# Chi-Square Tests

R Handout

*Derek H. Ogle*

---

## First Commands

```
> library(NCStats)
```

## Chi-Square Test

### From Summarized Observed Table

When Chinook Salmon (*Oncorhynchus tshawytscha*) were first introduced to Lake Superior there was concern that they would compete with native Lake Trout (*Salvelinus namaycush*) for Cisco (*Coregonus artedi*). Preliminarily, fisheries biologists classified the dominant food items (Cisco, Smelt (*Osmerus mordax*) (another type of fish), or *Mysis* (an invertebrate)) in the diets of 50 Lake Trout and 40 Chinook Salmon. They found that 32, 10, and 8 Lake Trout diets were dominated by Cisco, Smelt, and *Mysis*, respectively. Similarly, 18, 18, and 4 Chinook Salmon diets were dominated by Cisco, Smelt, and *Mysis*, respectively. Test (at the 10% level) if the distribution of dominant food items differs at the 5% level?

```
> freq <- c(32,10,8,18,18,4)
> ( obs <- matrix(freq,nrow=2,byrow=TRUE) )
```

```
     [,1] [,2] [,3]
[1,]   32   10    8
[2,]   18   18    4
```

```
> rownames(obs) <- c("Lake Trout","Chinook Salmon")
> colnames(obs) <- c("Cisco","Smelt","Mysis")
> obs
```

```
              Cisco Smelt Mysis
Lake Trout       32    10     8
Chinook Salmon   18    18     4
```

```
> chi1 <- chisq.test(obs,correct=FALSE)
> chi1$expected
```

```
                 Cisco    Smelt    Mysis
Lake Trout     27.77778 15.55556 6.666667
Chinook Salmon 22.22222 12.44444 5.333333
```

```
> chi1$expected >= 5
```

```
              Cisco Smelt Mysis
Lake Trout     TRUE  TRUE  TRUE
Chinook Salmon TRUE  TRUE  TRUE
```

```
> chi1
```

```
Pearson's Chi-squared test with obs
X-squared = 6.5083, df = 2, p-value = 0.03861
```

```
> chi1$residuals
```

```
                   Cisco     Smelt      Mysis
Lake Trout      0.8011103 -1.408590  0.5163978
Chinook Salmon -0.8956686  1.574852 -0.5773503
```

```
> percTable(obs,margin=1,digits=1)
```

```
               Cisco Smelt Mysis Sum
Lake Trout        64    20    16 100
Chinook Salmon    45    45    10 100
```

```
> ( obs2 <- obs[,-2] )
```

```
               Cisco Mysis
Lake Trout        32     8
Chinook Salmon    18     4
```

```
> ( chi2 <- chisq.test(obs2,correct=FALSE) )
```

```
Warning in chisq.test(obs2, correct = FALSE): Chi-squared approximation may be incorrect
```

```
Pearson's Chi-squared test with obs2
X-squared = 0.0301, df = 1, p-value = 0.8624
```

### From Raw Data

The General Social Survey (GSS) is a nationwide survey that has been administered since 1972 to gather data on contemporary American society in an attempt to monitor and explain trends in attitudes, behaviors, and attributes of American society. One part of that survey asked respondents to state their opinion on how true the following statement was – "*All radioactivity is made by humans.*" Respondents were also categorized by their highest educational degree. The results from this portion of the GSS is in SciTest1.csv. Use these data to determine, at the 5% level, if the distribution of responses to this statement differs among levels of education.

```
> setwd("C:/aaaWork/Web/GitHub/NCMTH107/modules/")
> ST1 <- read.csv("SciTest1.csv")
> str(ST1)
```

```
'data.frame':   2549 obs. of  2 variables:
 $ degree : Factor w/ 5 levels "bach","grad",..: 5 5 5 5 5 5 5 5 5 5 ...
 $ scitest: Factor w/ 4 levels "def.not","def.true",..: 2 2 2 2 2 2 2 2 2 2 ...
```

```
> levels(ST1$degree)
```

```
[1] "bach"  "grad"  "hs"    "jc"    "lt.hs"
```

```
> ST1$fdegree <- factor(ST1$degree,levels=c("lt.hs","hs","jc","bach","grad"))
> levels(ST1$scitest)
```

```
[1] "def.not"   "def.true"  "prob.not"  "prob.true"
```

```
> ST1$fscitest <- factor(ST1$scitest,levels=c("def.not","prob.not","prob.true","def.true"))
```

```
> ( freq.tbl <- xtabs(~fdegree+fscitest,data=ST1) )
```

```
       fscitest
```

```
fdegree def.not prob.not prob.true def.true
  lt.hs      52      112       155       70
  hs        366      451       437      114
  jc         60       44        36        9
  bach      214      135        78       12
  grad      123       57        18        6
```

```
> ST1.chi <- chisq.test(freq.tbl,correct=FALSE)
> ST1.chi$expected
```

```
       fscitest
fdegree   def.not  prob.not prob.true  def.true
  lt.hs 124.37623 121.93448 110.48882  32.20047
  hs    437.39506 428.80816 388.55708 113.23970
  jc     47.64025  46.70498  42.32091  12.33386
  bach  140.36289 137.60730 124.69047  36.33935
  grad   65.22558  63.94508  57.94272  16.88662
```

```
> ST1.chi$expected >= 5
```

```
       fscitest
fdegree def.not prob.not prob.true def.true
  lt.hs    TRUE     TRUE      TRUE     TRUE
  hs       TRUE     TRUE      TRUE     TRUE
  jc       TRUE     TRUE      TRUE     TRUE
  bach     TRUE     TRUE      TRUE     TRUE
  grad     TRUE     TRUE      TRUE     TRUE
```

```
> ST1.chi
```

```
Pearson's Chi-squared test with freq.tbl
X-squared = 288.2331, df = 12, p-value < 2.2e-16
```

```
> ST1.chi$residuals
```

```
       fscitest
fdegree    def.not   prob.not  prob.true   def.true
  lt.hs -6.4897392 -0.8996675  4.2345762  6.6612431
  hs    -3.4137460  1.0716718  2.4575530  0.0714471
  jc     1.7906992 -0.3958064 -0.9716327 -0.9492868
  bach   6.2154218 -0.2222643 -4.1813025 -4.0375730
  grad   7.1536326 -0.8685073 -5.2473280 -2.6492428
```

```
> percTable(freq.tbl,margin=1,digits=1)
```

```
       fscitest
fdegree def.not prob.not prob.true def.true   Sum
  lt.hs    13.4     28.8      39.8     18.0 100.0
  hs       26.8     33.0      31.9      8.3 100.0
  jc       40.3     29.5      24.2      6.0 100.0
  bach     48.7     30.8      17.8      2.7 100.0
  grad     60.3     27.9       8.8      2.9  99.9
```