

Stat/Math 415 Homework

Due Friday Dec 6; Joseph Sepich (jps6444)

1 Problem 9.1-3

Michigan Lottery digit drawing follows a multinomial distribution where p_i is the probability of drawing digit i . Problem constraints are below:

- $H_0: p_0 = p_1 = \dots = p_9 = 0.1$
- $\alpha = 0.05$

```
# Data
digits <- c(1, 6, 9, 9, 3, 8, 5, 0, 6, 7, 4, 7, 5, 9, 4, 6, 5, 6, 4, 4, 4, 8, 0,
           9, 3, 2, 1, 5, 4, 5, 7, 3, 2, 1, 4, 6, 7, 1, 3, 4, 4, 8, 8, 6, 1, 6,
           1, 2, 8, 8)
```

To test this hypothesis we can use a Chi-squared test statistic:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

First let's get the observed counts:

```
observed <- 0:9
for (i in 1:10) {
  observed[i] <- sum(digits == i-1)
  print(paste0(i-1, ' has a count of ', observed[i]))
}
```

```
## [1] "0 has a count of 2"
## [1] "1 has a count of 6"
## [1] "2 has a count of 3"
## [1] "3 has a count of 4"
## [1] "4 has a count of 9"
## [1] "5 has a count of 5"
## [1] "6 has a count of 7"
## [1] "7 has a count of 4"
## [1] "8 has a count of 6"
## [1] "9 has a count of 4"
```

Now for the expected counts:

$$E_i = n * \hat{p}_0$$

```
n <- 50
p <- 0.1

print(paste0('Every expected value is ', n * p))
```

```
## [1] "Every expected value is 5"
```

```
# Test statistic
E <- n * p
print(paste0('The value of the test statistic is ', sum((observed - E)^2 / E)))
```

```
## [1] "The value of the test statistic is 7.6"
```

Now if we look in the chi-squared table for $c - 1 = 9$ degrees of freedom, we get the critical region to be $\chi^2 > \chi^2(c - 1)_\alpha = \chi^2(9)_{0.05} = 16.92$. Since our test statistic of 7.6 is less than the critical value of 16.92 we do **not** reject the null hypothesis.

2 Problem 9.1-7

Problem constraints

- $n = 150$
- $c = 5$
- $\bar{x} = 0.5$
- $\alpha = 0.01$

We want to test if X has a poisson distribution. Let us first create our data.

```
observed <- c(0, 1, 2, 3, 4) # 5 groups
X <- c(92, 46, 8, 3, 1) # number of dishes
```

To test this against a poisson distribution, we must get the expected values using the pmf. We can find the expected number by finding the probability of that number occurring and multiply by the total experiments n .

$$E_i = p_i * n$$
$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

```
E <- 0:4 # 5 groups
lambda <- 0.5
n <- 150

for (i in 0:4) {
  prob <- ((lambda ^ i) * exp(-lambda)) / factorial(i)
  E[i+1] <- prob * n
  print(paste0('Expected value of ', i, ' colonies is ', E[i+1], ' prob of ', prob))
}
```

```
## [1] "Expected value of 0 colonies is 90.979598956895 prob of 0.606530659712633"
## [1] "Expected value of 1 colonies is 45.4897994784475 prob of 0.303265329856317"
## [1] "Expected value of 2 colonies is 11.3724498696119 prob of 0.0758163324640792"
## [1] "Expected value of 3 colonies is 1.89540831160198 prob of 0.0126360554106799"
## [1] "Expected value of 4 colonies is 0.236926038950247 prob of 0.00157950692633498"
```

We can test this using our chi-squared test statistic:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Let's combine the last three groups to have a reasonable frequency

```
# Test statistic
```

```
X <- c(X[1], X[2], sum(X[3:5]))
```

```
E <- c(E[1], E[2], sum(E[3:5]))
```

```
paste0('O: ', X)
```

```
## [1] "O: 92" "O: 46" "O: 12"
```

```
paste0('E: ', E)
```

```
## [1] "E: 90.979598956895" "E: 45.4897994784475" "E: 13.5047842201641"
```

```
print(paste0('The value of the test statistic is ', sum((X - E)^2 / E)))
```

```
## [1] "The value of the test statistic is 0.184838888809246"
```

Now if we look in the chi-squared table for $k - 1 - d = 1$ degrees of freedom, we get the critical region to be $\chi^2 > \chi^2(k - 1 - d)_\alpha = \chi^2(1)_{0.01} = 6.635$. Since our test statistic of 0.185 is less than the critical value of 6.635 we do **not** reject the null hypothesis.

