BST 140.652 Final Exam

Notes:

- You may use your one 8.5 by 11 formula sheet.
- Please use only the basic mathematical functions on your calculator.
- 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!

____

signature and **printed name**

1. A matched retrospective case/control study was conducted to investigate an airborne environmental toxicant's effect on lung cancer. The data are given below

	Cases		
Controls	Exposed	Unexposed	
Exposed	243	54	297
Unexposed	189	153	342
	432	207	639

a. Estimate the marginal odds ratio and the subject specific odds ratio. (Do not form CIs, just give the estimates.)

b. Consider testing whether the proportion of exposed cases is the same as the proportion of exposed controls. State the relevant hypotheses defining any notation that you use, perform the relevant test and interpret your results.

c. As in the HW, consider stratifying by matched pair and representing each pair as one of the four tables:

	Case	Ctrl
E	1	1
U	0	0

	Case	Ctrl
E	1	0
U	0	1

	Case	Ctrl
E	0	1
U	1	0

	Case	Ctrl
E	0	0
U	1	1

a b c d

labels

Show that the CMH estimate of the odds ratio is the same as the subject specific estimate of the odds ratio for matched pairs data. Hint, the CHM OR estimate given in the notes is:

$$\frac{\sum_k n_{11k}n_{22k}/n_{++k}}{\sum_k n_{12k}n_{21k}/n_{++k}}.$$

Both est ↓ $O(E|C)/O(E|\bar{C})$

a. Conditional OR $\frac{189}{54} \approx 3.5$

Marginal OR

cases	432	207
ctrl's	297	342

OR $\frac{432 \cdot 342}{297 \cdot 207} \approx 2.4$

b. McNemar's test $H_0\ P_{EC} = P_{EL}$ $H_a: P_{EC} \neq P_{EL}$

P_{EC} = Prob. of exp. given being a case

P_{EL} = Prob. of exp given being a ctrl

$\frac{(189-54)^2}{189+54} = 75$ reject by any reasonable criterion

c. $n_{11k} \cdot n_{22k} = 1$ only for group b and 0

o/w

$n_{12k} \cdot n_{21k} = 1$ only group c and 0

o/w $\quad n_{++k} = 2$ always

$$\therefore \quad \frac{\sum n_{11k} n_{22k} / n_{++k}}{\sum n_{12k} n_{21k} / n_{++k}} = \frac{\sum \text{people of type b} / 2}{\sum \text{people of type c} / 2}$$

$$= \frac{189/2}{54/2} = \frac{189}{54}$$

2. A nuclear power plant pump company claims that the failure rate for their pump is below the industry standard of $.15$ failures per day. It is known not to have a failure rate above $.15$. A test run of their product yielded 2 failures in 20 days. Give and interpret the relevant *exact* P-value; state your hypotheses and the assumptions that you use.

$H_0: \lambda = .15$ $\quad$ $H_a: \lambda \le .15$

pop mean failures per/day

$x = 2 \quad t = 20$

Pvalue $= P(X \le 2 \mid t = 20, \lambda = .15)$

($20 \times .15 = 3$ note)

$$= \frac{3^0 e^{-3}}{0!} + \frac{3^1 e^{-3}}{1!} + \frac{3^2 e^{-3}}{2!}$$

$$= e^{-3} \cdot (1 + 3 + 4.5) = e^{-3}\, 8.5$$

$$= .17$$

Fail to reject H_0 & conclude that there is insufficient evidence to support the claim of a failure rate below .15.

Assume $X \sim \text{Poisson}(\lambda 20)$

3. Refer to the previous problem. The company created a second product that they claim has **one half** the failures rate of the pump from question 1. Suppose a test was conducted using both pumps. Let Y_1 and t_1 be the number of failures and monitoring time of their first product and Y_2 and t_2 be the number of failures and the monitoring time for their second product. Let λ_1 and λ_2 be the failure rates. **Derive** the maximum likelihood estimates of λ_1 and λ_2 under the relevant null hypothesis. State the assumptions that you use. (Hint the contribution to the likelihood for the first pump is $\lambda_1^{y_1} e^{-t_1\lambda_1}$.)

$$H_0: \lambda_2 \leq \lambda_1 \tfrac{1}{2}$$
$$H_a: \lambda_2 > \lambda_1 \, 1/2$$

est. $\hat{\lambda}_1$ under H_0

likelihood $\lambda_1^{y_1} e^{-t_1\lambda_1} \times (\tfrac{1}{2}\lambda_1)^{y_2} e^{-t_2 \lambda_1/2}$

$$= \lambda_1^{y_1+y_2} \exp\{-\lambda_1(t_1 + \tfrac{t_2}{2})\}$$

log like $= (y_1+y_2)\log\lambda_1 - \lambda_1(t_1 + \tfrac{t_2}{2})$

$$\frac{d\ell}{d\lambda} = \frac{y_1+y_2}{\lambda_1} - (t_1 + t_2/2) = 0$$

$\Rightarrow \hat{\lambda}_1 = (y_1+y_2)/(t_1 + \tfrac{1}{2}t_2)$

$\Rightarrow \hat{\lambda}_2 = \tfrac{1}{2}\hat{\lambda}_1$

Checking 2nd deriv.

$$\frac{d^2\ell}{d\lambda^2} = -\frac{(y_1+y_2)}{\lambda_1^2} < 0$$

4. The Departments of Biostatistics and Epidemiology are competing in a chili cook-off. Both departments entered two dishes. A panel of blinded, independent, judges agreed on a consensus ranking of the four dishes from best (1) to worst (4). Rules of the competition stipulated that no ties were allowed. Biostatistics' two dishes won first and second. Does this suggest that Biostatistics chili is preferable to Epidemiology chili? Report and interpret a relevant P-value. (Show your work.)

1	2	3	4	w	
B	B	E	E	3	observed
B	E	B	E	4	
B	E	E	B	5	
E	B	B	E	5	
E	B	E	B	6	
E	E	B	B	7	

w	3	4	5	6	7
p(w)	1/6	1/6	2/6	1/6	1/6

p = 1/6 Not significant at .05 level. However, with this small of a sample a larger α might be warranted. Besides everyone knows Biostat chili is better than Epi chili.