STAT 426 — Sections 1GR, 1UG — Spring 2017

Sample Exam 1

Full Name: Key ID/Email: ______________

- This is an 80 minute exam. There are 5 problems, one of which is for the graduate section only.

  The exam is worth a total of 42 points for the undergraduate section and 52 points for the graduate section.
- You may use *one* physical page of personal notes and a standard scientific calculator. (You may *not* share these items with anyone else.)
- *Write all answers in the spaces provided.* If you require more space to write your answer, you may use the back side of the page.
- You are not allowed to communicate with anyone except the instructor or proctors before you submit this exam.

Useful Abbreviations:

CI = confidence interval
SE = standard error
E = expected value   Var = variance   Cov = covariance
df = degrees of freedom
ML = maximum likelihood   LRT = likelihood ratio test L = log-likelihood
RR = relative risk
H_0 = the null hypothesis of a test H_a = the alternative hypothesis of a test

1. The relationship between coffee drinking and myocardial infarction (MI) was studied in young women, aged 30–49. This retrospective study included 487 cases hospitalized for the occurrence of a MI. Nine hundred eighty patients hospitalized for some other acute condition were selected as controls. Data for consumption of five or more cups of coffee, stratified by smoking status were:

Nonsmokers		
Cups per day	MI	Controls
≥ 5	14	49
< 5	75	381

Smokers		
Cups per day	MI	Controls
≥ 5	138	134
< 5	260	416

(a) Compute both conditional odds ratios as well as the marginal odds ratio. Give approximate 95% confidence intervals for the conditional odds ratios. [10 pts]

$$\hat{\theta}_N = \frac{14 \cdot 381}{75 \cdot 49} \approx 1.451 \qquad \hat{\theta}_S = \frac{138 \cdot 416}{260 \cdot 134} \approx 1.648$$

$$\hat{\theta} = \frac{(14+138)\cdot(381+416)}{(75+260)\cdot(49+134)} \approx 1.98$$

$$\hat{\sigma}(\log \hat{\theta}_N) \approx \sqrt{\frac{1}{14} + \frac{1}{49} + \frac{1}{75} + \frac{1}{381}} \approx 0.328$$

$$\hat{\sigma}(\ln \hat{\theta}_S) \approx \sqrt{1/138 + 1/134 + 1/260 + 1/416} \approx 0.1448$$

Nonsmokers:

$$\exp(\ln 1.451 \pm 1.96 \cdot 0.328) \approx (0.76, 2.76)$$

Smokers:

$$\exp(\ln 1.648 \pm 1.96 \cdot 0.1448) \approx (1.24, 2.19)$$

(b) Do you think that stratification on smoking status was important in this study? Explain. [2 pts]

It does not seem essential.

The conditional odds ratios are in the same direction as the marginal: more cups of coffee is associated with higher prob. of MI.

2. The Chinese Mini-Mental Status Test (CMMS) is a test consisting of 114 items to identify people with Alzheimer's disease and senile dementia among people in China. An extensive clinical evaluation was performed, whereby participants were interviewed by psychiatrists and nurses and a definitive diagnosis of dementia was made. The table below shows the results.

CMMS score	Non-demented	Demented	Total
0-5	0	2	2
6-10	0	1	1
11-15	3	4	7
16-20	9	5	14
21-25	16	3	19
26-30	18	1	19
Total	46	16	62

Suppose that a cutoff value of ≤ 20 on the test is used to identify people with dementia.

(a) Estimate the sensitivity of the test. [3 pts]

$$\hat{P}(\text{score} \leq 20 \mid \text{Demented}) = \frac{2+1+4+5}{16} = 0.75 = \hat{p}_1$$

(b) Estimate the specificity of the test. [3 pts]

$$\hat{P}(\text{score} > 20 \mid \text{Non-demented}) = \frac{16+18}{46} \approx 0.739 = \hat{p}_2$$

(c) The *positive predictive value (PPV)* is the probability that someone who tests positive actually has the disease. Suppose the prevalence of dementia is 5% in this population. Use this, together with the estimated sensitivity and specificity, to estimate the PPV. [4 pts]

$$\hat{P}(\text{Demented} \mid \text{score} \leq 20)$$

$$= \frac{\hat{p}_1 \cdot 0.05}{\hat{p}_1 \cdot 0.05 + (1-\hat{p}_2) \cdot (1-0.05)}$$

$$\approx \frac{0.0375}{0.0375 + 0.261 \cdot 0.95} \approx 0.131$$

3. The following table gives results for a study comparing surgery and radiation therapy for treating cancer of the larynx.

