Stat 579 Midterm Exam 1

1. Suppose $n_1 \sim BIN(n, \pi)$.

(a) Derive the maximum likelihood estimate $\widehat{\pi}$.

(b) The mle $\widehat{\pi}$ has an asymptotic normal distribution $\widehat{\pi} \sim N(\pi, \pi(1-\pi)/n)$.
Use the delta method to determine $\sigma^2(\log(\widehat{\pi}/(1-\widehat{\pi})))$, the asymptotic variance of the sample log odds.

(a) $\ell(\pi) = \binom{n}{n_1} \pi^{n_1} (1-\pi)^{n-n_1}$, $\quad L(\pi) = n_1 \log \pi + (n-n_1) \log(1-\pi)$

$$\frac{dL}{d\pi} = \frac{n_1}{\pi} - \frac{n-n_1}{1-\pi} \overset{set}{=} 0, \quad \frac{\pi}{n_1} = \frac{1-\pi}{n-n_1},$$

$$\pi(n-n_1) = (1-\pi)n_1, \quad n\pi = n_1, \quad \hat{\pi} = \frac{n_1}{n}$$

[redacted]

(b) $\sigma^2(g(\hat{\pi})) = [g'(\pi)]^2 \frac{\pi(1-\pi)}{n}$, [redacted]

$$g(t) = \log\left(\frac{t}{1-t}\right) = \log t - \log(1-t)$$

$$g'(t) = \frac{1}{t} + \frac{1}{1-t} = \frac{1}{t(1-t)}$$

$$\sigma^2\left(\log \frac{\hat{\pi}}{1-\hat{\pi}}\right) = \left[\frac{1}{\pi(1-\pi)}\right]^2 \cdot \frac{\pi(1-\pi)}{n} = \frac{1}{\pi(1-\pi)} \cdot \frac{1}{n}$$

[redacted]

2. Consider data from a prospective study on the relationship between daily aspirin use and the onset of heart disease.

	Disease	No Disease	Total
Non-Aspirin	28	656	684
Aspirin	18	658	676

(a) Compute an estimate of the relative risk.

(b) Provide an interpretation of your result, stated in the context of the problem.

(c) Explain the difference between a prospective study and a retrospective study. What parameters can be estimated from a prospective study? What parameters can be estimated from a retrospective study?

(a) $\hat{RR} = \frac{\hat{\pi}_1}{\hat{\pi}_2} = \frac{28/684}{18/676} = \frac{.0409}{.0266} = 1.537$

(b) We estimate that non Aspirin users are 1.537 times more likely to develop heart disease than Aspirin users.

(c) In prospective sampling, the input variable is fixed by the sampling design so only the response variable is observed from experimentation.

In retrospective sampling, the response variable is fixed by the sampling design so only the ~~[illegible]~~ input variable is observed from experimentation.

From a ~~[illegible]~~ prospective study ~~sampling~~, we can estimate the difference in proportions, the relative risk, and the odds ratio.

From a retrospective study, we can only estimate the odds ratio.

3. Consider data from a prospective study on the relationship between a new treatment and complications from respiratory illness.

	Complications	Recovery	Total
Control	29	21	50
Treatment	14	36	50

$\hat{\pi}_1 = .58$

$\hat{\pi}_2 = .28$

(a) Provide an equation for $\sigma^2\left(\log \widehat{RR}\right)$.

(b) Provide an equation for $\widehat{\sigma}\left(\log \widehat{RR}\right)$.

(c) Compute a 95% confidence interval for RR.

$$(a) \quad \sigma^2(\log \hat{RR}) = \left(\frac{1-\pi_1}{\pi_1}\right)\frac{1}{n_{1+}} + \left(\frac{1-\pi_2}{\pi_2}\right)\frac{1}{n_{2+}}$$

$$(b) \quad \hat{\sigma}(\log \hat{RR}) = \left(\frac{1}{n_{11}} - \frac{1}{n_{1+}} + \frac{1}{n_{21}} - \frac{1}{n_{2+}}\right)^{\frac{1}{2}}$$

$$(c) \quad \log \hat{RR} = \log\left(\frac{.58}{.28}\right) = \log(2.0714) = 0.7282$$

$$\hat{\sigma}(\log \hat{RR}) = \left(\frac{1}{29} - \frac{1}{50} + \frac{1}{14} - \frac{1}{50}\right)^{\frac{1}{2}} = 0.2567$$

$$95\% \text{ CI for } \log RR = 0.7282 \pm 1.96(0.2567)$$

$$= 0.7282 \pm 0.5032$$

$$= [0.2250, 1.2314] = [l, u]$$

$$95\% \text{ CI for } RR = [e^{l}, e^{u}] = [1.25, 3.43]$$

4. Consider data from a retrospective study on the relationship between daily alcohol consumption and the onset of esophagus cancer.

	cancer	no cancer
daily alcohol consumption > 80g	71	82
daily alcohol consumption < 80g	60	441
Total	131	523

(a) Compute an estimate of the odds ratio.
(b) Interpret the direction of the association, stated in the context of the problem.
(c) What assumption is necessary for the sample odds ratio to serve as an estimate of the relative risk?

(a) $\hat{\Theta} = \dfrac{n_{11} n_{22}}{n_{12} n_{21}} = \dfrac{71(441)}{82(60)} = 6.364$

(b) Since $\hat{\Theta} > 1$, we estimate that those with higher daily alcohol consumption have a greater probability of esophagus cancer.

(c) When the probability of disease in the study population is small for both input groups, the odds ratio approximates the relative risk.

5. Consider data from a retrospective study on the relationship between smoking and myocardial infarction.

	m.i.	no m.i.
smoker	17	18
nonsmoker	9	34
Total	26	52

(a) Provide an equation for $\widehat{\sigma}\left(\log \widehat{\theta}\right)$.
(b) Does your answer to (a) depend on the sampling scheme? Explain.
(c) Compute a 95% confidence interval for $\log \theta$.
(d) Compute an estimate of the correlation γ. Provide an interpretation of the effect size, stated in the context of the problem.

(a) $\hat{\sigma}(\log \hat{\theta}) = \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}\right)^{\frac{1}{2}}$

(b) The equation for $\hat{\sigma}(\log \hat{\theta})$ does not depend on the sampling scheme. When asymptotic normality holds, replacing parameters with statistics invokes the likelihood principle.

(c) $\log \hat{\theta} = \log\left(\frac{17 \cdot 34}{18 \cdot 9}\right) = \log(3.5679) = 1.27$

$\hat{\sigma}(\log \hat{\theta}) = \left(\frac{1}{17} + \frac{1}{18} + \frac{1}{34} + \frac{1}{9}\right)^{\frac{1}{2}} = 0.50$

95% CI for $\log \theta = 1.27 \pm 1.96(0.50) = [0.27, 2.27]$

(d) $\hat{\gamma} = \frac{\hat{\theta}-1}{\hat{\theta}+1} = 0.562$, $\left(\hat{\gamma} = \frac{C-D}{C+D} = \frac{17(34)-18(9)}{17(34)+18(9)} \right)$

We estimate that there is a medium/large size positive association between smoking and myocardial infarction.