Social Studies 201
Notes for November 24, 2006

Introduction to hypothesis testing – Chapter 9

These notes introduce hypothesis testing, a topic that will occupy the remainder of the semester. The notes describe the steps involved in testing an hypothesis about a population mean and some principles of hypothesis testing.

Introduction. The aim of hypothesis testing is to bring data to bear on a suspicion or claim, in an attempt to decide whether the suspicion or claim are supported by the data. Using the principles of hypothesis testing, the data obtained from a sample may negate a claim, support a claim, or be inconclusive about the claim. For each hypothesis test there is an associated probability or significance level that the conclusion is either correct or incorrect. While such a test can never provide definitive proof or negation of an hypothesis, the conclusion obtained from conducting a test can provide strong support for the hypothesis, or they can provide evidence casting doubt on the hypothesis. While the findings from an hypothesis test are never definitive, the conclusion of a test has a particular probability of being correct or incorrect.

When conducting a formal hypothesis test, the hypothesis is a specific statement about a characteristic of a population. Such a statement emerges from more general suspicions or claims about the population characteristic.

1. **Suspicions**. Hypotheses can be at a very general level, emerging from suspicions or doubts that individuals or groups have about some social issue or from a theoretical model. One current example is related to environmenal issues – there is a suspicion, and a fair degree of evidence, that there is global warming. Exactly what is the cause or consequence of this is unclear but one suspicion that people on the Prairies may have is that such global warming could result in productive agricultural regions becoming desert-like. Another suspicion often expressed is that some of the foods or liquids we consume are unsafe – in past years there was suspicion that Regina water was unsafe to drink.

   Such suspicions are at a fairly general level and may not be associated with a specific hypothesis about a population mean or proportion. One aim of researchers is to translate some of these suspicions into testable hypotheses, so the data researchers produce can be used to either determine whether these suspicions can be supported or whether they are unfounded.

2. **Claims**. A more focussed type of statement can be considered a claim. Claims might be distinguished from suspicions in that the individual making the claim has done some investigation of an issue and has some evidence to make such a claim. But evidence may be incomplete or the claim may remain at a fairly general level.

   As examples, consider claims about effects of NAFTA, the North America Free Trade Agreement. Some in the business community and government may claim that NAFTA has expanded the number of jobs in Canada and has been of benefit to Canadians. Detractors of NAFTA may claim that it has led to a loss of jobs, reduced sovereignty, and resulted in a worsening of the living standard of Canadians. Each of the participants in such a debate is likely to have some evidence to support their claim.

   A researcher interested in testing these claims would find it necessary to establish particular propositions about the effect of NAFTA, adopt some assumptions about how to examine these issues, and translate the claims into specific hypotheses that can be tested using data the researcher has access to or can obtain.

3. **Research findings**. A third, and more specific level is examination or re-examination of a research finding. Very specific research findings may have been obtained by other researchers, and you may wish to test these findings by replicating or questioning them. Other researchers may have conducted well constructed surveys or experiments, but you may wish to re-examine the same issue using data from another population or by using a different theoretical perspective or methodology. These lead to constructing specific hypotheses about the characteristics of the populations being examined – hypotheses that can be tested using using data obtained from the population.

Each of these levels of research involves construction of specific hypotheses about characteristics of populations – characteristics such as the value of the mean of a population. Researchers use general statements about these characteristics and translate them into specific statements about a population mean, μ, or the proportion of a population with a characteristic, p. The method of examining such specific hypotheses is the subject of these notes.

Summary of the method of hypothesis testing – pp. 559-563

The method of hypothesis testing involves constructing an hypothesis concerning the value of a population parameter such as the population mean, standard deviation, or proportion. If the statistic from a sample of the population is relatively close to the hypothesized value of the population parameter, then the hypothesis is not rejected. But if the sample value, or statistic, is distant from the hypothesized value, the hypothesis is rejected and regarded as incorrect. Attached to each of these conclusions is a probability that the conclusion is correct or incorrect – conclusions about hypotheses are made with certainty but with a particular probability.

Hypothesis testing can be ordered into a series of six steps. These are as follows, with a test of a population mean used to illustrate the method.

Before conducting the test, it is necessary to identify and define the population parameter that is to be tested. This is often a parameter such as the true mean, μ, or proportion, $\hat{p}$, of a population. To illustrate the method, the following steps are aimed at obtaining information about the true mean, μ of some population.

