HW 3

Andrew Jowe

1

   a. $\phi = 3.3599$

   b. Using $\phi = 3.3599$, $d.f.\{SSA\} = 2$, and $d.f.\{SSE\} = 21$ from HW2, we use the table to find that the power is 0.99.

   c. $\frac{\hat{\Delta}}{\sigma_\epsilon} = 3.1429$

   d. Using the table, we find that the minimum sample size for each group is 4.

2

   a. For all $Y_{ij} = \mu_i + \gamma_i + \epsilon_{ij}$, we are interested to know the factor effect (γ_i) and confidence interval $(\mu_i - \mu_.)$ of each group. The assumptions are:

      i. Y_{ij} were randomly sampled (independent)
      ii. The i groups are independent for all $i = 1, ..., a$
      iii. Errors are independent and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$

   b. $\gamma_1 = 6.6667, \gamma_2 = 0.6667, \gamma_3 = -7.3333$

   c. $\gamma_2 = 0$. Therefore the mean for group 2 is equal to the overall mean. γ_2 is the factor effect, or how the overall mean is effected by a subject being in group 2.

   d. $\sum \gamma_i = 0$ Proof: $\gamma_i = \bar{Y}_{i.} - \bar{\bar{Y}} \sum \gamma_i = \sum(\bar{Y}_{i.} - \bar{\bar{Y}}) \to \sum(\bar{Y}_{i.} - \frac{1}{a}\sum \bar{Y}_{i.}) \to \sum \bar{Y}_{i.} - \sum \frac{1}{a}\sum \bar{Y}_{i.} \to \sum \bar{Y}_{i.} - \frac{a}{a}\sum \bar{Y}_{i.} \to \sum \bar{Y}_{i.} - \sum \bar{Y}_{i.} \to 0$

   e. If the ANOVA model holds, the distribution of $\hat{\gamma}_k$ is normal where $\hat{\gamma}_k \sim N(\gamma_k, \sigma^2[\hat{\gamma}_k])$. $\sigma^2[\hat{\gamma}_k] = \sigma^2[\bar{Y}_{k.} - \bar{\bar{Y}}] \to \sigma^2[\bar{Y}_{k.}] + \sigma^2[\bar{\bar{Y}}] \to \sigma^2[\bar{Y}_{k.}] + \sigma^2[\frac{\sum \bar{Y}_{i.}}{a}] \to \sigma^2[\bar{Y}_{k.}] + \sigma^2[\frac{\sum \bar{Y}_{i.}}{3}] \to \sigma^2[\bar{Y}_{k.}] + \frac{\sum \sigma^2[\bar{Y}_{i.}]}{9} \to \frac{\sigma^2[Y_k]}{n_k} + \frac{1}{9}\sum \frac{\sigma^2[Y_i]}{n_i} \to \frac{\sigma^2[Y_k]}{n_k} + 0.9135 \therefore \hat{\gamma}_k \sim N(\gamma_k, \frac{\sigma^2[Y_k]}{n_k} + 0.9135)$??

3

   a. The 95 confidence interval for μ_1 is $[34.7248, 41.2752]$. This means that we are 95% sure that the true value of $\mu_1 \in [34.7248, 41.2752]$.

   b. $\mu_1 - \mu_3 \in [7.1886, 20.8114]$ μ_1 is between 7.1886 and 20.8114 more than μ_3. Since all values in the confidence interval is positive, we believe that the two means are different.

   c. Yes, my confidence interval agrees with that statement. We expect the recovery time for "Below Average" patient to be at least 34.7248 days, which is greater than 30 days.

d. The largest difference we expect an above average patient to have compared to a below average in average recovery time is 20.8114 as it is our upper bound of our confidence interval for the difference between the two means.

4

a. The power is the probability that we correctly reject the null hypothesis where age does affect the cash offer.

b. $\phi = 6.4653$

c. Using $\phi = 6.4653$, $d.f.\{SSA\} = 2$, and $d.f.\{SSE\} = 33$ from HW2, we use the table to find that the power is 1.00.

d. The power is super high because our factor effect is high relative to the error (i.e. SSA is a lot higher than MSE).

5

a. For all $Y_{ij} = \mu_i + \gamma_i + \epsilon_{ij}$, we are interested to know the factor effect (γ_i) and confidence interval ($\mu_i - \mu_.$) of each group. The assumptions are:

    i. Y_{ij} were randomly sampled (independent)
    ii. The i groups are independent for all $i = 1, ..., a$
    iii. Errors are independent and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$

b. $\mu_M - \mu_Y = 6.25$

c. $\sigma^2[\bar{M} - \bar{Y}] \rightarrow \sigma^2[\bar{M}] + \sigma^2[\bar{Y}] \rightarrow 1.29^2 + 1.73^2 \rightarrow 4.657$

d. $\epsilon_{ij} \sim N(0, \sqrt{MSE}) \rightarrow \epsilon_{ij} \sim N(0, 1.5888)$ This means that 2.5 is 1.5735 standard deviations above the mean. Using the z-table, $p \in [0.0571, 0.0582]$.

