BMI 713: Computational Statistics for Biomedical Sciences

Assignment 3

September 23, 2010 (due Sept. 30)

1. Was Tyrannosaurus Rex Warm-Blooded?

   It is known that the oxygen isotopic composition of vertebrate bone phosphate is related to the body temperature at which the bone forms. Differences in means at different bone sites would indicate nonconstant temperatures throughout the body. Minor temperature differences would be expected in warm-blooded animals. The dataset bone_oxygen.txt on the course website shows several measurements of the oxygen isotopic composition of bone phosphate in each of 6 bone specimens from a single Tyrannosaurus rex skeleton. (Data modified from R. E. Barrick, and W. J. Showers, Thermophysiology of *Tyrannosaurus* rex: Evidence from Oxygen Isotopes, *Sciences* 265 (1994): 222-224.)

   Read the data file bone_oxygen.txt from the course website, and save it as a data frame in R:

   ```
   t.rex<-read.table("http://compbio.med.harvard.edu/BMI713/bone\_oxygen.txt",
   header=T, sep="\t")
   ```

   a. Is there any evidence that the oxygen isotopic composition of proximal caudal bone is different from that of mid-caudal bone? Please state the null hypothesis H_0 and the alternative hypothesis H_1, choose a proper test for the comparison, and calculate the test statistic, p-value, and the 95% confidence interval for the true mean difference. Can the observed difference be attributed to chance?

   b. Suppose the samples are from normal populations, use F-test to test if the oxygen isotopic compositions of proximal and mid-caudal bones have equal variances. What is the degree of freedom?

   c. Is there evidence that the means are different for all the 6 different bones? Calculate the sum of squares and mean squares for within and between groups, construct an ANOVA table, perform the F-test and find the p-value. What inference can you make? (R hints: `aov` function can be used to fit an ANOVA model, and `summary` function can be used to summarize the results of model fitting functions.)

2. True or false. If the answer is false, please explain your reason briefly.

   a. The significance level α of a test is equal to the probability that the null hypothesis is true.

   b. If the significance level α of a test is decreased, the power would be expected to increase.

   c. The power of a test is equal to the probability that the null hypothesis is rejected.

   d. A type I error occurs when the test statistic falls in the rejection region of the test.

   e. The power of a test is determined by the null distribution of the test statistic.

3. Let $X_1, ..., X_{25}$ be a random sample from a normal distribution with mean μ_1 and standard deviation 10.

   a. What is the power of the test of H_0: $\mu_0 = 0$ versus H_A: $\mu_1 = 1.5$ at level $\alpha = 0.05$ of H_0: $\mu_0 = 0$?

   b. What is the power of the test of H_0:$\mu_0 = 0$ versus H_A: $\mu_1 = 3$ at level $\alpha = 0.05$?

   c. Graph the power of the test of H_0 :$\mu_0 = 0$ versus H_A: $\mu_1 = 1.5$ as a function of α.

   d. How large should the sample size be, so that the test of H_0: $\mu_0 = 0$ versus H_A: $\mu_1 = 1.5$ has a power greater than 0.5 if $\alpha = 0.05$?

   e. How do the factors n, σ, α, $|\mu_1 - \mu_0|$ affect the power in general?

R hints

i. You can create a function to calculate power, like:

```
power <- function(n, mu0, mu1, alpha, sigma) {
    # your expression to calculate power
                                            }
```

then, `power(25, 0, 1.5, 0.05, 10)` will give the power in a.

ii. An example of `plot` function for c.:
To plot the curve for $y = x^2$, $2 \leq x \leq 2$, in R we can do:

```
x <- seq(from=-2, to=2, by=0.01)
y <- x^2
plot(x, y, type="l")
```