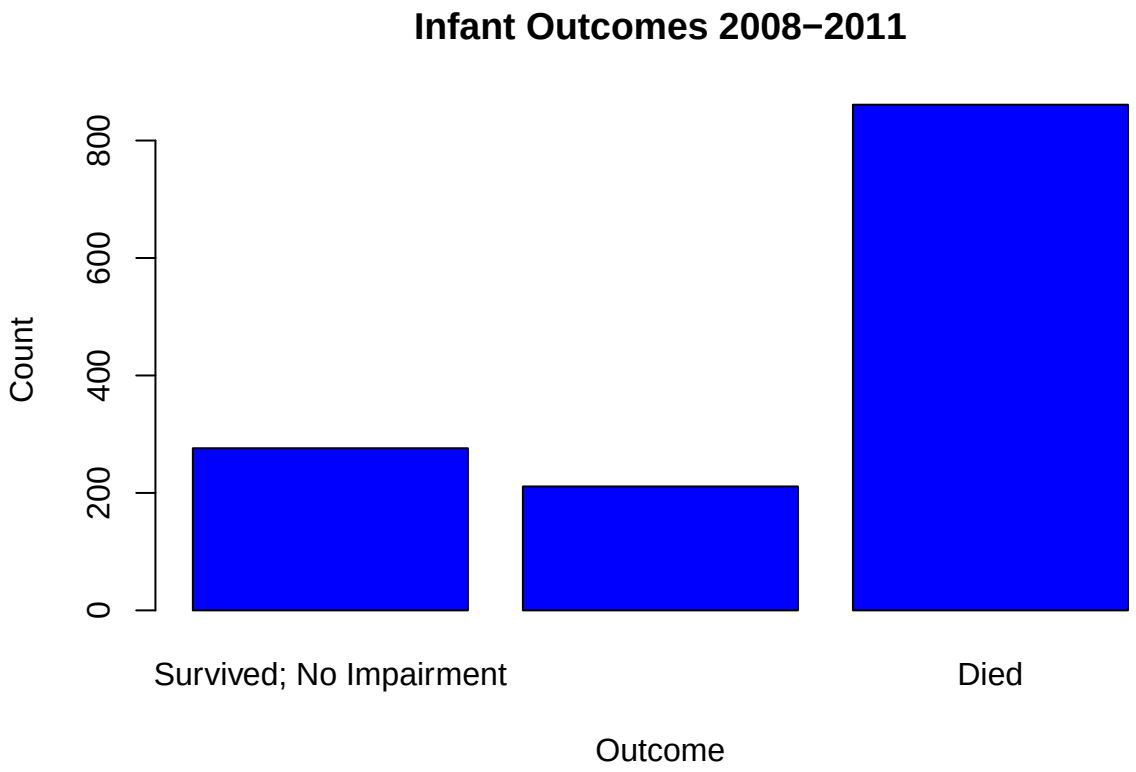
Homework 1, Due September 13 at 11:45am by uploading .Rmd and .pdf files on Sakai

Provide clearly documented code in an R markdown document where appropriate. Points will be deducted in each question if results are not reproducible. Whenever possible, provide statistical evidence to support your answers.

1. Analysis of periviable neonates (8 points)

In a 2017 New England Journal of Medicine paper, Duke researchers examined outcomes of infants born at 22-24 weeks of gestation. Of 1348 infants in the study born at these gestational ages in 2008-2011, 861 died, 211 survived with neurodevelopmental impairment, and 276 survived without neurodevelopmental impairment.



Suppose historical rates of all three outcomes (from the early 2000's) are 15% survival without impairment, 15% survival with impairment, and 70% death. Estimate the proportions of infants surviving without impairment, surviving with impairment, and dying, providing point and interval estimates. Evaluate the hypothesis that the current rates are equal to the historical rates of 15, 15, and 70%, respectively, using the multinomial distribution. Display results using informative graphics.

2. Agresti 1.7 (12 points total)

Complete the given parts and add part f, given below. In a crossover trial comparing a new drug to a standard, π denotes the probability that the new one is judged better. It is desired to estimate π and to test $H_0 : \pi = 0.50$ against $H_a : \pi \neq 0.50$. In 20 independent observations, the new drug is better each time.

2.a. (2 points)

Plot the likelihood function. Is it close to the quadratic shape that large-sample normal approximations utilize?

