

```
# 以下から、奇数偶数の好み実験結果の一部のデータをインポートしてください。
# http://lab.kenrikodaka.com/_download/class/2022_AppliedMedia/oddeven.csv
```

```
#Environmentパネルの「Import Dataset」から
#oddeven.csvのデータを、expdatという変数に読み込みます。
```

```
#Name : expdat (入力してください)
#Encoding:Automatic
#Heading:Yes
#Row names:Automatic
#Separator:Comma
#Decimal:Period
#Quote : Double
#Coomen:None
#na.stfings:NA
#Strings as factors:Unchecked (チェックは要りません)
```

```
#誕生日と奇数偶数の好みに関するアンケートの架空のデータ (1000人分) です。
#month (誕生月) ・ name (誕生日) ・ preference (奇数が好き : 1, 偶数が好き : 0)
#sex (男性 : 1, 女性 : 0) ・ $domhand (利き手が左 : 1, 利き手が右 : 0)
#age (年齢)
```

```
#最初の6行をちょっと出し
head(expdat)
```

```
#  month day preference sex domhand age
#1      8 29           1   1         0 19
#2      1 29           0   0         1 18
#3      3 19           0   1         0 18
#4     10  6           0   0         0 18
#5     10 10           1   0         0 18
#6      1 13           0   1         0 19
```

```
# 翻訳すると以下のようになります。
# 8月1日、奇数好き、男性、右利き、19歳
# 1月29日、偶数好き、女性、左利き、18歳
# 3月19日、偶数好き、男性、右利き、18歳
# 10月6日、偶数好き、女性、右利き、18歳
# 10月10日、奇数好き、女性、右利き、18歳
# 1月13日、偶数好き、男性、右利き、19歳
```

```
expdat$gender = factor(expdat$sex, levels=0:1, labels=c("FEMALE", "MALE"))
expdat$domhand =
factor(expdat$domhand, levels=0:1, labels=c("RIGHT", "LEFT"))
expdat$preference =
factor(expdat$preference, levels=0:1, labels=c("EVEN", "ODD"))
```

```
#####
#####
##### [KAI] カイ二乗分布 #####
#####
#####
```

```

#-----
# [KAI.5] 独立性検定：「性別」と「数字の好み」
#-----

# 偶数・奇数の好みは、性別によって差があるか？

# 1000人分の性別のサンプル
expdat$gender
#[1] MALE   FEMALE MALE   FEMALE FEMALE MALE   FEMALE MALE   MALE   FEMALE
MALE
#[12] FEMALE FEMALE FEMALE FEMALE FEMALE FEMALE MALE   FEMALE MALE
FEMALE MALE
#...

# (性別に対応する) 1000人分の好み
expdat$preference
#[1] ODD   EVEN EVEN EVEN ODD   EVEN EVEN EVEN EVEN EVEN EVEN EVEN EVEN EVEN
EVEN
#[16] ODD   EVEN EVEN EVEN ODD   EVEN ODD   ODD   ODD   ODD   ODD   EVEN EVEN
EVEN EVEN
#...

