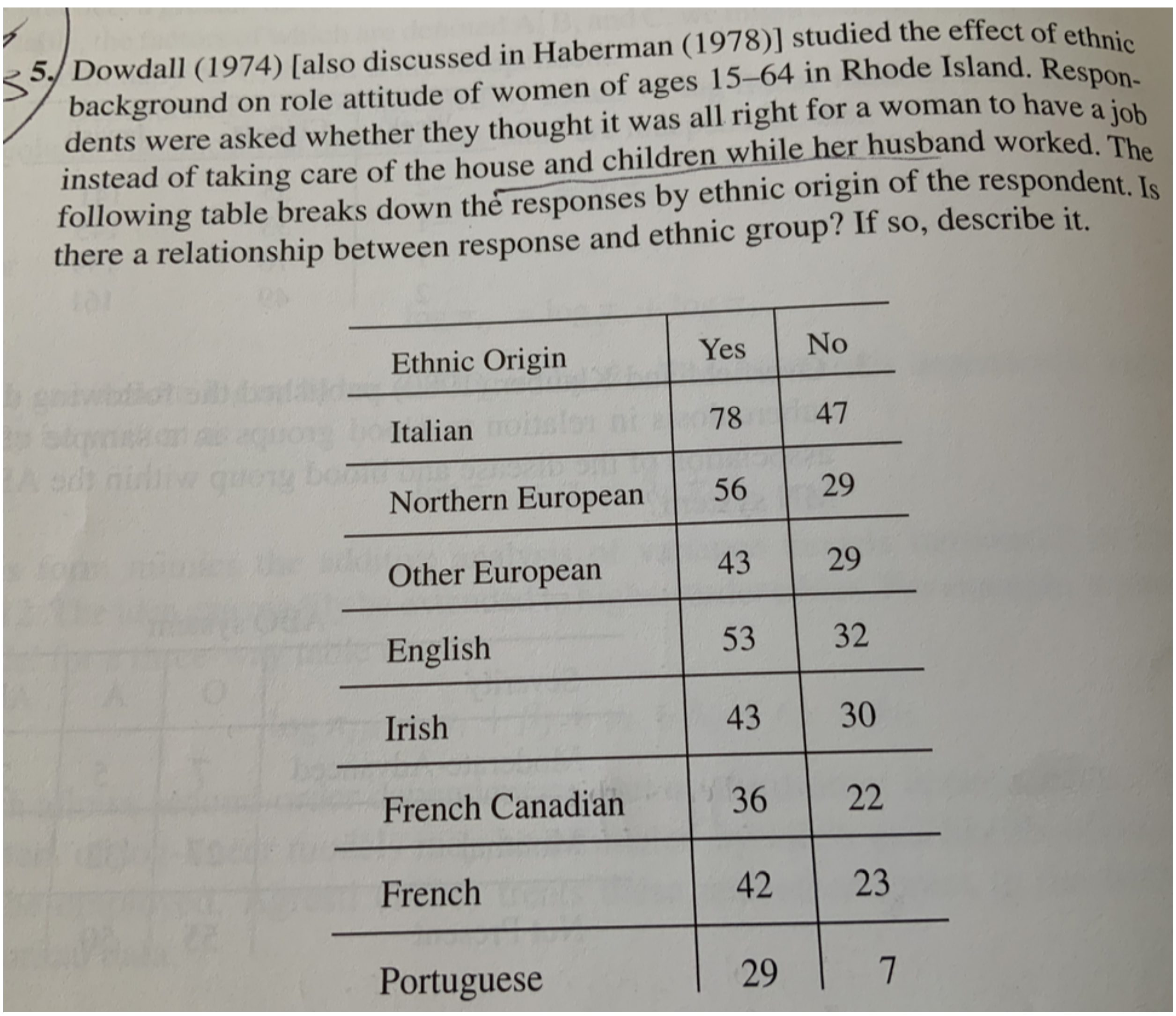
stats100b_hw7

Takao

5/30/2022

1. Test of independence

```r
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.56.11 PM.png")
```



5. Dowdall (1974) [also discussed in Haberman (1978)] studied the effect of ethnic background on role attitude of women of ages 15–64 in Rhode Island. Respondents were asked whether they thought it was all right for a woman to have a job instead of taking care of the house and children while her husband worked. The following table breaks down the responses by ethnic origin of the respondent. Is there a relationship between response and ethnic group? If so, describe it.