2.b. (2 points)

Give the ML estimate of π. Conduct a Wald test and construct a 95% Wald CI for π. Are these sensible?

2.c. (2 points)

Conduct a score test, reporting the p-value. Construct a 95% score confidence interval. Interpret.

2.d. (2 points)

Conduct a likelihood-ratio test and construct a likelihood-based 95% confidence interval. Interpret.

2.e.(2 points)

Construct an exact binomial test. Interpret.

2.f. (2 points)

Estimate a 95% HPD interval for π when a U[0,1] prior is used. Interpret.

3. Agresti 1.41 (10 points)

For the Dirichlet prior for multinomial probabilities, show the Bayesian estimator of the posterior expectation is a weighted average of the prior mean and sample proportions as in Agresti (1.19).

4. Agresti 2.26 (10 points)

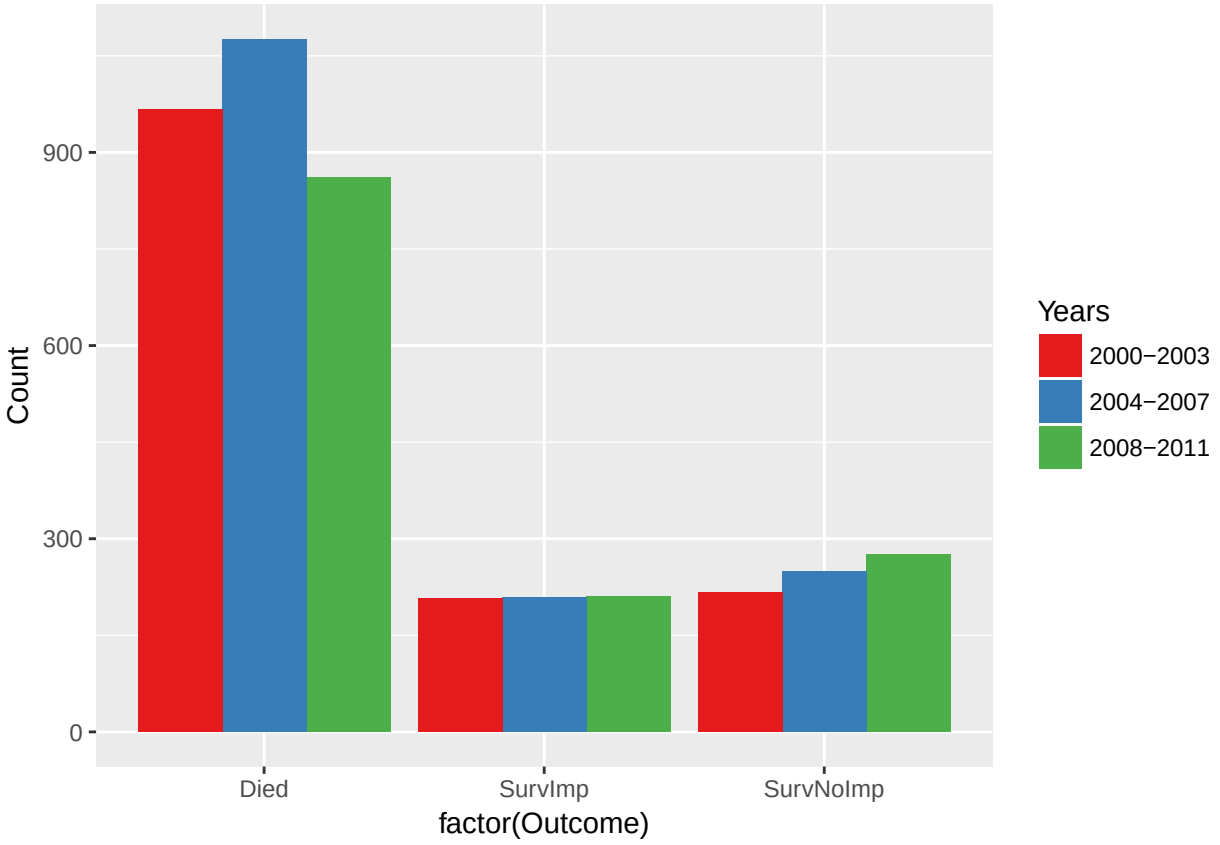
Show that the OR and RR need not be similar when π_i is close to 1.0 for both groups.