# 「性別」と「好み」の対応に関する2x2の集計表
obs = table(expdat$gender, expdat$preference)
#      EVEN ODD
#FEMALE  370 203
#MALE    230 197

# 戻り値のクラスは「table」です。
class(obs)
# [1] "table"

# tableはtableはdata.frame等のデータを
# クロス集計した頻度値を格納するのに使用されます。

# 「table」の各セルの値は以下のような書式で参照できます。
obs[1,1]
obs["FEMALE","EVEN"]
#370

obs[,2]
obs[, "ODD"]
#FEMALE   MALE
#203     197

# addmarginsを使うと自動的に合計欄を作ってくれます。
obs = addmargins(obs)
#      EVEN  ODD  Sum
#FEMALE  370  203  573
#MALE    230  197  427
#Sum     600  400 1000

# 観測度数のテーブルを期待度数のテーブルにコピー
expected = obs

# 合計欄をうまく使って、一行で期待度数を計算

```

```

for(i in 1:2){
  for(j in 1:2){
    expected[i,j] = obs[i,"Sum"] * (obs["Sum",j] / obs["Sum","Sum"])
    #expected[i,j] = obs[i,3] * (obs[3,j] / obs[3,3]) #同じです。
  }
}
}
expected
#      EVEN    ODD    Sum
#FEMALE 343.8 229.2 573.0
#MALE   256.2 170.8 427.0
#Sum    600.0 400.0 1000.0

# 合計欄をカット
obs = obs[1:2,1:2]
expected = expected[1:2,1:2]

# 各セルのχ二乗値の計算
kai2_table = (obs-expected)^2 / expected
#      EVEN    ODD
#FEMALE 1.996626 2.994939
#MALE   2.679313 4.018970

# 全てのセルの誤差項を足したものがχ二乗値
kai2 = sum(kai2_table)
#[1] 11.68985

# chisq.test関数を使うと、上記の計算を自動的に行ってくれます。
# (correct=Fは、イエーツ補正の有無です。ここではFALSEにします)
result = chisq.test(expdat$gender, expdat$preference, correct=F)
#X-squared = 11.69, df = 1, p-value = 0.0006284

result$statistic
#X-squared
#11.68985

# 既にp値も見えていますが、これを実際に確かめてみましょう。

#-----
# [KAI.6] カイ二乗値の分布を調べる
#-----

# データと条件を揃えて、
# 427人の男性と573人の女性、
# 偶数対奇数の好みが「60:40」のとき

set.seed(1017)
getKai2 = function(){

  # 1:5を人数分だけランダムに生成
  sample.female = sample(1:5,573,replace=TRUE);
  sample.male = sample(1:5,427,replace=TRUE);

  # 1:3を0に4:5を1に変換 (floor関数を使っている)
  sample.female = floor(sample.female / 4)
  sample.male = floor(sample.male / 4)
}

```

```

# 0を573、1を427並べた性別サンプル
gender = c(rep(0,times=573),rep(1,times=427))
# 前半 (1:573) に女性の好み、後半 (574:1000) に男性の好みを対応させる
pref = c(sample.female,sample.male)

# カイ二乗値を計算
result = chisq.test(gender,pref,correct=F)
result$statistic
}

# getKai2()を10000回繰り返す
sample = replicate(10000,getKai2())
summary(sample)
#Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
#0.000002  0.094335  0.426451  0.982526  1.278155 18.543110

# ヒストグラムを見る。
par(cex=2) #フォントサイズを挙げる (デフォルトの2倍)
h = hist(sample,breaks = seq(0,20,by=0.2)) #縦軸はFrequency (頻度)
h = hist(sample,breaks = seq(0,20,by=0.2),probability = T) #縦軸は確率密度

# 横軸が11.69以上の確率密度を全て加算
sum(h$density[h$breaks>11.69],na.rm = T)
# 0.0025

# これがp値に対応。
# 女性と男性との間で、好みが、観測値以上の偏りを示す確率は、
# 0.0025 (5%以下)

# すなわち「性別」と「好み」は、違いに独立ではない。
# より具体的に、女性の方が偶数を有意に好む傾向にある。

# ちなみに、chisq.testを使うと、正確なカイ二乗分布を使った
# (近似ではない) p値がわかる。
result = chisq.test(expdat$gender, expdat$preference, correct=F)
#X-squared = 11.69, df = 1, p-value = 0.0006284

result$p.value
#0.0006284205

# 自由度は1 (2-1) * (2-1)
result$parameter
#df
#1

#-----
# [KAI.5] 独立性検定の練習 (誕生日の数字の偶数と奇数の影響)
#-----

#####
# 男性のケース
#####
#-----
# 誕生月と好みの関係 (男性)

```

```

#-----

# 男性の誕生月のベクトルを2で割った余り
tmp = expdat$month[expdat$gender=="MALE"] %% 2
#[1] 0 1 1 0 1 0 1 0 1 1 0 0 0 0 0 0 1 1 0 1 1 0 0 1 0 0 0 1 1 0 1 1 0 0 1
0 0 1 0 0
#[41] 1 0 0 0 1 0 0 0 0 1 1 1 1 0 1 1 0 0 1 1 0 1 1 0 1 0 1 0 0 1 1 1 1 1
0 0 0 1 0 1
#...

