Chapter 8. Testing Hypotheses

Part 1. Introduction and One-Sample Tests

J.C. Wang

Western Michigan University
Department of Statistics

Goal and Objectives

- **Goal**: To learn about Hypotheses Testing Procedures. To learn hypothesis-testing methodology as a technique for analyzing differences and making decisions.
- **Objectives**:
  - H_0 vs. H_1
  - Type I vs. Type II errors
  - One-tailed vs. Two-tailed tests
  - *p*-Value
  - Understand relationship among α, β, and n
  - Apply methodology

Outline

Definitions

Four parts of statistical testing

- Hypotheses
- Test statistic
- *p*-value
- Conclusion

Statistical Hypotheses
one-tailed test

- Hypotheses has two parts: Null hypothesis, H_0 and Alternate hypothesis, H_1
- Keyword: *less than* $\Longrightarrow$ *left-tailed test*

$$H_0 : \mu \geq \mu_0 \text{ vs. } H_1 : \mu < \mu_0, \text{ where } \mu_0 \text{ is a constant.}$$

- Keywords: *greater than* $\Longrightarrow$ *right-tailed test*

$$H_0 : \mu \leq \mu_0 \text{ vs. } H_1 : \mu > \mu_0, \text{ where } \mu_0 \text{ is a constant.}$$

Statistical Hypotheses
two-tailed test

- Hypotheses has two parts: Null hypothesis, H_0 and Alternate hypothesis, H_1
- Keywords: *not equal to* $\Longrightarrow$ *two-tailed test*

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0, \text{ where } \mu_0 \text{ is a constant.}$$

Test Statistics

- The test statistic is equal to z if the SD is from the population or a process.

$$z = \frac{pt.est - H_0\ value}{SE},\ \text{where } SE = \frac{\sigma}{\sqrt{n}}$$

- The test statistic is equal to t if the SD is computed from the sample.

$$t = \frac{pt.est - H_0\ value}{SE},\ \text{where } SE = \frac{s}{\sqrt{n}}$$

P Value

one-tailed test

p-value comes in two distributions:

- Normal distribution

$$p\text{-value} = \textbf{normalCDF}(|TS|, 9999)$$

- Student t distribution

$$p\text{-value} = \textbf{tCDF}(|TS|, 9999, df)$$

where $|TS|$ is the absolute value of the test statistic

P Value

two-tailed test

p-value comes in two distributions:

- Normal distribution

$$p\text{-value} = 2 \times \textbf{normalCDF}(|TS|, 9999)$$

- Student t distribution

$$p\text{-value} = 2 \times \textbf{tCDF}(|TS|, 9999, df)$$

where $|TS|$ is the absolute value of the test statistic

Drawing Conclusion

in a hypothesis test

If the p-value $< \alpha$ (Significance level), then reject the Null Hypothesis (H_0): there is evidence that ...; otherwise, do not reject the Null Hypothesis (H_0): there is no evidence that ....

Note: Significance level = maximum allowable risk of committing type I decision error; p-value = observed risk of committing type I decision error

iClicker Question 8.1 Pre-lecture

iClicker Question 8.1 Pre-lecture

One-Sample Z Test

cereal box packaging example

Let's consider our cereal box example. You are the manager of the packaging process at a cereal manufacturing plant. You want to determine if the cereal filling process is in control. The process requires no corrective action if the correct amount of cereal per box is 368 grams. To study this, you decide to take a **random sample of 25 boxes**, weigh each one, and then evaluate the difference between the sample statistic and the hypothesized population parameter by comparing the mean weight from the sample to the **expected mean of 368 grams** specified by the company. The **sample mean is 372.5** and the **process standard deviation is 15**. **Is there evidence** that the weight is **different from 368 grams**? You have selected $\alpha = 0.05$ as your significance level.

Cereal Box Packing Example

continued

- Given: $\mu_0 = 368$; $\overline{x} = 372.5$; $\sigma = 15$; $n = 25$.
- What are the key words in this problem?
  Process SD indicates that we will use z for the test statistic. **Is there evidence** means that we use a hypothesis test and **different from 368 grams** means two-tailed alternative.
- You should use this approach to answer the following question.
- Use TI calculator z-test because of a *process SD*.

