STAT4015HW1

Mingming Xu

9/13/2019

* *

1. Large-sample z-test for difference in proportions

$H_0 : p_{boy} - p_{girl} = 0$ vs. $H_a : p_{boy} - p_{gir} > 0$

2. Two-samples independent t-test for difference in polpulation means

$H_0 : \mu_{male} - \mu_{female} = 0$ vs. $H_a : \mu_{male} - \mu_{female} \neq 0$

3. Paired t-test for diffrence in means.

$H_0 : \mu_1 - \mu_2 = 0$ vs. $H_a : \mu_1 - \mu_2 \neq 0$

4. One-sample t-test for population mean.

$H_0 : \mu = 5$ vs. $H_a : \mu > 5$

5.Large-sample z-test for population proportion.

$H_0 : p = 0.70$ vs.$H_a : p \geq 0.70$

6.One-sample t-test for popluation mean.

$H_0 : \mu = 0.75$ vs.$H_a : \mu < 0.75$

7.Paired t-test for a diffrence of means.

$H_0 : \mu_f - \mu m = 10$ vs. $H_a : \mu_f - \mu_m \neq 10$

8. Large-sample z-test for popluation mean.

$H_0 : p = 0.80$ vs.$H_a : p > 0.80$

* *

1.p_1 is the proportion of Americans who favor the death penalty today; p_2 is the proportion of Americans who favor the death penalty in 1990.

a.$\hat{p_1} = 0.621$ and $\hat{p_2} = 1125/1500 = 0.75$,n1=1000,n2=1500

Assumptions: These observations form two random samples.And,$n_1\hat{p_1} \geq 10$,$n_1(1-\hat{p_1}) \geq 10$,$n_2\hat{p_2} \geq 10$,$n_2(1-\hat{p_2}) \geq 10$

Hypotheses: $H_0 : p_1 - p_2 = 0$ vs.$H_a : p_1 - p_2 < 0$

Test statistic:

$z = \frac{\hat{p_1}-\hat{p_2}-0}{\sqrt{\hat{p}(1-\hat{p})(1/n_1+1/n_2)}}$ where $\hat{p} = (n_1\hat{p_1} + n_2\hat{p_2})/(n_1 + n_2)$

```
phat=(621+1125)/(1000+1500)
z=(0.621-0.75)/sqrt(phat*(1-phat)*(1/1000+1/1500))
z
```

```
## [1] -6.884899
```

P-value:

```
pnorm(z)
```

```
## [1] 2.891435e-12
```

Conclusion: Because p-value is less than $\alpha = 0.05$, we reject the null hypothesis.

b.

a 95% confidence interval is (-1,$(\hat{p_1} - \hat{p_2}) + z_{1-\alpha}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}$).

```
(0.621-0.75)+qnorm(1-0.05)*sqrt(0.621*(1-0.621)/1000+0.75*(1-0.75)/1500)
```

```
## [1] -0.09777554
```

So, the confidence interval is (-1,-0.09777554).

Confirm: prop.test() in R

```
favor=c(621,1125)
sample=c(1000,1500)
prop.test(favor,sample,alternative = "less")
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  favor out of sample
## X-squared = 46.791, df = 1, p-value = 3.948e-12
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1.00000000 -0.09694221
## sample estimates:
## prop 1 prop 2
##  0.621  0.750
```

2.

```
unemployment=read.csv("unemployment.csv")
xbar_1=mean(unemployment$HiSchool)
xbar_2=mean(unemployment$College)
xbar_1;xbar_2
```

```
## [1] 6.566667
```

```
## [1] 2.458333
```

```
s_1=sd(unemployment$HiSchool)
s_2=sd(unemployment$College)
s_1;s_2
```

```
## [1] 1.542921
```

```
## [1] 0.4888918
```

```
##Because s_1>2s_2, we use Two-samples independent t-test for difference in polpulation means with unequ
```

μ_1 is the sample mean of the percent unemployment for high school,μ_2 is the sample mean of the percent unemployment for college.

a.

Assumption: The observations are from two independent random samples

Hypothesis: $\mu_1 - \mu_2 = 0$ vs.$\mu_1 - \mu_2 \neq 0$

Test statistic:

$t = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

```
n1=12
n2=12
t=(xbar_1-xbar_2)/sqrt(((s_1^2)/n1)+((s_2^2)/n2))
t
```

```
## [1] 8.793
```

P-value:

```
2*(1-pt(abs(t),11))
```

```
## [1] 2.630535e-06
```

Conclusion: Because p-value is less than $\alpha = 0.05$, we reject the null hypothesis.

b. A 95% confidence interval is $(\bar{x_1} - \bar{x_2}) \pm t_{1-\alpha/2,n-1}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

```
po.est=xbar_1-xbar_2
se=sqrt(((s_1^2)/n1)+((s_2^2)/n2))
t.p=qt(1-0.05/2,n1-1)
ci=c(po.est-se*t.p,po.est+se*t.p)
ci
```

```
## [1] 3.079972 5.136695
```

Confirm:

```
t.test(unemployment$HiSchool,unemployment$College,conf.level = 0.95,var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  unemployment$HiSchool and unemployment$College
## t = 8.793, df = 13.187, p-value = 7.039e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.100401 5.116266
## sample estimates:
## mean of x mean of y
##  6.566667  2.458333
```

3.

