STA 138 Discussion 7 – solutions

Fall 2020

	$\frac{2}{15}$	$\frac{13}{15}$		
	C	D		
A	16	110	126	$\frac{63}{75}$
B	4	20	24	$\frac{4}{25}$
	20	130	150	

1. In the table above, you can find the results of an experiment by Kahneman and Tversky on peoples' risk preferences. The experiment consisted of eliciting choices by people for two decisions:

   (i) Choose between:

   E_1 $\frac{252}{15}$ $\frac{1638}{15}$

   $\frac{48}{15}$ $\frac{312}{15}$

      A. a sure gain of \$240

      B. 25% chance for \$1000 gain, 75% chance for \$0 gain

   (ii) Choose between:

      C. a sure loss of \$750

      D. 75% chance to lose \$1000, 25% chance to lose \$0

   When people choose B over A, we take them to be "risk loving" in choice (i); similarly, we take people who choose D over C to be "risk loving" in choice (ii). For the moment you can assume that our large sample approximation holds for this contingency table.

   (a) What do you estimate to be the odds ratio of choosing B for D vs. C? What does this tell you about risk preferences between the two choices?

   The odds ratio here is
   $$\frac{20/110}{4/16} \approx 0.73 \; .$$
   From this we can conclude that, among people sampled, people are *less* likely to be risk loving in choice (i) if they are risk loving in (ii) than otherwise.

   (b) Can you, using Pearson's χ^2 test of independence, conclude that risk lovingness for (i) has an "effect" on risk livingess for (ii), at significance level $\alpha = 0.1$?

   We have here a Pearson χ^2 test of independence with statistic

   H_0: $I-1+J-1$

   H_A: $IJ-1$ $\quad - \quad (I-1)(J-1)$

   $df = (2-1)(2-1)$
   $$\sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \approx 0.2747$$
   From the χ^2_1 null distribution, then, we get p-value 0.6002, and fail to reject the null hypothesis of independence at $\alpha = 0.1$. So, we conclude that this is not sufficient evidence to conclude definitively that the population from which these people are sampled has the risk preferences obtained here.

   (c) Can you, using a likelihood ratio test, conclude that risk lovingness for (i) has an "effect" on risk livingess for (ii), at significance level $\alpha = 0.1$?

   In this case we have a L-R test of independence with statistic
   $$\Lambda = \frac{\max L(H_0)}{\max L(H_A)} \qquad -2\log\Lambda = -2\sum_{ij} O_{ij} \log\left({}^{E_{ij}}\!/_{O_{ij}}\right) \approx 0.2609$$
   From the χ^2_1 null distribution, then, we get p-value 0.6095, and fail to reject the null hypothesis of independence at $\alpha = 0.1$. So, we conclude that this is not sufficient evidence to conclude definitively that the population from which these people are sampled has the risk preferences obtained here.

I:	<2000 B.C.	V:	1200-1500 A.D.
II:	2000-500 B.C.	VI:	1500-1700 A.D.
III:	500 B.C.-500 A.D.	VII:	1700-1850 A.D.
IV:	500-1200 A.D.	VIII:	>1850 A.D.

Table 1: Time periods

	I	II	III	IV	V	VI	VII	VIII
right	79	222	169	97	107	134	133	151
left	11	19	7	8	11	9	7	16

Table 2: Contingency table

2. A study counted the numbers of depictions of right- and left- handed people in artwork (Table 2) over different time periods (Table 1). Assuming that the artworks were randomly sampled, can you conclude that the relative frequencies of left- and right- handedness change over time?

   If we choose to use Pearson's χ^2 test here, we will get

$$\sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \approx 9.3079$$

(2-1)×(8-1)

   From the χ^2_1 null distribution, then, we get p-value 0.2313. Not given a significance level, we choose one: let's say 0.01. With this significance level, we fail to reject H_0, and conclude that there is not sufficient evidence to conclude definitively that there is an "effect" of time period on depictions of handedness.

   If instead we used a LR test here, we would get statistic 9.4917 and p-value 0.2193, yielding the same result.

Appendix: R Script

```
O <- matrix(c(16,110,
              4,20),
            byrow=TRUE,
            ncol=2) # observed
E <- rowSums(O)%*%t(colSums(O))/sum(O) # expected
pearsonStatistic <- sum((O-E)^2/E)
pearsonpVal <- 1-pchisq(pearsonStatistic,1) # df = 1 !
LRstatistic <- -2*sum(O*log(E/O))
LRpVal <- 1-pchisq(LRstatistic,1)

counts <- matrix(c(79, 11,
                   222, 19,
                   169, 7,
                   97, 8,
                   107, 11,
                   134, 9,
                   133, 7,
                   151, 16),
                 nrow=2)
timePeriods <- c("<2000 B.C.",
                      "2000-500 B.C.",
                      "500 B.C.-500 A.D.",
                      "500-1200 A.D.",
                      "1200-1500 A.D.",
                      "1500-1700 A.D.",
                      "1700-1850 A.D.",
                      ">1850 A.D.")
colnames(counts) <- c(as.roman(1:8))
rownames(counts) <- c("right","left")
O <- counts # observed
E <- rowSums(O)%*%t(colSums(O))/sum(O) # expected
pearsonStatistic <- sum((O-E)^2/E)
pearsonpVal <- 1-pchisq(pearsonStatistic,7) # df = 7 !
LRstatistic <- -2*sum(O*log(E/O))
LRpVal <- 1-pchisq(LRstatistic,7)
```