STAT502 Lab #2

1. An experiment was conducted to test the effects of different diets on turkeys. Thirty turkeys were randomly assigned to five diet groups, six per group. The measured response is weight gain (in pounds). We are going to compute all the quantities needed for the one-way ANOVA F-test and put them in the ANOVA table. We will also make a conclusion from the F-test.

Group	Weight gains					
Diet 1	4.1	3.3	3.1	4.2	3.6	4.4
Diet 2	5.2	4.8	4.5	6.8	5.5	6.2
Diet 3	6.3	6.5	7.2	7.4	7.8	6.7
Diet 4	6.5	6.8	7.3	7.5	6.9	7.0
Diet 5	9.5	9.6	9.2	9.1	9.8	9.1

Use the following commands to enter the data for the five diets. The `as.factor()` function tells R to treat the diet numbers as categorical labels.

```
Diet_1 = c( 4.1, 3.3, 3.1, 4.2, 3.6, 4.4 ); n1 = length(Diet_1)
Diet_2 = c( 5.2, 4.8, 4.5, 6.8, 5.5, 6.2 ); n2 = length(Diet_2)
Diet_3 = c( 6.3, 6.5, 7.2, 7.4, 7.8, 6.7 ); n3 = length(Diet_3)
Diet_4 = c( 6.5, 6.8, 7.3, 7.5, 6.9, 7.0 ); n4 = length(Diet_4)
Diet_5 = c( 9.5, 9.6, 9.2, 9.1, 9.8, 9.1 ); n5 = length(Diet_5)
y = c(Diet_1, Diet_2, Diet_3, Diet_4, Diet_5)
x = as.factor(c(rep(1,n1), rep(2,n2), rep(3,n3), rep(4,n4), rep(5,n5)))
y; x
```

(a) Use `boxplot(y~x)` to create boxplots of the weight gain for the five diets. Considering the difference between variation *among* groups and variation *within* groups, comment on the evidence for different population means.

(b) Compute the total sum of squares SSTO. As a hint, the R function `var(y)` will calculate (what are r, n_i, and n_T?)

$$\frac{1}{n_T}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_{..})^2$$

(c) Find the individual means $\overline{Y}_{i\cdot}$ and the overall mean $\overline{Y}_{..}$. Then, compute the treatment sum of squares SSTR. As a hint, SSTR can be written as

$$\text{SSTR} = n_1(\overline{Y}_{1\cdot}-\overline{Y}_{..})^2 + n_2(\overline{Y}_{2\cdot}-\overline{Y}_{..})^2 + \cdots + n_r(\overline{Y}_{r\cdot}-\overline{Y}_{..})^2$$

(d) Find the individual variances s_i^2, and then compute the error (residual) sum of squares SSE. As a hint, SSE can be written as

$$\text{SSE} = (n_1-1)s_1^2 + (n_2-1)s_2^2 + \cdots + (n_r-1)s_r^2$$

(e) Verify from your answers in parts (a), (b), and (c) that SSTO = SSTR + SSE.

(f) Compute the mean squares for treatment MSTR and error MSE. Finally, compute the F-statistic for testing the hypotheses

$$H_0: \mu_1 = \cdots = \mu_r \quad \text{versus} \quad H_a: \text{at least one} \neq$$

(g) With $\alpha = .05$, what is the critical value for rejecting H_0? What is the p-value? Use the functions `qf()` and `pf`. Based on your answers here, state your conclusion.

(h) Repeat the above analysis using the R functions below. Do you get the same answers?

```
mod1 = lm(y ~ x)
anova(mod1)
```

2. You may do this problem by hand or in R. It is recommended you use R to gain more familiarity with it.

   The following table summarizes some statistics for $r = 4$ groups.

i	n_i	$\bar{Y}_{i.}$	s_i^2
1	10	10.69	0.830
2	12	12.17	0.894
3	8	11.95	0.510
4	9	11.82	0.741

   Use it to complete the ANOVA table.

Source	df	SS	MS	F	p value
Treatment	--	-----	-----	-----	-----
Error	--	-----	-----		
Total	--	-----			