Chapter 6: Hypothesis Testing with R - part B

Alan T. Arnholt
Appalachian State University
arnholt@math.appstate.edu
Testing $\pi$: The $z$-Test

Consider Example 6.6 from your Book. A recent report claimed that 20% of all college graduates find a job in their chosen field of study. A survey of a random sample of 500 graduates found that 110 obtained work in their field. Is there statistical evidence to refute the claim?
Step 1:

$H_O : \pi = 0.20$

$H_A : \pi \neq 0.20$
Step 2:

Let $p$ represent the proportion of college graduates that find work in their chosen field of study. In this case $p_{obs} = 110/500 = 0.22$. 
Step 3:

Since \( n \times \pi > 5 \) and \( n \times \pi \times (1 - \pi) > 5 \), the sampling distribution of \( p \) is approximately normal with \( \mu_p = \pi \) and \( \sigma_p = \sqrt{\frac{\pi \times (1 - \pi)}{n}} \). The standardized test statistic is

\[
Z = (p - \pi_0) / \sqrt{\pi_0 \times (1 - \pi_0) / n} \sim N(0, 1)
\]

The observed value of the standardized test statistic is

\[
z_{obs} = (0.22 - 0.2) / \sqrt{0.2 \times 0.8 / 500} = 1.118034.
\]

R-commands

```r
> n <- 500
> X <- 110
> p.obs <- X/n
> z.obs <- (p.obs - .2)/sqrt(.2*.8/n)
> z.obs
[1] 1.118034
```
Step 4:

We are interested in determining how likely it is to observe a value as extreme as 1.118 in a standard normal distribution. To do this, we find the area to the right of 1.118 in a standard normal distribution and multiply the answer times 2 since the alternative hypothesis is non-directional.

\[
2 \times P(Z \geq 1.118) = 2 \times (1 - P(Z \leq 1.118)) = 2 \times (1 - 0.868) = 2 \times (0.132) = 0.264. \text{ In other words, the } p\text{-value is 0.264. This value is not small enough to suggest the data is inconsistent with the null hypothesis. Consequently, we fail to reject the null hypothesis.}
\]

R-cmds

```r
> p.value <- 2*(1-pnorm(z.obs))
> p.value
[1] 0.2635525
```
Step 5:

We fail to find evidence to suggest the proportion of college graduates that find work in their chosen field is not 20%.
Using the function `prop.test()`

```r
> prop.test(x=110,n=500,p=.2,correct=FALSE)

1-sample proportions test without continuity correction

data: 110 out of 500, null probability 0.2
X-squared = 1.25, df = 1, p-value = 0.2636
alternative hypothesis: true p is not equal to 0.2
95 percent confidence interval:
 0.1859009 0.2583687
sample estimates:
  p
0.22

> z.obs <- (prop.test(x=110,n=500,p=.2,correct=FALSE)$statistic)^.5
> z.obs
X-squared
1.118034
```

For this type of problem, `prop.test()` has two required arguments `x`, and `n`. For additional help, see the help file by typing `?prop.test`. Note that `correct=FALSE` is chosen since we did not apply a continuity correction when computing $z_{obs}$. 
Testing $\pi$: Using the Binomial Distribution

Consider Example 6.6 from your Book. A recent report claimed that 20% of all college graduates find a job in their chosen field of study. A survey of a random sample of 500 graduates found that 110 obtained work in their field. Is there statistical evidence to refute the claim?

Using the function

\texttt{prop.test()}

\textbf{Testing $\pi$: Using the Binomial Distribution}

\texttt{binom.test()}

\textbf{Review of Bivariate Data Analysis}

\texttt{R Code}

\textbf{Additional Resources Related to R}

\textbf{Best of Skill on Your Finals!}
Step 1:

\[ H_O : \pi = 0.20 \]

\[ H_A : \pi \neq 0.20 \]
Step 2:

Let $X$ represent the number of college graduates that find work in their chosen field of study.
Step 3:

Under reasonable assumptions, (independence, constant probability of success) the distribution of $X$ is approximately $Bin(n = 500, \pi = 0.20)$. In this case, $x_{\text{obs}} = 110$. 
Step 4:

We are interested in determining how likely it is to observe a value as extreme as 110. To do this, we add up all of the probabilities smaller than or equal to $P(X = 110)$. Clearly we do not want to compute all 501 possible probabilities for $X$ by hand.

