ISQA 8160 Exam III

Brian Detweiler

Tuesday, August 9, 2016

```
## Warning: package 'nortest' was built under R version 3.2.3
```

Chapter 4.1

4.1 Problem 7.) Exposure to nitrous oxide, an anesthetic, is suspected as a cause for miscarriages among pregnant nurses and dental assistants who sustained prolonged periods of exposure in their occupation. Data are collected from three different groups of pregnant females and it is recorded how many have miscarriages and how many full-term deliveries.

	Dental Miscarriage	Assistants Full Term		O.R. Miscarriage	Nurses Full Term		Out-Patient Miscarriage	Nurses Full Term	
Exposed	8	32	40	3	18	21	0	7	7
Not Exposed	26	210	236	3	21	24	10	75	85
Totals	34	242	276	6	39	45	10	82	92

(a) Use T_4, with a correction for continuity when finding the p-value, to investigate this theory.

From the data, we have

$$H_0 : P_{1i} \leq P_{2_i}$$
$$H_a : P_{1i} \geq P_{2_i} \text{ for all } i \text{ and } P_{1i} > P_{2_i} \text{ for some } i$$
$$\alpha = 0.05$$
$$z_{1-\alpha} = z_{0.95} = 1.645$$
$$c = \{T_4 : T_4 > z_{1-\alpha}\}$$
$$x_1 = 8$$
$$x_2 = 3$$
$$x_3 = 0$$
$$r_1 = 40$$
$$r_2 = 21$$
$$r_3 = 7$$
$$c_1 = 34$$
$$c_2 = 6$$
$$c_3 = 10$$
$$N_1 = 276$$
$$N_2 = 45$$
$$N_3 = 92$$

Using $T4$ and substituting in our values, we get

$$T_4 = \frac{\sum x_i - \sum \frac{r_i c_i}{N_i}}{\sqrt{\sum \frac{r_i c_i (N_i - r_i)(N_i - c_i)}{N_i^2 (N_i - 1)}}}$$

$$= \frac{(8 + 3 + 0) - \left(\frac{(40)(34)}{276} + \frac{(21)(6)}{45} + \frac{(7)(10)}{92}\right)}{\sqrt{\frac{(40)(34)(276-40)(276-34)}{276^2(276-1)} + \frac{(21)(6)(45-21)(45-6)}{45^2(45-1)} + \frac{(7)(10)(92-7)(92-10)}{92^2(92-1)}}}$$

$$= \frac{11 - \left(\frac{340}{69} + \frac{14}{5} + \frac{35}{46}\right)}{\sqrt{\frac{88264}{23805} + \frac{364}{275} + \frac{17425}{6348}}}$$

$$= \frac{11 - 8.4884057971}{\frac{1}{7590} \cdot \sqrt{447982931}}$$

$$= \frac{2.5115942029}{2.78861755829}$$

$$= 0.900659251547 \ngtr 1.645$$

Verifying in R,

```
x_1 <- 8
x_2 <- 3
x_3 <- 0
r_1 <- 40
r_2 <- 21
r_3 <- 7
```

2

```r
c_1 <- 34
c_2 <- 6
c_3 <- 10
N_1 <- 276
N_2 <- 45
N_3 <- 92

xi <- c(x_1, x_2, x_3)
ri <- c(r_1, r_2, r_3)
ci <- c(c_1, c_2, c_3)
Ni <- c(N_1, N_2, N_3)

T4 <- function(xi, ri, ci, Ni) {
  numerator <- sum(xi) - sum(mapply(function(r, c, N) { (r*c)/N }, ri, ci, Ni))
  denominator <- sqrt(sum(mapply(function(r, c, N) {
      (r * c * (N - r) * (N - c)) / (N^2 * (N - 1))
  }, ri, ci, Ni)))

  numerator/denominator
}

T4.pval <- function(xi, ri, ci, Ni) {
  numerator <- sum(xi) - sum(mapply(function(r, c, N) { (r*c)/N }, ri, ci, Ni))
  denominator <- sqrt(sum(mapply(function(r, c, N) {
      (r * c * (N - r) * (N - c)) / (N^2 * (N - 1))
  }, ri, ci, Ni)))

  (numerator - 0.5)/denominator
}

T4.val <- exact(T4(xi, ri, ci, Ni))
```

```
## [1] 1.055246885439115
```

```r
print(paste('T4 = ', T4.val))
```

```
## [1] "T4 =  1.05524688543911"
```

```r
z.val <- exact(T4.pval(xi, ri, ci, Ni))
```

```
## [1] 0.84517176975735209
```

```r
p.val <- exact(1 - pnorm(q=z.val))
```

```
## [1] 0.19900747083791315
```

```r
print(paste('p-value = ', p.val))
```

```
## [1] "p-value =  0.199007470837913"
```

With a p-value of 0.207, We cannot reject the null hypothesis. Anesthesia appears to have no affect on pregnancy.

