# Chi-Square Tests in R Handout

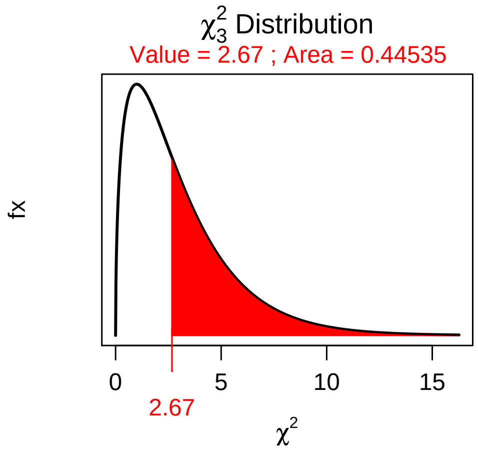
## 1 Initialization

```
> library(NCStats)
> setwd("C://aaaWork//Class Materials//MTH107//Lecture//HOs//")
```

## 2 t Distribution Calculations

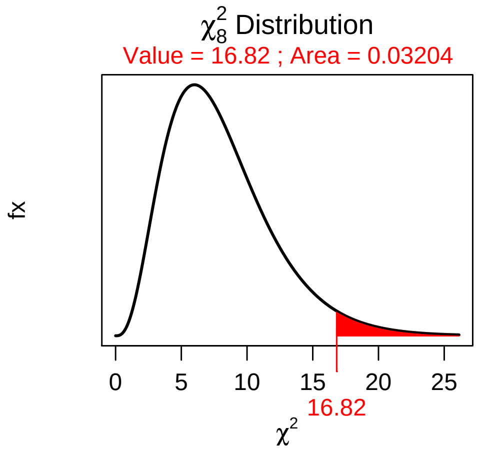
An example of computing the p-value $\chi^2$=2.67 and df=3.

```
> distrib(2.67,distrib="chisq",df=3,lower.tail=FALSE)
```




An example of computing the p-value if $\chi^2$=16.82 and df=8.

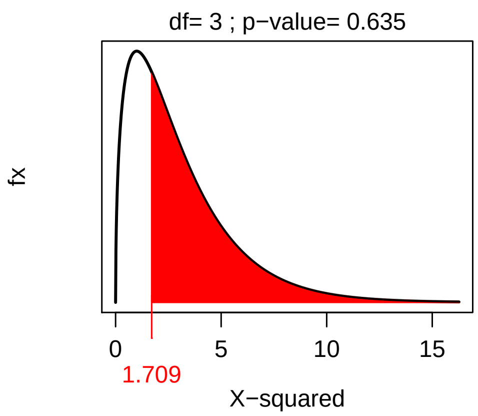
```
> distrib(16.82,distrib="chisq",df=8,lower.tail=FALSE)
```



# 3 Goodness-of-Fit Test

A particular type of corn is known to have one of four types of kernels: purple-smooth, purple-wrinkled, yellow-smooth, and yellow-wrinkled. The purple (P) and smooth (S) alleles are dominant. The cross between heterozygous individuals (i.e., PpSs) should produce a 9:3:3:1 ratio of purple-smooth, purple-wrinkled, yellow-smooth, and yellow-wrinkled individuals. Of the kernels on a randomly selected cob – 32 were purple-smooth, 14 were purple-wrinkled, 8 were yellow-smooth, 4 were yellow-wrinkled. Use the results to determine, at the 5% level, if the theoretical 9:3:3:1 ratio is upheld with these data.

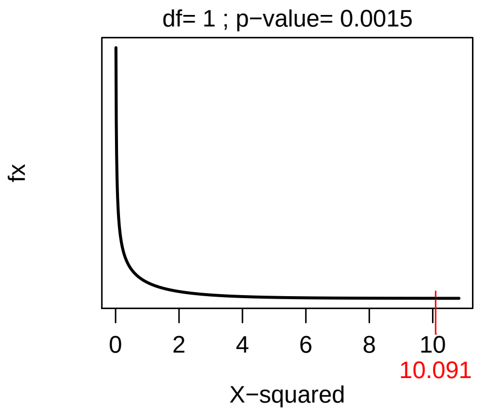
```
> corn <- c(32,14,8,4)
> p.mendel <- c(9,3,3,1)
> corn.chi <- chisq.test(corn,p=p.mendel,rescale=TRUE,correct=FALSE)
> corn.chi$expected

[1] 32.625 10.875 10.875  3.625

> corn.chi$observed

[1] 32 14  8  4

> corn.chi

        Chi-squared test for given probabilities

data:  corn
X-squared = 1.7088, df = 3, p-value = 0.635

> plot(corn.chi)
```




