Hypothesis Testing

Chung-I Li and Yu Shyr

Center for Quantitative Sciences
Vanderbilt University School of Medicine

July 22, 2013

Learning Goals

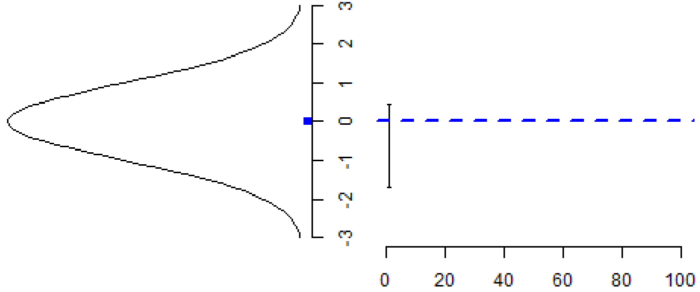
1. Learn the concepts of interval estimating and testing hypothesis via simulation
2. Learn how to conduct parametric and nonparametric tests
   - Single Population
     - testing one mean
     - testing one proportion
   - Two Populations
     - testing two means
     - testing two proportions
   - Paired Data
   - 3^+ groups

Concept of Interval Estimating

Let $X_1, \ldots, X_{30}$ be a random sample for normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. The 95% confidence interval estimator for μ is

$$[\bar{X} - 1.96\frac{1}{\sqrt{30}}, \bar{X} + 1.96\frac{1}{\sqrt{30}}]$$

where $\bar{X} = \sum_{i=1}^{30} X_i/30$.

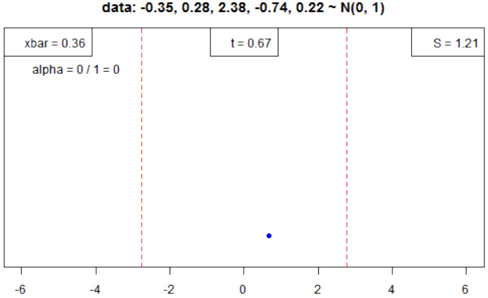


Concept of testing hypothesis

Let $X_1, \ldots, X_n$ be a random sample for a normal distribution $N(\mu, \sigma)$. Consider the testing

$$H_0 : \mu = 0 \text{ versus } H_1 : \mu = 2.$$

H_0 will be rejected if $|t = \frac{\bar{X}-0}{S/\sqrt{n}}| > 2.78$ where $\bar{X} = \sum_{i=1}^{n} X_i/n$ and $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$.

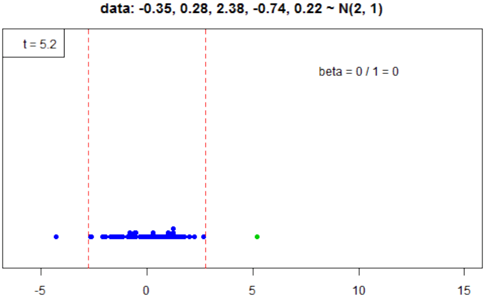


Concept of testing hypothesis

Let $X_1, \ldots, X_n$ be a random sample for a normal distribution $N(\mu, \sigma)$. Consider the testing

$$H_0 : \mu = 0 \text{ versus } H_1 : \mu = 2.$$

H_0 will be rejected if $|t = \frac{\bar{X}-0}{S/\sqrt{n}}| > 2.78$ where $\bar{X} = \sum_{i=1}^{n} X_i/n$ and $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$.



Learning Goals

parametric tests

- Single Population
  - testing one mean
  - testing one proportion
- Two Populations
  - testing two means
  - testing two proportions
- Paired Data
- 3^+ groups

Hypothesis Testing For Single Population Mean

```
> t.test(mydata$base_score, mu = 8)


One Sample t-test

data:  mydata$base_score
t = -1.748, df = 128, p-value = 0.0829
alternative hypothesis: true mean is not equal to 8
95 percent confidence interval:
 7.157 8.052
sample estimates:
mean of x
    7.605
```

Because the p-value 0.0829 is greater than 0.05, we cannot conclude that mean is not equal to 8.

Hypothesis Testing For Single Population Proportion

```
> prop.test(table(mydata$smoker), p = 0.6)


1-sample proportions test with continuity correction

data:  table(mydata$smoker), null probability 0.6
X-squared = 10.35, df = 1, p-value = 0.001295
alternative hypothesis: true p is not equal to 0.6
95 percent confidence interval:
 0.3702 0.5471
sample estimates:
     p
0.4574
```

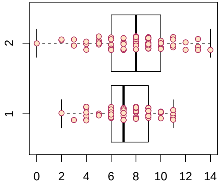
Because the p-value is less than 0.05, we conclude that the proportion is not equal to 0.6.

Hypothesis Testing For Two Population Means

```
> t.test(base_score ~ randomization, data = mydata)

Welch Two Sample t-test
data:  base_score by randomization
t = -1.46, df = 124.4, p-value = 0.147
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.5302  0.2313
sample estimates:
mean in group 1 mean in group 2
          7.262           7.912
```

Because p-value is greater than 0.05, we can't conclude that the mean difference bewteen two groups are different.