(b) Use T_5 to test the hypothesis of no miscarriage effect due to exposure to nitrous oxide. Compare the p-value with part (a).

```
T5 <- function(xi, ri, ci, Ni) {
  numerator <- sum(xi) - sum(mapply(function(r, c, N) { (r*c)/N }, ri, ci, Ni))
  denominator <- sqrt(sum(mapply(function(r, c, N) {
      (r * c * (N - r) * (N - c)) / (N^3)
  }, ri, ci, Ni)))

  numerator / denominator
}

T5.pval <- function(xi, ri, ci, Ni) {
  numerator <- sum(xi) - sum(mapply(function(r, c, N) { (r*c)/N }, ri, ci, Ni))
  denominator <- sqrt(sum(mapply(function(r, c, N) {
      (r * c * (N - r) * (N - c)) / (N^3)
  }, ri, ci, Ni)))

  (numerator) / denominator
}

exact(T5(xi, ri, ci, Ni))
```

```
## [1] 1.059909759293027
```

```
z.val <- exact(T5.pval(xi, ri, ci, Ni))
```

```
## [1] 1.059909759293027
```

```
exact(1 - pnorm(q=z.val))
```

```
## [1] 0.14459282766611858
```

Notably, the p-value is 0.14459282766611858, which is smaller than the p-value of T_4, but still not significant.

(c) Which analysis, using T_4 or T_5, seems more appropriate in this case?

T_5 is more appropriate, because both the samples and the outcomes are random.

■

Chapter 4.2

4.2 Problem 4.) Three professors are teaching large classes in introductory statistics. At the end of the semester, they compare grades to see if there are significant differences in their grading policies.

				Grade			
Professor	**A**	**B**	**C**	**D**	**F**	**WP**	**WF**
Smith	12	45	49	6	13	18	2
Jones	10	32	43	18	4	12	6
White	15	19	32	20	6	9	7

Are these differences significant? Which test are you using? Are the grades assigned by Professors Jones and White significantly different? How would the results be interpreted?

We will use the *Chi-squared Test for Differences in Probablilities, r x c*, with the following hypothesis:

$$H_0 : p_{ij} = p_{2j} = p_{3j}$$
$$H_a : \text{ Not all are equal}$$

```r
Grades <- matrix(data = c(12, 10, 15,
                          45, 32, 19,
                          49, 43, 32,
                           6, 18, 20,
                          13,  4,  6,
                          18, 12,  9,
                           2,  6,  7),
                 nrow=3,
                 ncol=7)
rownames(Grades) <- c('Smith', 'Jones', 'White')
colnames(Grades) <- c('A', 'B', 'C', 'D', 'F', 'WP', 'WF')
Grades
```

```
##        A  B  C  D  F WP WF
## Smith 12 45 49  6 13 18  2
## Jones 10 32 43 18  4 12  6
## White 15 19 32 20  6  9  7
```

We obtain our test statistic with

$$T = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \text{ where } E_{ij} = \frac{n_i C_j}{N}$$

```r
rxcT <- function(Matr) {
  N <- sum(Matr)
  T_ <- 0
  for (i in c(1:nrow(Matr))) {

    ni <- sum(Matr[i,])

    for (j in c(1:ncol(Matr))) {
      Ci <- sum(Matr[,j])
      Eij <- (ni * Ci) / N

      Oij <- Matr[i, j]
      T_ <- T_ + (((Oij - Eij)^2) / Eij)
    }
  }

  T_
}

rxcT(Grades)
```

```
## [1] 28.91509
```

The Chi-squared statistic is given by

```r
exact(qchisq(p = 0.95, df = ((nrow(Grades) - 1) * (ncol(Grades) - 1))))
```

```
## [1] 21.02606981748306
```

And since $T = 28.91509 > 21.026$, we reject the null hypothesis.