	Outcome	
Therapy	Successful	Unsuccessful
Radiation	3	15
Surgery	2	21

Use an appropriate test for $H_0 : \theta = 1$ against $H_a : \theta > 1$, where θ is the odds ratio for success of radiation versus surgery. Give the P-value. [10 pts]

Some small counts, so use Fisher's Exact Test.

$$P_{H_0}(n_{11}=3) = \frac{\binom{18}{3}\binom{23}{2}}{\binom{41}{5}} \approx 0.275$$

$$P_{H_0}(n_{11}=4) = \frac{\binom{18}{4}\binom{23}{1}}{\binom{41}{5}} \approx 0.094$$

$$P_{H_0}(n_{11}=5) = \frac{\binom{18}{5}\binom{23}{0}}{\binom{41}{5}} \approx 0.011$$

$$\begin{aligned} P\text{-value} &= P_{H_0}(n_{11} \geq 3) \\ &\approx 0.275 + 0.094 + 0.011 \\ &= 0.38 \end{aligned}$$

4. The following table classifies a sample of psychiatric patients according to their diagnosis and whether they are prescribed drugs as part of their treatment.

Diagnosis	Drugs	No Drugs
Schizophrenia	150	13
Neurosis	18	19
Personality Disorder	47	52

Compute an appropriate test statistic of the null hypothesis that drug prescription is independent of diagnosis. State the asymptotic distribution of this statistic under the null hypothesis of independence. [10 pts]

For example, the Pearson statistic:

$n = 299$ $\quad \hat{\pi}_{1+} \approx 0.545 \quad \hat{\pi}_{+1} \approx 0.719$

$\hat{\pi}_{2+} \approx 0.124 \quad \hat{\pi}_{+2} \approx 0.281$

$\hat{\pi}_{3+} \approx 0.331$

$\hat{\mu}_{11} = n\hat{\pi}_{1+}\hat{\pi}_{+1} \approx 117.2 \quad \hat{\mu}_{12} \approx 45.8$

$\hat{\mu}_{21} \approx 26.7 \quad \hat{\mu}_{22} \approx 10.4$

$\hat{\mu}_{31} \approx 71.2 \quad \hat{\mu}_{32} \approx 27.8$

$$X^2 \approx \frac{(150-117.2)^2}{117.2} + \frac{(13-45.8)^2}{45.8} + \frac{(18-26.7)^2}{26.7} + \frac{(19-10.4)^2}{10.4} + \frac{(47-71.2)^2}{71.2} + \frac{(52-27.8)^2}{27.8} \approx 71.9$$

$$\underset{H_0}{\dot{\sim}} \chi^2_{(3-1)(2-1)} = 2 \text{ df}$$

GRADUATE SECTION ONLY

5. Let $X_1, X_2, \ldots, X_n$ be independent random variables. Suppose X_t has a Poisson distribution with mean $t\lambda$, where λ is unknown.

(a) Find the ML estimator $\hat{\lambda}$. [3 pts]

$$L(\lambda) = \text{constant} + \sum_t x_t \ln t\lambda - \sum_t t\lambda$$

$$L'(\lambda) = \sum_t x_t \frac{1}{t\lambda} t - \sum_t t = \frac{1}{\lambda}\sum_t x_t - \binom{n+1}{2}$$

$$L'(\hat{\lambda}) = 0 \implies \hat{\lambda} = \sum_t x_t \Big/ \binom{n+1}{2}$$

(b) Find an expression for approximating $\text{Var}(\hat{\lambda}^2)$. [3 pts]

$$\sum_t X_t \sim \text{Poisson}\left(\binom{n+1}{2}\lambda\right)$$

Normal approx. to Poisson: $\dfrac{\sum X_t - \binom{n+1}{2}\lambda}{\sqrt{\binom{n+1}{2}}} \,\dot{\sim}\, N(0,\lambda)$

$$\Rightarrow \sqrt{\binom{n+1}{2}}\,(\hat{\lambda}-\lambda) \,\dot{\sim}\, N(0,\lambda)$$

so, with $g(\lambda) = \lambda^2 \Rightarrow g'(\lambda) = 2\lambda$

$$\sqrt{\binom{n+1}{2}}\,(\hat{\lambda}^2-\lambda^2) \,\dot{\sim}\, N(0,(2\lambda)^2\lambda)$$

so $\text{Var}(\hat{\lambda}^2) \approx 4\hat{\lambda}^3 \Big/ \binom{n+1}{2}$

(c) Give an approximate 95% confidence interval for the parameter $\theta = \lambda^2$ for a data set in which $n = 15$ and $\sum_{t=1}^{n} X_t = 22$. [4 pts]

$$\hat{\lambda} = 22 \Big/ \binom{16}{2} \approx 0.183$$

(Approx.) Wald 95% CI by delta method:

$$\hat{\lambda}^2 \pm 1.96\sqrt{4\hat{\lambda}^3 \Big/ \binom{15+1}{2}}$$

$$\approx (0.183)^2 \pm 1.96\sqrt{0.000204}$$

$$\approx (0.0055,\ 0.0615)$$

(Alternative: Transform Wald interval for λ.)