Cereal Box Packing Example

continued

- What type of test this is? One-tailed test or two-tailed test?
  A: Two-tailed
- What are the hypotheses?
  A: $H_0 : \mu = 368$ vs. $H_1 : \mu \neq 368$
- What is the significance level α?
  A: $\alpha = 0.05$ (and critical value = $z_{.025} = 1.96$)
- What is the sample size?
  A: $n = 25$
- What is the standard error (SE)?

$$\textbf{A}: \ SE = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{25}} = 3$$

Cereal Box Packing Example

continued

- What is the test statistic?

$$z = \frac{pt.est - H_0 \text{ value}}{SE} = \frac{372.5 - 368}{3} = \frac{4.5}{3} = 1.5$$

- What is the p-value?
  A : p-value $= 2 \times$ **normalCDF**$(1.5, 9999) = 2 \times .0668 =$ .1336
- What is the conclusion, i.e., *is there evidence that the weight is different from 368 grams*?
  A: No, since p-value $\not< \alpha = 0.05$. Do not reject H_0 and conclude that there is not enough evidence that the true average weight differs from 368 grams.

Cereal Box Packing Example

using TI calculators

- **Do this**: STAT $\rightarrow$ TESTS $\downarrow$ Z-TEST $\rightarrow$ STATS $\downarrow$ μ:368 $\downarrow$ σ:15 $\downarrow$ $\bar{x}$:372.5 $\downarrow$ n: 25 $\downarrow$ μ:$\neq \mu_0$ $\downarrow$ CALCULATE
- **Readout**:

  Z-Test
  $\mu \neq 368$
  z=1.5
  p=.1336
  $\bar{x} = 372.5$
  n=25

One-Sample Z Test
summary

- **Assumptions**
  - Large sample (sample size $n \geq 30$, say) or normal data
  - 'Known' population standard deviation σ
- **Test statistic** (TS): (note: μ_0 = hypothesized value)

$$z = \frac{pt.est - \mu_0}{SE} = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

- **Hypothesis test**: one of
  - *Left-tailed*—$H_0 : \mu \geq \mu_0$ vs. $H_1 : \mu < \mu_0$.
  - *Right-tailed*—$H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$.
  - *Two-tailed*—$H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$.

One-Sample Z Test
summary, continued

- **p-value**:
  $\#ofTails \times P(Z > |TS|) =$
  $\#ofTails \times$ **normalCDF**$(|TS|, 9999)$
  Reject H_0 if p-value $< \alpha$.
- **Critical value**:
  $CVal = c \times z_{\alpha/\#ofTails} = c \times$ **invNORM**$(1 - \alpha/\#ofTails)$,
  where $c = -1$ for left-tailed and $c = 1$, otherwise.
- **Rejection rule using critical value**: reject H_0
  - (*Left-tailed*) if $TS < CVal$
  - (*Right-tailed*) if $TS > CVal$
  - (*Two-tailed*) if $|TS| > CVal$. That is, reject H_0 if $TS > CVal$ or $TS < -CVal$. And we write $\pm CVal$ for critical value.

One-Sample t Test

TV violence example

Now let's consider the situation where we must compute the standard deviation from the data.
Suppose you are concerned about the amount of violence on TV. For a Stat 2160 project, you decide to randomly select 10 TV programs, watch them, and count the number of "violent scenes" in each. Here is the data you collected:

```
32   12   20   10    4   18   25   26   17   14
```

TV Violence Example

continued

Let's review the given information before we answer the following question. Is there evidence that the (claim) number of "violent scenes" is *at least 21 scenes*? What is your conclusion? How many tails do we have for this test? Assume $\alpha = 0.05$.
You're testing against the claim (null hypothesis).

TV Violence Example

continued

- How many tails is this test?
  A: One; Note the keyword, *at least* or *less than*.
- What are the hypotheses?
  A: $H_0 : \mu \geq 21$ vs. $H_1 : \mu < 21$
- What is the significance level?
  A: $\alpha = 0.05$
- What is the sample size?
  A: $n = 10$

TV Violence Example

continued, using TI calculator

- **Do this**: (assumed data in L_1) `STAT` $\rightarrow$ `TESTS` $\downarrow$ `T-TEST` $\rightarrow$ `DATA` $\downarrow$ μ_0 : 21 $\downarrow$ `List:`L_1 $\downarrow$ `Freq:`1 $\downarrow$ μ :< μ_0 $\downarrow$ `CALCULATE`
- **Readout**:

  T-Test
  $\mu < 21$
  $t = -1.2137$
  p=0.1279
  $\overline{x} = 17.8$
  Sx=8.3373
  n=10

TV Violence Example
continued

- What is the standard error?

$$SE = \frac{s}{\sqrt{n}} = \frac{8.3373}{\sqrt{10}} = 2.6365.$$

- What is the test statistic (TS)?

$$t = \frac{pt.est - H_0\text{-}value}{SE} = \frac{17.8 - 21}{2.6365} = -1.2137.$$

- How many tails in this test?
  A: One. Note: keyword=*at least*.