Ethnic Origin	Yes	No
Italian	78	47
Northern European	56	29
Other European	43	29
English	53	32
Irish	43	30
French Canadian	36	22
French	42	23
Portuguese	29	7

For $i = 1, \dots, 8$ ethnic origin, and $j = 1,2$ response,

$$H_0 : \pi_{ij} = \pi_{i.} \times \pi_{.j} \text{ (independence)}$$

$$H_1 : \pi_{ij} \neq \pi_{i.} \times \pi_{.j} \text{ for some i,j pair}$$

Generating the function, we have that

```
Q1 <- function(data){
  holder <- data;chi_stats <- 0
  Tot <- sum(data);row_sum <- rowSums(data);col_sum <- colSums(data)
  for(i in 1:dim(data)[1]){
    for(j in 1:dim(data)[2]){
      holder[i, j] <- row_sum[i]*col_sum[j]/Tot
      chi_stats <- chi_stats + (data[i,j] - holder[i,j])^2/holder[i,j]
    }
  }
  df <- (dim(data)[1] - 1)*(dim(data)[2] - 1)
  cat("The test statistics = ", chi_stats, "\n")
  cat("df = ", df, "\n")
  cat("P-value = ", pchisq(chi_stats, df, lower.tail = FALSE))
}
```

```
data <- matrix(c(78,56,43,53,43,36,42,29,47,29,29,32,30,22,23,7), ncol = 2, nrow = 8, byrow = FALSE)
Q1(data)
```

```
## The test statistics =  6.029392
## df =  7
## P-value =  0.5363217
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis

2. Test of Independence

```
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.56.24 PM.png")
```

2. Grades in an elementary statistics class were classified by the students' majors. Test to see if grade and major are independent
(this is problem 7, Rice, page 532-533. Data attached below)

13.8 Problems 533

7.

Grade	Major: Psychology	Biology	Other
A	8	15	13
B	14	19	15
C	15	4	7
D-F	3	1	4

8. A randomized double-blind experiment compared the effectiveness of several drugs in ameliorating postoperative nausea. All patients were anesthetized with nitrous oxide and ether. The following table shows the incidence of nausea during the first four postoperative hours for each of several drugs and a placebo (Beecher 1959). Compare the drugs to each other and to the placebo.

	Number of Patients	Incidence of Nausea
Placebo	165	95
Chlorpromazine	152	52
Dimenhydrinate	85	52
Pentobarbital (100 mg)	67	35
Pentobarbital (150 mg)	85	37

For i = 1,2,3,4 and j = 1,2,3

$$H_0 : \pi_{ij} = \pi_{i.} \times \pi_{.j} \text{ (independence)}$$

$$H_1 : \pi_{ij} \neq \pi_{i.} \times \pi_{.j} \text{ for some i,j pair}$$

```
data <- matrix(c(8,15,13,14,19,15,15,4,7,3,1,4), ncol = 3, nrow = 4, byrow = TRUE)
Q1(data)
```

```
## The test statistics =  12.18346
## df =  6
## P-value =  0.05799901
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis

3. Test of Homogeneity

```r
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.56.29 PM.png")
```

3.Use data from problem 8 Rice, page 533. Ignore the data for placebo. Conduct a test to compare the drugs to each other. [hint : what is the sample size for each drug ?]

For j = 1,2,3,4 groups (different drugs) and k = 1,2 categories (nausea, no nausea),

$$H_0 : \pi_{k1} = \pi_{k2} = \pi_{k3} = \pi_{k4}$$
$$H_1 : \pi_{kj} \neq \pi_{kj'} \text{ for some } j \neq j'$$

All the samples are from the same categorical distribution

```r
data <- matrix(c(152-52, 52, 85 - 52, 52, 67 - 35, 35, 85 - 37, 37), nrow = 4, ncol = 2, byrow = TRUE)
Q1(data)
```

```
## The test statistics =  17.60278
## df =  3
## P-value =  0.000531108
```

Since the p-value is less than 0.05, we have significant evidence to reject the null hypothesis

4. Test of homogeneity

```r
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.56.33 PM.png")
```

4. Do Problem 2, Rice, page 530-531. Problem and data are copied below.
Test the null hypothesis that the mortality patterns are the same between Chinese and Jewish.

Week	Chinese	Jewish	
-2	55	141	
-1	33	145	
1	70	139	
2	49	161	

Phillips and Smith (1990) conducted a study to investigate whether people could briefly postpone their deaths until the occurrence of a significance occasion. The senior woman of the household plays a central ceremonial role in the Chinese Harvest Moon Festival. Phillips and Smith compared mortality patterns of old Jewish women and old Chinese women who dies of natural caused for the weeks immediately preceding and following the festival, using records from California for the years 1960-1984. Week -1 is the week preceding the festival, week 1 is the week following, etc.

For j = 1,2,3,4 groups (-2,-1, 1, 2 week) and k = 1,2 categories (Chinese, Jewish),

$$H_0 : \pi_{k1} = \pi_{k2} = \pi_{k3} = \pi_{k4}$$
$$H_1 : \pi_{kj} \neq \pi_{kj'} \text{ for some } j \neq j'$$

```
data <- matrix(c(55,141,33,145,70,139,49,161), nrow = 4, ncol = 2, byrow = TRUE)
Q1(data)
```

```
## The test statistics =  12.42079
## df =  3
## P-value =  0.006072305
```

Since the p-value is less than 0.05, we have significant evidence to reject the null hypothesis.

