

Hypothesis Testing Procedures for One-sample Data

Ryan Miller

1. The one-sample Z-test
 - ▶ rationale, steps, examples
2. The one-sample T-test
 - ▶ rationale, steps, examples

Proceduralized hypothesis testing

So far we've used the following procedure for hypothesis testing:

- 1) Propose null and alternative hypotheses
- 2) Use StatKey to simulate outcomes under the null hypothesis (ie: create a null distribution)
- 3) Compare the outcome observed in the real data against the null distribution to find the p -value

This procedure can be made more general using *test statistics*, or Z -scores, that describe how many standard errors (SE) the observed outcome is above or below what we'd expect under H_0

Proceduralized hypothesis testing

Randomization test:

- ▶ Propose H_0

The Z -test:

- ▶ Propose H_0

Proceduralized hypothesis testing

Randomization test:

- ▶ Propose H_0
- ▶ Use StatKey to simulate outcomes under the null hypothesis

The Z-test:

- ▶ Propose H_0
- ▶ Use CLT formulas to calculate an SE , then find
$$Z = \frac{\text{Observed} - \text{Null}}{SE}$$

Proceduralized hypothesis testing

Randomization test:

- ▶ Propose H_0
- ▶ Use StatKey to simulate outcomes under the null hypothesis
- ▶ Locate the observed outcome in the null distribution and count the simulated outcomes that are at least as extreme

The Z-test:

- ▶ Propose H_0
- ▶ Use CLT formulas to calculate an SE , then find $Z = \frac{\text{Observed} - \text{Null}}{SE}$
- ▶ Locate Z in the Standard Normal curve and find the area outside it

Example (single proportion)

For a single proportion, Central Limit theorem suggests:

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

- ▶ In the infant toy choice study, $H_0 : p = 0.5$, which reflected infants choosing randomly between the two toys
 - ▶ The observed outcome was $\hat{p} = 14/16 = 0.875$, or 14 of 16 infants chose the “helper”

Example (single proportion)

For a single proportion, Central Limit theorem suggests:

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

- ▶ In the infant toy choice study, $H_0 : p = 0.5$, which reflected infants choosing randomly between the two toys
 - ▶ The observed outcome was $\hat{p} = 14/16 = 0.875$, or 14 of 16 infants chose the “helper”

1) Under H_0 , $SE = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5(1-0.5)}{16}} = 0.125$

Example (single proportion)

For a single proportion, Central Limit theorem suggests:

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

- ▶ In the infant toy choice study, $H_0 : p = 0.5$, which reflected infants choosing randomly between the two toys
 - ▶ The observed outcome was $\hat{p} = 14/16 = 0.875$, or 14 of 16 infants chose the “helper”

- 1) Under H_0 , $SE = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5(1-0.5)}{16}} = 0.125$
- 2) Then, $Z = \frac{\hat{p}-p_0}{SE} = \frac{0.875-0.5}{0.125} = 3$

Example (single proportion)

For a single proportion, Central Limit theorem suggests:

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

- ▶ In the infant toy choice study, $H_0 : p = 0.5$, which reflected infants choosing randomly between the two toys
 - ▶ The observed outcome was $\hat{p} = 14/16 = 0.875$, or 14 of 16 infants chose the “helper”
- 1) Under H_0 , $SE = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5(1-0.5)}{16}} = 0.125$
- 2) Then, $Z = \frac{\hat{p}-p_0}{SE} = \frac{0.875-0.5}{0.125} = 3$
- 3) Comparing $Z = 3$ against a Standard Normal curve, the one-sided p -value is 0.0013 (two-sided is 0.0026)

Practice #1

We've previously discussed a study conducted by Johns Hopkins University that found 31 of 39 babies born 15 weeks early went on to survive. According to Wikipedia, the survival rate for babies born this early is 70%. Does the Johns Hopkins University study provide compelling evidence to refute Wikipedia's claim?

- 1) Propose a null hypothesis and an alternative hypothesis
- 2) Use CLT to find a standard error, then calculate a Z -value measuring how standard errors the observed outcome is from the value specified in H_0
- 3) Compare Z against the Standard Normal distribution to find the p -value

Practice #1 (solution)

- 1) $H_0 : p = 0.7$ vs. $H_a : p \neq 0.7$
- 2) $SE = \sqrt{\frac{0.7(1-0.7)}{39}} = 0.073$, then notice we observed $\hat{p} = 31/39 = 0.795$, so $Z = \frac{0.795-0.7}{0.073} = 1.29$
- 3) The two-sided p -value corresponding to $Z = 1.29$ is 0.198, indicating that these data do not provide sufficient evidence to refute Wikipedia's claim.

