Professor Notes About Linear Models Foundations Homework 2

- Tables (and figures) should be labeled as described in the homework format description. Table labels go ABOVE the table and figure labels go BELOW the table. Tables (and figures) should be referred to in your answers. See the key below for a model of this.
- Use complete sentences to answer questions.
- Keep "many" decimals in intermediate calculations ... i.e., don't round until the final answer.
- Note explanations in the key below.

Fish Oil and Blood Pressure

1. Results from `t.test()` are in Table 1.

Table 1. Results from 2-sample t-test of diastolic blood pressure by diet type.

```
 Two Sample t-test with DBP by diet
t = 3.0621, df = 12, p-value = 0.009861
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  2.225174 13.203398
sample estimates:
    mean in group Fish mean in group Standard
              6.571429              -1.142857
```

2. The ANOVA table (from `anova()`) is in Table 2.

Table 2. Analysis of variance table for the diastolic blood pressure by diet type.

```
           Df Sum Sq Mean Sq F value   Pr(>F)
diet        1 208.29 208.286  9.3762 0.009861
Residuals  12 266.57  22.214
```

3. The coefficients table (from `summary()`) is in Table 3.

Table 3. Coefficient results from the one-way ANOVA fit of diastolic blood pressure by diet type.

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)     6.571      1.781   3.689  0.00310
dietStandard   -7.714      2.519  -3.062  0.00986
---
Residual standard error: 4.713 on 12 degrees of freedom
Multiple R-squared: 0.4386,Adjusted R-squared: 0.3918
F-statistic: 9.376 on 1 and 12 DF,  p-value: 0.009861
```

4. The overall p-value from the ANOVA table ($p = 0.0099$; Table 2), for the slope coefficient ($p = 0.0099$; Table 3), and for the two-sample t-test ($p = 0.0099$; Table 1) are equal. These p-values are all equivalent because the 2-sample t-test hypothesis of equal means (or difference in means equals zero) is the same as the hypothesis for the slope (see below about the slope representing the difference in means) which is the same as the hypothesis for the ANOVA table (i.e., simple model of one mean representing both groups).
5. With these p-values, very strong evidence to reject the null hypothesis exists. Thus, the group means appear to be different.

6. The intercept coefficient for the linear model (6.57; Table 3) is the same as the mean of the first (Fish) group in the 2-sample t-test (6.57; Table 1). This occurs because an intercept is defined as the "value of Y when X=0, on average" and the fish group is coded with a zero in `lm()`.
7. The difference in the means (i.e., -1.14-6.57 = -7.71; Table 1) is the same as the slope coefficient in the linear model coefficients (i.e., -7.71; Table 3). This is because the slope coefficient shows the change in Y (DBP) for a one "unit" change in X ("diet group"), which is a change from the "Fish" group (coded with a zero) to the "Standard" group (code with a one) because of the coding in `lm()`.
8. The df from the two-sample t-test (12; Table 1) and the within-group df from the ANOVA table(12; Table 2) are identical. The within-group df are equal to the total number of individuals ($n = n_1 + n_2$) minus the number of groups ($I = 2$), which is the same as for the 2-sample t-test.
9. The F test statistic (9.38) is equal to the square of the t test statistic (3.06^2=9.38). This relationship occurs when the numerator df for the ANOVA is equal to one (i.e., there are only two groups).
10. The SE for the difference in means is equal to $\frac{\bar{x}_1 - \bar{x}_2}{t} = \frac{6.571429 + 1.142857}{3.0621} = 2.519280$. The pooled variance is then equal to this value divided by the sum of the reciprocals of the sample sizes – i.e., $\frac{SE^2_{\bar{x}_1 - \bar{x}_2}}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2.519280^2}{\frac{1}{7} + \frac{1}{7}} = 22.214$. This is the same as MS_{within} (Table 1).

R Appendix.

```
library(NCStats)
setwd("T:/Biometry/")
d <- read.csv("FishOil.csv")
fish.t <- t.test(DBP~diet,data=d,var.equal=TRUE)
lm1 <- lm(DBP~diet,data=d)
anova(lm1)
summary(lm1
```