TV Violence Example
continued

- What is the p-value? *p*-value =
  $\#\text{ofTails} \times \textbf{tCDF}(|TS|, 9999, df)$
  $= 1 \times \textbf{tCDF}(1.2137, 9999, 9) = 0.1279$
- What is the critical value?

$$CVal = -t_{\alpha,df} = -\textbf{invT}(1-\alpha, df) = -\textbf{invT}(.95, 9) = -1.8331.$$

Reject H_0 if $TS < CVal$. Here, $-1.2137 \not< -1.8331$, so do not reject H_0. Note also: for left-tailed test, $CVal = -\textbf{invT}(1-\alpha, df) = \textbf{invT}(\alpha, df)$.

TV Violence Example

continued

- What is the conclusion?
  - Is p-value $< \alpha$? No, since $0.1279 \nless 0.05$.
  - Conclusion: do not reject H_0. There is not enough evidence to conclude that the true average number of violence scenes is less than 21.

Slow Wave Sleep Example

Exercise 1, page 123

```
21 20 22 7 9 14 23 9 10 25 15 17 11
```

- $\overline{x} = 15.6$, $s = 6.13$
- The sample average is how far below 20%?
  A: $15.6 - 20 = -4.4$
- $$SE = \frac{s}{\sqrt{n}} = \frac{6.13}{\sqrt{13}} = 1.7.$$
- The sample average is how many SE's below 20%?

$$\frac{-4.4}{1.7} = -2.5785.$$

Slow Wave Sleep Example

continued

- Is the sample average significantly below 20%, or is it just chance variation?
  A: Significantly below (reasoned later).
- If you conduct a test of significance on the following hypothesis: 'Does the data provide scientific evidence that elderly males spend less than 20% of their sleep time in REM?', how would you write the null and alternative hypotheses?
  A: $H_0 : \mu \geq 20$ vs. $H_1 : \mu < 20$.
- What is the test statistic you would use?

$$t = \frac{-4.4}{1.7} = -2.5785.$$

Slow Wave Sleep Example

continued

- What distribution curve would you use to compute the p-value?
  A: t-curve.
- Calculate a (one-tailed) p-value for your test.
  p-value $= \textbf{tCDF}(|-2.5785|, 9999, 12) = 0.0121$
- What is the conclusion of your test?
  A: Since p-value $< \alpha = 0.05$, we reject H_0 and conclude that elderly males spend less than 20% of their sleep time in REM.

One-Sample t Test

summary

- **Assumptions**
  - Large sample (sample size $n \geq 25$, say) or normal data
  - Unknown population standard deviation σ
- **Test statistic** (TS): (note: μ_0 = hypothesized value)

$$t = \frac{pt.est - \mu_0}{SE} = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

- **Hypothesis test**: one of
  - *Left-tailed*—$H_0 : \mu \geq \mu_0$ vs. $H_1 : \mu < \mu_0$.
  - *Right-tailed*—$H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$.
  - *Two-tailed*—$H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$.

One-Sample t Test

summary, continued

- **p-value**:
  $\#ofTails \times P(t > |TS|) = \#ofTails \times \textbf{tCDF}(|TS|, 9999, df)$
  Reject H_0 if p-value $< \alpha$.
- **Critical value**:
  $CVal = c \times t_{\alpha/\#ofTails,df} = c \times \textbf{invT}(1 - \alpha/\#ofTails, df)$,
  where $c = -1$ for left-tailed and $c = 1$, otherwise.
- **Rejection rule using critical value**: reject H_0
  - (*Left-tailed*) if $TS < CVal$
  - (*Right-tailed*) if $TS > CVal$
  - (*Two-tailed*) if $|TS| > CVal$. That is, reject H_0 if $TS > CVal$ or $TS < -CVal$. And we write $\pm CVal$ for critical value.

One-Sample t Test

summary, continued

Using MATH-SOLVER in TI-83 for *CVal*: set equation to

$$0 = tcdf(L, U, D) - A/T$$

Then

- **Left-tailed**: Set $L = -9999$ (i.e. $-\infty$); $D = df$ (here $n - 1$); $A = \alpha$; and $T = 1$. Solve U for *CVal*.
- **right-tailed**: Set $U = 9999$ (i.e. ∞); $D = df$ (here $n - 1$); $A = \alpha$; and $T = 1$. Solve L for *CVal*.
- **two-tailed**: Set $U = 9999$ (i.e. ∞); $D = df$ (here $n - 1$); $A = \alpha$; and $T = 2$. Solve L for *CVal*.

iClicker Question 8.2

iClicker Question 8.2