5

```
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.56.38 PM.png")
```

5. Do Problem 6 , Rice, page 532. Problem and data are copied below. Formulation of hypothesis testing is the key . There are two questions asked in this problem. A different formulation may be needed for answering the second question.

	Offspring is female	Offspring is male
Flying Fighters	51	38
Flying Transports	14	16
Not flying	38	46

It is conventional wisdom in military squadrons that pilots tend to father more girls than boys. Snyder(1961) gathered data for military fighter pilots. The sex of the pilots' offspring were tabulated for three kinds of flight duty during the month of conception.
(a) Is there any significant difference between three groups?
(b) In the United States in year 1950, 105.37 males were born for every 100 females. Are the data consistent with this population sex ratio ?

a. Test of Independence

For i = 1,2,3 (3 flight duties) and j = 1,2 (female, male),

$$H_0 : p_{ij} = p_{i.} \times p_{.j}$$

$$H_1 : p_{ij} \neq p_{i.} \times p_{.j} \text{ for some i,j pair}$$

```
data <- matrix(c(51,38,14,16,38,46), nrow = 3, ncol = 2, byrow = TRUE)
Q1(data)
```

```
## The test statistics =  2.75038
## df =  2
## P-value =  0.2527915
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis

b. Simple Pearson's Chi-Square Test

$$H_0 : p_{female} = \frac{100}{100 + 105.73} p_{male} = \frac{105.73}{100 + 105.73}$$

$$H_1 : p_{female} \neq \frac{100}{100 + 105.73} p_{male} \neq \frac{105.73}{100 + 105.73}$$

```
obs <- colSums(data);
exp <- c(100/(100 + 105.37) * sum(data) , 105.37 / (100 + 105.37) * sum(data))
chi_stats <- sum((obs - exp)^2 / exp)
df <- 1
cat("The test statistics = ", chi_stats, "\n")
```

```
## The test statistics =  0.3402491
```

```r
cat("df = ", df, "\n")
```

```
## df =  1
```

```r
cat("p-value = ", pchisq(chi_stats, df, lower.tail = FALSE))
```

```
## p-value =  0.5596854
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis

6. Test of Homogeneity

```r
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.56.41 PM.png")
```

6. Conduct an appropriate test to compare the literal style pattern between Sanditon I and Sanditon II. State the null hypothesis and the alternative hypothesis. Use $\alpha = 5\,\%$.

For j = 1,2 groups (Sanditon I, Sandition II) and k = 1,2,3,4,5,6 categories (a, an, this, that, with, without),

$$H_0 : \pi_{k1} = \pi_{k2}$$
$$H_1 : \pi_{kj} \neq \pi_{kj'} \text{ for some } j \neq j'$$

```r
data <- matrix(c(101, 83, 11, 29, 15, 15, 37, 22, 28, 43, 10, 4), nrow = 6, ncol = 2, byrow = TRUE)
Q1(data)
```

```
## The test statistics =  19.32881
## df =  5
## P-value =  0.00166902
```

Since the p-value is less than 0.05, we reject the null hypothesis

7. McNemar Test

```r
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.57.04 PM.png")
```

7.Do Problem 23 , Rice, page 538-539. Problem and data are copied below.

Does heavy exercise increase the risk of myocardial infarction ? Mittleman et al. (1993) studies this question by examining the activities of 1228 patients who had suffered myocardial infarctions. It was determined whether each patient has participated in heavy exertion in the hour before the onset of the infarction and also whether each had participated in heavy exertion at the same time the previous day. The results are given in this table. Does the study demonstrate that heavy exertion is associated with myocardial infarction? Conduct a suitable hypothesis testing.

	The day of myocardial infarction	
Previous Day	Heavy Exertion	No heavy Exertion
Heavy Exertion	4	9
No heavy Exertion	50	1165

Let p1. be the probability of Heavy Exertion (previous day), and p.1 be the probability of Heavy Exertion (the day of myocardial infarction)

$$H_0 : p_{1.} = p_{.1}$$

Defining the Q2 function, we have that

```r
Q2 <- function(data){
  chi_stats <- (data[1,2] - data[2,1])^2 / (data[1,2] + data[2,1])
  df = 1
  cat("test statistics = ", chi_stats, "\n")
  cat("df = ", df, "\n")
  cat("p-value = ", pchisq(chi_stats, df, lower.tail = FALSE))
}
```

```r
data <- matrix(c(4, 9, 50, 1165), nrow = 2, byrow = TRUE)
Q2(data)
```

```
## test statistics =  28.49153
## df =  1
## p-value =  9.410951e-08
```

Since the p-value is less than 0.05, we have significant evidence to reject the null hypothesis

8. McNemar Test

```r
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.57.08 PM.png")
```

8. Copied from Problem 26 on page 540, Rice.

.

