

Quantitative Genomics and Genetics - Spring 2016
BTRY 4830/6830; PBSB 5201.01

Homework 4

Assigned March 1; Due 11:59PM March 7

Problem 1 (Easy)

- Describe the type of cases where a ‘one-sided’ test would be justified over a ‘two-sided’ test and explain why such tests would be justified in these cases.
- We never know the true parameter value for a system / experiment such that the chances that any null hypothesis we would select to test would be correct would be vanishingly small. Given that this is the case, why is it worthwhile testing a null hypothesis?

Problem 2 (Medium)

Many of the following questions will require R code (!!)

provide a separate text file with your R code used to generate your answers!

For the questions a-j below, consider a heights in the US system, a ‘measure one person’ experiment, a random variable $X \sim N(\mu, \sigma^2)$ that reflects heights after scaling (i.e., the true heights after subtracting a constant value and dividing by a constant).

For questions a-h below, consider a test statistic $T = \text{mean}(\mathbf{x})$ (where recall that the sampling distribution of this statistic is $T \sim N(\mu, \frac{\sigma^2}{n})$) and consider the unrealistic case where we know the value of σ^2 (assigned to different values depending on the question).

- Assume that $\sigma^2 = 1$. Consider $H_0 : \mu = 0$ and $H_A : \mu > 0$. For an iid sample of size $n = 10$, make use of the R function ‘qnorm()’ to determine c_α the critical value of the test statistic when the Type 1 error $\alpha = 0.05$.
- Assume that $\sigma^2 = 1$. Consider $H_0 : \mu = 0$ and $H_A : \mu \neq 0$. For an iid sample of size $n = 10$, make use of the R function ‘qnorm()’ to determine c_α the critical value of the test statistic when the Type 1 error $\alpha = 0.05$.
- Code a function that simulates M different iid samples of size n assuming each experimental trial within a sample is $X \sim N(\mu, \sigma^2)$, where the function also calculates the test statistic

$T = \text{mean}(\mathbf{x})$ for each sample, and the function outputs the number of times the value of the test statistic is greater than or equal to the critical value c_α given a ‘one-sided’ or ‘two-sided’ test (i.e., the inputs to your function should be $\mu, \sigma^2, M, n, c_\alpha$, and an indicator of ‘one-sided’ or ‘two-sided’ test, and note that you will use the R function ‘`rnorm()`’ within your function). Simulate $M = 1000$ samples of size $n = 10$ assuming that $\sigma^2 = 1$ and that the null hypothesis $H_0 : \mu = 0$ is correct, and choose a critical value corresponding to $\alpha = 0.05$ for a one-sided test. Do the same when considering a two-sided test. What was the number you expected as output in each case? Why did it not output exactly the number you expected (unless you were extremely lucky...)?

- d. Assume that the true values of the parameters for the system are $\mu = 0.1$ and $\sigma^2 = 1$. Make use of the R function ‘`pnorm()`’ to calculate the power of the test described in part ‘a’.
- e. Assume that the true values of the parameters for the system are $\mu = 0.1$ and $\sigma^2 = 1$. Make use of the R function ‘`pnorm()`’ to calculate the power of the test described in part ‘b’.
- f. Repeat part ‘d’ but assume that $\sigma^2 = 10$ for the distribution of the test statistic under the null hypothesis and that the true $\sigma^2 = 10$.
- g. Repeat part ‘e’ but assume that $\sigma^2 = 10$ for the distribution of the test statistic under the null hypothesis and that the true $\sigma^2 = 10$.
- h. Make use of your function in part ‘c’ to simulate the number of times out of $M = 1000$ that your test statistic exceeds the critical value corresponding to $\alpha = 0.05$ for a one-sided test with $\sigma^2 = 1$ and null hypothesis $H_0 : \mu = 0$ but when assuming the TRUE parameter values are $\mu = 0.1$ and $\sigma^2 = 1$ (you should not have to alter your function, just the inputs!). Also do this for the same case but for a two-sided test. What was the number you expected as output in each case? Why did it not output exactly the number you expected (unless you were extremely lucky...)? Note that you have just performed a power analysis!!
- i. Assume that you have observed the following sample $\mathbf{x} = [1.19, 0.33, 0.19, -1.89, 0.49, 0.08, 0.98, 2.92, 1.31, 1.39]$. Calculate the likelihood ratio test statistic $LRT = -2\ln(\Lambda)$ for this sample for $H_0 : \mu = 0$ and $H_A : \mu \neq 0$ and calculate the p-value by substituting the value of this test statistic into the R function ‘`pchisq(LRT, 1)`’.
- j. The sampling distribution of the LRT is exactly chi-square as $n \rightarrow \infty$. In part ‘i’ (or any realistic case) your sample size does not approach infinite, so does this mean that what you calculated in part ‘i’ is not the correct p-value? If not, why might we treat it as if it is the correct p-value (i.e., why is this justified)?

Problem 3 (Difficult)

Consider a ‘one flip’ experiment, a random variable $X = \text{‘number of Heads’}$, a bernoulli probability model $X \sim \text{bern}(p)$ with true value p . We want to test the null hypothesis $H_0 : p = c$ and $H_A : p > c$ when considering the test statistic $T = \text{mean}(\mathbf{x})$ and an iid sample produced by n experimental trials. Write down two closed form equations for this case, the first that will allow you to calculate the Type I error and the second that will allow you to calculate the power, when assuming we have selected the critical value c_α of the test statistic T to be an integer between 0 and 10 (hint: they

will be the same except for swapping c_α and p !). Use these two equations to calculate the Type 1 error of $H_0 : p = 0.5$ and $H_A : p > 0.5$ for $n = 10$ when using the critical value $c_\alpha = 7$ and to calculate the power of this test when assuming the true parameter value is $p = 0.6$ (note that you may do the actual calculations using your equations in R but you do not need to include your R code, i.e., just provide your equations and the answers).