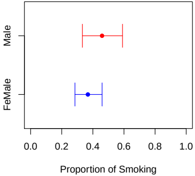
Hypothesis Testing For Two Population Proportions

Chi-Square Test

```
> prop.test(table(mydata$sex, mydata$smoker))

2-sample test for equality of proportions with continuity
correction
data:  table(mydata$sex, mydata$smoker)
X-squared = 0.6757, df = 1, p-value = 0.4111
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.3049  0.1094
sample estimates:
prop 1 prop 2
0.4286 0.5263
```

Because the p-value is greater than 0.05, we can't conclude that proportions are different.



Hypothesis Testing For Two Population Proportions

Fisher's Exact Test

```
> fisher.test(table(mydata$sex, mydata$smoker))


Fisher's Exact Test for Count Data

data:  table(mydata$sex, mydata$smoker)
p-value = 0.3374
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.2939 1.5480
sample estimates:
odds ratio
    0.6771
```

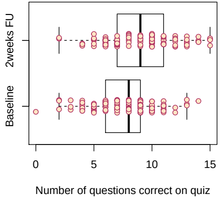
Because the p-value is greater than 0.05, we can't conclude that proportions are different.

Hypothesis Testing For Paired Data

```
> t.test(mydata$base_score, mydata$score1, paired = TRUE)

Paired t-test
data:  mydata$base_score and mydata$score1
t = -6.061, df = 128, p-value = 1.411e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.6041 -0.8145
sample estimates:
mean of the differences
                 -1.209
```

Because the p-value is less than 0.05, we conclude that the mean difference is not equal to 0.



Hypothesis Testing For Paired Data

McNemar's Chi-squared Test

```
> mytable <- table(mydata$base_score >= 6, mydata$score1 >= 6, dnn =
c("Baseline",
+      "2weeks FU"))
> mytable

        2weeks FU
Baseline FALSE TRUE
   FALSE    10   15
   TRUE      3  101

> mcnemar.test(mytable)


McNemar's Chi-squared test with continuity correction

data:  mytable
McNemar's chi-squared = 6.722, df = 1, p-value = 0.009522
```

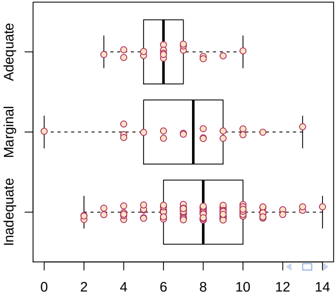
Because the p-value is less than 0.05, we conclude that proportions are not equal.

Hypothesis Testing For 3^+ Groups

```
> fit <- aov(base_score ~ pt_literacy, data = mydata)
> summary(fit)

             Df Sum Sq Mean Sq F value Pr(>F)
pt_literacy   1     42    42.3    6.69  0.011 *
Residuals   127    803     6.3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Because the p-value is less than 0.05, we conclude that the means among three groups are not equal.



Quick Review

Parametric tests

- Single Population
  - testing one mean, t.test()
  - testing one proportion, prop.test()
- Two Populations
  - testing two means, t.test()
  - testing two proportions, prop.test()
- Paired Data
  - testing two means, t.test(,paired=TRUE)
  - testing two proportions, mcnemar.test()
- 3^+ groups, aov()

Exercise

Excerise

- Compare difference in chew_score (raw score on the 3 subjective health literacy questions) between male and female (sex).
- Compare difference in the number of questions correct on quiz between 2-week follow-up (score1) and 6-week follow-up (score2).
- Compare difference in chew_score among three levels of health literacy (pt_literacy).

Learning Goals

Nonparametric tests

- Two Populations means
- Paired data
- 3^+ groups

Hypothesis Testing For Two Population Mean

Wilcoxon rank sum test

```
> wilcox.test(base_score ~ randomization, data = mydata)


Wilcoxon rank sum test with continuity correction

data:  base_score by randomization
W = 1737, p-value = 0.1097
alternative hypothesis: true location shift is not equal to 0
```

Because the p-value is greater than 0.05, we can't conclude that the mean bewteen two groups are different.

Hypothesis Testing For Paired Data

```
> wilcox.test(mydata$base_score, mydata$score1, paired = TRUE)


Wilcoxon signed rank test with continuity correction

data:  mydata$base_score and mydata$score1
V = 1122, p-value = 1.472e-07
alternative hypothesis: true location shift is not equal to 0
```

Because the p-value is less than 0.05, we conclude that the mean difference bewteen two groups are different is not equal to 0.

Hypothesis Testing For 3^+ Groups

Kruskal-Wallis rank sum test

```
> kruskal.test(base_score ~ pt_literacy, data = mydata)


Kruskal-Wallis rank sum test

data:  base_score by pt_literacy
Kruskal-Wallis chi-squared = 7.877, df = 2, p-value = 0.01948
```

Because the p-value is less than 0.05, we conclude that the means among three groups are not equal.

Overview

Parametric and Nonparametric tests

- Single Population
  - testing one mean, t.test(), wilcox.test()
  - testing one proportion, prop.test()
- Two Populations
  - testing two means, t.test(), wilcox.test()
  - testing two proportions, prop.test()
- Paired Data,
  - two means,t.test(,paired=TRUE), wilcox.test(,paired=TRUE)
  - two proportions, mcnemar.test()
- 3^+ groups, aov(), kruskal.test()