1. **State hypotheses**. When testing an hypothesis about a population mean, μ, the statement is a claim that μ equals some specific value. This is termed the null hypothesis and given the symbol H_0. An alternative hypothesis is given the symbol H_1 and is a statement that the mean is not equal to the specific value claimed in H_0. This alternative hypothesis is sometimes referred to as the research hypothesis. If the specific, hypothesized value of the population mean is given the symbol M, then the hypotheses to be tested are:

$$\text{Null hypothesis} \quad H_0 : \mu = M$$

$$\text{Alternative hypothesis} \quad H_1 : \mu \neq M$$

The alternative, or research, hypothesis could intead state that $\mu < M$ or $\mu > M$. These latter options are examined later.

In conducting an hypothesis test, a researcher assumes that the null hypothesis, H_0, is correct and the whole test is conducted assuming this null hypothesis is true. At the conclusion of the test, the null hypothesis is either rejected (when data are inconsistent with the null hypothesis) or nor rejected (when data are consistent with the null hypothesis).

2. **Test statistic and data**. If an hypothesis test about the mean of a population is conducted, the test statistic is ordinarily the sample mean, $\bar{X}$.

   In order to conduct the test, data about the sample mean $\bar{X}$, standard deviation s, and sample size n must be available, or produced by the researcher. These provide the data or evidence that is used to decide whether the null or alternative hypothesis is correct.

3. **Distribution of the test statistic**. For a test of a population mean, the sample mean is the test statistic. The sampling distribution of the sample means $\bar{X}$ must be known in order to proceed with the test. If the samples are large random samples, then the central limit theorem can be used to describe the distribution of these sample means. If the sample size is small, and the samples are random samples from a normally distributed population, then the t-distribution describes the distribution of sample means.

4. **Significance level**. Each hypothesis test is associated with a probability that is termed the level of significance. The level of significance is a small probability and is the complement of the confidence level – for example, 95% confidence level corresponds to 0.05 (or 5%) significance. The level of significance denotes an area in the tail or tails of the distribution of the statistic. The confidence level is a large area in the middle of a normal distribution, denoting a large area within which it is hoped that the sample mean lies. The significance level is an area in the tail or extreme portion of the distribution of the statistic. If the sample is in this extreme area, then H_0 is rejected.

As a result, the significance level specifies a probability that the sample value will be far from the hypothesized mean μ. It indicates a probability that the hypothesis could be rejected in error. Further explanation of this is provided later in these notes.

5. **Critical region or region of rejection of H_0**. Consistent with the hypothesized value for the mean, the distribution of the sample mean, and the significance level, a region of rejection for the null hypothesis is constructed. This is a region in the tail or tails of the sampling distribution. In the case of a 0.05 significance level, the critical region is all Z-values less than -1.96 or greater than +1.96. That is, there is an area of 0.05 in the two tails of the distribution, or 0.025 of the area in each tail, consistent with a Z value of 1.96 (B area of 0.025 in the table of the normal distribution). It is this disjointed region that is the critical region or region of rejection of H_0. The area in the middle of the distribution is the area of nonrejection of H_0.

6. **Conclusion**. Using the 0.05 level of significance, if the sample mean $\bar{X}$ is farther than 1.96 standard deviations from the mean hypothesized in H_0, then this null hypothesis is rejected. But if the sample mean $\bar{X}$ is not farther than 1.96 standard deviations from the hypothesize mean, then the null hypothesis H_0 is not rejected. In order to test this, the number of standard deviations that the sample mean lies from the hypothesize mean is calculated.

   Remember that a Z-value is also a number of standard deviations, and a Z-value can always be obtained with the formula for standardization of a variable:

$$Z = \frac{\text{variable} - \text{mean of variable}}{\text{standard deviation of variable}}.$$

   In this case, when the null hypothesis is that $\mu = M$, the Z-value associated with the sample mean $\bar{X}$ is

$$Z = \frac{\bar{X} - M}{\sigma/\sqrt{n}}.$$

   That is, the variable is $\bar{X}$ – its mean is hypothesized to be M and the standard deviation of $\bar{X}$ is $\sigma/\sqrt{n}$. This Z-value represents the number of standard deviations the sample mean $\bar{X}$ (with standard deviation $\sigma/\sqrt{n}$) is from M. Since the true standard deviation, σ, is generally unknown, the sample standard deviation, s, is used as an estimate of σ in the formula for Z. Thus

$$Z = \frac{\bar{X} - M}{s/\sqrt{n}}.$$

The final stage of the test is to compute the value of Z from the data (see item 2). If the value of Z is negative and below -1.96, or positive and above +1.96, then it is in the critical region and the null hypothesis is rejected. But if this Z-value is between -1.96 and +1.96, the null hypothesis that $\mu = M$ is not rejected. That is, the sample mean is considered sufficiently close to the hypothesized population mean that the null hypothesis is not rejected.