6

a. The 95 confidence interval for $\mu_M - \frac{\mu_Y + \mu_O}{2}$ is $[5.1472, 7.4328]$.

b. We are 95 confident that $\mu_M - \frac{\mu_Y + \mu_O}{2} \in [5.1472, 7.4328]$. This means that the mean of M is between 5.1472 and 7.4328 more than the average of means of Y and O.

c. The 95 confidence interval for $\mu_O - \mu_Y$ is $[-1.3996, 1.2396]$.

d. We are 95 confident that $\mu_O - \mu_Y \in [-1.3996, 1.2396]$. This means that the mean of O is between -1.3996 and 1.2396 different compared to the mean of Y.

e. I believe the means from 6A are significantly different because the confidence interval does not contain zero. The means from 6C are not significantly different because the confidence interval does not contain 0.

7

a. $\phi = 1.7496 \approx 1.5, d.f.[num] = 3, d.f.[den] = 20, \alpha = 0.05$. Using the F-table, our power is 0.62.

b. $\phi = 1.7496 \approx 1.5, d.f.[num] = 3, d.f.[den] = 20, \alpha = 0.01$. Using the F-table, our power is 0.34.

c. As we decrease α, the power goes down.

d. $\frac{\Delta}{\sigma_\epsilon} = 1.4286 \approx 1.25, a = 4, \alpha = 0.05, 1 - \beta = 0.80$. Using the sample size table, we find that $n_i = 15$ is our minimum sample size for all groups.

e. $\frac{\Delta}{\sigma_\epsilon} = 1.4286 \approx 1.25, a = 4, \alpha = 0.05, 1 - \beta = 0.90$. Using the sample size table, we find that $n_i = 20$ is our minimum sample size for all groups.

f. As we aim for a higher power, we need a larger sample size for each group.

8

a. False, we can see on the F-table as α increases, the power increases.
b. True, because the confidence interval does not contain 0, which means that we are $(1 - \alpha)100$ certain that the means are not equal.
c. False, we can see that on the F-table.
d. True, by definition of a $(1 - \alpha)100$ confidence interval for μ_i, we are $(1 - \alpha)100$ confident that μ_i is in the confidence interval.

I

a. The power is 0.9854.

b. The smallest sample size we need for each group is $n_i = 14$.

c. The smallest sample size we need for each group is $n_i = 18$.

d. Large, because the power of our test increases as our ϕ increases, which is positively correlated with our F-statistic due to SSA being in the numerator and MSE being in the denominator of ϕ.

II

a. The confidence interval is $[-419.1962, 568.0197]$.

b. This interval does not suggest a significant difference between the two means because zero is in the interval.

c. The confidence interval is $[-199.4183, 616.6536]$.

d. The results do not suggest a significant difference between the mean time of colon and the average of mean time of both stomach and bronchus because zero is in the interval.

III

a. Our p-value is 0.0159. Since $p < \alpha$ when $\alpha = 0.10$, we reject the null hypothesis. Therefore, the average weight is affected by each treatment.

b. $\gamma_1 = -0.041, \gamma_2 = -0.412, \gamma_3 = 0.453$

c. The power is the probability that we correctly reject the null hypothesis where treatment does affect the weight.

d. $1 - \alpha$ is the probability that we correctly accept the null hypothesis where the null hypothesis is true (i.e. treatment does not affect the weight).

IV

a. The power is 0.8503.

b. The smallest sample size we need for each group is $n_i = 24$.

c. Increase the sample size of each group.

d. The 95% confidence interval for $\frac{\mu_{ctrl} + \mu_{trt1}}{2} - \mu_{trt2}$ is $[-1.1749, -0.1841]$.

e. This means that the average of means of control and treatment 1 group is between 0.1841 and 1.1749 lower than mean of treatment 2 group.

