

## Practice Final Exam

1 In the population of healthy adult females, the mean platelet count (in units of  $10^9 / L$ ) is  $\mu=300$ . The standard deviation of platelet count (in units of  $10^9 / L$ ) is  $\sigma=125$ .

1a. If platelet count has a Gaussian distribution, what is the 95% prediction interval? (5 points)

In the population of adult female leukemia patients, the mean platelet count is  $\mu=600$  with a standard deviation of  $\sigma=100$ .

1b. If we use platelet count to diagnose leukemia (ie if platelet count is too high, we diagnose leukemia) what “threshold” platelet count value must be used in order to have 97.5% specificity? (10 points)

1c. Using this threshold value, what is the sensitivity of using platelet count as a test for leukemia? (10 points)

You have been assigned to a scientific advisory panel that is looking into whether recent exposure to a single pelvic x ray is associated with an increased risk of urinary tract infections (UTI) in women aged 30-40. Rentgen and Cathodsky report a relative risk (**RR**) of **3.0** in women exposed to x ray versus non exposed women with a corresponding **95% confidence interval of (0.7 to 18.0)**. Stratification was used to control for possible confounding variables.

2a. Since a RR is quoted, the study design probably is (circle all that could be true - 10 points)

- I. A retrospective study
- II. A prospective study
- III. A case control study
- IV A randomized clinical trial
- V An observational study
- VI. None of the above

2b. The incidence of UTI is 7% in the population of women who have **no** pelvic x ray exposure. **Based on the above**, what is the best estimate of the true UTI incidence (percent) in a comparable population of women who have had a pelvic x ray exposure? If this cannot be determined, briefly explain why. (5 points)

2c. For hypothesis testing to address the relation between x ray exposure and UTI infection, what is the null hypothesis value of a relative risk? (give a number). (5 points)

2d. The Z statistic ( $Z_{obs}$ ) for testing this null hypothesis is  $Z_{obs}=1.46$

Based on this  $Z_{obs}$ , what is the two sided p value? (Give a number – it may be approximate – 5 points)

Is this p value consistent with the 95% confidence interval above (yes or no)? (10 points)

Do we reject the null hypothesis based on the two sided  $p < 0.05=\alpha$  criterion? Briefly explain. (5 points)

2e. What is the power for this study (15 points)?

- i. The power is 97.5%
- ii. The power is 95%
- iii. The power is 70%
- iv. The power is 50%
- v. The power is 30%
- vi. The power is 5%
- vii. The power is 2.5%
- viii. The power can't be computed exactly but it is high
- ix. The power can't be computed exactly but it is low
- x. More information is needed to compute the power

Based on the power (or inability to compute power), is the study finding conclusively positive, conclusively negative, inconclusive or cannot be determined?

3. A group of 500 prostate cancer patients are being followed after surgical treatment. The **mean** follow up is one year although not every patient necessarily has one year follow up. So far, 10 deaths have been observed.

3a. Compute the mortality rate (hazard) rate for this sample.

3b. Assuming that the hazard rate is constant and that the relation between log survival and time is linear, compute the survival probability at 1 and 2 years.

4. Single live birth, full term infant birth weight in grams was measured in 189 mother-infant pairs. An equation giving predicted birth weight as a function of the mothers last pre pregnant weight in lbs (lwt), smoking status (smoke: 1=smoker, 0=non smoker) and whether the mother has a history of premature labor (ptl:1=yes, 0=no) is given by

$$\text{Predicted birth weight} = 2500 + 3.9 \text{ lwt} - 200 \text{ smoke} - 300 \text{ ptl} + 50 \text{ smoke} * \text{ptl}$$

All the terms in this equation are statistically significant at  $p < 0.01$

R square= 0.18,  
mean birthweight=2900 gm, birthweight SD = 750 gm , residual SD =  $SD_e = 680$  gm

4a. According to this equation, what is the mean birthweight in 120 lb non smokers with a history of premature labor. What is an (at least approximate) 95% prediction interval for birthweight in this group

4b. Circle all that are true

- i. This model is additive
- ii This model is linear in the regression coefficients (ie in  $b_0, b_1, b_2, b_3, b_4$ ).
- iii The mean birthweight difference between smokers and non smokers does NOT depend on mothers last weight
- iv The mean birth weight difference between smokers and non smokers does NOT depend on history of premature labor

4c. If the interaction were removed from the model, the regression (b) coefficient for smokers would be (circle one)

- i. -200, ii -150, iii -500, iv -450 v cannot tell without more information

4d. In order to be clinically useful, researchers would like a model that predicts birth weight to within +/- 300 gms in approximately 95% of subjects. What R square value would be needed to accomplish this goal.

5 Investigators in Germany compared IgG levels in children who were immunized for Pertussis versus those who were not. In addition to reporting the descriptive statistics, they report a two sided  $p = 0.04$  for their comparison.

The same study was carried out independently in the US using different children. This study reported a two sided  $p$  value of  $p = 0.01$ .

5a. If both studies used  $p < \alpha = 0.05$  as their significance criterion, what is the chance of a type I error in **at least** one of the two studies. Is this the same as the chance of a type I error in **both** studies?

5b. Using the conventional criterion of unadjusted  $p < 0.05$ , which of the two studies are significant, the German study, the US study, or both?

5c. Using the bonferroni criterion, which of the two studies are significant at 0.05, the German study, the US study or both?