3 Problem 9.2-3

We want to use a chi-squared test with $\alpha = 0.05$ to perform a homogeneity test on two Classes and the distribution of their test scores. In this case it would also be an independence test. We will combine the given sample into three parts: low, middle, high. This gives us the hypothesis:

$$H_0: p_L = p_M = p_H$$

Using 55 and 75 as our cutoff points for ranges of scores we get the following table.

Class	L	M	H	Totals
U	7	4	4	15
V	3	6	6	15

These are our observed values. Now we must get our expected values. To do this we recognize that $\hat{p}_j = \frac{1}{3}$, since each test group (L, M, H) contains 10 of 30 total samples. We then use the number of samples in each group to find the expected values as

$$n_i * p_j = E_{ij}$$

. This yields our test statistic:

$$\chi^2 = \sum_{i=1}^h \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

```
n <- 15 # sample size of each group
p <- 1 / 3
E <- n * p
L <- c(7, 3)
M <- c(4, 6)
H <- c(4, 6)

observed <- data.frame(L = L, M = M, H = H)
chi <- 0
for (i in 1:2) {
  for (j in 1:3) {
    chi <- chi + (observed[i, j] - E)^2 / E
  }
}
# Test statistic
print(paste0('The value of the test statistic is ', chi))
```

```
## [1] "The value of the test statistic is 2.4"
```

Now if we look in the chi-squared table for $(h-1)(k-1) = (2-1)(3-1) = 2$ degrees of freedom, we get the critical region to be $\chi^2 > \chi^2((h-1)(k-1))_\alpha = \chi^2(2)_{0.05} = 5.99$. Since our test statistic of 2.4 is less than the critical value of 5.99 we do **not** reject the null hypothesis.

4 Problem 9.2-9

We again use the same chi-squared test that we used in the last problem to test the independence/homogeneity of the two variables. This time we test whether those who had older brothers participated in sports at the same rate as those without.

$$H_0: p_S = p_{NS}$$

The observed values for this case are in the table. To find the expected value we can merely multiply the row and column totals and divide by the sample size:

```
Y <- c(12, 13, 25)
N <- c(8, 27, 35)
Tot <- c(20, 40, 60)

observed <- data.frame(Y = Y, N = N, Tot = Tot)

E_1 <- c(observed[1,3] * observed[3,1] / observed[3,3],
         observed[2,3] * observed[3,1] / observed[3,3])
E_2 <- c(observed[1,3] * observed[3,2] / observed[3,3],
         observed[2,3] * observed[3,2] / observed[3,3])

E <- data.frame(Yes = E_1, No = E_2)

print('Expected values')
```

```
## [1] "Expected values"
```

```
E
```

```
##           Yes           No
## 1  8.333333 11.66667
## 2 16.66667 23.33333
```

This yields our test statistic:

$$\chi^2 = \sum_{i=1}^h \sum_{j=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

```
chi <- 0
for (i in 1:2) {
  for (j in 1:2) {
    chi <- chi + (observed[i, j] - E[i,j])^2 / E[i,j]
  }
}
# Test statistic
print(paste0('The value of the test statistic is ', chi))
```

```
## [1] "The value of the test statistic is 4.14857142857143"
```

Now if we look in the chi-squared table for $(h-1)(k-1) = (2-1)(2-1) = 1$ degrees of freedom, we get the critical region to be $\chi^2 > \chi^2_{((h-1)(k-1))\alpha} = \chi^2_{(1)0.05} = 3.841$. Since our test statistic of 4.15 is greater than the critical value of 3.841 we **do** reject the null hypothesis.

Since we would not reject for the value in the chi-square table at a significance level of 0.025, we can say our p-value is between 0.005 and 0.025.