BST 140.652 Final Exam

Notes:

- You may use your one 8.5 by 11 formula sheet.
- No calculator should be necessary for this exam.
- Show your work on all questions. Simple "yes" or "no" answers will be graded as if blank.
- Please be neat and write legibly. Use the back of the pages if necessary.
- Good luck!

____

Printed name

1. The relative risk comparing the probability of success of a headache medicine to that of a placebo is believed to be confounded by age, which was categorized into three categories. Let $\hat{p}_{ij}$ for $i = 1, 2$ and $j = 1, 2, 3$ and n_{ij} be the sample proportion and sample sizes for treatment i (1 = headache medicine, 2 = placebo) and age j (1 = below 50, 2 = between 50 and 70, 3 = over seventy). Assume that the n_{ij} are all large. Propose an estimator for the assumed common log relative risk across age categories.

$$\hat{RR}_j = \frac{\hat{p}_{1j}}{\hat{p}_{2j}}$$

Let $\hat{\sigma}_j^2$ be the δ-method SE for the log of the RR. That is

$$\mathrm{Var}(\hat{RR}_j) \approx \hat{\sigma}_j^2$$

Proposed estimator of the log RR

$$\frac{\sum \log(\hat{RR}_j) \frac{1}{\hat{\sigma}_j^2}}{\sum \frac{1}{\hat{\sigma}_j^2}}$$

2. Researchers conducted a blind taste test of Coke versus Pepsi. Each of four people was asked which of two blinded drinks given in random order that they preferred. The data was such that 3 of the 4 people chose Coke. Assuming that this sample is representative, test the hypothesis that Coke is preferred to Pepsi. Symbolically state relevant hypothesis defining any notation that you use. Set up the calculations for the relevant P-value, but do not solve for the final number. Assume that the P-value is $.3125$ and interpret your results in the language of the problem.

Let P be the pop. portion of people who prefer coke. Then we want to test

$H_0: p = .5 \quad H_a: p > .5$

Let X = # out of 4 that prefer Coke

Assume $X \sim Bin(4, p)$

$$\text{Pvalue} = P(X \geq 3 \mid p = .5) = \binom{4}{3}.5^3 .5^1 + \binom{4}{4}.5^4 .5^0$$

Assume Pvalue = .3125 then at an $\alpha = .05$ Type I error rate there is insufficient evidence to conclude that more people prefer Coke. Note the small sample size.

3. At a party a friend brags that they can tell the difference between cheap and expensive wine. You ask him to taste eight glasses wine for which you have randomized as 4 expensive and 4 cheap. Set up the calculations for a P-vlaue for the relevant hypothesis test. Do not solve for a final number. Assume that the P-value is .2429, interpret the results in the context of the problem.

	Actual wine		
Friend's guess	Expensive	Cheap	
Expensive	3	1	4
Cheap	1	3	4
	4	4	

H_0: Actual wine status is indep of friend's guess

As or more extreme tables $\begin{matrix} 3 & 1 \\ 1 & 3 \end{matrix}$ $\quad \begin{matrix} 4 & 0 \\ 0 & 4 \end{matrix}$

$$\text{P value} = \frac{\binom{4}{3}\binom{4}{1}}{\binom{8}{4}} + \frac{\binom{4}{4}\binom{4}{0}}{\binom{8}{4}}$$

Assume Pvalue = .2429 then there is insufficient evidence to reject the null hypothesis. That is, this number of correct guesses or more is consistent with what could be seen by chance.

Note again, the sample size is small making it difficult to truly evaluate his claim.

4. Let $\bar{X}$ be the sample mean from n iid observations from a population with mean μ and variance σ^2. Let s be the sample variance. Using this notation, derive a confidence interval for $\mu^{1/3}$.

Use the delta method

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1) \qquad f = \mu^{1/3} \qquad f' = \frac{1}{3}\mu^{-2/3}$$

δ-method std. error est $f'(\bar{x})\hat{SE}_{\bar{x}} = \frac{1}{3\bar{x}^{2/3}}\frac{s}{\sqrt{n}}$

Hence $1-\alpha$ CI $(\alpha \in [0,1])$

$$(\bar{x})^{1/3} \pm z_{1-\frac{\alpha}{2}}\left(\frac{1}{3\bar{x}^{2/3}}\frac{s}{\sqrt{n}}\right)$$

5. Consider the table below comparing the self reported job stress levels of three occupations. 100 people from each of three occupations were surveyed.

Occupation	High stress	Low stress	Total
Executive	65	35	100
Professor	70	30	100
Lion tamer	15	85	100
Total	150	150	300

Researchers are interested in whether or not the perception of job stress differs by occupation. Symbolically relevant hypotheses, defining the notation used. Set up calculations for the relevant chi squared statistic, but do not solve for a final number. Assume that the Chi squared test statistic is 74, interpret this in the context of the problem.

