

STAT 222 HW 1

Matthew Zhao

Problem 1

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

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

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

(c)

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

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

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

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

We can conclude that the treatment effects are not equal at the 95% confidence level since $0.05 > p = 0.01014922$.

Problem 2

```
rats = read.table("http://users.stat.umn.edu/~gary/book/fcdade.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
```

```
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 $\bar{y}_{k.} \pm t_{n_k-1, \frac{\alpha}{2}} * \frac{s_k}{\sqrt{n_k}}$. Specifically, we can find $t_{n_k-1, \frac{\alpha}{2}}$ for the 95% confidence level and group $k = 1$ using the R code below (where $n_1 = 7$ and $\frac{\alpha}{2} = 0.025$):

```
qt(1-0.025,6)
```

```
## [1] 2.446912
```

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

$\mu_1 = 3.7457 \pm 2.446912 * \frac{0.2840}{\sqrt{7}} = 3.7457 \pm 0.2626562$, and so we can write that the 95% confidence interval for μ_1 is $\mu_1 \in (3.483044, 4.008356)$.

Problem 4

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

```
#tidyverse initialization 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
```

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

```
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
```

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

```
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(pt(
      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))
  }
}
```

```
## [1] 0.06400576
## [1] 0.102324
## [1] 0.9472324
## [1] 0.5656035
## [1] 0.998772
## [1] 0.9966053
```

diet pair	mean diff	Standard Error (SE)	t-stat	p-value
2-1	-0.1657143	0.1052685	-1.574205	0.06400576
3-1	-0.147381	0.1131603	-1.302409	0.102324
4-1	0.1767857	0.1052685	1.679379	0.9472324
3-2	0.01833333	0.1098475	0.166898	0.5656035
4-2	0.3425	0.1016991	3.367779	0.998772
4-3	0.3241667	0.1098475	2.95106	0.9966053

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
