Statistics 3858b Assignment 4

Handout March 24, 2015; Due date: April 6, 2015

These problems are all from the course text unless otherwise stated.

1. 8.10.4 (e)
2. 8.10.5 (d)
3. 9.11.19
4. 11.6.3
5. Use the data from problem 11.6.40. Test the hypothesis that the two population distributions for the bearing measurements are the same using the Mann Whitney test. In particular use the form of the test statistic needed for table A8.
6. In the two sample normal case (see handout from class) consider the hypothesis test of

$$H_0 : \mu_X - \mu_Y = \delta \text{ versus } H_A : \mu_X - \mu_Y > \delta$$

   (a) Derive the GLR (generalized likelihood ratio) test.

   (b) Show the rejection region is of the form

$$R = \left\{ \mathbf{x} : \frac{(\bar{x} - \bar{y} - \delta_0)}{\sqrt{S_n^2 \left(\frac{1}{n} + \frac{1}{m}\right)}} > c \right\}$$

   for some appropriate constant c.

7. 13.8.1. In addition state the statistical model and the parameter space for this problem, and the null hypothesis that you are testing. You may use either Pearson's chi squared statistic or the generalized likelihood ratio.

8. (non parametric bootstrap)

   See the R script with this assignment A4-scripthandout-2015.r

   It will analyze some data from North Carolina births. You will read in the data but just use the gestation period of 39 weeks.

   There are also some additional covariates in the data, but for this analysis we will only study the gestation length 39 weeks.

   You will use the nonparametric bootstrap to construct confidence intervals for population means (as discussed in class) and also for a population variance based on

$$W = \frac{S^2}{\mathrm{Var}_F(X)} \ .$$

   In particular you will have to work with the r.v.

$$W^* = \frac{S^{*2}}{\mathrm{Var}_{F_n}(X)}$$

   Use the bootstrap method to obtain a 95% confidence interval for $\sigma^2 = \mathrm{Var}_F(X)$.

   The data set is fairly large, with a sample of size n = 327. You will also analyze a random subsample of size n = 30 from this data. Note that it is almost impossible for 2 students to have the same random subset of size 30 for this part of the problem.

   The details of this question are in the R script.