Chi-Squared

Ken Ye

2023-09-29

```r
# Create a data frame with the given data
exam_data <- data.frame(Year = c(1996, 1997, 1998, 1999, 2000, 2001,
                                 2002, 2003, 2004, 2005, 2006, 2007,
                                 2008, 2009, 2010, 2011, 2012, 2013,
                                 2014, 2015),
                        N = c(6964, 7173, 7348, 7311, 7048, 6802,
                              7074, 6751, 7056, 7051, 7006, 7090,
                              7194, 7226, 7335, 7337, 7303, 7482,
                              7601, 7839),
                        PassRate = c(0.82, 0.85, 0.84, 0.85, 0.86,
                                     0.88, 0.87, 0.92, 0.92, 0.92,
                                     0.91, 0.94, 0.91, 0.88, 0.87,
                                     0.84, 0.85, 0.86, 0.87, 0.89) )

head(exam_data)
```

```
##   Year    N PassRate
## 1 1996 6964     0.82
## 2 1997 7173     0.85
## 3 1998 7348     0.84
## 4 1999 7311     0.85
## 5 2000 7048     0.86
## 6 2001 6802     0.88
```

```r
# Calculate Pass Count and Fail Count
exam_data$PassCount <- round(exam_data$N * exam_data$PassRate)
exam_data$FailCount <- exam_data$N - exam_data$PassCount

head(exam_data)
```

```
##   Year    N PassRate PassCount FailCount
## 1 1996 6964     0.82      5710      1254
## 2 1997 7173     0.85      6097      1076
## 3 1998 7348     0.84      6172      1176
## 4 1999 7311     0.85      6214      1097
## 5 2000 7048     0.86      6061       987
## 6 2001 6802     0.88      5986       816
```

```r
# Create Contingency Table 1
table1 <- matrix(c(sum(exam_data$PassCount[exam_data$Year < 2003]),
```

```
                  sum(exam_data$FailCount[exam_data$Year < 2003]),
                  sum(exam_data$PassCount[exam_data$Year >= 2003 & exam_data$Year < 2011]),
                  sum(exam_data$FailCount[exam_data$Year >= 2003 & exam_data$Year < 2011])),
                nrow = 2, byrow = TRUE,
                dimnames = list(c("Pass", "Fail"), c("Before 2003", "2003-2010")))

table1
```

```
##      Before 2003 2003-2010
## Pass       42394      7326
## Fail       51517      5192
```

```
# Chi-Squared test
chisq.test(table1)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table1
## X-squared = 793.98, df = 1, p-value < 2.2e-16
```

```
# Create Contingency Table 2
table2 <- matrix(c(sum(exam_data$PassCount[exam_data$Year < 2011]),
                  sum(exam_data$FailCount[exam_data$Year < 2011]),
                  sum(exam_data$PassCount[exam_data$Year >= 2011]),
                  sum(exam_data$FailCount[exam_data$Year >= 2011])),
                nrow = 2, byrow = TRUE,
                dimnames = list(c("Pass", "Fail"), c("Before 2011", "2011-2015")))

table2
```

```
##      Before 2011 2011-2015
## Pass       93911     12518
## Fail       32396      5166
```

```
# Chi-Squared test
chisq.test(table2)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table2
## X-squared = 102.02, df = 1, p-value < 2.2e-16
```