Stat/Math 415 Homework 8

Due Friday Nov 22; Joseph Sepich (jps6444)

1 Problem 9.3-1

This problem relates to the analysis of variacne (ANOVA). We need to use various error sources to determine if we should reject the null hypothesis. First of all let's define three error sources (where n and m are the sample size and number of groups respectively):

$$SS_E = SS_{TO} - SS_T = \Sigma_{i=1}^m \Sigma_{j=1}^n (x_{ij} - \overline{x_{..}})^2; MS_E = \frac{SS_E}{n-m}$$

$$SS_T = \Sigma_{i=1}^m (\overline{x_{i.}} - \overline{x_{..}})^2 n_i; MSTE = \frac{SS_T}{m-1}$$

$$SS_{TO} = \Sigma_{i=1}^m \Sigma_{j=1}^n (x_{ij} - \overline{x_{i.}})^2$$

We can then use these error sources to define the F Distribution:

$$F = \frac{MS_T}{MS_E}$$

And we reject the null hypothesis if $F \geq F_\alpha(m-1, n-m)$. Let's now calculate the required values:

```
x_1 <- c(5, 9, 6, 8)
x_2 <- c(11, 13, 10, 12)
x_3 <- c(10, 6, 9, 9)

x_matrix <- data.frame(x_1 = x_1, x_2 = x_2, x_3 = x_3)

n <- 4
m <- 3

mean_1 <- sum(x_1) / n
mean_2 <- sum(x_2) / n
mean_3 <- sum(x_3) / n

means <- c(mean_1, mean_2, mean_3)
total_mean <- sum(means) / m

SS_T <- 0
SS_E <- 0
SS_TO <- 0

for (i in 1:m) {
  SS_T <- SS_T + n * (means[i] - total_mean)^2
}

for (i in 1:m) {
  for (j in 1:n) {
```

```r
    SS_E <- SS_E + (x_matrix[j, i] - means[i])^2
  }
}

for (i in 1:m) {
  for (j in 1:n) {
    SS_TO <- SS_TO + (x_matrix[j, i] - total_mean)^2
  }
}

MS_T <- SS_T / (m-1)
MS_E <- SS_E / (n * m - m)

F_var <- MS_T / MS_E

print(paste0('SS(TO) is ', SS_TO))
```

```
## [1] "SS(TO) is 66"
```

```r
print(paste0('SS(T) is ', SS_T))
```

```
## [1] "SS(T) is 42"
```

```r
print(paste0('SS(E) is ', SS_E))
```

```
## [1] "SS(E) is 24"
```

```r
print(paste0('MS(T) is ', MS_T))
```

```
## [1] "MS(T) is 21"
```

```r
print(paste0('MS(E) is ', MS_E))
```

```
## [1] "MS(E) is 2.66666666666667"
```

```r
print(paste0('F test statistic is ', F_var))
```

```
## [1] "F test statistic is 7.875"
```

Using a significance level of $\alpha = 0.05$ we can reference the value of $F_\alpha(m-1, n-m) = F_{0.05}(2, 9) = 4.2565$ in the F distribution table reference. Since our F test statistic is $7.875 > 4.2565$, we can **reject** our null hypothesis (H_0) of $\mu_1 = \mu_2 = \mu_3$.

2 Problem 9.3-15

This problem relates to the analysis of variacne (ANOVA). We need to use various error sources to determine if we should reject the null hypothesis. First of all let's define three error sources (where n and m are the sample size and number of groups respectively):

$$SS_E = SS_{TO} - SS_T = \Sigma_{i=1}^{m}\Sigma_{j=1}^{n}(x_{ij} - \overline{x_{..}})^2; MS_E = \frac{SS_E}{n-m}$$

$$SS_T = \Sigma_{i=1}^{m}(\overline{x_{i.}} - \overline{x_{..}})^2 n_i; MSTE = \frac{SS_T}{m-1}$$

$$SS_{TO} = \Sigma_{i=1}^{m}\Sigma_{j=1}^{n}(x_{ij} - \overline{x_{i.}})^2$$

We can then use these error sources to define the F Distribution:

$$F = \frac{MS_T}{MS_E}$$

And we reject the null hypothesis if $F \geq F_\alpha(m-1, n-m)$. Let's now calculate the required values:

```
x_1 <- c(500, 650, 530, 680, NA)
x_2 <- c(700, 620, 780, 830, 860)
x_3 <- c(500, 520, 400, 580, 410)

x_matrix <- data.frame(x_1 = x_1, x_2 = x_2, x_3 = x_3)

n <- c(length(x_1) - 1, length(x_2), length(x_3))
m <- 3

mean_1 <- sum(x_1, na.rm = TRUE) / n[1]
mean_2 <- sum(x_2, na.rm = TRUE) / n[2]
mean_3 <- sum(x_3, na.rm = TRUE) / n[3]

means <- c(mean_1, mean_2, mean_3)

total_mean <- 0
for (i in 1:m) {
  total_mean <- total_mean + sum(x_matrix[,i], na.rm = TRUE)
}
total_mean <- total_mean / sum(n)

SS_T <- 0
SS_E <- 0
SS_TO <- 0

for (i in 1:m) {
  SS_T <- SS_T + n[i] * (means[i] - total_mean)^2
}

for (i in 1:m) {
  for (j in 1:n[i]) {
    if (is.na(x_matrix[j, i])) {
      break
    }
    SS_E <- SS_E + (x_matrix[j, i] - means[i])^2
  }
}

for (i in 1:m) {
```

```r
  for (j in 1:n[i]) {
    if (is.na(x_matrix[j, i])) {
      break
    }
    SS_TO <- SS_TO + (x_matrix[j, i] - total_mean)^2
  }
}

MS_T <- SS_T / (m-1)
MS_E <- SS_E / (sum(n) - m)

F_var <- MS_T / MS_E

print(paste0('SS(TO) is ', SS_TO))
```

```
## [1] "SS(TO) is 278171.428571429"
```

```r
print(paste0('SS(T) is ', SS_T))
```

```
## [1] "SS(T) is 193011.428571429"
```

```r
print(paste0('SS(E) is ', SS_E))
```

```
## [1] "SS(E) is 85160"
```

```r
print(paste0('MS(T) is ', MS_T))
```

```
## [1] "MS(T) is 96505.7142857143"
```

```r
print(paste0('MS(E) is ', MS_E))
```

```
## [1] "MS(E) is 7741.81818181818"
```

```r
print(paste0('F test statistic is ', F_var))
```

```
## [1] "F test statistic is 12.4655102999396"
```

Using a significance level of $\alpha = 0.05$ we can reference the value of $F_{\alpha}(m-1, n-m) = F_{0.05}(2, 11) = 3.9823$ in the F distribution table reference. Since our F test statistic is $12.47 > 3.9823$, we can **reject** our null hypothesis (H_0) of $\mu_1 = \mu_2 = \mu_3$. This implies that different feed supplements do **not** have the same affect on cow weight.

3 Problem 8.4-3

3.1 Part a

3.2 Part b

3.3 Part c

3.4 Part d

4 Problem 8.4-7

4.1 Part a

4.2 Part b

4.3 Part c

5 Problem 8.4-15