ISQA 8160 Exam II

Brian Detweiler

Thursday, July 28, 2016

```
## Warning: package 'vcdExtra' was built under R version 3.2.5

## Loading required package: vcd

## Warning: package 'vcd' was built under R version 3.2.5

## Loading required package: grid

## Loading required package: gnm

## Warning: package 'gnm' was built under R version 3.2.5
```

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	85

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

From the data, we have

$$
\begin{aligned}
H_0 &: P_{1i} \le P_{2_i} \\
H_a &: P_{1i} \ge P_{2_i} \text{ for all } i \text{ and } P_{1i} > P_{2_i} \text{ for some } i \\
\alpha &= 0.05 \\
z_{1-\alpha} &= z_{0.95} = 1.645 \\
c &= \{T_4 : T_4 > z_{1-\alpha}\} \\
x_1 &= 8 \\
x_2 &= 3 \\
x_3 &= 0 \\
r_1 &= 40 \\
r_2 &= 21 \\
r_3 &= 7 \\
c_1 &= 34 \\
c_2 &= 6 \\
c_3 &= 10 \\
N_1 &= 276 \\
N_2 &= 45 \\
N_3 &= 85
\end{aligned}
$$

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

$$
\begin{aligned}
T_4 &= \frac{\sum x_i - \sum \frac{r_i c_i}{N_i}}{\sqrt{\sum \frac{r_i c_i (N_i - r_i)(N_i - c_i)}{N_i^2 (N_i - 1)}}} \\
&= \frac{(8+3+0) - \left(\frac{(40)(34)}{276} + \frac{(21)(6)}{45} + \frac{(7)(10)}{85}\right)}{\sqrt{\frac{(40)(34)(276-40)(276-34)}{276^2(276-1)} + \frac{(21)(6)(45-21)(45-6)}{45^2(45-1)} + \frac{(7)(10)(85-7)(85-10)}{85^2(85-1)}}} \\
&= \frac{11 - \left(\frac{340}{69} + \frac{14}{5} + \frac{14}{17}\right)}{\sqrt{\frac{88264}{23805} + \frac{364}{275} + \frac{195}{289}}} \\
&= \frac{11 - 8.55106564364}{\sqrt{5.70616932864}} \\
&= \frac{2.44893435636}{2.38875895156} \\
&= 1.02519107454 \not> 1.645
\end{aligned}
$$

Verifying in R,

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

```r
c_1 <- 34
c_2 <- 6
c_3 <- 10
N_1 <- 276
N_2 <- 45
N_3 <- 85

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)) - 0.5
  denominator <- sqrt(sum(mapply(function(r, c, N) {
      (r * c * (N - r) * (N - c)) / (N^2 * (N - 1))
  }, ri, ci, Ni)))

  numerator/denominator
}

exact(T4(xi, ri, ci, Ni))
```

```
## [1] 1.02519107453538
```

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

```
## [1] 0.81587736388475
```

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

```
## [1] 0.2072851399254297
```

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)) - 0.5
  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] 1.02978398087389
```

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

```
## [1] 0.81953253452475094
```

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

```
## [1] 0.20624132377879
```

Notably, the p-value is 0.206, slightly smaller than the 0.207 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:

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

```
 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

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

```
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

```
  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

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

rxcT(Grades)
```

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

```
# 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.

■