As for the grades assigned by Jones and White, we have

```r
# Remove Agent Smith from the Matrix
Grades <- Grades[-1,]

rxcT(Grades)
```

```
## [1] 5.727965
```

```r
# Compute Chi-Square statistic
exact(qchisq(p = 0.95, df = ((nrow(Grades) - 1) * (ncol(Grades) - 1))))
```

```
## [1] 12.591587243744
```

From this result, we can see that there is not a significant difference between Mr. White and Mr. Jones.

∎

Chapter 4.3

4.3, Problem 1): Test the hypothesis that the following samples were obtained from populations having the same medians.

$$\text{Sample 1}: 35, 42, 42, 30, 15, 31, 29, 29, 17, 21$$
$$\text{Sample 2}: 34, 38, 26, 17, 42, 28, 35, 33, 16, 40$$
$$\text{Sample 3}: 17, 29, 30, 36, 41, 30, 31, 23, 38, 30$$
$$\text{Sample 4}: 39, 34, 22, 27, 42, 33, 24, 36, 29, 25$$

Perform pairwise comparison for Population Medians.

$$H_0 : \text{ All populations have the same median}$$
$$H_a : \text{ At least two populations have different medians}$$
$$T = \frac{N^2}{ab} \sum_{i=1}^{c} \frac{(O_{1i} - \frac{n_i a}{N})^2}{n_i}$$
$$\alpha = 0.05$$
$$c = 10$$
$$\chi^2_{0.95,(c-1)} = \chi^2_{0.95,9} = 16.91898$$
$$\text{pairwise comparisons } = \binom{4}{2} = 6$$

```
Median.Data <- matrix(data = c(35, 34, 17, 39,
                               42, 38, 29, 34,
                               42, 26, 30, 22,
                               30, 17, 36, 27,
                               15, 42, 41, 42,
                               31, 28, 30, 33,
                               29, 35, 31, 24,
                               29, 33, 23, 36,
                               17, 16, 38, 29,
                               21, 40, 30, 25),
                  nrow=4,
                  ncol=10)
rownames(Median.Data) <- c('Sample 1', 'Sample 2', 'Sample 3', 'Sample 4')
Median.Data
```

```
##          [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## Sample 1   35   42   42   30   15   31   29   29   17    21
## Sample 2   34   38   26   17   42   28   35   33   16    40
## Sample 3   17   29   30   36   41   30   31   23   38    30
## Sample 4   39   34   22   27   42   33   24   36   29    25
```

```
med <- median(Median.Data)
med
```

```
## [1] 30
```

```
# Get the count in our samples that are greater than or less-than-or-equal-to our median
s1_gtr <- length(unlist(lapply(Median.Data[1, ], function(x) { if (x > med) x})))
s1_lte <- length(unlist(lapply(Median.Data[1, ], function(x) { if (x <= med) x})))
s2_gtr <- length(unlist(lapply(Median.Data[2, ], function(x) { if (x > med) x})))
s2_lte <- length(unlist(lapply(Median.Data[2, ], function(x) { if (x <= med) x})))
s3_gtr <- length(unlist(lapply(Median.Data[3, ], function(x) { if (x > med) x})))
s3_lte <- length(unlist(lapply(Median.Data[3, ], function(x) { if (x <= med) x})))
s4_gtr <- length(unlist(lapply(Median.Data[4, ], function(x) { if (x > med) x})))
s4_lte <- length(unlist(lapply(Median.Data[4, ], function(x) { if (x <= med) x})))

Median.Test.Data <- t(matrix(data = c(s1_gtr, s2_gtr, s3_gtr, s4_gtr,
                                      s1_lte, s2_lte, s3_lte, s4_lte),
                            nrow=4,
                            ncol=2))

colnames(Median.Test.Data) <- c('Sample 1', 'Sample 2', 'Sample 3', 'Sample 4')
rownames(Median.Test.Data) <- c('> Median', '<= Median')

Median.Test.Data
```

```
##           Sample 1 Sample 2 Sample 3 Sample 4
## > Median         4        6        4        5
## <= Median        6        4        6        5
```

```
medianT <- function(Matr) {
  N <- sum(Matr)

  # sum first row
  a <- sum(Matr[1,])

  # sum second row
  b <- sum(Matr[2,])


  T_ <- 0
  for (i in c(1:ncol(Matr))) {
    ni <- sum(Matr[, i])

    O1i <- Matr[[1, i]]


    T_ <- T_ + ( ((O1i - ((ni * a) / N))^2) / ni)

  }

  T_ <- (N^2 / (a * b)) * T_
  T_
}

exact(qchisq(p = 0.95, df = (ncol(Median.Test.Data) - 1)))
```

```
## [1] 7.814727903251179
```

```
T_ <- medianT(Median.Test.Data)
p.val <- pchisq(T_, df=1, lower = FALSE)

print(paste('T = ', T_))
```

