Twoway_Tables.R

monca016

Mon Oct 08 18:23:54 2018

```
# H0 for Shopping at X, Y, Neither
# and Saw Ad, Did Not See Ad

##        Shopped   Shopped   Neither   Sum
##           at X      at Y
```

```
# Saw Ad   |   100      80     220     400
#          |
# Did Not  |
  # See Ad |    25      20      55      100
#          |
# Sum      |   125      100      275     500

# Multinomial: Testing K proportions for a single variable
  # An employer wants to know which days of the week employees are absent in a five day work week
  # The day of the week for a random sample of 60 absences was recorded

  # actual number of absences per day M-F
ObsCounts <- c(15, 12, 9, 9, 15)
ObsCounts
```

```
## [1] 15 12  9  9 15
```

```
  # expected number if all days are the same
  # This is just 60 / 5 = 12
ExpCounts <- rep(12, times = 5)
ExpCounts
```

```
## [1] 12 12 12 12 12
```

```
  # expected probabilities
ExpProb <- ExpCounts/sum(ExpCounts)
ExpProb
```

```
## [1] 0.2 0.2 0.2 0.2 0.2
```

```
  # Multinomial test
chisq.test(ObsCounts, p = ExpProb)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  ObsCounts
## X-squared = 3, df = 4, p-value = 0.5578
```

```
  # cell by cell contributions to observed Chi-Square value: ((Obs-Exp)^2/E)
  # Higher values indicate where the largest contributions to the
  # X^2 value come from as an indication of where the
  # relationship is, if there is one
(chisq.test(ObsCounts, p = ExpProb)$residuals)^2
```

```
## [1] 0.75 0.00 0.75 0.75 0.75
```

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 3.4.4
```

```
# Contigency Table
  # Transit Railroads is interested in the relationshpi between travel distances and the ticket class purchased.
  # A random sample of 200 passengers is taken.
Transit <- read_excel("Transit.xlsx", na = "NA", col_names = TRUE)

  # create and save contigency table
TransitTable <- table(Transit)
TransitTable
```

```
##         Distance
## Class    1-100 101-200 201-300 301-400 401-500
##   First      6       8      15      21      10
##   Second    14      16      17      14       6
##   Third     21      18      16      12       6
```

```
  # Test of independence
chisq.test(TransitTable)
```

```
##
##  Pearson's Chi-squared test
##
## data:  TransitTable
## X-squared = 15.923, df = 8, p-value = 0.0435
```

```
  # cell-by-cell contributions to observed Chi-Square value: ((Obs-Exp)^2/Exp)
(chisq.test(TransitTable)$residuals)^2
```

```
##         Distance
## Class         1-100    101-200    201-300    301-400    401-500
##   First  3.22682927 1.67936508 0.02500000 3.37659574 1.75151515
##   Second 0.00511285 0.26474058 0.05263682 0.19339632 0.25466757
##   Third  2.43376044 0.46502935 0.13187215 1.54905421 0.51318804
```

```
  # cell-by-cell expected values
chisq.test(TransitTable)$expected
```

```
##         Distance
## Class     1-100 101-200 201-300 301-400 401-500
##   First  12.300   12.60   14.40  14.100    6.60
##   Second 13.735   14.07   16.08  15.745    7.37
##   Third  14.965   15.33   17.52  17.155    8.03
```

```
# R 5.1 Clinic
# A health clinic with 5 offices is open 7 days a week.
# Ops manager is currently maintainng the previous staffing level that is balanced on weekdays and reduced on Sat/Sun.
# Her predecessor claimed that patient demand is fairly level during the week and about 25% on Sat/Sun.
# Current manager suspects that the staff wants their weekends free
# She decides to study patient demand to see whether assumed demand pattern holds
# Ops manager takes random sample of 10 days for each day of the week over past year.
# For 70 days observed, actual and predicted patient counts by day are:

#           Obs         Exp
# Sunday    450         645
# Monday    662         862
# Tuesday   831         862
# Wednesday 1042        862
# Thursday  1103        862
# Friday    1075        862
# Saturday  437         645
# Total     5600        5600

Clinic_ObsCounts <- c(450, 662, 831, 1042, 1103, 1075, 437)
Clinic_ObsCounts
```

```
## [1]  450  662  831 1042 1103 1075  437
```

```
Clinic_ExpCounts <- c(645, 862, 862, 862, 862, 862, 645)
Clinic_ExpCounts
```

```
## [1] 645 862 862 862 862 862 645
```

```
Clinic_ExpProb <- Clinic_ExpCounts/sum(Clinic_ExpCounts)
Clinic_ExpProb
```

```
## [1] 0.1151786 0.1539286 0.1539286 0.1539286 0.1539286 0.1539286 0.1151786
```

```
# A.  State the appropriate hypothesis

# There is no relationship between the number of patients and the day of the week;
# Patient demand is fairly level during the week and 25% reduced on Saturday and Sunday

# B.  Perform a goodness-of-fit test to determine whether the patient demand follows the predicted pattern.

chisq.test(Clinic_ObsCounts, p = Clinic_ExpProb)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  Clinic_ObsCounts
## X-squared = 331.15, df = 6, p-value < 2.2e-16
```

```
(chisq.test(Clinic_ObsCounts, p = Clinic_ExpProb)$residuals)^2
```

```
## [1] 58.953488 46.403712  1.114849 37.587007 67.379350 52.632251 67.075969
```

```
# C.  What are the assumptions for the test?  Check the assumptions as needed and possible.

  # Random sampling
  # Process stability
  # np >= 5, the sample size has to be big enough that our proportion size is at least 5 for all of my days of the week

# R 5.2 Gasoline
# An investigation of brand loyalty, members of a random sample of car owners were asked about the brand of
# gasoline in their last two purchases.

Gasoline <- read_excel("Gasoline.xlsx", na = "NA", col_names = TRUE)

GasolineTable <- table(Gasoline$`Second-last`, Gasoline$Last)
#GasolineTable <- table(Gasoline)
GasolineTable
```

```
##
##      1  2  3  4
##   1 47 14 27 14
##   2 18 63 33 13
##   3 26 48 59 16
##   4 30 23 29 53
```

```
chisq.test(GasolineTable)
```

```
##
##  Pearson's Chi-squared test
##
## data:  GasolineTable
## X-squared = 114.12, df = 9, p-value < 2.2e-16
```

```
(chisq.test(GasolineTable)$residuals)^2
```

```
##
##              1          2          3          4
##   1 21.8764183  8.0874729  0.2001518  1.3561017
##   2  4.7713302 18.9654562  0.3614979  4.8770563
##   3  2.3792598  0.5847585  5.9655403  5.0642490
##   4  0.1065681  6.5298009  2.5406117 30.4527412
```

```
# A.  Is there evidence of a relationship between the brands of gasoline in successive fillups?

# B.  What are the assumptions for the test?  Check the assumptions as needed and possible.

  # Random sampling
  # Process stability
  # H0 is true


# C.  What is the nature of the relationship, if it exists?
```