Review for Chapter 4, 5, 6

Huamei Dong

03/24/16

1. Practice examples
2. Quick review of Z test or Test for population means
3. Quick review of Z test for population proportions
4. Quick review of χ^2 test
5. About paired sample data

Example 1 Using the data (smoke.txt) to test the whether there is a difference between the proportion of men who smoke and the proportion of women who smoke at significance level of 0.05. What is a 95% confidence interval for the difference between two proportions?

Answer:

```
>smoke<-read.table('smoke.txt', as.is=T, header=T, sep="\t")
>smoke1<-smoke[,c("gender", "smoke")]
>table(smoke1)
```

	smoke	
gender	No	Yes
Female	731	234
Male	539	187

(1) $H_0 : p_m = p_w$ $H_A : p_m \neq p_w$
(2) Check the condition:
(3) Calculate Z value:
(4) Calculate p-value:
(5) Conclusion:

Example 2 Using the data (smoke.txt) to test whether there is no difference in proportions of smoking between marital status.

Answer:

```
>smoke2<-smoke[, c("smoke", "maritalStatus")]
>table(smoke2)
```

	maritalStatus				
smoke	Divorced	Married	Separated	Single	Widowed
No	103	669	46	269	183
Yes	58	143	22	158	40

(1) State hypotheses:

(2) Check conditions:

(3) Calculate χ^2 value:

(4) Find p-value:

(5) Conclusion:

2. Z test or T test and confidence intervals for population means

(a) When population distribution is nearly normal, population mean is μ and population standard deviation is σ. No matter sample size big or small, we have

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = Z \sim N(0,1)$$ that is $$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$ or Z test.

(b) When population distribution is nearly normal, population mean is μ, usually in reality population standard deviation σ is unknown, we have to use sample standard deviation *S* to estimate σ. No matter sample size big or small, we have $$\frac{\bar{X}-\mu}{s/\sqrt{n}} = t_{n-1}$$

(c) When population distribution is nearly normal, population mean is μ, standard deviation σ is unknown. If sample size is large enough (usually larger than 30) , we think *S* will be close enough to σ. Then $$\frac{\bar{X}-\mu}{s/\sqrt{n}} \text{ is almost } Z$$

(d) When population distribution is not nearly normal, population mean is μ, standard deviation is σ. When sample size is large, as long as it is not too skewed, by central limited theorem,

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \text{ is almost } Z$$

(e) When population distribution is not nearly normal, population mean is μ, standard deviation σ is unknown. When sample size is large, we think *S* will be close enough to σ. Also because sample size is large, as long as it is not too skewed, by central limited theorem,

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} \text{ is almost } Z$$

3. Z test and confidence intervals for population proportions

$\hat{p}_1$: Point estimate for p_1, $\hat{p}_2$: Point estimate for p_2	
95% confidence interval for p_1 $\left(\hat{p}_1-1.96\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}},\ \hat{p}_1+1.96\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}}\right)$	
Hypothesis test for H_0: $p_1=0.5$ Using $SE=\sqrt{\frac{0.5(1-0.5)}{n_1}}$ to calculate z score $Z=\frac{\hat{p}_1-null\ \ value}{SE}=\frac{\hat{p}_1-0.5}{SE}$ and p-value	
$\hat{p}_1-\hat{p}_2$ Point estimate for p_1-p_2	
95% confidence interval for p_1-p_2 : Here $Z^*=1.96$ $\left(\hat{p}_1-\hat{p}_2-Z^*\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}},\ \hat{p}_1-\hat{p}_2+Z^*\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$	
Hypothesis test for H_0: p_1-p_2=0.2 using $SE\approx\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ and $Z=\frac{point\ \ estimate-null\ \ value}{SE}=\frac{\hat{p}_1-\hat{p}_2-0.2}{SE}$ to get p-value	
Hypothesis test for H_0: p_1-p_2=0 or p_1=p_2 using $SE=\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1}+\frac{\hat{p}(1-\hat{p})}{n_2}}$ Here $\hat{p}$ is pooled proportion estimate. $Z=\frac{point\ \ estimate-null\ \ vaule}{SE}=\frac{\hat{p}_1-\hat{p}_2-0}{SE}$	