**STAT 40001/MA59800 Statistical Computing Fall 2017**

**Lab-15**

1. Suppose we want to test if online instruction in a math course results in a different grade distribution than a traditional class. From each group, 50 students are randomly selected. At the end of the semester each student is assigned a grade of A,B,C,D, or F. We have the following data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Instruction | A | B | C | D | F |
| Online | 10 | 13 | 16 | 9 | 2 |
| Traditional | 4 | 12 | 15 | 9 | 10 |

Test whether the two instructional methods give about the same grades? Use 0.05 level of significance.  
Online = scan()

Traditional = scan()

data = rbind(Online,Traditional)  
dimnames(data) <- list(Instruction=c("Online","Traditional"),Category=c("A","B", "C","D","F"))

data  
 Category

Instruction A B C D F

Online 10 13 16 9 2

Traditional 4 12 15 9 10  
chisq.test(data)  
 Pearson's Chi-squared test

data: data

X-squared = 7.977, df = 4, p-value = 0.09242  
p-value > 0.05, do not have evidence to reject null hypothesis, the two instructional methods give about different grades.

1. Thirteen Honda Accord are chosen to study the gas mileage. Below is the mpg for these vehicles.

27 26 31 30 30 28 26 24 30 30 23 30 23  
Perform a test to see whether the data is coming from a normal distribution.  
data = scan()  
1: 27 26 31 30 30 28 26 24 30 30 23 30 23  
14:   
Read 13 items  
shapiro.test(data)

Shapiro-Wilk normality test

data: data

W = 0.86289, p-value = 0.04207

P-value < 0.05, reject null hypothesis of normality.

1. An article at Mobilize.org reported that the median credit-card balance for undergraduate students was $1770 for those who carried a balance from month to month. A professor at a community college believes that the median credit-card balance of students at his college is different than $1770. To test this hypothesis, he obtains a random sample of 20 students enrolled at the college who carry a credit-card balance from month to month and asks them to disclose their credit-card debt. The results of the survey are presented in Table 3 in dollars. Do the data indicate that the median credit-card debt of students at the professor's college differs from $1770 at the α=0.05 level of significance?

|  |  |  |  |
| --- | --- | --- | --- |
| 6000 | 870 | 1530 | 1660 |
| 1060 | 1790 | 1630 | 3180 |
| 2180 | 2370 | 1800 | 2170 |
| 1210 | 410 | 1720 | 1270 |
| 570 | 1050 | 2320 | 1120 |

install.packages("PASWR")

library(PASWR)

data = scan()  
1: 6000 870 1530 1660

5: 1060 1790 1630 3180

9: 2180 2370 1800 2170

13: 1210 410 1720 1270

17: 570 1050 2320 1120

21:

Read 20 items

> SIGN.test(data,md=1770)

One-sample Sign-Test

data: data

s = 8, p-value = 0.5034

alternative hypothesis: true median is not equal to 1770

95 percent confidence interval:

1130.482 2126.906

sample estimates:

median of x

1645

Conf.Level L.E.pt U.E.pt

Lower Achieved CI 0.8847 1210.000 1800.000

Interpolated CI 0.9500 1130.482 2126.906

Upper Achieved CI 0.9586 1120.000 2170.000  
p-value > 0.05, do not have enough evidence to reject null hypothesis. So the credit-card debt of students at the professor's college differs from $1770

1. The exec.pay data in UsingR library contains data on salries of CEOs at 199 top companies in the United States. The amount are in $10,000s . Do a sign test to determine whether the median pay is more than $220,000.  
   library(UsingR)  
   SIGN.test(exec.pay,md = 22, alt='greater')

One-sample Sign-Test

data: exec.pay

s = 113, p-value = 0.008506

alternative hypothesis: true median is greater than 22

95 percent confidence interval:

23 Inf

sample estimates:

median of x

27

Conf.Level L.E.pt U.E.pt

Lower Achieved CI 0.9407 23 Inf

Interpolated CI 0.9500 23 Inf

Upper Achieved CI 0.9557 23 Inf  
p-value < 0.05, reject the null hypothesis, the median pay is more than $220,000

1. The data frame *galton* data set in UsingR library has data collected by Francis Galton in 1885. Each data point contains a child’s height and an average of his or her parent’s height. Do an appropriate test if there is a difference in the mean height.  
   wilcox.test(galton$child,galton$parent)  
    Wilcoxon rank sum test with continuity correction

data: galton$child and galton$parent

W = 388360, p-value = 0.0002425

alternative hypothesis: true location shift is not equal to 0

p-value < 0.05, reject null hypothesis, there is a difference in the mean height