a. Let μ is the population average murder rate arrest rate

A 95% cinfidence interval is $(\bar{x} - t_{1-\alpha,n-1}\frac{s}{\sqrt{n}},\infty)$

```
xbar=mean(USArrests$Murder)
s=sd(USArrests$Murder)
n=50
alpha=0.05
xbar-qt(1-alpha,n-1)*s/sqrt(n)
```

```
## [1] 6.755308
```

So, A 95% cinfidence interval is (6.755308,∞), and it doesn't contain 5.

```
t.test(USArrests$Murder,alternative = "greater",mu=5)
```

```
##
##  One Sample t-test
##
## data:  USArrests$Murder
## t = 4.5263, df = 49, p-value = 1.921e-05
## alternative hypothesis: true mean is greater than 5
## 95 percent confidence interval:
##  6.755308      Inf
## sample estimates:
## mean of x
##     7.788
```

b.

A 95% cinfidence interval is $(\bar{x} - t_{1-\alpha,n-1}\frac{s}{\sqrt{n}},\infty)$

```
xbar=mean(USArrests$Murder)
s=sd(USArrests$Murder)
n=50
alpha=0.07
xbar-qt(1-alpha,n-1)*s/sqrt(n)
```

```
## [1] 6.863984
```

So, A 95% cinfidence interval is $(6.755308,\infty)$, and it does contain 7.

```
t.test(USArrests$Murder,alternative = "greater",mu=7,var.equal = FALSE)
```

```
##
##  One Sample t-test
##
## data:  USArrests$Murder
## t = 1.2793, df = 49, p-value = 0.1034
## alternative hypothesis: true mean is greater than 7
## 95 percent confidence interval:
##  6.755308      Inf
## sample estimates:
## mean of x
##     7.788
```

4.

Test procedure:Two-samples independent t-test for difference in polpulation means

```
dominos=c(18,20,22,24,25,25)
papa=c(15,21)
x_1bar=mean(dominos)
x_2bar=mean(papa)
s_1=sd(dominos)
s_2=sd(papa)
x_1bar;x_2bar;s_1;s_2
```

```
## [1] 22.33333

## [1] 18

## [1] 2.875181

## [1] 4.242641
```

```
##Because the s_1 is not greater than 2*s_2 , we use Two-samples independent t-test for difference in po
```

Assumptions: These observations form a random sample and are independent.

Hypothesis: $\mu_1 - \mu_2 = 0$ vs.$\mu_1 - \mu_2 \neq 0$

Test statistic:

$t = \frac{\bar{x_1} - \bar{x_2}}{s_p / \sqrt{1/n_1 + 1/n_2}}$ where $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

```
n1=6
n2=2
s_p=sqrt(((n1-1)*s_1^2+(n2-1)*s_2^2)/(n1+n2-2))
t=(x_1bar-x_2bar)/(s_p*sqrt((1/n1+1/n2)))
t
```

```
## [1] 1.687695
```

P-value:

```
2*(1-pt(abs(t),n1+n2-2))
```

```
## [1] 0.1424382
```

Conclusion: Because p-value is greater than $\alpha = 0.05$, we fail to reject the null hypothesis.

So, we do not have enough evidence to explain that there is any diffrence in the delivery time between Dominos and Papa Murphy's.

```
t.test(dominos,papa,var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  dominos and papa
## t = 1.6877, df = 6, p-value = 0.1424
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.949368 10.616035
## sample estimates:
## mean of x mean of y
##  22.33333  18.00000
```

5.This Large_sample z-test for a population proportion

Let p be the population proportion of firefighters who worked on the site for less than six months and had cardiovascular issues.So the sample mean $\hat{p} = \frac{3000}{9697}, p_0 = 0.3$

Assumption: These observations form a random sample and the sample size n is large enough to ensure that $np_0 \geq 15$ and $n(1-p_0) \geq 15$.

Hypothesis:$H_0 : p = 0.3$ vs.$H_a : p < 0.3$

Test statistic:

$z = \frac{\hat{p} - 0.3}{\sqrt{\frac{0.3(1-0.3)}{n}}}$

```
n=9697
p_hat=3000/9697
z=(p_hat-0.3)/sqrt(0.3*0.7/n)
z
```

```
## [1] 2.014353
```

P-value:

```r
pnorm(z)
```

```
## [1] 0.9780137
```

Conclusion:

Bacuase the p-value is larger that $\alpha = 0.1$, we fail to reject the null hypothesis.

```r
prop.test(x=3000,n=9697,p=0.3,alternative = "less",conf.level = 0.99)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  3000 out of 9697, null probability 0.3
## X-squared = 4.0131, df = 1, p-value = 0.9774
## alternative hypothesis: true p is less than 0.3
## 99 percent confidence interval:
##  0.0000000 0.3204498
## sample estimates:
##        p
## 0.309374
```

Bonus:

1.A study on whether warm light is more suitable for people to read. 20 randomle selected adults between the ages 20 and 30. in a fully enclosed ,no sunlight room, eachone s given an identical book. On the first day, using warm light, we record each person's reading time. And the next day, using cold light ,we record each person's reading time.

2. A study on whether warm light is more suitable for people to read.20 randomly selected adults between the ages 20 and 30 were randomly divided into 2 groups of ten people each. After that, the two groups were placed into two completely enclosed rooms without sunlight.Each person is given an identical book.One room is used warm light and the second room is used cool light.Later, we record each person's reading time.