Hypothesis Testing Procedures for One-sample Data

Ryan Miller

Outline

1. Hypothesis testing for one-sample categorical data
   - Exact binomial test, one-sample Z-test
2. Hypothesis testing for one-sample quantitative data
   - One sample t-test, permutation tests

Review

Hypothesis testing involves the following steps:

1) Propose null and alternative hypotheses
2) Evaluate how much evidence the sample data provide against the null hypothesis
3) Make a decision

Steps #1 and #3 are straightforward, but Step #2 will vary depending upon the situation

One-sample categorical data

The phrase "one-sample categorical data" describes a *single categorical outcome* for a *single group of cases*. We've described this type of data in two ways:

1) Frequencies - how many cases belong to each category
2) Proportions - what fraction of the total sample is each category

Either of these two descriptive statistics serve as the basis of a hypothesis test

The exact binomial test

- The **exact binomial test** involves an *exact model* for the *number of observed successes* under a given null hypothesis
  - More specifically, the *null distribution* is the binomial distribution defined by p (specified in H_0) and n
  - For the infant toy-choice example, $X \sim \text{Binomial}(n = 16, p = 0.5)$

```r
## Calculating the 1-sided p-value using d-binom
dbinom(14, size = 16, prob = 0.5) + dbinom(15, size = 16, prob = 0.5) +
  dbinom(16, size = 16, prob = 0.5)
```

```
## [1] 0.002090454
```

The exact binomial test in R

The `binom.test` function makes performing the exact binomial test very quick and easy:

```r
binom.test(x = 14, n = 16, p = 0.5, alternative = "greater")
```

```
##
##  Exact binomial test
##
## data:  14 and 16
## number of successes = 14, number of trials = 16, p-value = 0.00209
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.6561748 1.0000000
## sample estimates:
## probability of success
##                  0.875
```

The exact binomial test and CLT

- When learning about Central Limit theorem, an important caveat was that the sample size was sufficiently larger
  - For a single proportion, the "success-failure" condition was $np \geq 10$ and $n(1-p) \geq 10$

The exact binomial test and CLT

- When learning about Central Limit theorem, an important caveat was that the sample size was sufficiently larger
  - For a single proportion, the "success-failure" condition was $np \geq 10$ and $n(1-p) \geq 10$
- Previously, we looked at a study published in *Nature* where 14 of 16 infants chose the "friendly" toy
  - Notice that only 2 "failures" were observed, so the Normal distribution suggested by CLT might be inappropriate

```r
## Comparing with the p-value found using CLT
pnorm(14/16, mean = 0.5, sd = sqrt(0.5*0.5/16), lower.tail = FALSE)
```

```
## [1] 0.001349898
```

While it doesn't change our conclusion, the exact p-value was roughly *double* that of the CLT approach

The one-sample Z-test

For sufficiently large samples (at least 10 “successes” and 10 “failures”), Central Limit theorem provides a probability model for the null distribution:

$$\hat{p} \sim N\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$$

This model can then be *standardized*, resulting in the Z-test:

$$Z = \frac{\text{observed} - \text{expected}}{SE} = \frac{\hat{p} - p_0}{p_0(1-p_0)/n} \sim N(0, 1)$$

The Z-test expresses the observed outcome as a Z-score that can be compared against the Standard Normal distribution.

Steps of the Z-test

The main advantage of the standardized approach is that it can be applied to a variety of applications involving the same set of steps:

1) State null and alternative hypotheses
2) Calculate a Z-value that relates the observed outcome with what you'd expect under the null hypothesis
3) Compare this Z-value against the Standard Normal curve to obtain a p-value
4) Use the p-value to reach a conclusion

Practice

In the infant toy choice example, recall that 14 of 16 infants chose the "helper toy" and the researchers were interested in disproving the hypothesis that infants choose randomly between toys.

1) Use a Z-test to evaluate the hypothesis $H_0 : p = 0.5$. Clearly list all four steps of this test.
2) Compare the p-value of this test with the results of the two tests performed earlier in this presentation.

Practice (solution)

1) Step 1: $H_0 : p = 0.5$ vs. $H_A : p \neq 0.5$, Step 2: $Z = \frac{0.875-0.5}{0.5(1-0.5)/16} = 3$, Step 3: using `pnorm`, the one-sided p-value is 0.00135, Step 4: There is strong evidence to reject H_0 and conclude that infants do not choose randomly (suggesting they can detect friendly behavior)
2) The Z-test is identical to using CLT directly to perform a hypothesis test. However, notice that only 2 "failures" were observed in the sample data, which indicates an insufficient sample size to trust the Normal model that is used in the Z-test. For this reason, the exact binomial p-value should be preferred.

The one-sample T-test

For a single mean, CLT suggests:

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

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

The one-sample T-test

For a single mean, CLT suggests:

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

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

$$T = \frac{\text{observed} - \text{expected}}{SE} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

When testing $H_0 : \mu = \mu_0$ this T-value is compared against a t-distribution with $df = n - 1$ to find the p-value.

Practice

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 significantly higher than 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 that can be used to test whether action is necessary.
2) For these data, $\bar{x} = 0.525$ 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 (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.

Alternatives to the t-test

- The t-test is only statistically valid for small, Normally distributed samples, or for large samples (of any distributional shape)
- For small samples that do not follow a Normal distribution, an alternative approach is to perform a test of the *median* of the population ($H_0 : m = m_0$) using the *sample median*
    - This test is known as the Wilcoxon signed rank test, available through `wilcox.test` in R

```
radon = c(.2,.7,.3,.9,.5,.3,.7,.6)
wilcox.test(x = radon, mu = 0.4, alternative = "greater")
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  radon
## V = 26, p-value = 0.1462
## alternative hypothesis: true location is greater than 0.4
```

Summary

- The exact binomial test should be used evaluate $H_0 : p = p_0$ using sample data
  - The Z-test will produce a result similar to that of the exact binomial test when the "success-failure" condition is met
- The T-test should be used to evaluate $H_0 : \mu = \mu_0$
  - This test involves calculating a T-value defined as $T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ and comparing it to a t-distribution with $df = n - 1$
- Wilcoxon's signed rank test is an alternative to the T-test that uses the median, but it's less powerful and should only be used for small, non-Normal samples