[1] "T = 1.10275689223058"

```
print(paste('p-value = ', p.val))
```

[1] "p-value = 0.293661876843241"

Clearly, $T \not> 1.10275689223058$, so we cannot reject the null hypothesis. All populations appear to have the same median.

■

Chapter 4.4

4.4, Problem 1): One hundred married couples were interviewed, and the husband and wife were asked separately for their first choice for the next U.S. president, with the following results.

		Wife's Choice		
		A	B	Other
	A	12	22	6
Husband's Choice	B	25	21	4
	Other	3	7	0

Compute the following.

(a) T

```
 Presidents <- matrix(data = c(12, 25, 3,
                              22, 21, 7,
                               6, 4, 0),
              nrow=3,
              ncol=3)
rownames(Presidents) <- c('A', 'B', 'Other')
colnames(Presidents) <- c('A', 'B', 'Other')
Presidents
```

```
##          A  B Other
## A       12 22     6
## B       25 21     4
## Other    3  7     0
```

```
exact(rxcT(Presidents))
```

```
## [1] 6.34
```

$$T = 6.34$$

(b) Cramer's coefficient

```
R1 <- function(Matr) {
  T_ <- rxcT(Matr)
  N <- sum(Matr)
  q <- min(dim(Matr))

  R1 <- T_ / (N * (q - 1))
  R1
```

```
}

cramers <- function(Matr) {
  sqrt(R1(Matr))
}

cramers(Presidents)
```

```
## [1] 0.1780449
```

(c) R_1

```
R1(Presidents)
```

```
## [1] 0.0317
```

(d) R_2

```
R2 <- function(Matr) {
  T_ <- rxcT(Matr)
  N <- sum(Matr)
  sqrt(T_ / (N + T_))
}

R2(Presidents)
```

```
## [1] 0.2441722
```

(e) R_3

```
R3 <- function(Matr) {
  T_ <- rxcT(Matr)
  N <- sum(Matr)
  T_/N
}

R3(Presidents)
```

```
## [1] 0.0634
```

(f) R_4

```
R4 <- function(Matr) {
  T_ <- rxcT(Matr)
  N <- sum(Matr)
  r <- dim(Matr)[1]
  c <- dim(Matr)[2]
  sqrt(T_ / (N * sqrt( (r - 1) * (c - 1) ) ))
}

R4(Presidents)
```

```
## [1] 0.1780449
```

∎

Chapter 4.5

4.5 Problem 1.) Test the following data to see if they could have come from a popuilation whose values are uniformly distributed between 0.0000 and 0.9999.

0.4755	0.2186	0.5112	0.3826	0.5758
0.4274	0.4295	0.5233	0.7500	0.5484
0.6454	0.8620	0.5482	0.5926	0.5440
0.2484	0.5758	0.9145	0.6687	0.3007
0.6521	0.5456	0.5101	0.3607	0.3943
0.3979	0.4438	0.6328	0.9056	0.8283
0.4352	0.5381	0.5646	0.4102	0.5689
0.7297	0.8757	0.8230	0.8432	0.4396
0.5337	0.5498	0.9096	0.3768	0.4403
0.8522	0.4004	0.2595	0.8008	0.3686
0.4995	0.8403	0.4993	0.8312	0.4295
0.3003	0.4887	0.4067	0.2172	0.2925
0.3900	0.7979	0.9763	0.3003	0.2172
0.5274	0.6793	0.2113	0.5166	0.4632
0.5590	0.5836	0.9329	0.4579	

