STAT 222 HW 1

Matthew Zhao

Problem 1

(a) $\bar{y}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{g} n_i \bar{y}_{i\cdot} = \frac{1}{7+8+6+8}(7*3.7457 + 8*3.5800 + 6*3.5983 + 8*3.9225) = \frac{1}{29} * 107.8297 \approx 3.718266$

$SS_{trt} = \sum_{i=1}^{g} n_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = 7*(3.7457 - 3.718266)^2 + 8*(3.5800 - 3.718266)^2 + 6*(3.5983 - 3.718266)^2 + 8*(3.9225 - 3.718266)^2 \approx 0.5782515$

(b) $SSE = \sum_{i=1}^{g}(n_i - 1)s_i^2 = (7-1)*0.2840^2 + (8-1)*0.1821^2 + (6-1)*0.0962^2 + (8-1)*0.1971^2 \approx 1.03427$

(c)

Source	df	SS	MS	F-stat
Treatment	$\text{df}_{\text{trt}} = 4 - 1 = 3$	$\text{SS}_{\text{trt}} = 0.5782515$	$\text{MS}_{\text{trt}} = \frac{\text{SS}_{\text{trt}}}{\text{df}_{\text{trt}}} = \frac{0.5782515}{3} \approx 0.1927505$	$\text{F} = \frac{\text{MS}_{\text{trt}}}{\text{MSE}} = \frac{0.1927505}{0.0413708} \approx 4.659095$
Error	$\text{df}_{\text{E}} = 29 - 4 = 25$	$\text{SSE} = 1.03427$	$\text{MSE} = \frac{\text{SSE}}{N-g} = \frac{1.03427}{25} \approx 0.0413708$	

(d) We run the code below to find the p value.

```
pf(4.659095,3,25, lower.tail = FALSE)
```

```
## [1] 0.01014922
```

Since $0.05 > p = 0.01014922$, we can reject the null hypothesis that all the means are equal at the 95% confidence level, potentially indicating that the treament effects are statistically significant.

Problem 2

```
rats = read.table("http://users.stat.umn.edu/~gary/book/fcdae.data/ex3.1", h=T)
anova(lm(y ~ diet, data = rats))
```

```
## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq  Mean Sq F value Pr(>F)
## diet       1 0.14903 0.149030  2.7493 0.1089
## Residuals 27 1.46358 0.054207
```

```r
anova(lm(y ~ as.factor(diet), data = rats))
```

```
## Analysis of Variance Table
##
## Response: y
##                 Df  Sum Sq  Mean Sq F value  Pr(>F)
## as.factor(diet)  3 0.57821 0.192736  4.6581 0.01016 *
## Residuals       25 1.03440 0.041376
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 3

The 95% confidence interval for a group mean for group k is denoted $\bar{y}_{k\cdot} \pm t_{N-g,\frac{\alpha}{2}} * \frac{\sqrt{MSE}}{\sqrt{n_k}}$. Specifically, we can find $t_{N-g,\frac{\alpha}{2}}$ for the 95% confidence level and group $k = 1$ using the R code below (where $N = 29$, $g = 4$, and $\frac{\alpha}{2} = 0.025$):

```r
qt(0.05/2,29-4,lower.tail = F)
```

```
## [1] 2.059539
```

Plugging the values back in, we can construct a confidence interval for the group mean μ_1:

$\mu_1 = 3.7457 \pm 2.059539 * \frac{\sqrt{0.0413708}}{\sqrt{7}} = 3.7457 \pm 0.1583317$, and so we can write that the 95% confidence interval for μ_1 is $\mu_1 \in (3.587368, 3.904032)$.

Problem 4

Mean diff $= \bar{y}_{k\cdot} - \bar{y}_{l\cdot}$:

```r
#tidyverse initalization hidden
rats_sum <- rats %>% group_by(as.factor(diet)) %>% summarise(group_mean = mean(y))
for (x in 1:(nrow(rats_sum)-1)) {
  for (y in (x+1):nrow(rats_sum)) {
    print(rats_sum$group_mean[y]-rats_sum$group_mean[x])
  }
}
```

```
## [1] -0.1657143
## [1] -0.147381
## [1] 0.1767857
## [1] 0.01833333
## [1] 0.3425
## [1] 0.3241667
```

$\text{SE}(\bar{y}_{k\cdot} - \bar{y}_{l\cdot}) = \sqrt{\text{MSE}(\frac{1}{n_k} + \frac{1}{n_l})}$

```
rats_count <- rats %>% count(as.factor(diet))
for (x in 1:(nrow(rats_count)-1)) {
  for (y in (x+1):nrow(rats_count)) {
    print(sqrt((0.0413708)*((1/rats_count$n[x])+(1/rats_count$n[y]))))
  }
}
```

```
## [1] 0.1052685
## [1] 0.1131603
## [1] 0.1052685
## [1] 0.1098475
## [1] 0.1016991
## [1] 0.1098475
```

$\text{t-stat} = \frac{\bar{y}_{k\cdot} - \bar{y}_{l\cdot}}{\sqrt{\text{MSE}(\frac{1}{n_k} + \frac{1}{n_l})}}$

```
rat_comb <- left_join(rats_sum,rats_count,by = "as.factor(diet)")
for (x in 1:(nrow(rat_comb)-1)) {
  for (y in (x+1):nrow(rat_comb)) {
    print((rat_comb$group_mean[y]-rat_comb$group_mean[x])/(sqrt((0.0413708)*((1/rat_comb$n[x])+(1/rat_c
  }
}
```

```
## [1] -1.574205
## [1] -1.302409
## [1] 1.679379
## [1] 0.166898
## [1] 3.367779
## [1] 2.95106
```

Finding p-values (N-g $= 29 - 4 = 25$):

```
for (x in 1:(nrow(rat_comb)-1)) {
  for (y in (x+1):nrow(rat_comb)) {
    print(2*pt(abs
      (as.numeric((rat_comb$group_mean[y]-rat_comb$group_mean[x])/
        (sqrt(
          (0.0413708)*((1/rat_comb$n[x])+(1/rat_comb$n[y]))
              )
         ))),25,lower.tail = FALSE))
  }
}
```

```
## [1] 0.1280115
## [1] 0.2046479
## [1] 0.1055351
## [1] 0.8687929
## [1] 0.002456
## [1] 0.006789461
```

diet pair	mean diff	Standard Error (SE)	t-stat	p-value
2-1	-0.1657143	0.1052685	-1.574205	0.1280115
3-1	-0.147381	0.1131603	-1.302409	0.2046479
4-1	0.1767857	0.1052685	1.679379	0.1055351
3-2	0.01833333	0.1098475	0.166898	0.8687929
4-2	0.3425	0.1016991	3.367779	0.002456
4-3	0.3241667	0.1098475	2.95106	0.006789461

4-2 and 4-3 are significant at the 5% significance level.

Underline Diagram

```
  2      3       1      4
 3.58   3.60   3.75   3.92
-------------------
              ------------
```