

# Poisson\_fit2\_gof.R

John

Thu Jul 26 14:05:11 2018

```
# Poisson_fit2_gof.R
# Fitting the Poisson to frequency data

# Load necessary libraries
library(ggplot2)
library(MASS)

# Read in data set
poisdata <- read.table(header=T,colClasses=c("numeric","numeric"),text="
y obsfreq
0 24
1 16
2 16
3 18
4 15
5 9
6 6
7 5
8 3
9 4
10 3
11 0
12 1
")

# Generate offset y values for plot
poisdata <- transform(poisdata,yexp=y-0.1)
poisdata <- transform(poisdata,yobs=y+0.1)

# Print data
poisdata
```

```
##      y obsfreq yexp yobs
## 1  0      24 -0.1  0.1
## 2  1      16  0.9  1.1
## 3  2      16  1.9  2.1
## 4  3      18  2.9  3.1
## 5  4      15  3.9  4.1
## 6  5       9  4.9  5.1
## 7  6       6  5.9  6.1
## 8  7       5  6.9  7.1
## 9  8       3  7.9  8.1
## 10 9       4  8.9  9.1
## 11 10      3  9.9 10.1
## 12 11      0 10.9 11.1
## 13 12      1 11.9 12.1
```

```

# Convert tabulated data to raw form
ydata <- rep(poisdata[, "y"], poisdata$obsfreq)

# Calculate mean and variance
mean(ydata)

## [1] 3.166667

var(ydata)

## [1] 7.770308

# Fit Poisson distribution
fitout <- fitdistr(ydata, "poisson")
fitout

##      lambda
## 3.166667
## (0.1624466)

# Calculate expected frequencies
poisdata <- transform(poisdata, poisprob=dpois(y, fitout$estimate))
poisdata <- transform(poisdata, expfreq=fitout$n*poisprob)

# Calculate test values for each cell
poisdata <- transform(poisdata, cellchi2=(obsfreq-expfreq)^2/expfreq)
poisdata <- transform(poisdata, sumchi2=cumsum(cellchi2))
poisdata <- transform(poisdata, olnoe=obsfreq*log(obsfreq/expfreq))
poisdata <- transform(poisdata, sumlike=cumsum(olnoe))

# Print revised data
poisdata

##      y obsfreq yexp yobs      poisprob      expfreq      cellchi2      sumchi2
## 1  0      24 -0.1  0.1 4.214384e-02  5.05726122 7.095290e+01 70.95290
## 2  1      16  0.9  1.1 1.334555e-01 16.01466053 1.342091e-05 70.95291
## 3  2      16  1.9  2.1 2.113045e-01 25.35654584 3.452558e+00 74.40547
## 4  3      18  2.9  3.1 2.230437e-01 26.76524284 2.870494e+00 77.27597
## 5  4      15  3.9  4.1 1.765763e-01 21.18915058 1.807792e+00 79.08376
## 6  5       9  4.9  5.1 1.118316e-01 13.41979537 1.455655e+00 80.53941
## 7  6       6  5.9  6.1 5.902225e-02  7.08266978 1.654989e-01 80.70491
## 8  7       5  6.9  7.1 2.670054e-02  3.20406490 1.006653e+00 81.71157
## 9  8       3  7.9  8.1 1.056896e-02  1.26827569 2.364525e+00 84.07609
## 10 9       4  8.9  9.1 3.718710e-03  0.44624515 2.830098e+01 112.37707
## 11 10      3  9.9 10.1 1.177591e-03  0.14131096 5.783064e+01 170.20770
## 12 11      0 10.9 11.1 3.390036e-04  0.04068043 4.068043e-02 170.24838
## 13 12      1 11.9 12.1 8.945928e-05  0.01073511 9.116299e+01 261.41137
##      olnoe      sumlike
## 1 37.37349010 37.37349
## 2 -0.01465382 37.35884
## 3 -7.36717109 29.99167
## 4 -7.14118282 22.85048
## 5 -5.18158627 17.66890
## 6 -3.59555675 14.07334
## 7 -0.99534873 13.07799
## 8  2.22508814 15.30308

```

```
## 9  2.58286211 17.88594
## 10 8.77272470 26.65867
## 11 9.16621406 35.82488
## 12      NaN      NaN
## 13 4.53423530      NaN
```

```
# Generate plot showing frequencies
ggplot(poisdata,aes(yobs,obsfreq))+
  geom_bar(stat="identity",width=0.05,fill="blue")+
  geom_point(color="blue",size=3)+
  geom_bar(aes(yexp,expfreq),stat="identity",width=0.05,fill="red")+
  geom_point(aes(yexp,expfreq),color="red",size=3)+
  ggtitle("Fitting the Poisson to frequency data")
```