Practice #2

A 2009 study investigated the behavior of 120 players in a rock-paper-scissors tournament. In their first move, 66 players chose “rock”, 39 chose “paper”, and 15 chose “scissors”. Do these data provide statistically compelling evidence that players in this tournament were not choosing their first move randomly?

- 1) Propose a null hypothesis and an alternative hypothesis
- 2) Use CLT to find a standard error, then calculate a Z -value measuring how standard errors the observed outcome is from the value specified in H_0
- 3) Compare Z against the Standard Normal distribution to find the p -value

Practice #2 (solution)

- 1) $H_0 : p = 0.33$, where p is the proportion that choose “rock”,
vs. $H_0 : p \neq 0.33$
- 2) $SE = \sqrt{\frac{0.33(1-0.33)}{120}} = 0.043$, then notice we observed
 $\hat{p} = 66/120 = 0.55$, so $Z = \frac{0.55-0.33}{0.043} = 5.12$
- 3) The p -value is nearly zero, indicating overwhelming evidence
that players in this tournament are not choosing randomly,
they're playing “rock” significantly more often than expected
(55% of the time).

Modifications for testing a single mean

For a single mean, recall that CLT suggests:

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

However, estimating σ (an unknown population parameter) via s (the sample standard deviation) introduces additional uncertainty that necessitates the t -distribution

Modifications for testing a single mean

Using $SE = \frac{s}{\sqrt{n}}$, the one-sample T -test calculates a T -value:

$$T = \frac{\bar{x} - \mu}{SE}$$

Then compares this T -value against a t -distribution with $df = n - 1$ to find the p -value

Practice #1

Radon is a toxic gas and the second leading cause of lung cancer. The EPA's federal action limit is 4 pCi/L, but they advise residents to take action if levels are above 0.4 pCi/L. In the basement of an Iowa home, radon levels are tested on 8 randomly selected dates over the course a month, the measurements are:

$$\{.2, .7, .3, .9, .5, .3, .7, .6\}$$

- 1) Propose a null hypothesis and an alternative hypothesis
- 2) The mean of these data is $\bar{x} = 0.525$ and $s = 0.243$, use this information to find a T -value measuring how standard errors the observed outcome is from the value specified in H_0
- 3) Compare T against the proper t -distribution to find the p -value

Practice #1 (solution)

- 1) $H_0 : \mu = 0.4$ vs. $H_a : \mu > 0.4$
- 2) $T = \frac{0.525 - 0.4}{0.243/\sqrt{8}} = 1.455$
- 3) The one-sided p -value using $df = 7$ is 0.094, so we conclude there's borderline evidence of Radon levels that exceed 0.4. Since a Type II error is more costly than a Type I error in application, we might still choose to reject H_0 .

Practice #2

In 2010, new international rules were created to regulate swimsuit coverage and material after an inordinate amount of records were set at the 2008 Olympics by swimmers wearing a suit known as the LZR Racer. The wetsuits data, which is available on our course website the contains 1500m swim velocities of 12 competitive swimmers with and without a scientifically designed wetsuit.

- 1) Use StatKey to find the sample mean and sample standard deviation of the variable “difference”
- 2) Propose null and alternative hypotheses these researchers should evaluate using the variable “difference”
- 3) Find the T pertaining to these data, then compare it against the proper t -distribution to find the p -value.

Practice #2 (solution)

- 1) Using the Descriptive Statistics and Graphs section of StatKey, $\bar{x} = 0.077$ and $s = 0.022$
- 2) $H_0 : \mu = 0$, or the average improvement in velocity when wearing the wetsuit is zero, vs. $H_a : \mu \neq 0$
- 3) $T = \frac{0.077 - 0}{0.022/\sqrt{12}} = 12.12$, the p -value is nearly zero, indicating overwhelming evidence that the wetsuit improves swim velocity

Comments (paired designs)

- ▶ The wetsuits study is an example of a **paired design**, a type of design where each subject serves as their own control
- ▶ Paired designs have a number of statistical advantages over other designs, with an important one being the elimination of confounding variables

Summary

This presentation streamlines hypothesis testing into a procedure involving a consistent set of steps that can be applied to nearly any scenario:

- 1) Propose a null and alternative hypothesis
- 2) Calculate a test statistic (a Z -value or a T -value)
- 3) Compare the test statistic against a null distribution (a Standard Normal curve or a t -distribution) to find the p -value
- 4) Use the p -value to make a decision

We focused on two specific applications:

- ▶ Testing $H_0 : p = p_0$ using $Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$
- ▶ Testing $H_0 : \mu = p_0$ using $T = \frac{\bar{x} - \mu}{s/\sqrt{n}}$