$$H_0 : P(X \text{ is in class } j) = p_j^* \text{ for } 1, \ldots, c$$
$$H_a : P(X \text{ is in class } j) \neq p_j^* \text{ for at least 1 class} 1, \ldots, c$$

```
ch45 <- c(0.4755, 0.2186, 0.5112, 0.3826, 0.5758,
          0.4274, 0.4295, 0.5233, 0.7500, 0.5484,
          0.6454, 0.8620, 0.5482, 0.5926, 0.5440,
          0.2484, 0.5758, 0.9145, 0.6687, 0.3007,
          0.6521, 0.5456, 0.5101, 0.3607, 0.3943,
          0.3979, 0.4438, 0.6328, 0.9056, 0.8283,
          0.4352, 0.5381, 0.5646, 0.4102, 0.5689,
          0.7297, 0.8757, 0.8230, 0.8432, 0.4396,
          0.5337, 0.5498, 0.9096, 0.3768, 0.4403,
          0.8522, 0.4004, 0.2595, 0.8008, 0.3686,
          0.4995, 0.8403, 0.4993, 0.8312, 0.4295,
          0.3003, 0.4887, 0.4067, 0.2172, 0.2925,
          0.3900, 0.7979, 0.9763, 0.3003, 0.2172,
          0.5274, 0.6793, 0.2113, 0.5166, 0.4632,
          0.5590, 0.5836, 0.9329, 0.4579)
```

Let's break this down into buckets based on the tenths place value and generate a frequency table.

```
tens <- table(unlist(lapply(ch45, function(x) { floor (x * 10)})))
tens
```

```
##
##  2  3  4  5  6  7  8  9
##  7 10 16 19  5  3  9  5
```

The test stastic we'll use is given by the following formula:

$$T = \sum_{j=1}^{c} \frac{(O_j - E_j)^2}{E_j}$$

where c is the number of classes with observed frequencies, the expected value, $E_j = p_j^* N$, and N is the total count of observations. We'll use this to create our goodness-of-fit function.

```
# Takes a vector of observations and the test population probability
goodness.of.fit <- function(data, p) {
  # Create a DataFrame from our observations
  data.df <- as.data.frame(table(data))

  # Test statistic
  T_ <- 0

  # Total number of observations
  N <- length(data)

  # Total number of classes (unique observation values)
  C <- length(data.df$Freq)

  for (j in c(1:C)) {
    Ej <- p * N
    Oj <- as.numeric(levels(data.df$data)[j])

    T_ <- T_ + ((Oj - Ej)^2 / Ej)
  }

  print(paste('N = ', N))
  print(paste('c = ', C))
  print(paste('T = ', exact(T_)))
}
```

(a) Do test of Uniform and Normal distribution

We can now execute our function and compare it with $\chi^2_{(1-\alpha),(c-1)}$

```
# Probability of a uniform is 1/(d-c). In our case, c = 0, d = 9.
p <- 1 / (9 - 0)

goodness.of.fit(tens, p)
```

14

```
## [1] "N =   8"
## [1] "c =   7"
## [1] 859.3472222222223
## [1] "T =   859.347222222222"
```

```
qchisq(p=0.95, df=length(tens))
```

```
## [1] 15.50731
```

$T > \chi^2$, so we reject H_0. The data are not distributed uniformly.

To test for normality, we can use the *nortest* package, which contains several tests for normality, including Pearson's Chi-squared.

$$H_0 : \text{ The data originate from a normal population}$$
$$H_a : \text{ The data are not normally distributed}$$

```
pearson.test(ch45)
```

```
##
##  Pearson chi-square normality test
##
## data:  ch45
## P = 20.703, p-value = 0.01404
```

