Statistics 140 Winter 17

Exam #1 Part 2

Sarah Ruckman

Last 4 Digits of SID: 7194

1. A premium dog food company was interested in marketing a new puppy food for puppies that are under nourished and not gaining weight at an appropriate rate. The company claims a median weight gain of 3 ounces on this new formula within a prescribed period of time. AmyG and Minh have decided to test the claim using 12-week-old Labrador Retreiver puppies. They select eight healthy Lab puppies at random. After the appropriate amount of time, the following data was recorded: Preform the appropriate test.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Weight Gain | 3.78 | 3.35 | 2.45 | 4.27 | 4.87 | 3.14 | 3.98 | 2.47 |

**H­0: The median weight gain is 3 ounces.**

**Ha: The median weight gain is not 3 ounces.**

R Code for Quantile Test:

> #there are 2 obs. that are below 3 and none that are 3. Sample size stays 8

> binom.test(x=2,n=8,alternative="two.sided")

Exact binomial test

data: 2 and 8

number of successes = 2, number of trials = 8, p-value = 0.2891

alternative hypothesis: true probability of success is not equal to 0.5

95 percent confidence interval:

0.03185403 0.65085579

sample estimates:

probability of success

0.25

**TS: n=8, number of observations below 3 = 2, p-value = 0.2891**

**Since the p-value of 0.2891 is greater than α of 0.05, We do not reject H0**

**There is insufficient evidence to conclude that the median weight gain is significantly different from 3 ounces.**

1. A large number of probationary employees are working in a bioengineering company under the direction of four managers. At the end of the probationary period, each manager was asked to evaluate and rate each of the probies, either 0 = not acceptable or 1 = acceptable as permanent employees. The question is whether there is a significant difference in the ratings among the four managers. Zi-yue, Andy, and Pedro conduct a study addressing the question. They obtain a random sample of 8 probies and their ratings from each of the managers.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Probationary Employee | Manager 1 | Manager 2 | Manager 3 | Manager 4 | Ri Totals |
| PE 1 | 1 | 1 | 0 | 1 | 3 |
| PE 2 | 1 | 1 | 0 | 1 | 3 |
| PE 3 | 1 | 1 | 0 | 1 | 3 |
| PE 4 | 1 | 1 | 0 | 1 | 3 |
| PE 5 | 1 | 1 | 1 | 1 | 4 |
| PE 6 | 1 | 1 | 0 | 1 | 3 |
| PE 7 | 1 | 1 | 0 | 1 | 3 |
| PE 8 | 1 | 0 | 1 | 0 | 2 |
| Cj Totals | 8 | 7 | 2 | 7 | 24 |

**H0: The probationary employees are equally rated by the four managers.**

**Ha: The probationary employees are not equally rated by the four managers.**

Cochran’s Test was performed.

Blocks: Probationary employees

N = 24

c = 4

r = 8

**TS = X2 with c-1 df = 4-1 = 3 df**

**= = 12**

**RR = Reject H0 if T > X20.05, 3 = 7.815**

**Since T = 12 is greater than X20.05, 3 = 7.815, we reject H0**

**There is sufficient evidence to indicate at least one manager is not equally rating the probationary employees.**

1. Christin, Kevin and SarahK believe that at least 80% of students do not like the new Banner registration system. They obtain a random sample of 25 students and find that 19 do not like the new Banner. Preform the appropriate test.

**P = the proportion of students that do not like the new Banner registration system**

**H0: p ≤ 0.80**

**Ha: p > 0.80**

R Code for Binomial Test:

> prop.test(x=19,n=25,p=0.80,alternative="greater")

1-sample proportions test with continuity correction

data: 19 out of 25, null probability 0.8

X-squared = 0.0625, df = 1, p-value = 0.5987

alternative hypothesis: true p is greater than 0.8

95 percent confidence interval:

0.5775027 1.0000000

sample estimates:

p

0.76

**TS: n=25, T=19, α = 0.05**

**X2 = 0.0625 with p-value = 0.5987**

**Since the p-value of 0.5987 is greater than α = 0.05, we do not reject H0**

**There is insufficient evidence to indicate that the proportion of students that do not like the new banner system is significantly higher than 80%.**

