Lecture 34

Example Day

STAT 8010 Statistical Methods I
November 15, 2019

Whitney Huang
Clemson University

Example

The following contingency table contains enrollment data for a random sample of students from several colleges at Purdue University during the 2006-2007 academic year. The table lists the number of male and female students enrolled in each college. Use the two-way table to conduct a χ^2 test from beginning to end. Use $\alpha = .01$

(Observed)	Female	Male	Total
Liberal Arts	378	262	640
Science	99	175	274
Engineering	104	510	614
Total	581	947	1528

CLEMSON UNIVERSITY

Example Cont'd

(Expected)	Female	Male
Liberal Arts	$\frac{640\times581}{1528}=243.35$	$\frac{640\times947}{1528}=396.65$
Science	$\frac{274\times581}{1528}=104.18$	$\frac{274\times947}{1528}=169.82$
Engineering	$\frac{614\times581}{1528}=233.46$	$\frac{614\times947}{1528}=380.54$

partial χ^2	Female	Male
Lib Arts	$\frac{(378-243.35)^2}{243.35}=74.50$	$\frac{(262-396.65)^2}{396.65}=45.71$
Sci	$\frac{(99-104.18)^2}{104.18}=0.26$	$\frac{(175-169.82)^2}{169.82}=0.16$
Eng	$\frac{(104-233.46)^2}{233.46}=71.79$	$\frac{(510-380.54)^2}{380.54}=44.05$

$$\chi^2 = 74.50 + 45.71 + 0.26 + 0.16 + 71.79 + 44.05 = \boxed{236.47}$$

The $df = (3-1)\times(2-1) = 2 \Rightarrow$ Critical value $\chi^2_{\alpha=.01,df=2} = \boxed{9.21}$

Therefore we **reject** H_0 (at .01 level) and conclude that there is a relationship between gender and major.

R Code & Output

```
table <- matrix(c(378, 99, 104,
                  262, 175, 510), 3, 2)
colnames(table) <- c("Female", "Male")
rownames(table) <- c("Liberal Arts", "Science",
"Engineering")
table
```

```
             Female Male
Liberal Arts    378  262
Science          99  175
Engineering     104  510
```

```
chisq.test(table)
```

```
	Pearson's Chi-squared test

data:  table
X-squared = 236.47, df = 2, p-value <
2.2e-16
```


CLEMSON UNIVERSITY

Take Another Look at the Example

(Proportion)	Female	Male	Total
Liberal Arts	.59 (.65)	.41 (.28)	(.42)
Science	.36 (.17)	.64 (.18)	(.18)
Engineering	.17 (.18)	.83 (.54)	(.40)
Total	.38	.62	1

Rejecting $H_0 \Rightarrow$ conditional probabilities are not consistent with marginal probabilities

Example: Comparing Two Population Proportions

Let $p_1 = \mathbb{P}(Female|Liberal\,Arts)$ and $p_2 = \mathbb{P}(Female|Science)$.

$n_1 = 640, X_1 = 378$, $n_2 = 274, X_2 = 99$

- $H_0 : p_1 - p_2 = 0$ vs. $H_a : p_1 - p_2 \neq 0$
- $z_{obs} = \frac{.59 - .36}{\sqrt{\frac{.52\times.48}{640} + \frac{.52\times.48}{274}}} = 6.36 > z_{0.025} = 1.96$
- We do have enough statistical evidence to conclude that $p_1 \neq p_2$ at .05% significant level.

R Code & Output

```r
prop.test(x = c(378, 99), n = c(640, 274),
          correct = F)
```

```
        2-sample test for equality of
        proportions without continuity
        correction

data:  c(378, 99) out of c(640, 274)
X-squared = 40.432, df = 1, p-value =
2.036e-10
alternative hypothesis: two.sided
95 percent confidence interval:
 0.1608524 0.2977699
sample estimates:
   prop 1    prop 2
0.5906250 0.3613139
```

CLEMSON UNIVERSITY

Example: Test for Homogeneity

Let $p_1 = \mathbb{P}(Liberal\,Arts)$, $p_2 = \mathbb{P}(Science)$, $p_3 = \mathbb{P}(Engineering)$

- The Hypotheses:

  $H_0 : p_1 = p_2 = p_3 = \frac{1}{3}$

  H_a : At least one is different

- The Test Statistic:

$$\chi^2_{obs} = \frac{(640 - 509.33)^2}{509.33} + \frac{(274 - 509.33)^2}{509.33} + \frac{(614 - 509.33)^2}{509.33}$$
$$= 33.52 + 108.73 + 21.51 = 163.76 > \chi^2_{.05,df=2} = 5.99$$

- Rejecting H_0 at .05 level

R Code & Output

```r
chisq.test(x = c(640, 274, 614), p = rep(1/3, 3))
```

```
        Chi-squared test for given
        probabilities

data:  c(640, 274, 614)
X-squared = 163.76, df = 2, p-value
< 2.2e-16
```

The Lady Tasting Tea Experiment

A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup. We will consider the problem of designing an experiment by means of which this assertion can be tested. [...] [It] consists in mixing eight cups of tea, four in one way and four in the other, and presenting them to the subject for judgment in a random order. The subject has been told in advance of that the test will consist, namely, that she will be asked to taste eight cups, that these shall be four of each kind [...]. — Fisher, 1935.

Milk poured first (4 cups) Tea poured first (4 cups)

R Code & Output

```
TeaTasting <-
matrix(c(3, 1, 1, 3), nrow = 2,
       dimnames = list(Guess = c("Milk", "Tea"),
                       Truth = c("Milk", "Tea")))
TeaTasting
```

```
      Truth
Guess  Milk Tea
  Milk    3   1
  Tea     1   3
```

```
fisher.test(TeaTasting, alternative = "greater")
```

```
        Fisher's Exact Test for Count Data

data:  TeaTasting
p-value = 0.2429
alternative hypothesis: true odds ratio is greater
than 1
```

The Lady Tasting Tea