## 3.1 Only Two Levels

In a randomly selected national sample of 1,007 adults, aged 18 and older, conducted Aug. 22-25, 2005, Gallup polls found that that 403 respondents approved of the way that George W. Bush was handling his presidency. In a previous sample (Aug. 8-11, 2005), 45% of the respondents approved of George W. BushŠs handling of the presidency. Assuming that this earlier value was true for the entire population, determine, at the 5% level, if the approval rating has changed by the Aug. 22-25, 2005 sample.

```
> apprv.w <- c(403,1007-403)
> bo.chi <- chisq.test(apprv.w,p=c(0.45,0.55),rescale=TRUE,correct=FALSE)
> bo.chi$expected

[1] 453.15 553.85

> bo.chi$observed

[1] 403 604

> bo.chi

        Chi-squared test for given probabilities

data:  apprv.w
X-squared = 10.0911, df = 1, p-value = 0.00149

> plot(bo.chi)
```




```
> prop.test(403,1007,p=0.45,alt="two.sided",conf.level=0.95,correct=FALSE)

        1-sample proportions test without continuity correction

data:  403 out of 1007, null probability 0.45
X-squared = 10.0911, df = 1, p-value = 0.00149
alternative hypothesis: true p is not equal to 0.45
95 percent confidence interval:
 0.3703726 0.4307831
sample estimates:
        p
0.4001986
```

# 4 Chi-Square Test

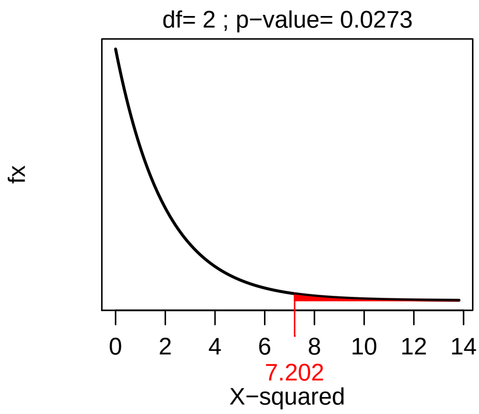
## 4.1 From Summarized Observed Table

Do the dominant food items in lake trout and chinook salmon differ at the 5% level? Summarized observed table given in class.

```
> obs <- matrix(c(38,12,10,18,18,4),nrow=2,byrow=TRUE)
> obs

     [,1] [,2] [,3]
[1,]   38   12   10
[2,]   18   18    4

> rownames(obs) <- c("lake trout","chinook salmon")
> colnames(obs) <- c("lake herring","smelt","mysis")
> obs

               lake herring smelt mysis
lake trout               38    12    10
chinook salmon           18    18     4

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

               lake herring smelt mysis
lake trout             33.6    18   8.4
chinook salmon         22.4    12   5.6

> chi1$expected >= 5

               lake herring smelt mysis
lake trout             TRUE  TRUE  TRUE
chinook salmon         TRUE  TRUE  TRUE

> chi1$observed

               lake herring smelt mysis
lake trout               38    12    10
chinook salmon           18    18     4

> chi1

        Pearson's Chi-squared test

data:  obs
X-squared = 7.2024, df = 2, p-value = 0.02729

> chi1$residuals^2

               lake herring smelt     mysis
lake trout        0.5761905     2 0.3047619
chinook salmon    0.8642857     3 0.4571429

> chi1$residuals
```

```
               lake herring     smelt       mysis
lake trout        0.7590721 -1.414214  0.5520524
chinook salmon   -0.9296697  1.732051 -0.6761234
```

```
> plot(chi1)
```




## 4.2 From Raw Data

On the GSS, respondents were asked 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. Use the results in the **SciTest1.txt** data file to determine, at the 5% level, if the response to the question differs among levels of education

```
> ST1 <- read.table("SciTest1.txt",head=TRUE)
> 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 <- table(ST1$fdegree,ST1$fscitest)
> freq.tbl

        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

         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$observed >= 5

       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$observed

       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

        Pearson's Chi-squared test

data:  freq.tbl
X-squared = 288.2331, df = 12, p-value < 2.2e-16

> ST1.chi$residuals

          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
```

## 4.3 Alternative Method of Receiving Results

One can use `chisq.detail()` from the `TeachingDemos` package to provide an alternative presentation of the results. This is demonstrated below with the results from above.

```
> chisq.detail(freq.tbl)

observed
expected
```

```
      def.not prob.not prob.true def.true Total
lt.hs   52      112       155       70      389
       124.38  121.93    110.49    32.20

hs     366      451       437      114     1368
       437.40  428.81    388.56   113.24

jc      60       44        36        9      149
        47.64   46.70     42.32    12.33

bach   214      135        78       12      439
       140.36  137.61    124.69    36.34

grad   123       57        18        6      204
        65.23   63.95     57.94    16.89

Total  815      799       724      211     2549


Cell Contributions
      def.not   prob.not   prob.true   def.true
lt.hs   42.12 +     0.81 +     17.93 +    44.37 +
hs      11.65 +     1.15 +      6.04 +     0.01 +
jc       3.21 +     0.16 +      0.94 +     0.90 +
bach    38.63 +     0.05 +     17.48 +    16.30 +
grad    51.17 +     0.75 +     27.53 +     7.02 = 288.23

df = 12  P-value = 0
```

## 5 Class Exercise

- Review Exercises 11.10 and 11.12.
- Review Exercises 11.20 and 11.22.