1. Brandon and Linda were interested in determining whether there is a relationship between dog gender and training method. They went to a professional dog training facility and obtained the following data. Preform the appropriate test.

|  |  |  |  |
| --- | --- | --- | --- |
| Treatment | Clicker | Snack Reward | Total |
| Male | 3 | 7 | 10 |
| Female | 6 | 5 | 11 |
| Total | 9 | 12 | 21 |

**H0: Dog gender and training method are independent (There is no relationship)**

**Ha: Dog gender and training method are not independent (There is a relationship)**

R Code for Fisher’s Exact Test:

> a<-matrix(c(3,7,6,5),2,2,byrow=TRUE)

> a

[,1] [,2]

[1,] 3 7

[2,] 6 5

> fisher.test(a)

Fisher's Exact Test for Count Data

data: a

p-value = 0.387

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

0.0395915 2.8869759

sample estimates:

odds ratio

0.3757631

**TS: p-value = 0.387**

**Since the p-value = 0.387 is greater than α = 0.05, we do not reject H0**

**There is sufficient evidence to indicate that dog gender and training method are independent. (There is no relationship)**

1. Affordable health insurance is a major concern in the country. RebeccaL and Cindy were interested in examining whether there is a relationship between gender and whether a person is insured. They obtain a random sample of individuals and recorded the following data. Preform the appropriate test.

|  |  |  |
| --- | --- | --- |
| Coverage | Male | Female |
| Insured | 86 | 93 |
| Uninsured | 16 | 15 |

**H0: Gender and whether a person is insured are independent.**

**(There is no relationship)**

**Ha: Gender and whether a person is insured are not independent.**

**(There is a relationship)**

R Code for Chi Squared Test for Independence:

> coverage<-read.table("C:\\Users\\Sarah\\Downloads\\insure1a.dat", header=TRUE)

> coverage

male female

1 86 93

2 16 15

> chisq.test(coverage)

Pearson's Chi-squared test with Yates' continuity correction

data: coverage

X-squared = 0.029713, df = 1, p-value = 0.8631

**TS: X2 = 0.029713 with p-value = 0.8631**

**Since the p-value of 0.8631 is greater than α = 0.05, we do not reject H0**

**There is sufficient evidence to indicate that gender and whether a person is insured are independent. (There is no relationship)**

1. A manufacturing process produces 10,000 64Gb flash drives per day. To ensure that the process is functioning in control, each day, for 20 days, Derick and Daniel select a random sample of 200 flash drives and test them. They record the following data:

Defective

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 14 | 11 | 25 | 8 | 15 | 6 | 16 | 4 | 17 | 13 | 9 | 22 | 17 | 18 | 22 | 12 | 13 | 17 | 21 | 16 |

They are interested in testing whether the 80th percentile is 20.

**H0: The 80th quantile is 20.**

**Ha: The 80th quantile is not 20.**

**TS: Using the quantile test:**

**T1: # of observations ≤ 20 = 15**

**T2: # of observations < 20 = 15**

**T = T1 = T2 = 15 so α1 = α2 = α/2 = 0.05/2 =0.025**

**RR:**

**T ≤ t1 where P(Y ≤ t1) = α1 = 0.025**

**P(Y ≤ 11) = 0.010 P(Y ≤ 12) = 0.032**

**T > t2 where P(Y ≤ t2) = 1 - α2 = 1 – 0.025 = 0.975**

**P(Y ≤ 18) = 0.931 P(Y ≤ 19) = 0.988**

**Reject H0 if T ≤ 12 or T > 19**

**Since T = 15 is not ≤ 12 or > 19, Do not reject H0**

**There is sufficient evidence to indicate that the 80th quantile is 20.**

1. Jirashuddhi, Siddarth, and Nathan were interested in determining whether there is a significant difference in the number of videos streamed on their phones between males and females. They selected a random sample of males and females (who stream videos on their phones) and recorded the following data (measured in hours, in a month period of time). Perform the appropriate test.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gender | Hours of Videos Streamed | | | | | | | | |  |
| Male | 1.8 | 3.7 | 2.8 | 5.6 | 4.6 | 0.3 | 5.3 | 4.3 | 6.6 | 7.9 |
| Female | 1.8 | 3.6 | 1.8 | 5.4 | 2.6 | 0.3 | 3.5 | 2.3 | 6.6 | 3.2 |