Notation and conventions of hypothesis testing

1. **Null and alternative hypothesis**. The hypothesis that is being tested is termed the null hypothesis and given the symbol H_0. This represents a specific value for a population parameter, such as $\mu = 50$ or, in the case of proportion, $p = 0.4$. A null hypothesis is a statement that the population parameter has a specific value – as a result the null hypothesis is an equality.

   The alternative hypothesis is a statement concerning other possibilities, for example that $\mu \neq 50$. The alternative hypothesis is sometimes termed the research hypothesis and is given the symbol H_1. It may be a two-directional alternative, for example, that $\mu \neq 50$; another option is that it is a one-directional alternative, for example, $\mu < 50$ or $\mu > 50$. Since the alternative hypothesis is that the population characteristic is different than the specific value hypothesized in H_0, the alternative hypothesis is ordinarily an inequality.

2. **Level of significance**. This is usually given the symbol α, or alpha, the first letter of the Greek alphabet. This is a small value or probability, perhaps 0.10, 0.05, 0.01, or some other small value. When asking questions about hypothesis testing, a significance level may, or may not, be given. If such a level is not given, the 0.05 level is the most commonly adopted level.

3. **Test statistic**. This is a statistic obtained from a sample – statistics such as $\bar{X}$ or $\hat{p}$. These are often converted into Z-values or t-values.

4. **Critical values and critical region**. This is the region where the null hypothesis is rejected, and is in the extremes of the distribution. In the case of a normal distribution for the distribution of a sample mean, this is the Z-value or values so that there is exactly α of the area under the normal curve in the tail or tails of the normal distribution. For example, if $H_0 : \mu = M$ and $H_1 : \mu \neq M$, and if $\alpha = 0.05$, then the critical values are $Z = -1.96$ and $Z = +1.96$. In this case, the critical region is a two-tailed region of all Z-values less than -1.96 or greater than +1.96. If the sample mean is more than this number of Z-values from the hypothesized mean, the null hypothesis is rejected. If the sample mean is not in the critical region (i.e. not more than 1.96 standard deviations from the hypothesized mean) then the null hypothesis is not rejected.

Example – mean age of University of Regina undergraduates

Following is a typical question about testing an hypothesis. A specific claim is stated and sample data are used to test the claim. The data cited here come from the *Survey of Student Attitudes and Experiences* conducted in 1998 in Social Studies 306 and available in the file ssae.sav at

`t:\students\public\201\`

Question. Test the hypothesis that the mean age of University of Regina undergraduates is 23 years of age. Use the 0.05 level of significance. In order to conduct this test, use the data from the survey. Assume this was a random sample of 698 undergraduates, with a sample mean age of 22.46 years and a standard deviation of 5.60 years.

Answer. The above question requests a test of the hypothesis that the population mean is 23. In this case the population mean to be tested is μ, the true mean age of all University of Regina undergraduates in 1998. This mean age is unknown and the data from the sample is used to conduct the test.
The steps involved in testing this hypothesis are as follows.

1. **Hypotheses**. The claim is that $\mu = 23$. No direction concerning an alternative is stated in the question so the alternative is that $\mu \neq 23$. The null and alternative hypotheses are

$$H_0 : \mu = 23$$

$$H_1 : \mu \neq 23$$

   In words, the null hypothesis, H_0, is that the mean age of U of R undergraduates is 23 years of age. The alternative hypothesis, H_1, is that the mean age is not 23 years of age. That is, the alternative hypothesis is a two-directional inequality, that the null hypothesis is incorrect.

   At the conclusion of the test, the data will lead the researcher to either supporting the null hypothesis that μ is 23 years or rejecting this hypothesis. Together, the significance level, the sampling distribution of the test statistic, and the value of the test statistic determine which of these two conclusions is the correct result.