5. Agresti 2.28 (10 points)

In comparing new and standard treatments with success probabilities π_1 and π_2, the number needed to treat (NNT) is the number of patients that would need to be treated with the new treatment instead of the standard in order for one patient to benefit. Explain why a natural estimate of this is $\frac{1}{\widehat{\pi}_1 - \widehat{\pi}_2}$.

6. Periviable infants (12 points total)

Recall the data on outcomes of periviable infants. Of 1348 infants in the study born at these gestational ages in 2008-2011, 861 died, 211 survived with neurodevelopmental impairment, and 276 survived without neurodevelopmental impairment. In 2000-2003, of 1391 infants born at these gestational ages, 217 survived with no impairment, 207 survived with impairment, and 967 died. In 2004-2007, of 1535 total infants, 250 survived without impairment, 209 survived with impairment, and 1076 died.



6.a. Nominal variables (6 points)

Provide a formal test of the hypothesis that time period is not associated with outcome. Interpret results of your test in clear language that could be used in a scientific publication. If there is a change over time, clearly describe the nature of that change.

6.b. Ordinal variables (6 points)

Now assume that the outcomes are ordered so that death is the worst outcome and survival without impairment is the best outcome. Conduct a linear trend test using scores (1,2,3) for time (1=2000-2003) and outcome (3=dead, 2=impaired, 1=survived and not impaired) and interpret the results. Do the results change if we use the outcome weights 100=dead, 10=impaired, 1=survived not impaired? If so, explain how and why the results differ. Comment on the appropriateness of the scores for both variables.

7. Oral contraceptives and thromboembolism (8 points)

We consider a retrospective matched pair case-control study of thromboembolism (formation of a blood clot that breaks loose and is carried by the bloodstream to plug another vessel) and oral contraceptive use. The cases were 175 women of reproductive age, discharged alive from 43 hospitals in 5 cities after experiencing idiopathic thrombophlebitis, pulmonary embolism, or cerebral thrombosis or embolism. Controls were females matched with their cases for hospital, residence, time of hospitalization, race, age, marital status, parity, and insurance pay status. The question of interest is whether cases are more likely than controls to use oral contraceptives.

Of the 175 matched pairs, in 10 pairs both cases and controls used oral contraceptives (OC's), in 95 pairs neither used OC's, in 13 pairs only the control used OC's, and in 57 pairs only the case used OC's.

Using the appropriate test, evaluate the null hypothesis that the proportion of cases exposed to OC's is the same as the proportion of controls who are exposed, and provide the p-value for this test. Calculate the OR and exact 95% CI comparing the odds of thromboembolism in OC users to those in non-OC users. Interpret your results.

8. Passive smoking and cancer risk (15 points)

A landmark study examined the association between passive smoking and cancer in a group of 509 individuals with cancer and 489 cancer-free (unmatched) controls. Researchers defined passive smoking as exposure to the cigarette smoke of a partner who smoked at least one cigarette per day for at least 6 months. One potential confounding variable is smoking by the participants (i.e., personal smoking) because personal smoking is related to both cancer risk and partner smoking. Therefore, it is important to consider personal smoking when examining the relationship between passive smoking and cancer risk.

Among cases, 120 were exposed to passive smoke but did not smoke themselves; 161 were smokers also exposed to passive smoke; 111 had no exposure from passive or active smoking; and 117 were active smokers who were not exposed to passive smoke. Among controls, 80 were exposed to passive smoke but did not smoke themselves; 130 were smokers also exposed to passive smoke; 155 had no exposure from passive or active smoking; and 124 were active smokers who were not exposed to passive smoke.

Calculate the conditional odds ratios comparing the odds of being a case by passive smoking exposure status, conditional on active smoking status (active smoker or nonsmoker). Evaluate whether case status and passive smoking are conditionally independent given active smoking status. Is the association between passive smoking and case status homogeneous across categories of active smoking status? Provide statistical evidence to support your answers.

9. Exercise without pain! (15 points)

An athletics company produces a new model of running shoe that includes a harder material in the insert to correct for overpronation. The company is worried that the material will cause heel tenderness due to some loss of cushioning with the strike of each step. The company enrolled 87 runners and asked them about occasional heel tenderness at study entry and then again after using the new shoe for one month. The company would like to know whether the proportion of runners with occasional heel tenderness is the same before the study and after using the study shoe for one month.

Analyze the data and produce a professional report (as part of this document) providing statistical evidence that addresses the question of interest to the company. Data are in the file dontbeaheel.xlsx.