# ファクタに変換
month.male = factor(tmp, levels=0:1, labels=c("EVEN", "ODD"))
#[1] EVEN ODD ODD EVEN ODD EVEN ODD EVEN ODD ODD EVEN EVEN EVEN EVEN
EVEN EVEN
#[17] ODD ODD EVEN ODD ODD EVEN EVEN ODD EVEN EVEN EVEN ODD ODD
EVEN ODD ODD
#...
pref.male = expdat$preference[expdat$gender=="MALE"]
#[1] ODD EVEN EVEN EVEN EVEN EVEN EVEN ODD ODD ODD ODD ODD EVEN EVEN
EVEN EVEN
#[17] EVEN EVEN EVEN EVEN EVEN ODD EVEN ODD ODD ODD EVEN ODD EVEN
EVEN EVEN EVEN

table(month.male, pref.male)
#           pref.male
#month.male EVEN ODD
#   EVEN  121  98
#   ODD   109  99

# 棒グラフで可視化
par(cex=2) #フォントサイズを挙げる (デフォルトの2倍)
barplot(table(pref.male, month.male),
        legend=c("PREFER TO EVEN", "PREFER TO ODD"),
        xlab="MONTH NUMBER",
        ylim=c(0, 400))

chisq.test(month.male, pref.male, correct=F)
#X-squared = 0.34802, df = 1, p-value = 0.5552

# 男性の「偶数-奇数の好み」に「誕生月の偶数・奇数」の影響はない。

#-----
# 誕生日 (日にち) と好みの関係 (男性)
#-----

# 同様に、
tmp = expdat$day[expdat$gender=="MALE"] %% 2
day.male = factor(tmp, levels=0:1, labels=c("EVEN", "ODD"))
pref.male = expdat$preference[expdat$gender=="MALE"]

table(day.male, pref.male)
#           pref.male
# day.male EVEN ODD
#   EVEN  113  97
#   ODD   117 100

# 棒グラフで可視化

```

```

par(cex=2) #フォントサイズを挙げる (デフォルトの2倍)
barplot(table(pref.male,day.male),
        legend=c("PREFER TO EVEN","PREFER TO ODD"),
        xlab="DAY NUMBER",
        ylim=c(0,400))

chisq.test(day.male, pref.male, correct=F)
#X-squared = 0.00049653, df = 1, p-value = 0.9822

# 男性の「偶数-奇数の好み」に「誕生日の偶数・奇数」の影響はない。

#-----
# 誕生日の2つの数字の交互作用と好みの関係 (女性)
#-----

month = expdat$month[expdat$gender=="MALE"] %% 2
day = expdat$day[expdat$gender=="MALE"] %% 2

# 誕生日の月も日にちも偶数
even2.male = month==0 & day==0
#[1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE
TRUE TRUE
#[14] TRUE TRUE TRUE FALSE FALSE TRUE FALSE FALSE FALSE TRUE FALSE
FALSE FALSE
#...

# 誕生日の月も日にちも奇数
odd2.male = month==1 & day==1
#[1] FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE TRUE TRUE FALSE
FALSE FALSE
#[14] FALSE FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE FALSE FALSE
FALSE FALSE
#...

# それ以外
etc.male = (month==1 & day==0) | (month==0 & day==1)
#[1] TRUE FALSE FALSE TRUE TRUE TRUE TRUE FALSE FALSE FALSE FALSE
FALSE FALSE
#[14] FALSE FALSE FALSE TRUE FALSE FALSE TRUE FALSE TRUE FALSE TRUE
TRUE TRUE

# これでも同じ結果となります。
# etc.male = even2.male==F & odd2.male==F

# これらをラベルづけしたベクトルdouble.maleをつくります。
# etc.maleは使いません、。
double.male = vector("character",427)
for(i in 1:427){
  if(even2.male[i]){
    double.male[i] = "EVEN2"
  }else if(odd2.male[i]){
    double.male[i] = "ODD2"
  }else{
    double.male[i] = "ETC"
  }
}
}

```

```
double.male
#[1] "ETC" "ODD2" "ODD2" "ETC" "ETC" "ETC" "ETC" "EVEN2"
"ODD2" "ODD2"
#[11] "EVEN2" "EVEN2" "EVEN2" "EVEN2" "EVEN2" "EVEN2" "ETC" "ODD2"
"EVEN2" "ETC"
#...