E_{ij} = expected count in row i column j

$$= \frac{n_{+j}}{n} * n_{i+} = \frac{100 * 150}{300} \text{ for all cells} = 50$$

$$\chi^2 \text{ stat} = \frac{(65-50)^2}{50} + \frac{(35-50)^2}{50} + \frac{(70-50)^2}{50} + \frac{(30-50)^2}{50} + \frac{(15-50)^2}{50} + \frac{(85-50)^2}{50} = \text{Assumed } 74 \quad df = 2$$

Note $P(\chi^2_2 \geq 74) < .012$

So we reject H_0 and conclude that there is evidence to suggest that stress and job type are related. In this case, perhaps surprisingly, lion tamers have a very low stress job relative to the other two professions.

6. Consider the density given by $f(x) = 2x$ for $0 \leq x \leq 1$. A friend has some data that she claims was drawn from this density. You would like to test that hypothesis. The data was summarized via this frequency counts given below:

	Range of values for x				
	$\left[0,\ \sqrt{\frac{1}{4}}\right]$	$\left(\sqrt{\frac{1}{4}},\ \sqrt{\frac{2}{4}}\right]$	$\left(\sqrt{\frac{2}{4}},\ \sqrt{\frac{3}{4}}\right]$	$\left(\sqrt{\frac{3}{4}},\ 1\right]$	Total
Observed Counts	245	255	235	265	1000
Expected Counts					

That is, for example, 245 of her observations were between 0 and $\sqrt{\frac{1}{4}}$. State the kind of test you are performing. Symbolically state relevant hypotheses defining any notation that you use. Calculate the expected cell counts **showing your work**. Set up calculations for the relevant chi-squared statistic but do not solve for a final number. Suppose that the chi-squared statistic value in this problem was 2. Interpret this value in the context of this problem.

Upper tail chi-square probabilities for 1 df.

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	1.000	0.752	0.655	0.584	0.527	0.480	0.439	0.403	0.371	0.343
1	0.317	0.294	0.273	0.254	0.237	0.221	0.206	0.192	0.180	0.168
2	0.157	0.147	0.138	0.129	0.121	0.114	0.107	0.100	0.094	0.089
3	0.083	0.078	0.074	0.069	0.065	0.061	0.058	0.054	0.051	0.048
4	0.046	0.043	0.040	0.038	0.036	0.034	0.032	0.030	0.028	0.027
5	0.025	0.024	0.023	0.021	0.020	0.019	0.018	0.017	0.016	0.015
6	0.014	0.014	0.013	0.012	0.011	0.011	0.010	0.010	0.009	0.009

Upper tail chi-square probabilities for 2 df.

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	1.000	0.951	0.905	0.861	0.819	0.779	0.741	0.705	0.670	0.638
1	0.607	0.577	0.549	0.522	0.497	0.472	0.449	0.427	0.407	0.387
2	0.368	0.350	0.333	0.317	0.301	0.287	0.273	0.259	0.247	0.235
3	0.223	0.212	0.202	0.192	0.183	0.174	0.165	0.157	0.150	0.142
4	0.135	0.129	0.122	0.116	0.111	0.105	0.100	0.095	0.091	0.086
5	0.082	0.078	0.074	0.071	0.067	0.064	0.061	0.058	0.055	0.052
6	0.050	0.047	0.045	0.043	0.041	0.039	0.037	0.035	0.033	0.032
7	0.030	0.029	0.027	0.026	0.025	0.024	0.022	0.021	0.020	0.019
8	0.018	0.017	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.012

Upper tail chi-square probabilities for 3 df.

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	1.000	0.992	0.978	0.960	0.940	0.919	0.896	0.873	0.849	0.825
1	0.801	0.777	0.753	0.729	0.706	0.682	0.659	0.637	0.615	0.593
2	0.572	0.552	0.532	0.513	0.494	0.475	0.457	0.440	0.423	0.407
3	0.392	0.376	0.362	0.348	0.334	0.321	0.308	0.296	0.284	0.272
4	0.261	0.251	0.241	0.231	0.221	0.212	0.204	0.195	0.187	0.179
5	0.172	0.165	0.158	0.151	0.145	0.139	0.133	0.127	0.122	0.117
6	0.112	0.107	0.102	0.098	0.094	0.090	0.086	0.082	0.079	0.075
7	0.072	0.069	0.066	0.063	0.060	0.058	0.055	0.053	0.050	0.048
8	0.046	0.044	0.042	0.040	0.038	0.037	0.035	0.034	0.032	0.031
9	0.029	0.028	0.027	0.026	0.024	0.023	0.022	0.021	0.020	0.019
10	0.019	0.018	0.017	0.016	0.015	0.015	0.014	0.013	0.013	0.012
11	0.012	0.011	0.011	0.010	0.010	0.009	0.009	0.008	0.008	0.008