2. **Test statistic**. The claim is about μ, the mean age of all U of R undergraduates. The sample mean of the age of undergraduates in the sample, $\bar{X}$, is the test statistic that will be used here. If $\bar{X}$ is close to $\mu = 23$, the null hypothesis will not be rejected; if $\bar{X}$ is distant from $\mu = 23$, the null hypothesis will be rejected and the alternative hypothesis that $\mu \neq 23$ will be accepted. The question is how far $\bar{X}$ is from the hypothesized mean of 23, and what is its probability of being this far from 23. This depends on the distribution of $\bar{X}$ and the significance level.

3. **Distribution of the test statistic**. The sample is said to be a random sample of U of R undergraduates, with a sample size of $n = 698$. This is a large random sample so the central limit theorem can be used. As a result,

$$\bar{X} \text{ is Nor} \left(\mu, \frac{\sigma}{\sqrt{n}} \right).$$

That is, the sampling distribution of $\bar{X}$ is normally distributed with mean μ and standard deviation $s/\sqrt{n}$, where s can be used as an estimate of the population standard deviation σ, since n is large.

4. **Significance level**. In this question, the level of significance requested is 0.05, that is, $\alpha = 0.05$. This is the area in the two tails of a normal distribution, split equally between each tail so that 0.025 of the area is in each tail.

5. **Critical region**. The critical region is the extreme area that is some distance from the hypothesized mean. In this case, the sample means have a normal distribution. The significance level is $\alpha = 0.05$, and with a two-directional alternative hypothesis, this means there is to be 0.025 of the area in each tail of the distribution. Checking the B column of the table of the normal distribution, an area of 0.025 in one tail of the distribution is associated was a Z of 1.96. The critical region is thus all $Z < -1.96$ or $Z > +1.96$, so that there is 0.025 of the area in the critical region at each end of the distribution. The region for rejecting H_0 is thus -1.96 standard deviations or more to the left of centre or +1.96 standard deviations or more to the right of centre.

   The area in middle of the distribution is closer to the hypothesized mean, so the null hypothesis is not rejected if the sample is in this middle area. This region of nonrejection of H_0 is all Z-values between -1.96 and +1.96.

   These two regions can be summarized as follows:

$$\text{Region of rejection of } H_0 \text{ is: } Z \leq -1.96 \text{ or } Z \geq +1.96$$

$$\text{Region of nonrejection of } H_0 \text{ is: } -1.96 \leq Z \leq +1.96$$

6. **Conclusion**. In order to determine whether the sample mean $\bar{X}$ is within the critical region or not, it is necessary to determine the distance $\bar{X}$ is from the hypothesized mean μ. This can be determined by obtaining the Z-value associated with the sample mean – that is, how many standard deviations is $\bar{X} = 22.46$ from the hypothesized mean of $\mu = 23$.

The standard deviation of the mean is $\sigma/\sqrt{n}$, but since σ is unknown, this is estimated by $s/\sqrt{n}$. The number of standard deviations, or Z-values, that $\bar{X}$ is from the mean is

$$\begin{aligned} Z &= \frac{\bar{X}-\mu}{s/\sqrt{n}} \\ &= \frac{22.46-23}{5.60/\sqrt{698}} \\ &= \frac{-0.54}{5.60/26.420} \\ &= \frac{-0.54}{0.2120} \\ &= -2.548 < -1.96. \end{aligned}$$

That is, the sample mean is 2.548 standard deviations below the hypothesized mean of $\mu = 23$. This is below the critical cut-off point of -1.96, so this Z-value is in the critical region for the test. That is, the sample mean is -2.548 standard deviations from the hypothesized mean of 23, a great distance, and one that is extreme enough to be in the left 0.025 of the distribution.

Since this Z-value is in the critical region, the conclusion of the test is to reject the null hypothesis H_0 and accept the alternative hypothesis H_1. This conclusion is made at the $\alpha = 0.05$ level of significance.

The conclusion is that the mean age of U of R undergraduates is not 23 years of age. At the 0.05 level of significance, there is evidence that the mean age is not 23 years, but a different age. The sample yielded a mean age of 22.46, less than 23, so from this the evidence is that the mean age of U of R undergraduates is likely below 23 years of age.