**H0: We cannot conclude a difference in hours of video streamed between gender.**

**Ha: We can conclude a difference in hours of video streamed between gender.**

R Code using a Mann and Whitney Test:

> data<-read.table("C:\\Users\\Sarah\\Downloads\\video1a.dat", header=TRUE)

> data

Male Female

1 1.8 1.8

2 3.7 3.6

3 2.8 1.8

4 5.6 5.4

5 4.6 2.6

6 0.3 0.3

7 5.3 3.5

8 4.3 2.3

9 6.6 6.6

10 7.9 3.2

> attach(data)

> names(data)

[1] "Male" "Female"

> Male

[1] 1.8 3.7 2.8 5.6 4.6 0.3 5.3 4.3 6.6 7.9

> Female

[1] 1.8 3.6 1.8 5.4 2.6 0.3 3.5 2.3 6.6 3.2

> wilcox.test(Male,Female)

Wilcoxon rank sum test with continuity correction

data: Male and Female

W = 68, p-value = 0.1849

alternative hypothesis: true location shift is not equal to 0

Warning message:

In wilcox.test.default(Male, Female) :

cannot compute exact p-value with ties

**TS: p-value = 0.1849**

**Since the p-value is 0.1849 is greater than α = 0.05, we do not reject H0**

**There is insufficient evidence to indicate a difference in the hours of videos streamed between genders.**

1. A common observation in ecology is that species diversity decreases as you get further from the equator. Sarah R and Patrick were interested in determining whether there is a significant correlation between the location (latitude) and the number of a particular bird species present. They select a random sample of locations and recorded the following information. Perform the appropriate test.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Location | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Latitude | 39.2 | 38.8 | 39.4 | 39.0 | 38.6 | 38.5 | 39.7 | 38.0 | 38.9 |
| Bird #’s | 128 | 137 | 108 | 109 | 135 | 94 | 113 | 118 | 96 |
| Location | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |  |
| Latitude | 39.5 | 39.1 | 38.2 | 38.3 | 38.4 | 37.2 | 37.9 | 37.6 |  |
| Bird #’s | 98 | 121 | 152 | 108 | 118 | 157 | 125 | 114 |  |

**H0: There is no significant correlation between location(latitude) and the number of a particular bird species present.**

**Ha: There is a significant correlation between location(latitude) and the number of a particular bird species present.**

R Code for Cox and Stuart Test:

> bird<-read.table("C:\\Users\\Sarah\\Downloads\\birds1a.dat", header = TRUE)

> bird #print as check

> attach(bird)

> names(bird)

[1] "Loc" "Lat" "BirdNumb"

> sorted\_bird = bird[order(Lat),]

> sorted\_bird #print as check

> attach(sorted\_bird)

The following objects are masked from bird:

BirdNumb, Lat, Loc

> names(sorted\_bird)

[1] "Loc" "Lat" "BirdNumb"

> library(randtests)

> cox.stuart.test(BirdNumb)

Cox Stuart test

data: BirdNumb

statistic = 2, n = 7, p-value = 0.4531

alternative hypothesis: non randomness

**TS: p-value = 0.4531**

**Since the p-value of 0.4531 is greater than α of 0.05, we do not reject H0**

**There is insufficient evidence to conclude a significant coerrlation between location(latitude) and bird numbers.**

1. Disc dog is the more generic name for what is commonly called Frisbee dog. In disc dog competitions, dogs and their human flying disc throwers compete in events such as distance catching and somewhat choreographed freestyle catching. Scores differ among judges, even when the same performance is being evaluated. Mirella (Judge1) and Karina (Judge2) have been selected to be judges in a local disc dog competition. The scores, repeated for 10 competitors are as follows. Preform the appropriate test to determine whether there is a significant difference in scoring between the two judges.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dog | Abby | Dexter | Kali | Cody | Dusty | Olive | Lakota | Luke | Mercedes | Shadow |
| Judge1 | 83.00 | 77.00 | 79.00 | 90.00 | 94.00 | 81.00 | 95.00 | 93.00 | 88.00 | 96.00 |
| Judge2 | 84.00 | 73.00 | 78.00 | 95.00 | 96.00 | 78.00 | 90.00 | 87.00 | 84.00 | 95.00 |
| Diff | - | + | + | - | - | + | + | + | + | + |

