

Homework Assignment 2
(Due Wednesday, September 30, 2015)

Please hand in a print-out of your answer and R code, and also email your R code to Zhiyuan (Jason) Xu <xu0284@umn.edu>. Note: John Verzani's simpleR notes is available on <https://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>

1. Problem 6.5-6.9 from John Verzani's simpleR notes. (20 points)
2. Problem 2.2 (page 62) from Foulkes's book. (10 points)
3. Problem 2.4 (page 62) from Foulkes's book. (10 points)
4. In example 2.6 (page 52) from Foulkes's book, write down the first 5 rows of the design matrix in the full model. [Hint: page 51] (10 points)
5. Departure from Hardy-Weinberg equilibrium is often a sign of potential genotyping problems. Among 1,000 study subjects, we see 649 AA, 300 AB, and 51 BB at one locus, and 640 AA, 360 AB and 0 BB at another locus. Use χ^2 test to find out whether or not the loci are in Hardy-Weinberg equilibrium. (5 points)
6. This exercise is for practicing central limit theorem.
 - (a) Draw 15 samples from Normal(mean=3, sd=1.5) and repeat this experiment 200 times, plot the distribution of sample means. Calculate the percentage of times when sample mean greater than or equal to 3.2. (5 points)
 - (b) Repeat (a) for the case the population SD is 1.5 but n=100 (2 points)
 - (c) Compare the sample distributions obtained in (a) and (b) to the population distribution, Normal(mean=3, sd=1.5), what do you observe? (3 points)
7. Consider data on the treatment response of 12 mice from strain A and 9 mice from strain B.

Strain A: 132 72 102 115 59 103 86 159 60 94 80 97

Strain B: 101 96 93 106 81 77 106 97 74

Assume that the measurements from strain A are independent draws from a normal distribution with mean μ_A and SD σ_A , and that the the measurements from strain B are independent draws from a normal distribution with mean μ_B and SD σ_B .

- (a) Calculate an approximate 95% confidence interval for the difference between the strain means, allowing for the possibility that the two strains have different SDs. Derive the answers using R function `t.test()`. (2 points)
 - (b) Test the hypothesis $H_0 : \mu_A = \mu_B$ versus the alternative $H_a : \mu_A \neq \mu_B$ using `t.test()`. (3 points)
 - (c) Use the `wilcox.test()` function to perform a rank-sum test on these data. (3 points)
 - (d) Implement a permutation test to test for a difference between the strains. (5 points)
 - (e) Give an interpretation of your results. (5 points)
8. Suppose we measure some treatment response on a set of 10 mice from strain A, and the same sort of treatment response on a set of 10 mice from strain B. We are interested in testing the hypothesis whether or not there is a difference in treatment responses of strains A and B.
- (a) Assume you did the test at a 5% significance level, and the p-value you obtained was 0.057. Interpret your findings. (3 points)
 - (b) Assume your friend in the lab did exactly the same experiment. Why would you not expect that he/she gets exactly the same p-value? Would you expect that he/she gets roughly the same p-value? (3 points)
9. Use the FAMuss data to test if there are differences in non-dominant arm muscle strength (NDRM.CH) between `actn3_r577x` genotype groups.
- (a) Plot the data [Hint: use **stripchart**] (5 points)
 - (b) Obtain the analysis of variance table, and comment on the results. (5 points)
 - (c) Remove two biggest observations in each genotype group, repeat (a) and (b). (5 points)
 - (d) Perform Kruskal-Wallis test and comment on the results from the three tests. (3 points)