R-cmands

```r
> p.obs <- dbinom(110,500,.2) # P(X=110)
> p <- dbinom(0:500,500,.2) # all 501 p-values
> p.val <- sum(p[p<=p.obs]) # adding p-values <= p.obs
> p.val
[1] 0.2636306
```

In other words, the $p$-value is 0.264. This value is not small enough to suggest the data is inconsistent with the null hypothesis. Consequently, we fail to reject the null hypothesis.
Step 5:

We fail to find evidence to suggest the proportion of college graduates that find work in their chosen field is not 20%.
Using the function `binom.test()`

```r
> binom.test(x=110, n=500, p=.2)

    Exact binomial test

data:  110 and 500
number of successes = 110, number of trials = 500, p-value = 0.2636
alternative hypothesis: true probability of success is not equal to 0.2
95 percent confidence interval:
  0.1844367 0.2589117
sample estimates:
probability of success
  0.22
```

For this type of problem, `binom.test()` has two required arguments `x`, and `n`. For additional help, see the help file by typing `?binom.test`. 
Review of Bivariate Data Analysis

Using the data frame `Name` in the `BSDA` package, produce a scatterplot of `revenue` versus `value`. Quantify the relationship by computing the sample correlation coefficient. Identify any influential values in the scatterplot. Fit a straight line to the data. Remove the influential observation `Marlboro` and refit a line to the data. Report the coefficient of determination for each fitted model.
• Model 1: \( \hat{Y}_i = -0.8889132 + 2.0243924 \times x_i \), \( R^2 = 0.8843339 \)

• Model 2: \( \hat{Y}_i = -0.7525967 + 1.9434180 \times x_i \), \( R^2 = 0.7669864 \)

R Code

```r
> plot(revenue,value,col="blue")
> cor(value,revenue)
[1] 0.9403903
> identify(revenue,value,labels=Brand,col="red")
[1] 1
> Model1 <- lm(value~revenue)
> abline(Model1,col="green")
> Model2 <- lm(value[-1]~revenue[-1])
> abline(Model2,col="orange",lty="dashed")
> coef(Model1)
(Intercept) revenue
-0.8889132 2.0243924
> coef(Model2)
(Intercept) revenue[-1]
-0.7525967 1.9434180
> summary(Model1)$r.square
[1] 0.8843339
> summary(Model2)$r.square
[1] 0.7669864
```
Additional Resources Related to R

- STT-2810 Home Page
- R Home Page
- Contributed Documentation for R

Additional Resources Related to R

- Alan’s Notes
- Testing $\pi$: The $\zeta$-Test
  - Step 1:
  - Step 2:
  - Step 3:
  - Step 4:
  - Step 5:
- Using the function
  `prop.test()`
- Testing $\pi$: Using the Binomial Distribution
  - Step 1:
  - Step 2:
  - Step 3:
  - Step 4:
  - Step 5:
- Using the function
  `binom.test()`
- Review of Bivariate Data Analysis
- R Code
  - Additional Resources Related to R
- Best of Skill on Your Finals!
Best of Skill on Your Finals!

- Alan’s Notes
- Testing $\pi$: The $z$-Test
  - Step 1:
  - Step 2:
  - Step 3:
  - Step 4:
  - Step 5:
- Using the function `prop.test()`
- Testing $\pi$: Using the Binomial Distribution
  - Step 1:
  - Step 2:
  - Step 3:
  - Step 4:
  - Step 5:
- Using the function `binom.test()`
- Review of Bivariate Data Analysis
- R Code
- Additional Resources Related to R
- Best of Skill on Your Finals!

Click here to pause the presentation.