**H0: There is no significant difference in scoring between the two judges.**

**Ha: There is a significant difference in scoring between the two judges.**

R Code for Sign Test:

> judge1<-c(83,77,79,90,94,81,95,93,88,96)

> judge2<-c(84,73,78,95,96,78,90,87,84,95)

> diff=judge1-judge2

> diff

[1] -1 4 1 -5 -2 3 5 6 4 1

> library(BSDA)

Loading required package: e1071

Loading required package: lattice

Attaching package: ‘BSDA’

The following object is masked from ‘package:datasets’:

Orange

> SIGN.test(diff,md=0)

One-sample Sign-Test

data: diff

s = 7, p-value = 0.3437

alternative hypothesis: true median is not equal to 0

95 percent confidence interval:

-1.675556 4.675556

sample estimates:

median of x

2

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

Lower Achieved CI 0.8906 -1.0000 4.0000

Interpolated CI 0.9500 -1.6756 4.6756

Upper Achieved CI 0.9785 -2.0000 5.0000

**TS: p-value = 0.3437**

**Since the p-value of 0.3437 is greater than α = 0.05, we do not reject H0**

**There is insufficient evidence to indicate a significant difference between the two judges.**

1. Nicholas, Amy J, and Yongjae were interested in whether there is a relationship between gender and admission to a university. They obtain independent random samples from each of three departments at a local university and recorded the following data.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Department A | | | | Department B | | | | Department C | | | |
| Gender | Admit | Deny | Total | Gender | Admit | Deny | Total | Gender | Admit | Deny | Total |
| Male | 65 | 122 | 187 | Male | 18 | 8 | 26 | Male | 50 | 98 | 148 |
| Female | 69 | 140 | 209 | Female | 118 | 69 | 187 | Female | 30 | 51 | 81 |
| Total | 134 | 262 | 396 | Total | 136 | 77 | 213 | Total | 80 | 149 | 229 |

**H0: Gender and admission to a university are independent. (No relationship)**

**Ha: Gender and admission to a university are not independent. (Relationship)**

SAS Code for Mantel Haenszel Test:

options ls = 78 ps =55 nocenter nodate;

data mhtest;

infile'C:\Users\Sarah\Downloads\university1a.dat' firstobs =2;

do department = 1 to 3;

if department = 1 then department1 = 'DepartmentA';

else if department = 2 then department1 = 'DepartmentB';

else department1 = 'DepartmentC';

do admit = 1 to 2;

if admit = 1 then admit1 = 'Admit';

else admit1 = 'Deny ';

do gender = 1 to 2;

if gender =1 then gender1 = 'Female';

else gender1 = 'Male ';

input wt @@;

output;

end;

end;

end;

proc print;

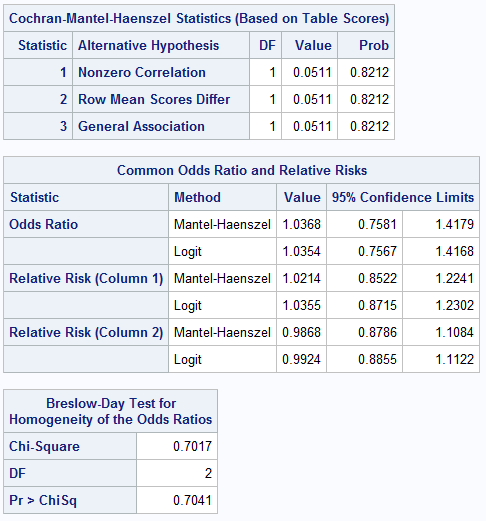
proc freq;

weight wt;

tables department1\*admit1\*gender1/norow nocol cmh;

run;

quit;



**TS: p-value = 0.8212**

**Since the p-value of 0.8212 is greater than α = 0.05, we do not reject H0**

**There is insufficient evidence to indicate a relationship between gender and admission to a university.**

1. Tina, Maher, and Amber were interested in determining whether a recent presentation, given to new pet owners, regarding the benefits of pet microchips altered people’s mind about having their new pet microchipped. They obtained information and summarized it as follows. Preform the appropriate test.

