Special Topics Comp Stat & Pro MAT5999 and Computational Stats & Prob. AIM 5002
Written Assignment 8 (25 points)
4/4/22

Solutions to be returned by the beginning of class on Monday, 4/11.

1. Consider the accompanying 2×3 table displaying the sample proportions that fell in the various combinations of categories (e.g., 13% of those in the sample were in the first category of both factors).

.13	.19	.28
.07	.11	.22

   (a) **(5 points)** Suppose that the sample consisted of $n = 100$ people. Use the chi-squared test for independence with significance level .10.

   (b) **(2 points)** Repeat part (a) assuming that the sample size was $n = 1000$.

   (c) **(3 points)** What is the smallest sample size n for which these observed proportions would result in rejection of the independence hypothesis?

2. **(5 points)** You can choose to solve this problem either on paper or in R (in the latter case you probably first have to see the documentation of **chisq.test** by typing **help(chisq.test)** as we only used goodness-of-fit test in class, and did not discuss the contingency tables).

   A random sample of smokers was obtained, and each individual was classified both with respect to gender and with respect to the age at which he/she first started smoking. The data in the accompanying table is consistent with summary results reported in the article "Cigarette Tar Yields in Relation to Mortality in the Cancer Prevention Study II Prospective Cohort" (British Med. J., 2004: 72-79).

	Males	Females
Age $<$16	25	10
Age 16 - 17	24	32
Age 18 - 20	28	17
Age $>$ 16	19	34

   Carry out a test of hypothesis to decide whether there is an association between the two factors.

3. **(5 points)** The article "The Gap Between Wine Expert Ratings andConsumer Preferences" (Intl. J. of Wine Business Res.,2008: 335-351) studied differences between expert and consumer ratings by considering medal ratings for wines, which could be gold (G), silver (S), or bronze (B). Three categories were then established: 1. Rating is the same [(G,G), (B,B), (S,S)]; 2. Rating differs by one medal [(G,S), (S,G), (S,B), (B,S)]; and 3. Rating differs by two medals [(G,B), (B,G)]. The observed frequencies for these three categories were 69, 102, and 45, respectively. On

the hypothesis of equally likely expert ratings and consumer ratings being assigned completely by chance, each of the nine medal pairs has probability 1/9. Carry out an appropriate chi-squared test using a significance level of .10 by first obtaining P-value information.

4. (**you can get 5 points for any of the parts (a), (b) or (c). It is enough to solve one of the 3 parts, whichever you prefer.**) Suppose that $X_1, ..., X_m$ is a random sample from exponential distribution with parameter λ_1 and $Y_1...Y_n$ is a random sample from exponential distribution with parameter λ_2 where X_i is also independent from Y_j for every i and j. We have **Theorem: The ratio** $F = \frac{\lambda_2 \bar{Y}}{\lambda_1 \bar{X}}$ **has F distribution with 2n numerator degrees of freedom and 2m denominator degrees of freedom.**

   (a) Prove the above theorem. Hint: you can use the results of the extra problem on HW3.

   To test $H_0 : \lambda_1 = \lambda_2$ versus $h_a : \lambda_1 > \lambda_2$ at level α, the decision rule is: reject H_0 if $\bar{y}/\bar{x} \geq F_{\alpha,2n,2m}$.

   (b) Find a formula for the P-value of the above test. Your answer should be a function of $\Psi(\bar{y}/\bar{x})$, where Ψ is the cdf of the F distribution with 2n numerator degrees of freedom and 2m denominator degrees of freedom.

   (c) Find the value of the type II error probability $\beta(\lambda')$.Here we are assuming that H_0 is false so that $\lambda_1/\lambda_2 = \lambda' > 1$. Your answer should be of the form $\Psi(c/\lambda')$ where c is a constant that you have to find.