Insulin pumps are used by diabetic patients to control blood glucose levels, but a side effect, diabetic ketoacidosis (DKA), may occur. Mecklenburg et al(1984) gathered data on incidence of DKA before and after pump therapy. Test whether the rate of DKA remains the same after therapy. Formulate the problem and use an appropriate test. Find the p-value.

	Before Therapy	
After Therapy	No DKA	DKA
No DKA	128	7
DKA	19	7

Let p1. be the probability of no DKA after therapy, and p.1 be the probability of no DKA before therapy.

$$H_0 : p_{1.} = p_{.1}$$

```r
data <- matrix(c(128, 7, 19, 7), nrow = 2, byrow = TRUE)
Q2(data)
```

```
## test statistics =  5.538462
## df =  1
## p-value =  0.01860293
```

Since the p-value is less than 0.05, we have significant evidence to reject the null hypothesis

9

```r
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.57.13 PM.png")
```

9. In a study to identify factors affecting physicians' decisions to advise or not to advise patients to stop smoking, the following two tables were given (data from problem 22, Rice, page 538). All patients in the study cohort are smokers. Some are advised to quit smoking; the others are not advised to quit.

	Advised	Not advised
Male physician	78	94
Female physician	50	89

	Advised	Not advised
Smoker physician	13	37
Nonsmoker physician	115	146

(i) What proportion of the male-physician's patients were advised to quit smoking? What proportion of the female-physician's patients were advised to quit smoking ?
(ii) Conduct a test to see if the physician's gender is independent of the physician's decision.
(iii) Conduct a test to see if the physician's smoking status is independent of the physician's decision.

i

```
FM <- matrix(c(78, 94, 50, 89), nrow = 2, byrow = TRUE)
SN <- matrix(c(13, 37, 115, 146), nrow = 2, byrow = TRUE)
cat("proportion of male physician who advised patients to quit smoking = ", FM[1,1] / rowSums(FM)[1])
```

```
## proportion of male physician who advised patients to quit smoking =  0.4534884
```

```
cat("proportion of female physician who advised patients to quit smoking", FM[2,1]/rowSums(FM)[2])
```

```
## proportion of female physician who advised patients to quit smoking 0.3597122
```

ii. Test of independence

For i = 1,2 (male, female) and j = 1,2 (advised, not advised),

$$H_0 : p_{ij} = p_{i.} \times p_{.j}$$

$$H_1 : p_{ij} \neq p_{i.} \times p_{.j} \text{ for some i,j pair}$$

```
Q1(FM)
```

```
## The test statistics =  2.791434
## df =  1
## P-value =  0.09476941
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis

iii. Test of independence

For i = 1,2 (smoker, non-smoker) and j = 1,2 (advised, not advised),

$$H_0 : p_{ij} = p_{i.} \times p_{.j}$$

$$H_1 : p_{ij} \neq p_{i.} \times p_{.j} \text{ for some i,j pair}$$

```
Q1(SN)
```

```
## The test statistics =  5.652076
## df =  1
## P-value =  0.01743472
```

Since the p-value is less than 0.05, we have significant evidence to reject the null hypothesis

10

```
knitr::include_graphics("/Users/takaooba/Desktop/Screen Shot 2022-11-28 at 4.57.18 PM.png")
```

10. (continued from Problem 9).

(i) Find the odds ratio for comparing $odds(advised|male)$ with $odds(advised|female)$

(ii) Find the odds ratio for comparing $odds(advised|smoker)$ with $odds(advised|nonsmoker)$

i

$$\frac{odds(advised|male)}{odds(advised|female)} = \frac{P(advised|male)}{1 - P(advised|male)} / \frac{P(advised|female)}{1 - P(advised|female)}$$

```
r <- ((FM[1,1] / rowSums(FM)[1]) / (1-FM[1,1] / rowSums(FM)[1])) / ((FM[2,1] / rowSums(FM)[2]) / (1-FM[2
cat("odds ratio = ", r)
```

```
## odds ratio =  1.477021
```

ii

$$\frac{odds(advised|smoker)}{odds(advised|nonsmoker)} = \frac{P(advised|smoker)}{1 - P(advised|smoker)} / \frac{P(advised|nonsmoker)}{1 - P(advised|nonsmoker)}$$

```
r <- ((SN[1,1] / rowSums(SN)[1]) / (1-SN[1,1] / rowSums(SN)[1])) / ((SN[2,1] / rowSums(SN)[2]) / (1-SN[2
cat("odds ratio = ", r)
```

```
## odds ratio =  0.4460635
```