**STAT 601 ~ Assignment #2 (57 points)**

**Review the following:**

***Lecture Handouts* :** Probability, Binomial Distribution, and the Normal Distribution, there are blank notes and annotated ones. You will also need the normal tables which are linked on the website.

***Narrated Powerpoint Lectures*** under heading 5 – Parts 1 - 4

***Non-narrated Powerpoint Lectures*** under heading 5 - Parts 1 – 4

**1. Low Birth Weight Risk Factors**  ([**Lowbirth.JMP**](http://course1.winona.edu/BDEPPA/STAT%20701%20Online/Data/JMP%20Data/Low%20Birth%20Weight%20%28small%29.JMP))

The purpose of this study was to identify potential risk factors for low birth weight.   The following categorical variables were measured and coded as follows:

Prev? = women has previous history of premature labor (History, None)

Hyper = hypertension during pregnancy (HT, Normal)

Smoke = smoked during pregnancy (Cig = smoker, No Cig = non-smoker)

Uterine = uterine irritability during pregnancy (Irritation, None)

Minority = minority status (Nonwhite, White)

Low = low birth weight indicator (Low, Normal) 🡨 Response/Outcome

a)  For each risk factor calculate P(Low|*risk factor present*) and P(Low|*risk factor absent*) for each of the four  potential risk factors. In general, what do these probabilities tell you about each of the potential risk factors? (5 pts.)

b)  Use your answers in part (a) to calculate the relative risk (RR) associated with each potential risk factor and put them in the table below. Interpret the RR associated with smoking during pregnancy. (5 pts.)

c)  As this was a case-control study the relative risks calculated in part (b) are not truly valid, it would be better to summarize the “risk” associated with these factors using the OR. Calculate the odds ratio (OR) associated with each characteristic and put them in the table below. Interpret the OR associated with smoking during pregnancy.  (5 pts.)

Problem #1 parts (d) and (e) are on the following page.

**Problem #1 parts (d) & (e)**

d) Use the results from part (b) and (c) above to complete the table below

|  |  |  |
| --- | --- | --- |
| **Risk Factor** | **Relative Risk (RR)** | **Odds Ratio (OR)** |
| Previous history of premature labor |  |  |
| Hypertension during pregnancy |  |  |
| Smoked during pregnancy |  |  |
| Uterine irritability during pregnancy |  |  |
| Minority status of mother |  |  |

e)  Which factor do has the greatest associated risk of having a child with low birth
weight?  the least? Explain your answers.  (2 pts.)

**2. Esophageal Cancer and Alcohol Consumption**

Tuyns et al. (1977) examined the potential relationship between regular alcohol consumption and esophageal cancer using a case-control study. A sample of 200 esophageal cancer patients was taken and 775 community-based controls were selected to form a comparison group. One potential risk factor examined in the study was whether or not the individual drank 80 or more grams of alcohol/day.

|  |  |  |  |
| --- | --- | --- | --- |
| **Daily Alcohol Use** | **Case** (Esophageal cancer) | **Control** | *Row Total* |
| > 80 g | 96 | 109 | 205 |
| < 80 g | 104 | 666 | 770 |
| *Column Total* | 200 | 775 | n = 975 |

**a)** Why is not valid to calculate P(Esophageal Cancer| Daily Alcohol Use > 80g)? (2 pts.)
 **b)** Use the results below to quantify the risk associated with consuming 80 or more
 grams of alcohol/day. Interpret your findings. (4 pts.)

**c)** Does this prove that regular alcohol drinking **causes** esophageal cancer?
 Explain your answer. (2 pts.)

**3. Assessment of Radiological Tests in the Detection of Coronary Artery Disease**

Begg et al. (1988) in their paper “Assessment of Radiological Tests: Control of Bias and Other Design Considerations” looked at the performance of radionuclide ventriculography as a diagnostic test for detecting coronary artery disease. The following results were obtained when using the test on 481 individuals known to have coronary artery disease and 452 individuals who do not have the disease.

|  |  |  |  |
| --- | --- | --- | --- |
| Test Result | Coronary Artery Disease (D+) | Disease Absent (D-) | Column Total |
| Positive (T+) | 302 | 80 | 382 |
| Negative (T-) | 179 | 372 | 551 |
| Row Total | 481 | 452 | n = 933 |

a) Calculate the sensitivity, specificity, false positive, and false negative probabilities using the results above. (4 pts.)

b) For a population in which the prevalence of coronary artery disease is .10 (or 10%) calculate the probability that an individual has coronary artery disease given that they test positive using radionuclide ventriculography, i.e. what is the positive predictive value of this test? (3 pts.)

c) For the same population, what is the probability that a person that tests negative does not have coronary artery disease, i.e. what is the negative predictive value of this test?
(3 pts.)

