example of testing for an unfair coin, the test statistic was the number of heads X out of 10 tosses, and the rejection region was $\{8, 9, 10\}$.

Errors in Hypothesis Testing

For essentially any test procedure, there is a chance that the test will give a misleading conclusion:

- When the null hypothesis H_0 is true but is rejected, this is called a *Type I error*.
- When the null hypothesis H_0 is false but is not rejected, this is called a *Type II error*.

We cannot eliminate the possibility of these errors. However, we can quantify their probability of occuring:

- The probability of a Type I error is denoted by α.
- The probability of a Type II error is denoted by β.

Typically, the choice of rejection region involves a tradeoff between the two types of errors. But by using larger samples, both error probabilities may be reduced.

Example

Suppose we test a coin by tossing it 10 times and rejecting it if we get 8 or more heads. If in reality the coin is fair, what is the probability α of a Type I error? If the coin is unfair with probability $p = .75$ of being heads, what is the probability β of a Type II error?

To find the Type I error, we assume $p = .5$ and calculate the probability that the number of heads X is at least 8:

$$\alpha = P(X \geq 8) = \sum_{x=8}^{10} \binom{10}{x} (.5)^x (1 - .5)^{10-x} = .055$$

To find the Type II error in the case $p = .75$, we calculate the probability that the number of heads X is less than 8:

$$\beta = P(X < 8) = \sum_{x=0}^{7} \binom{10}{x} (.75)^x (.25)^{10-x} = .474$$

Example

In the previous example, how do the error probabilities change if we instead reject the coin if we get 7 or more heads out of 10?

To find the Type I error, we assume $p = .5$ and calculate the probability that the number of heads X is at least 7:

$$\alpha = P(X \geq 7) = \sum_{x=7}^{10} \binom{10}{x} (.5)^x (1 - .5)^{10-x} = .171$$

To find the Type II error in the case $p = .75$, we calculate the probability that the number of heads X is less than 7:

$$\beta = P(X < 7) = \sum_{x=0}^{6} \binom{10}{x} (.75)^x (.25)^{10-x} = .224$$

By enlarging the rejection region, we increased α but decreased β.

Example

Suppose we test a coin by tossing it 10 times and rejecting it if we get 8 or more heads. If in reality the coin is fair, what is the probability α of a Type I error? If the coin is unfair with probability $p = .75$ of being heads, what is the probability β of a Type II error?

To find the Type I error, we assume $p = .5$ and calculate the probability that the number of heads X is at least 8:

$$\alpha = P(X \geq 8) = \sum_{x=8}^{10} \binom{10}{x} (.5)^x (1 - .5)^{10-x} = .055$$

To find the Type II error in the case $p = .75$, we calculate the probability that the number of heads X is less than 8:

$$\beta = P(X < 8) = \sum_{x=0}^{7} \binom{10}{x} (.75)^x (.25)^{10-x} = .474$$

Example

In the previous example, how do the error probabilities change if we instead reject the coin if we get 7 or more heads out of 10?

To find the Type I error, we assume $p = .5$ and calculate the probability that the number of heads X is at least 7:

$$\alpha = P(X \geq 7) = \sum_{x=7}^{10} \binom{10}{x} (.5)^x (1 - .5)^{10-x} = .171$$

To find the Type II error in the case $p = .75$, we calculate the probability that the number of heads X is less than 7:

$$\beta = P(X < 7) = \sum_{x=0}^{6} \binom{10}{x} (.75)^x (.25)^{10-x} = .224$$

By enlarging the rejection region, we increased α but decreased β.

Problem

A type of candy bar is labeled 60 grams. We decide to test the label's accuracy using a random sample $X_1, \ldots, X_5$ by rejecting the null hypothesis $H_0 : \mu = 60$ if $\overline{X} < 59$. If the actual mean mass is $\mu = 58.5$, and $\sigma = 0.8$, what is the Type II error probability β?

In this case, $Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - 58.5}{0.8/\sqrt{5}}$ is a standard normal random variable, so

$$\begin{aligned}
\beta &= P(\overline{X} \geq 59) \\
&= P\left(\frac{\overline{X} - 58.5}{0.8/\sqrt{5}} \geq \frac{59 - 58.5}{0.8/\sqrt{5}}\right) \\
&= P(Z \geq 1.40) \\
&= .0808
\end{aligned}$$

Significance Level

As we have seen, for a fixed sample size, selecting a rejection region for a test involves a tradeoff between the Type I error probabilities α and the Type II error probability β.

A common practice is to design the test to achieve a specified small value of α, such as $\alpha = .1, .05$, or $.01$. The choice for α is called the *significance level*.

P-Values

There is a major drawback of the hypothesis-testing procedures considered so far: The tests each only have two outcomes – reject or do not reject – even if the test statistic is near the border of the rejection region.

One way to remedy this is to report the so-called P-value:

The *P-value* is the smallest significance level α for which the test would reject the null hypothesis.

For example, in the candy bar example, at the $\alpha = .01$ we failed to reject the null hypothesis; however, if we had used a less strict significance level, $\alpha = .05$, then the test would have rejected. The P-value would provide a more nuanced summary of the test result by indicating precisely at what significance level the test changes from rejecting to failing to reject.

P-Values

Loosely speaking, another way to describe the P-value is as follows:

The *P-value* is the probability, calculated assuming that the null hypothesis is true, of obtaining a value of the test statistic at least as extreme as the value actually observed.

Here are some key points:

- The P-value is a probability.
- This probability is calculated assuming that the null hypothesis is true.
- Beware: The P-value is not the probability that H_0 is true.
- The interpretation of “as extreme as” depends on the alternative hypothesis:
  - Consider hypotheses $H_0 : \mu = 0$ and $H_A : \mu > 0$. With a test statistic of $\overline{X} = 0.5$. The p-value is $P(\overline{X} > 0.5 | \mu = 0)$. 'As extreme as' is interperted as 'as large as'. $\mu = 0$ can be implemented by conditioning.

Hypothesis Testing Framework

If the p-value is smaller than the significance level $(p < \alpha)$, we reject H_0. The data is too unlikely, as determined by α, given our assumption H_0.

If the p-value is larger than the significance level $(p \geq \alpha)$, we fail to reject H_0. The data is plausible, as determined by α, given our assumption H_0.

Note: smaller α is more likely reject to a p-value so we expect a larger rejection region and hypotheses will be defined on more complicated sets. For example:

- $H_0 : \mu \geq 2$ and $H_A : \mu < 2$
- $H_0 : \mu = 2$ and $H_A : \mu \neq 2$