STA 378 Assignment 3

Crystal

1. Please obtain the formula for `ncp` at the bottom of Page 6 in the main paper. The easiest way to do it is to use a regression model with no intercept and 0-1 indicator variables for group membership.
   (a) What does the $\mathbf{X}$ matrix look like? The number of rows is $n_1 + n_2$.
   (b) What is $\mathbf{X}'\mathbf{X}$?
   (c) What is $(\mathbf{X}'\mathbf{X})^{-1}$? See why this was a good dummy variable coding scheme?
   (d) What are $\mathbf{L}$ and $\mathbf{h}$ in $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{h}$?
   (e) Now you can just use Expression (1.2) in the technical supplement and simplify.
2. We are testing the null hypothesis $\mu_1 = \mu_2$ versus $\mu_1 \neq \mu_2$ at $\alpha = 0.05$. Suppose the effect size $d = \frac{1}{2}$ and $n_1 = n_2 = 84$. Use R to calculate the power of the test. The answer is a number.
3. Suppose $n_1 = n_2 = 50$. What value of d gives a value of exactly 0.25? the answer is a number. Use R. Look at `help(uniroot)`. If you do not know how to write your own function in R, you may have to learn. My `fstat` function (see Jessie's homework) is an example.
4. Let the test statistic T be continuous, with pdf $d(t;\lambda)$ and cdf $p(t;\lambda)$. The quantity λ is the non-centrality parameter. Following the notation in the technical supplement, $\lambda = f_1(n)f_2(\texttt{es})$. The null hypothesis is rejected if $T > c$. We have a random sample $T_1, \ldots, T_k$ from this distribution, but they are selected for statistical significance, so that $T_j > c$ for $j = 1, \ldots, k$.
   (a) What is the conditional probability $Pr\{T \leq t | T > c\}$ for $t > c$? Express the answer in terms of F.
   (b) Obtain the conditional density of T given $T > c$ by differentiating your answer to 4a.
   (c) Ordinarily, the likelihood function is a product of densities, but in the case of selection for significance it is a product of conditional densities like the one you just derived. Write the likelihood function, assuming that the sample sizes $n_1, \ldots, n_k$ could be different, but the effect size `es` is the same for all the tests. What is the parameter here?
   (d) Write the log likelihood function.

Jessie

1. In the usual regression model, the design matrix $\mathbf{X}$ contains only fixed constants. This is weird, but for now we will accept it. We'll use the $\mathbf{X}$ matrix from the SAT data. Get the data with

```
sat = read.table("http://www.utstat.toronto.edu/~brunner/data/legal/openSAT.data.txt")
```

   and type `sat` or maybe just `head(sat)` to see what you have.

   (a) Write the regression model in scalar form.

   (b) Fit the regression model with `lm`.

   (c) Use R to calculate $\mathbf{X}'\mathbf{X}$ and $(\mathbf{X}'\mathbf{X})^{-1}$. See `help(lm)` for an easy way to obtain the $\mathbf{X}$ matrix.

2. Now test the null hypothesis $H_0 : \beta_2 = 0$ two ways. Obtain the test statistic and p-value, both numbers.

   (a) Using the output of `summary`.

   (b) Using the general linear F-test. You can use my function: `source("http://www.utstat.utoronto.ca/~brunner/Rfunctions/ftest.txt")` and then `ftest` to see the function definition.

3. Suppose the true value of β_2 is 0.001 and the true value of σ^2 is 0.3.

   (a) Use R to calculate the value of the non-centrality parameter. The answer is a number. You can just use Expression (1.2) in the technical supplement.

   (b) Use R to calculate the power of the F-test for $H_0 : \beta_2 = 0$. The answer is a number.

4. What value of $\frac{\beta_2}{\sigma}$ is required for

   (a) A power of 0.25?

   (b) A power of 0.50?

   (c) A power of 0.75?

   The answers are numbers. Use R. Look at `help(uniroot)`. If you do not know how to write your own function in R, you may have to learn. My `fstat` function is an example.