We reject the null hypothesis with a p-value of $0.01404 < 0.05$. The data do not originate from a normal population. However, we cannot say what distribution, if any, the data follow.
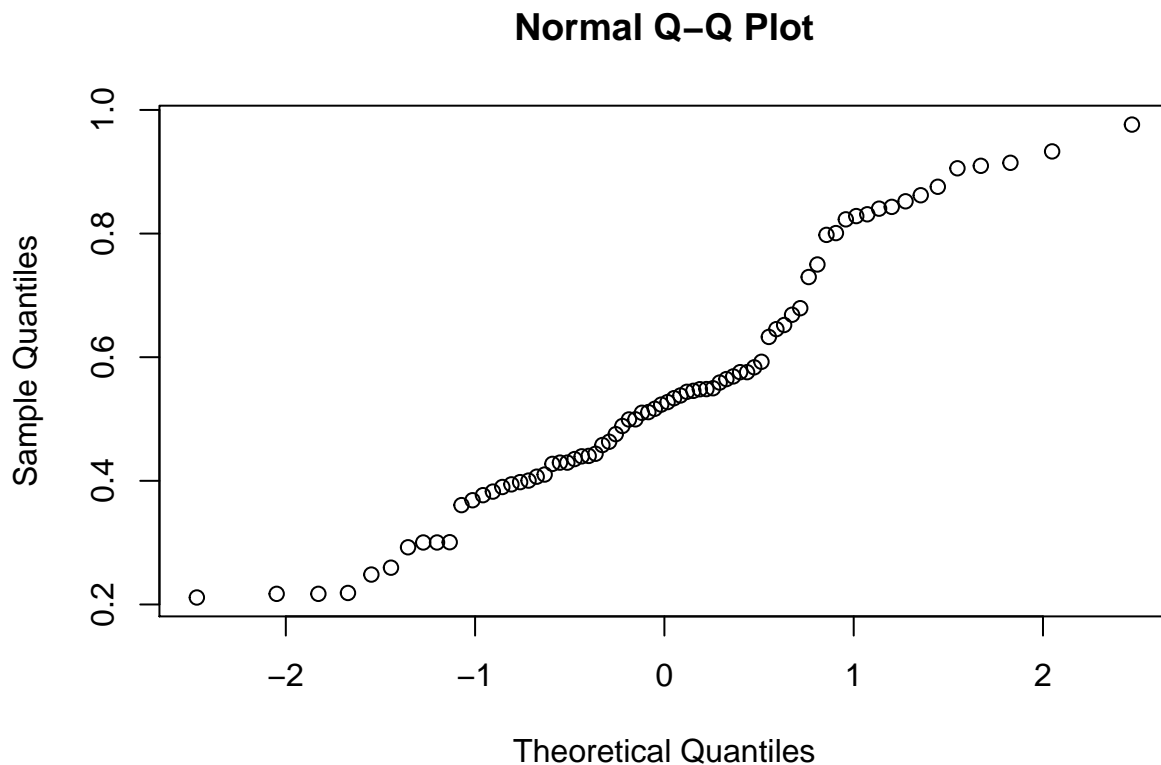
(b) Using Descriptive Statistics, Explore in SPSS perform K-S test (Lilliefors test) and Shapiro Wilk test for Normality.

$$H_0 : \text{The data originate from a normal population}$$
$$H_a : \text{The data are not normally distributed}$$

It can be useful to visualize the data on a Q-Q plot as a sanity check.

```
qqnorm(ch45)
```

Normal Q–Q Plot

Our sample does not look normal, but let's verify this with a few more tests.

```
lillie.test(ch45)
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  ch45
## D = 0.12056, p-value = 0.009591
```

```
shapiro.test(ch45)
```

```
##
##  Shapiro-Wilk normality test
```

```
##
## data:  ch45
## W = 0.95218, p-value = 0.007068
```

Both Lilliefors K-S test for normality and the Shapiro-Wilks test return p-values < 0.05. Therefore, we reject the null hypothesis that the data originate from a normal population. However, we cannot say what distribution, if any, the data follow.

■

Chapter 4.6

4.6, Problem 1). The relative effectiveness of two different sales techniques was tested on 12 volunteer housewives. Each housewife was exposed to each sales technique and asked to buy a certain product, the same product in all cases. At the end of each exposure, each housewife rated the technique with a 1 if she felt she would have agreed to buy the product and a 0 if she probably would not have bought the product.

<div align="center">Housewife</div>

	1	2	3	4	5	6	7	8	9	10	11	12
Technique 1	1	1	1	1	1	0	0	0	1	1	0	1
Technique 2	0	1	1	0	0	0	0	0	1	0	0	1

(a) Use Cochran's test.

We'll write our own function for Cochran's Test, using the following test statistic:

$$T = c(c-1) \frac{\sum_{j=1}^{c} \left(C_j - \frac{N}{c} \right)^2}{\sum_{i=1}^{r} R_i (c - R_i)}$$