|  |  |  |
| --- | --- | --- |
|  | After the Presentation | |
| Before Presentation | Plan to Microchip | Do Not Plan to Microchip |
| Plan to Microchip | 27 | 3 |
| Do Not Plan to Microchip | 30 | 12 |

**H0: The presentation did not change plans to microchip pets.**

**Ha: The presentation did change plans to microchip pets.**

R Code for McNemar’s Test:

> m<-matrix(c(27,3,30,12),2,2,byrow=TRUE)

> m

[,1] [,2]

[1,] 27 3

[2,] 30 12

> mcnemar.test(m)

McNemar's Chi-squared test with continuity correction

data: m

McNemar's chi-squared = 20.485, df = 1, p-value = 6.011e-06

**TS: X2 = 20.485 with p-value = 6.011 x 10-6**

**Since the p-value = 6.011 x 10-6 is less than α = 0.05, we reject H0**

**There is sufficient evidence to indicate that the presentation changed pet owners plans to microchip pets.**

1. Xin-Ru, Shahan, Zixiang were interested in determining whether there is a relationship between gender and game type college students play most frequently. They obtain a random sample of college students and record the following data.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Game Type | | |
| Gender | Video | Computer | Online |
| Female | 100 | 185 | 85 |
| Male | 305 | 110 | 75 |

**H0: Gender is independent of game type played most frequently.**

**Ha: Gender is not independent of game type played most frequently. (Relationship)**

R Code for Chi Sq. Test:

> game<-matrix(c(100,185,85,305,110,75),2,3,byrow=TRUE)

> game

[,1] [,2] [,3]

[1,] 100 185 85

[2,] 305 110 75

> chisq.test(game)

Pearson's Chi-squared test

data: game

X-squared = 108.83, df = 2, p-value < 2.2e-16

**TS: X2 = 108.83 with p-value = 2.2 x 10-16**

**Since the p-value = 2.2 x 10-16 is less than α = 0.05, we reject H0**

**There is a relationship between gender and game type played most frequently.**

1. Alfred and Bianca corporate multicultural affair officers, would like to access whether there are differences in median diversity awareness test scores for new employees training at three corporate locations (branch). They randomly select eight new employees from each of the three locations (branches) and record the following diversity test scores.

|  |  |  |
| --- | --- | --- |
| Branch A | Branch B | Branch C |
| 85 | 69 | 85 |
| 90 | 79 | 95 |
| 60 | 84 | 90 |
| 90 | 49 | 75 |
| 85 | 49 | 85 |
| 86 | 69 | 75 |
| 89 | 59 | 65 |
| 75 | 64 | 80 |

**H0: All the branches have the same median diversity awareness test scores.**

**Ha: At least one branch has a significantly different median diversity awareness test score.**

SAS Code for Median Test:

options nocenter nodate nonumber ls =**78** ps =**55** formdlim='#';

ods graphics off;

**data** diversity;

infile'C:\Users\Sarah\Downloads\diversity1a.dat' firstobs =**2**;

do rows = **1** to **8**;

do branch = **1** to **3**;

if branch = **1** then level = 'BranchA';

else if branch = **2** then level = 'BranchB';

else level = 'BranchC';

input scores @@;

output;

end;

end;

**proc** **print**;

**proc** **sort**;

by level;

**proc** **print**;

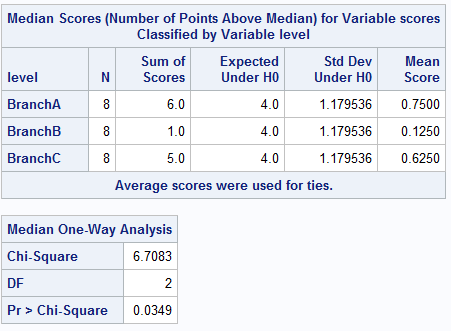
**proc** **npar1way** median;

class level;

var scores;

**run**;

**quit**;



**TS: X2 = 6.7083 with p-value = 0.0349**

**Since the p-value of 0.0349 is less than α = 0.05, we reject H0**

**There is sufficient evidence to suggest that at least one branch has a significantly different median diversity awareness test score.**