**4. Middle Ear Effusion in Breast-Fed and Bottle-Fed Infants (5 pts.)**A common symptom of otitus media in young children in the prolonged presence of fluid in the middle ear, known a middle-ear effusion. The presence of fluid may result in termporary hearing loss and interfere with normal learning skills in the first two years of life. One hypothesis is that babies who are breast-fed for at least 1 month build up some immunity against the effects of the disease and have less prolonged effusion than do bottle-fed babies. A small study of 24 pairs of babies is set up, where the babies are matched on a one-to-one basis according to age, sex, socioeconomic status, and type of medications taken. One member of the matched pair is a breast-fed baby, and other member is a bottle fed baby. The outcome variable is the duration of middle-ear effusion after the first episode of otitus media. The results are shown below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pair Number | Duration of effusion in breast-fed baby | Duration of effusion in bottle-fed baby | Difference |  Sign of Difference |
| 1 | 20 | 18 |  |  |
| 2 | 11 | 35 |  |  |
| 3 | 3 | 7 |  |  |
| 4 | 24 | 182 |  |  |
| 5 | 7 | 6 |  |  |
| 6 | 28 | 33 |  |  |
| 7 | 58 | 223 |  |  |
| 8 | 7 | 7 |  |  |
| 9 | 39 | 57 |  |  |
| 10 | 17 | 76 |  |  |
| 11 | 17 | 186 |  |  |
| 12 | 12 | 29 |  |  |
| 13 | 52 | 39 |  |  |
| 14 | 14 | 15 |  |  |
| 15 | 12 | 21 |  |  |
| 16 | 30 | 28 |  |  |
| 17 | 7 | 8 |  |  |
| 18 | 15 | 27 |  |  |
| 19 | 65 | 77 |  |  |
| 20 | 10 | 12 |  |  |
| 21 | 7 | 8 |  |  |
| 22 | 19 | 16 |  |  |
| 23 | 34 | 28 |  |  |
| 24 | 25 | 20 |  |  |

Do these data provide evidence that breast-fed babies have shorter durations of effusion when compared to bottle-fed babies that are the same age, sex, socioeconomic status, and on the same medications? Use the binomial distribution to answer this question and carefully justify your answer. Note: When the paired difference is zero, we drop that observation from the results and adjust *n* accordingly. (5 pts.)

**5. Use of Cyclosporine in Treatment of Aplastic Anemia Patients (5 pts.)**

Frickhofen et al. (1991) performed a study on the effect of using cyclosporine in addition to antilymphocyte globulin and methylprednisolone in the treatment of aplastic anemia patients. There was a sample of 43 patients that received the cyclosporine in addition to the other treatment. Historically, the use of antilymphocyte globulin and methylprenisolone without cyclosporine results in complete or partial remission in 40 percent of aplastic anemia patioents at the end of three months of treatment. We wish to determine if the use cyclosporine can increase significantly the percentage of patients experiencing complete or partial remission. In the clinical trial conducted by the researchers 28 of the 43 patients receiving cyclosporine in addition to the traditional treatment achieved complete or partial remission within three months.

Can we conclude on the basis of this result that the addition of cyclosporine to the treatment regimen is associated with an increase in the percent of aplastic anemia patients experiencing complete or partial remission? Again use the binomial distribution and carefully justify your answer. (5 pts.)

**6. Diabetes Screening Using Fasting Glucose Levels**

A standard test for diabetes is based on glucose levels in the blood after fasting for prescribed period. For healthy people the mean fasting glucose level is found to be 5.31 mole/liter with a standard deviation of 0.58mole/liter. For untreated diabetics the mean is 11.74, and the standard deviation is 3.50. In both groups the levels appear to be approximately Normally distributed.

To operate a simple diagnostic test based on fasting glucose levels we need to set a cutoff point, C, so that if a patient’s fasting glucose level is at least C we say they have diabetes. If it is lower, we say they do not have diabetes. Suppose we use C = 6.5.

**a)** What is the probability that a diabetic is correctly diagnosed as having diabetes, i.e. what is the sensitivity of the test? (2 pts.)

**b)** What is the probability that a nondiabetic is correctly diagnosed as not having diabetes, i.e. what is the specificity? (2 pts.)

Suppose we lower the cutoff value to C = 5.7.

**c)** What is the sensitivity now? (2 pts.)

**d)** What is the specificity now? (2 pts.)

In deciding what C to use, we have to trade off sensitivity for specificity. To do so in a reasonable way, some assessment is required of the relative “costs” of misdiagnosing a diabetic and misdiagnosing a nondiabetic. Suppose we required a 98% sensitivity.

**e)** What the value of C gives a sensitivity of .98 or 98%? Find the specificity of the test if this value of C was used? (4 pts.)