$$H_0 : \text{ The treatments are equally effective}$$
$$H_a : \text{ There are differences in effectiveness among the treatments}$$

```r
# Runs Cochran's test for related observations on a matrix
cochran <- function(Matr) {
  c <- ncol(Matr)
  r <- nrow(Matr)
  N <- sum(Matr)

  numerator <- 0
  for (j in 1:c) {
    Cj <- sum(Matr[, j])
    numerator <- numerator + (Cj - (N/c))^2
  }

  denominator <- 0
  for (i in 1:r) {
    Ri <- sum(Matr[i, ])
    denominator <- denominator + (Ri * (c - Ri))
  }

  T_ <- c * (c - 1) * (numerator / denominator)
  p.val <- pchisq(T_, df=(c - 1), lower = FALSE)

  print(paste('T: ', exact(T_)))
  print(paste('p-value: ', exact(p.val)))
}
```

```
Housewives <- matrix(data = c(1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 1,
                              0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1),
                     nrow=12,
                     ncol=2)
colnames(Housewives) <- c('Treatment 1', 'Treatment 2')
rownames(Housewives) <- c('1', '2', '3', '4', '5', '6', '7', '8', '9', '10', '11', '12')
Housewives
```

```
##    Treatment 1 Treatment 2
## 1            1           0
## 2            1           1
## 3            1           1
## 4            1           0
## 5            1           0
## 6            0           0
## 7            0           0
## 8            0           0
## 9            1           1
## 10           1           0
## 11           0           0
## 12           1           1
```

```
cochran(Housewives)
```

```
## [1] 4
## [1] "T:  4"
## [1] 0.045500263896358473
## [1] "p-value:  0.0455002638963585"
```

```
exact(qchisq(p = 0.95, df = (ncol(Housewives) - 1)))
```

```
## [1] 3.841458820694124
```

With $T = 4 > 3.841458820694124$, we reject the null hypothesis. The sales techniques do not seem to be equally effective.

(b) Rearrange the data and use McNemar's test in the large sample form suggested by Equation 3.5.1.

```
mcnemar <- function(Matr) {
  a <- Matr[1, 1]
  b <- Matr[1, 2]
  c <- Matr[2, 1]
  d <- Matr[2, 2]

  T_ <- ((b - c)^2)/(b + c)
  p.val <- pchisq(T_, df=1, lower = FALSE)

  print(paste('T = ', exact(T_)))
  print(paste('p-value = ', exact(p.val)))
}
```

```r
Housewives.McNemar <- t(matrix(data = c(4, 0,
                                        4, 4),
                        nrow=2,
                        ncol=2))
rownames(Housewives.McNemar) <- c('0', '1')
colnames(Housewives.McNemar) <- c('0', '1')
Housewives.McNemar
```

```
##   0 1
## 0 4 0
## 1 4 4
```

```r
mcnemar(Housewives.McNemar)
```

```
## [1] 4
## [1] "T =  4"
## [1] 0.045500263896358473
## [1] "p-value =  0.0455002638963585"
```

```r
exact(qchisq(p = 0.95, df = (ncol(Housewives.McNemar) - 1)))
```

```
## [1] 3.841458820694124
```

Notice, we get the same results as in *(a)*. This is because mathematically, Cochran's test is equal to McNemar's test for $c = 2$.

We once again reject the null hypothesis, and confirm that there is a difference in the techniques.

(c) Ignore the blocking effect in this experiment and treat the data as if 24 different housewives were used. Analyze the data using the test for differences in probabilities given in Section 4.1. Compare with Cochran's test and discuss.

$$H_0 : p_1 = p_2$$
$$H_a : p_1 \neq p_2$$
$$T_1 = \frac{\sqrt{N}(O_{11}O_{22} - O_{12}O_{21})}{\sqrt{n_1 n_2 C_1 C_2}}$$

```r
Housewives.ChiSq <- t(matrix(data = c(8, 4,
                                      4, 8),

                      nrow=2,
                      ncol=2))
rownames(Housewives.ChiSq) <- c('Success', 'Fail')
colnames(Housewives.ChiSq) <- c('Treatment 1', 'Treatment 2')
Housewives.ChiSq
```

```
##         Treatment 1 Treatment 2
## Success           8           4
## Fail              4           8
```

```
T_ <- rxcT(Housewives.ChiSq)
print(paste('T = ', T_))
```

```
## [1] "T =  2.66666666666667"
```

```
exact(qchisq(p = 0.95, df = (ncol(Housewives.ChiSq) - 1)))
```

```
## [1] 3.841458820694124
```

```
p.val <- pchisq(T_, df=1, lower = FALSE)
print(paste('p-value = ', p.val))
```

```
## [1] "p-value =  0.102470434859749"
```

Since $T = 2.6\overline{6} \not> 3.841458820694124$, we cannot reject the null hypothesis here. But this is the same data, right? No. By combining the data, we lose fidelity (information), and in a more mathematical sense, we lose power on this test due to the removal of the blocking.

∎