Lab-08

Nilanjana Dey

2024-01-26

Question-1: For example, we collected wild tulips and found that 71 were red, 40 were yellow and 29 were white. 1. Are these colors equally common? Find the expected values and residuals? 2. Suppose that, in the region where you collected the data, the ratio of red, yellow and white tulip is 4:3:1. Find the expected values and residuals?

Part 1:

HYPOTHESIS TESTING:

H0: THE COLORS ARE EQUALLY COMMON

VS

H1: THE COLORS ARE NOT COMMON

# Given data  
observed <- c(71, 40, 29)  
  
# Expected probabilities assuming equal distribution  
expected\_prob <- c(1/3, 1/3, 1/3)

# Expected values  
expected <- sum(observed) \* expected\_prob

# Chi-square test  
chisq.test(observed, p = expected\_prob)

##   
## Chi-squared test for given probabilities  
##   
## data: observed  
## X-squared = 20.329, df = 2, p-value = 3.852e-05

INTERPRETATION:

Here, p-value < 0.05. This indicates that H0 is rejected,i.e,the colors of the wild tulips are not equally common.

test=chisq.test(observed, p = expected\_prob)  
# Residuals  
residuals <- test$residuals  
residuals

## [1] 3.5620353 -0.9759001 -2.5861352

Part 2:

HYPOTHESIS TESTING:

H0:the ratio of red, yellow andwhite tulip is 4:3:1.

VS

H1: the ratio is not the same

# Given data  
observed <- c(71, 40, 29)  
  
# Expected probabilities assuming a 4:3:1 ratio  
expected\_prob <- c(4/8, 3/8, 1/8)

# Expected values  
expected <- sum(observed) \* expected\_prob  
expected

## [1] 70.0 52.5 17.5

# Chi-square test  
chisq.test(observed, p = expected\_prob)

##   
## Chi-squared test for given probabilities  
##   
## data: observed  
## X-squared = 10.548, df = 2, p-value = 0.005124

INTERPRETATION:

Here, p-value < 0.05. This indicates the rejection of H0.This means that the given ratio for the 3 colors is not applicable.

# Chi-square test  
chi\_square\_test <- chisq.test(observed, p = expected\_prob)  
# Residuals  
residuals <- chi\_square\_test$residuals  
residuals

## [1] 0.1195229 -1.7251639 2.7490258

Question -2: Avni is testing an octahedral die to see if it is biased. The observed results are given:

Observed:

7 10 11 9 12 10 14 7

Test the hypothesis that the die is fair.

HYPOTHESIS TESTING:

H0: DIE IS FAIR

VS

H1: DIE IS NOT FAIR

# Observed frequencies  
observed <- c(7, 10, 11, 9, 12, 10, 14, 7)  
  
# Expected probabilities assuming a fair die  
expected\_prob <- rep(1/8, 8)

# Expected values  
expected <- sum(observed) \* expected\_prob  
expected

## [1] 10 10 10 10 10 10 10 10

# Chi-square test  
chisq.test(observed, p = expected\_prob)

##   
## Chi-squared test for given probabilities  
##   
## data: observed  
## X-squared = 4, df = 7, p-value = 0.7798

INTERPRETATION:

The p-value > 0.05, which indicates the acceptance of H0. This denotes that the die is fair.

# Chi-square test  
chi\_square\_test <- chisq.test(observed, p = expected\_prob)  
# Residuals  
residuals <- chi\_square\_test$residuals  
residuals

## [1] -0.9486833 0.0000000 0.3162278 -0.3162278 0.6324555 0.0000000 1.2649111  
## [8] -0.9486833

Question- 3: When randomly selecting a card from a deck with replacement, are we equally likely to select a heart, diamond, spade, and club?I randomly selected a card from a standard deck 40 times with replacement. I pulled 13 hearts, 8 diamonds, 8 spades, and 11 clubs.

HYPOTHESIS TESTING:

H0:HEART,DIAMOND,SPADE,CLUB ARE EQUALLY LIKELY TO APPEAR

VS

H1: NOT EQUALLY LIKELY TO APPEAR

# Given data  
observed <- c(13, 8, 8, 11) # Observed frequencies for hearts, diamonds, spades, clubs  
  
# Expected probabilities assuming an equal distribution  
expected\_prob <- rep(1/4, 4)

# Expected values  
expected <- sum(observed) \* expected\_prob  
expected

## [1] 10 10 10 10

# Chi-square test  
chisq.test(observed, p = expected\_prob)

##   
## Chi-squared test for given probabilities  
##   
## data: observed  
## X-squared = 1.8, df = 3, p-value = 0.6149

INTERPRETATION:

The p-value > 0.05 which indicates the acceptance of H0. This means that the heart,spade,club and diamond is equally likely to appear at each trial.

# Chi-square test  
chi\_square\_test <- chisq.test(observed, p = expected\_prob)  
# Residuals  
residuals <- chi\_square\_test$residuals  
residuals

## [1] 0.9486833 -0.6324555 -0.6324555 0.3162278

Question 4:

A researcher is studying the distribution of the ages of customers in three different shopping malls in a city. The expected age distribution, based on census data, is as follows:

Mall A: 20% teenagers (ages 13-19), 50% adults (ages 20-40), 30% seniors (ages 41 and above). Mall B: 30% teenagers, 40% adults, 30% seniors. Mall C: 25% teenagers, 45% adults, 30% seniors. The researcher collects a random sample of 300 customers from each mall and records their age groups. The observed distribution is as follows:

Mall A: 60 teenagers, 180 adults, 60 seniors. Mall B: 90 teenagers, 100 adults, 110 seniors. Mall C: 70 teenagers, 120 adults, 110 seniors. Test the hypothesis that the observed age distribution in each mall is consistent with the expected distribution. Use a significance level of 0.05.

HYPOTHESIS TESTING:

H0: AGE DISTRIBUTION IN EACH MALL IS CONSISTENT

VS

H1: NOT CONSISTENT

# Given data  
observed\_mall\_A <- c(60, 180, 60)  
observed\_mall\_B <- c(90, 100, 110)  
observed\_mall\_C <- c(70, 120, 110)

# Expected proportions based on census data  
expected\_mall\_A <- c(0.20, 0.50, 0.30)  
expected\_mall\_B <- c(0.30, 0.40, 0.30)  
expected\_mall\_C <- c(0.25, 0.45, 0.30)

# Sample sizes  
sample\_size <- 300

# Expected frequencies  
expected\_freq\_mall\_A <- expected\_mall\_A \* sample\_size  
expected\_freq\_mall\_B <- expected\_mall\_B \* sample\_size  
expected\_freq\_mall\_C <- expected\_mall\_C \* sample\_size

# Combine observed and expected frequencies for each mall  
observed\_freqs <- rbind(observed\_mall\_A, observed\_mall\_B, observed\_mall\_C)  
expected\_freqs <- rbind(expected\_freq\_mall\_A, expected\_freq\_mall\_B, expected\_freq\_mall\_C)

chisq.test(observed\_freqs, p = expected\_freqs)

##   
## Pearson's Chi-squared test  
##   
## data: observed\_freqs  
## X-squared = 50.221, df = 4, p-value = 3.247e-10

INTERPRETATION:

The p-value is less than 0.05 which denotes the rejection of H0. This indicates that the age distribution in each mall is inconsistent.