# 2/25/2010 36-402/608 ADA-II H. Seltman
## Breakout #13 Results

Remember the bran study with 20 subjects given three diets (baseline, hi-fiber, and lo-fiber) in random order to see how they affect cholesterol. (We saw that order appeared to have no effect, and will not use it as a variable.) One way to express the scientific hypotheses is that we are simultaneously interested in testing $\mu_B = (\mu_L + \mu_H)/2$ and $\mu_H = \mu_L$. Although we could do two separate paired t-tests, we will do a single overall one-sample test of the two hypotheses.

```
BHL = bran$hifiber-bran$lofiber
BF = (bran$lofiber+bran$hifiber)/2-bran$baseline
means = c(mean(BF),mean(BHL))
n = nrow(bran)
T2 = n * as.numeric(means %*% solve(cov(cbind(BF,BHL))) %*% means)
T2 # 19.70
F = (n-2)/2/(n-1)*T2
F # 9.33
1-pf(F, 2, n-2) # 0.00166
```

**Question 1: How does the code relate to the formulas in the handout? What conclusion do you reach? What followup testing should be done?**

Recall the flea beetle study in which two different measurements are taken on two similar species of beetles. The question of interest is whether the collection of p=2 measurements are a distinguishing feature between the species (though not necessarily useful for distinguishing individuals).

```
> anova(aov(beet[,1:2]~species, data=beet), test="Hotelling")
# Error in model.frame.default(formula = beet[, 1:2] ~ species, data = beet,  :
#   invalid type (list) for variable 'beet[, 1:2]'

anova(aov(as.matrix(beet[,1:2])~species, data=beet), test="Hotelling")
# Analysis of Variance Table
#             Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
# (Intercept)  1           344.15   5678.4      2     33 < 2.2e-16 ***
# species      1             4.81     79.4      2     33 2.455e-13 ***
# Residuals   34
```

**Question 2: Why did the first attempt fail? How does the code relate to the formulas in the handout? How does `anova()` "know" to do MANOVA? What conclusion do you reach? How can you get a small p-value and then perhaps find that these measurements are not very useful for categorizing individual beetles?**

Recall the monkeys being tested for short and long term memory with and without brain surgery on the hippocampus.

```
mem$SL = with(mem,cbind(short=(week2+week4)/2,
                        long=(week8+week12+week16)/3))
anova(aov(SL~treatment, data=mem), test="Hotelling")
# Analysis of Variance Table
#             Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
# (Intercept)  1          250.795  1880.96      2     15 < 2.2e-16 ***
# treatment    1            1.643    12.32      2     15 0.0006831 ***
# Residuals   16
```

**Question 3: What is the first statement doing? What are the values of p and K? What conclusion do you reach? What could you do to check assumptions?**