MODULE 15

1-SAMPLE Z-TEST

Objectives:

1. Properly construct statistical hypotheses.
2. Understand the specifics of a 1-Sample Z-Test.
3. Perform the 11-steps of a significance test in a 1-Sample Z-Test situation.

Contents

A FOUNDATION FOR MAKING STATISTICAL INFERENCES was provided in Modules 12-14. Most of the material in Modules 13 and 14 is related to a 1-Sample Z-test, which is specifically formalized in this module. Other specific hypothesis tests are in Modules 16-19.

15.1 11-Steps of Hypothesis Testing

Hypothesis testing is a rigorous and formal procedure for making inferences about a parameter from a statistic. The 11 steps listed below will help make sure that all aspects important to hypothesis testing are completed. These steps should be used for all hypothesis tests in this and ensuing modules.

1. State the rejection criterion (α).
2. State the null and alternative hypotheses to be tested and define the parameter(s).
3. Identify (and explain why!) the hypothesis test to use (e.g., 1-Sample t, 2-sample t, etc.).
4. Collect the data (address study type and if randomization occurred).
5. Check all necessary assumptions (describe how you tested the validity).
6. Calculate the appropriate statistic(s).
7. Calculate the appropriate test statistic.
8. Calculate the p-value.
9. State your rejection decision about H_0.
10. Summarize your findings in terms of the problem.
11. **If** H_0 **was rejected**, compute and interpret an appropriate confidence region for the parameter.

The order of some of these steps is arbitrary, however Steps 1-3 **MUST** be completed before collecting the data (Step 4). Further note that we will perform Step 11 only to provide a more definitive statement about the value of the parameter when the H_0 as been rejected (i.e., if the parameter differs from the hypothesized value, then provide a range for which the actual parameter may exist).

◇ **A confidence region for a parameter articulates the direction and magnitude of the difference when H_0 is rejected.**

15.2 1-Sample Z-Test Specifics

A 1-Sample Z-Test tests $H_0 : \mu = \mu_0$, where μ_0 represents a specific value of μ, when σ is known. Other specifics of this test were discussed in the previous modules and are summarized in Table 15.1. The only test that can possibly be confused with a 1-Sample Z-Test is a 1-Sample t-Test (Module 16), which tests the same null hypothesis but when σ is unknown.

Table 15.1. Characteristics of a 1-Sample Z-Test.

- **Hypothesis:** $H_0 : \mu = \mu_0$
- **Statistic:** $\bar{x}$
- **Test Statistic:** $Z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$
- **Confidence Region:** $\bar{x} + Z^* \frac{\sigma}{\sqrt{n}}$
- **Assumptions:**
  1. σ is known
  2. $n > 30$, $n > 15$ and the **population** is not strongly skewed, OR the **population** is normally distributed.
- **When to Use:** Quantitative response, one population, σ is known.

15.2.1 Example - Intra-class Travel

Consider the following situation,

> *A dean is interested in the average amount of time it takes for students to get from one class to another. In particular, she wants to determine if it takes more than 10 minutes, on average, to go between classes. In an effort to test this hypothesis, she collected a random sample of 100 intra-class travel times and found the mean to be 10.12 mins. Assume that it is known from previous studies that the distribution of intra-class times is symmetric with a standard deviation of 1.60 mins. Test the dean's hypothesis with* $\alpha = 0.10$.

The 11-steps (Section 15.1) for completing a full hypothesis test for this example are as follows:

1. α=0.10.
2. $H_0 : \mu = 10$ mins vs. $H_A : \mu > 10$ mins, where μ is the mean time for ALL intra-class travel events.
3. A 1-Sample Z-Test is required because (i) a quantitative variable (intra-class travel time) was measured, (ii) individuals from one population were sampled (students at the Dean's school), and (iii) σ is thought to be known (=1.60 mins).
4. The data appear to be part of an observational study (the dean did not impart any conditions on the students) with a random selection of individuals.
5. (i) $n = 100 > 30$ and (ii) σ is thought to be known (=1.60 mins).
6. $\bar{x}$=10.12.
7. $Z = \frac{10.12-10}{\frac{1.60}{\sqrt{100}}} = \frac{0.12}{0.16}$=0.75.
8. p-value=0.2266.
9. H_0 is not rejected because the $p-value > \alpha = 0.10$.
10. It appears that the mean for **all** intra-class travel events is not more than 10 minutes.
11. The confidence region is not computed when H_0 is not rejected.