```

```
# 文字列ベクトルをファクタ化します。
# (もともと文字列なので見た目は変わりません)
# (インデックスは、EVEN2、ODD2、ETCの順となるようにします。)
double.male = factor(double.male, levels=c("EVEN2","ODD2","ETC"))
#[1] "ETC" "ODD2" "ODD2" "ETC" "ETC" "ETC" "ETC" "EVEN2"
"ODD2" "ODD2"
#[11] "EVEN2" "EVEN2" "EVEN2" "EVEN2" "EVEN2" "EVEN2" "ETC" "ODD2"
"EVEN2" "ETC"
#...
```

```
# 男性の好み (偶数・奇数)
pref.male = expdat$preference[expdat$gender=="MALE"]
```

```
table(double.male,pref.male)
#           pref.male
#double.male EVEN ODD
#   EVEN2    67  45
#   ODD2     63  47
#   ETC      100 105
```

```
# 棒グラフで可視化
par(cex=2) #フォントサイズを挙げる (デフォルトの2倍)
barplot(table(pref.male,double.male),
        legend=c("PREFER TO EVEN","PREFER TO ODD"),
        xlab="MONTH AND DAY NUMBER",
        ylim=c(0,300))
```

```
# 自由度は2になります。
chisq.test(double.male, pref.male, correct=F)
#X-squared = 4.2457, df = 2, p-value = 0.1197
```

```
# 誕生日の月日の交互作用は、奇数と偶数の好みに有意な影響は存在しない
```

```
#####
# 女性のケース
#####
```

```
#-----
# 誕生月と好みの関係 (女性)
#-----
```

```
tmp = expdat$month[expdat$gender=="FEMALE"] %% 2
month.female = factor(tmp, levels=0:1, labels=c("EVEN","ODD"))
pref.female = expdat$preference[expdat$gender=="FEMALE"]
table(month.female,pref.female)
#           pref.male
#month.female EVEN ODD
# MONTH.EVEN  189  79
# MONTH.ODD   181 124
```

```

# 棒グラフで可視化
par(cex=2) #フォントサイズを挙げる (デフォルトの2倍)
barplot(table(pref.female,month.female),
        legend=c("PREFER TO EVEN","PREFER TO ODD"),
        xlab="MONTH NUMBER",
        ylim=c(0,400))

chisq.test(month.female, pref.female, correct=F)
#X-squared = 7.7917, df = 1, p-value = 0.005249

# p<0.01 (誕生日の偶数・奇数が、数字の好みに有意に影響する)

#-----
# 誕生日 (日にち) と好みの関係 (女性)
#-----

tmp = expdat$day[expdat$gender=="FEMALE"] %% 2
day.female = factor(tmp,levels=0:1,labels=c("EVEN","ODD"))
pref.female = expdat$preference[expdat$gender=="FEMALE"]
table(day.female,pref.female)
#           pref.male
# day.female EVEN ODD
# DAY:EVEN   193  86
# DAY:ODD    177 117

# 棒グラフで可視化
barplot(table(pref.female,day.female),
        legend=c("PREFER TO EVEN","PREFER TO ODD"),
        xlab="DAY NUMBER",
        ylim=c(0,400))

chisq.test(day.female, pref.female, correct=F)
#X-squared = 5.0367, df = 1, p-value = 0.02482

# p<0.05 (誕生日の日にちの偶数・奇数が、数字の好みに有意に影響する)

#-----
# 誕生日の2つの数字の交互作用と好みの関係 (女性)
#-----

##宿題にします。

## 男性と同様に、
## 女性の誕生日の月と日にちの偶数・奇数の組み合わせが、
## 数字の偶数・奇数の好みに影響するか否かを検定してください。

## 最終的にchisq.testを実行し、
##  $\chi^2$ 乗値およびp値を求め、有意の有無を判定してください。

#作成したプログラムは、指定のリンク先に提出してください。
#ファイル名は「Class1020_Work_1950XX.R」をお願いします。
# (注) このファイルを上書き保存したものを提出すればよいです。

```