AMS 315
Data Analysis
Chapter Five Study Guide
Inferences about Population Central Values
Spring 2023

Context

The one sample tests discussed here are a basic tool of statistical analysis. Student's work developing a practical one-sample test is the foundation of applied statistics. The material of this chapter should be familiar to you from your prerequisite class.

Chapter Five

5.1. Introduction

This section is introductory. You should review the discussion of the vocabulary of testing a hypothesis.

5.2. Estimation of μ

This section specifies the one sample confidence interval for μ. The formula $\bar{y}_n \pm z_{\alpha/2}\sigma/\sqrt{n}$ is fundamental, and the values 1.645, 1.960, and 2.576 are commonly called for in statistics examinations.

5.3. Choosing the sample size for estimating μ

This section gives the solution to the problem of how large the sample size n should be so that the width of the confidence interval for μ meets a specification. Note the formulaa on page 231. This has great application in opinion sampling. A Bernoulli trail is the correct random variable to use for an opinion survey. The random variable Y_i takes the value 1 if respondent *i* meets a criterion (for example, supports candidate A in an upcoming election) and is 0 otherwise. Then, $E(Y_i) = p$, where *p* is the fraction in the population meeting the criterion, and $\text{var}(Y_i) = p(1-p) \le 0.25$. The estimate of *p* based on a random sample of size *n* is $\hat{p} = \frac{\sum_{i=1}^{n} Y_i}{n} = \bar{Y}_n$. Now, $\text{var}(\bar{Y}_n) = \frac{p(1-p)}{n} \le \frac{0.25}{n}$. Then, a random sample of size $n \ge \frac{0.25(z_{\alpha/2})^2}{E^2}$ will have a confidence interval of half-width *E* or less.

5.4. A statistical test for μ

This section contains the specification of the one sample test of the null hypothesis $H_0: \quad E(Y) = \mu$ when the value of $\mathrm{var}(Y)$ is explicitly stated in an examination problem. It also contains further basic vocabulary about testing hypotheses. There will be examination questions about this material. Pay particular attention to the discussion about calculating the probability of a Type II error.

5.5. Choosing the sample size for testing μ

I use a more general formula for the sample size *n* than the one given in the text. The null hypothesis is that Y is $N(E_0, \sigma_0^2)$. The alternative hypothesis is one-sided; for example, $H_1: \quad E(Y) > E_0$. The level of significance (probability of a Type I error) is set to α. We seek the sample size *n* that has probability of Type II error β when Y is $N(E_1, \sigma_1^2)$. Here, $E_0 < E_1$. Then, when $\alpha \le \frac{1}{2}$ and $\beta \le \frac{1}{2}$, $\sqrt{n} \ge \frac{|z_\alpha|\sigma_0 + |z_\beta|\sigma_1}{|E_0 - E_1|}$.

5.6. The level of significance of a statistical test

This section defines the p-value and shows how to use it. The p-value is almost universally reported in statistical output.

5.7. Inferences about μ for a Normal population, σ unknown

This is a fundamentally important section. Review the properties of Student's t distribution in the box in this section and the box giving the summary of the test.

5.8. **Inferences about μ when population is nonnormal and n is small: bootstrap methods**

Bootstrap procedures are increasing in popularity with the rapid improvement in computing power. It is extremely difficult to write an examination question on the bootstrap technique. Read the itemized steps of the procedures in this section.

5.9. Inferences about the median

I will not test you on this material. Later in the course, we will use monotonic transformations of the data to attempt to remove the effects of skewness.

5.10. Research study: percent calories from fat

Read this section for background.

5.11. Summary and key formulas

I will not test on the confidence interval of the median or testing a null hypothesis about the median of a random variable.

Past Examination Questions

1. A research team studied Y , the percentage of voters in favor of a candidate. The random variable Y had standard deviation $\sigma = 50\%$. A random sample of 800 voters was selected, and their average was $\bar{y}_{800} = 52.3\%$. What is the 95% confidence interval for $E(Y)$?

   Answer: $52.3\% \pm 3.46\%$. This problem and problem 2 are relatively realistic. As noted earlier in this chapter, the standard deviation of a random variable following the Bernouilli distribution is less than or equal to 50%. For example, when the probability of success is between 25% and 50%, the standard deviation is between 43.3% (for probability of success either 25% or 75%) and 50%.

2. A research team was conducting research about Y , the percentage of voters in favor of a candidate. The random variable Y had standard deviation $\sigma = 50\%$. They wished to test the null hypothesis $H_0 : E(Y) = 50\%$ against the alternative $H_1 : E(Y) \neq 50\%$. A random sample of 900 voters was selected, and their average score was $\bar{y}_{900} = 48.2\%$. What conclusions do you make using levels of significance 0.10, 0.05, and 0.01?

   Answer: Accept at the 0.10, 0.05, and 0.01 levels of significance.

3. A research team is conducting research about Y , a student's score on a national examination, which is normally distributed with standard deviation $\sigma = 150$. They wish to test the null hypothesis $H_0 : E(Y) = 1000$ at level of significance 0.005 against the alternative hypothesis that $H_1 : E(Y) > 1000$. How many observations n are necessary so that the probability of a Type II error is 0.01 when $E(Y) = 1050$ and $\sigma = 150$?

   Answer: 217 or more observations are necessary. When you work on a problem like this, rephrase the question in more general terms. For example, the question asks how large a sample size is necessary to detect a difference of 1/3 of a standard deviation; note the 50=1050-1000=150/3. For this alpha and beta, one needs at least 217 observations.

4. A research team took a sample of 8 observations from the random variable Y, which had a normal distribution $N(\mu, \sigma^2)$. They observed $\bar{y}_8 = 43.2$, where $\bar{y}_8$ is the average of the eight sampled observations and $s^2 = 517.5$ is the observed value of the unbiased estimate of σ^2 , based on the sample values. Test the null

hypothesis that $H_0 : E(Y) = 50$ against the alternative $H_1 : E(Y) \neq 50$ at the 0.10, 0.05, and 0.01 levels of significance.

Answer: Accept at the 0.10, 0.05, and 0.01 levels of significance.

5. A research team took a sample of 5 observations from the random variable Y, which had a normal distribution $N(\mu, \sigma^2)$. They observed $\bar{y}_5 = 56.2$, where $\bar{y}_5$ is the average of the five sampled observations and $s^2 = 53.6$ is the observed value of the unbiased estimate of σ^2, based on the sample values. Find the 99% confidence interval for $E(Y)$.

   Answer: 56.2 ± 15.07.

End of Guide