R Appendix:

```
( distrib(10.12,mean=10,sd=1.60/sqrt(100),lower.tail=FALSE) )
```

Review Exercises

15.1 A researcher is investigating the growth of a certain cactus under a variety of environmental conditions. He knows from previous research that the growth of this particular type of cactus is approximately normally distributed with a standard deviation of 1.40 cm. Under the current environmental conditions that he is investigating, however, he does not know the mean. He does hypothesize that it is no more than 4 cm. To test this hypothesis he used a preliminary sample of 10 randomly-selected cacti. He found the sample mean for these cacti to be 3.26 cm. Use this information to test his hypothesis with $\alpha = 0.05$. Answer

15.2 Owens and Pronin (2000) studied the age and growth of pike in Chivyrkui Bay on Lake Baikal. They found that the length of the sample of 30 pike in Lake Baikal was slightly right-skewed with a mean of 656.1 mm. Suppose that a recent article in an outdoor magazine reported the average length of all pike in this lake to be 600 mm long. It is known from previous studies that the standard deviation of pike length is about 130 mm. Perform a test, using a 95% confidence level, to determine if the mean length of pike reported by the researchers significantly differs from that reported in the outdoor magazine. Answer

15.3 1-Sample Z-Test in R

If raw data exist, the calculations for a 1-Sample Z-test can be efficiently computed with `z.test()`. This function requires the vector of quantitative data as the first argument, the hypothesized value for μ in `mu=`, and the known σ in `sd=`. Additionally, the type of alternative hypothesis may be declared in `alt=`, where `alt="two.sided"` (the default), `alt="less"`, and `alt="greater"` correspond to the "not equals", "less than", and "greater than" hypotheses, respectively. Finally, the level of confidence may be given as a proportion (between 0 and 1) in `conf.level=` (which defaults to 0.95). The results of `z.test()` mayd be assigned to an object and submitted to `plot()` to visualize the test statistic and p-value. Use of `z.test()` is illustrated in the following example.

15.3.1 Body Temperature

Consider the following situation,[1]

> Machowiak et al. *(1992) critically examined the belief that the mean body temperature is 98.6^{0}F by measuring the body temperatures in a sample of healthy humans. Their data are found in* BodyTemp.csv*. Use these data, with a supposedly known $\sigma = 0.63^0F$, and an $\alpha = 0.01$ to determine if the mean body temperature differs from 98.6^{0}F.*

The 11-steps (Section 15.1) for a hypothesis test for this example are as follows:

1. α=0.01.
2. $H_0 : \mu = 98.6^0$F vs. $H_A : \mu \neq 98.6^0$F, where μ is the mean body temperature for ALL healthy humans. [*Note that not equals was used because the researchers want to determine if the temperature is **different from** 98.6^{0}F.*]
3. A 1-Sample Z-Test is required because (i) a quantitative variable (i.e., body temperature) was measured, (ii) individuals from one population were sampled (i.e., healthy humans), and (iii) σ is thought to be known (= 0.63^{0}F).
4. The data appear to be part of an observational study although this is not made clear in the background information. There is also no evidence that randomization was used.
5. (i) $n = 130 > 30$ and (ii) σ is thought to be known (= 0.63^{0}F).
6. $\bar{x}$=98.25^{0}F (Table 15.2).
7. $Z = -6.35$ (Table 15.2).
8. p-value=$<$ 0.00005 (Table 15.2).
9. Reject H_0 because $p - value < \alpha = 0.01$.
10. It appears that the mean body temperature of ALL healthy humans is less than 98.6^{0}F. [*Note that the test was for a difference but because $\bar{x} < 98.6$ this more specific conclusion can be made.*]
11. I am 99% confident that the mean body temperature (μ) for ALL healthy humans is between 98.11 and 98.39^{0}F (Table 15.2).

R Appendix:

```
bt <- read.csv("data/BodyTemp.csv")
headtail(bt)
( bt.z <- z.test(bt$temp,mu=98.6,sd=0.63,conf.level=0.99) )
plot(bt.z)
```

[1] There is an interesting discussion of studies of body temperature at The Physics Factbook.

Table 15.2. Results from 1-Sample Z-Test for mean body temperature.

```
z = -6.3482, n = 130.000, Std. Dev. = 0.630, Std. Dev. of the sample mean =
0.055, p-value = 2.178e-10
99 percent confidence interval:
 98.10690 98.39156
sample estimates:
mean of bt$temp
       98.24923
```

Review Exercises

15.3 A study by Cheshire *et al.* (1994) reported on six patients with chronic myofascial pain syndrome (introduced in Review Exercise 13.3). The researchers determined the duration of pain for the six patients were 2.5, 2.7, 2.8, 2.8, 2.8, and 3.0. Test the hypothesis that the mean pain length was greater than 2.5 years at the 10% significance level. Assume that it is known that the distribution of duration of pain is normal with a standard deviation of 0.5 years. *Answer*

15.4 Suppose that it is known that cholesterol levels in women aged 21-40 in the U.S. has a mean of 190 mg/dl and standard deviation of 40 mg/dl. Suppose that we want to determine, at the 10% significance level, if the cholesterol level of Asian women is different from U.S. women as determined from 40 randomly selected Asian women aged 21-40 who had recently immigrated to the U.S. Assume that the Asian women have the same standard deviation as the U.S. women population. The data from this sample are found in Cholesterol.csv. *Answer*