Appendix

```
knitr::opts_chunk$set(echo = FALSE)

# 1-a
mu_T <- 32
mu_i <- c(38, 32, 24)
n_i <- c(8, 10, 6)
a <- 3
MSE <- 19.8424 # From HW2

phi <- round(1 / sqrt(MSE) * sqrt(sum(n_i * (mu_i - mu_T) ** 2) / a), 4)

# 1-c
mu_min <- min(mu_i)
mu_max <- max(mu_i)
effect_size <- round((mu_max - mu_min) / sqrt(MSE), 4)

# 2-b
Y_bar_bar <- sum(mu_i) / a
gamma_i <- round(mu_i - Y_bar_bar, 4)

# 2-d
gamma_T <- round(sum(gamma_i))

# 2-e
```

4

```r
sd_Y_i <- c(5.48, 3.46, 4.43)
var_Y_i <- sd_Y_i ** 2
var_Y_bar_bar <- round(sum(var_Y_i / n_i) / a ** 2, 4)

# 3-a
alpha <- 0.05
n_T <- 24
a <- 3
t_critical <- qt(1 - alpha / 2, n_T - a)
lower <- round(mu_i[1] - t_critical * sqrt(MSE * (1 / n_i[1])), 4)
upper <- round(mu_i[1] + t_critical * sqrt(MSE * (1 / n_i[1])), 4)

# 3-b
alpha <- 0.01
n_T <- sum(n_i)
critical_t <- qt(1 - alpha / 2, n_T - a)
lower_bound <- round(
  mu_i[1] - mu_i[3] - critical_t * sqrt(MSE * (1 / n_i[1] + 1 / n_i[3])),
  4
)
upper_bound <- round(
  mu_i[1] - mu_i[3] + critical_t * sqrt(MSE * (1 / n_i[1] + 1 / n_i[3])),
  4
)

# 4-b
n_i <- c(12, 12, 12)
mu_i <- c(27.75, 21.42, 21.50)
mu_T <- 23.56
a <- 3
MSE <- 2.5243 # from HW2

phi <- round(1 / sqrt(MSE) * sqrt(sum(n_i * (mu_i - mu_T) ** 2) / a), 4)

# 5-a
M <- 1
Y <- 3
O <- 2
mu_diff <- mu_i[M] - mu_i[Y]

# 5-c
sd_M <- 1.29
sd_Y <- 1.73
sd_diff_M_Y <- sd_M ** 2 + sd_Y ** 2

# 5-d
sd <- round(sqrt(MSE), 4)
sd_above_mean <- round(2.5 / sqrt(MSE), 4)

# 6-a
n_T <- sum(n_i)
alpha <- 0.05
CI_mid <- mu_i[M] - (mu_i[Y] + mu_i[O]) / 2
```

```r
t_critical <- qt(1 - alpha / 2, n_T - a)
lower <- round(
  CI_mid - t_critical * sqrt(
    MSE * (1 / n_i[M] + 1 / 2 ** 2 * 1 / n_i[Y] + 1 / 2 ** 2 * 1 / n_i[O])
  ),
  4
)
upper <- round(
  CI_mid + t_critical * sqrt(
    MSE * (1 / n_i[M] + 1 / 2 ** 2 * 1 / n_i[Y] + 1 / 2 ** 2 * 1 / n_i[O])
  ),
  4
)

# 6-c
CI_mid <- mu_i[O] - mu_i[Y]
lower <- round(
  CI_mid - t_critical * sqrt(
    MSE * (1 / n_i[O] + 1 / n_i[Y])
  ),
  4
)
upper <- round(
  CI_mid + t_critical * sqrt(
    MSE * (1 / n_i[O] + 1 / n_i[Y])
  ),
  4
)

# 7-a
mu_i <- c(13, 13, 18, 18)
n_i <- 6
a <- 4
sd_err <- 3.5

n_T <- n_i * a
mu_T <- sum(mu_i * n_i) / n_T
df_num <- a - 1
df_den <- n_T - a

phi <- round(1 / sd_err * sqrt(sum(n_i * (mu_i - mu_T) ** 2) / a), 4)

# 7-d
delta_over_sd_err <- round((max(mu_i) - min(mu_i)) / sd_err, 4)

# I-a
cancer <- read.csv("cancer.csv")
library(pwr)

give.me.power = function(ybar,ni,MSE,alpha){
  a = length(ybar) # Finds a
  nt = sum(ni) #Finds the overall sample size
  overall.mean = sum(ni*ybar)/nt # Finds the overall mean
```

```r
    phi = (1/sqrt(MSE))*sqrt( sum(ni*(ybar - overall.mean)^2)/a) #Finds the books value of phi
    phi.star = a *phi^2 #Finds the value of phi we will use for R
    Fc = qf(1-alpha,a-1,nt-a) #The critical value of F, use in R's function
    power = 1 - pf(Fc, a-1, nt-a, phi.star)# The power, calculated using a non-central F
    return(power)
}

group.means = by(cancer$Survival,cancer$Organ,mean)
group.nis = by(cancer$Survival,cancer$Organ,length)
the.model = lm(Survival ~ Organ, data = cancer)
anova.table = anova(the.model)
MSE = anova.table[2,3]
alpha = 0.05
the.power = round(give.me.power(group.means,group.nis,MSE,alpha), 4)

# I-b
overall.mean = sum(group.means*group.nis)/sum(group.nis)
effect.size = sqrt(
  sum( group.nis/sum(group.nis) *(group.means -overall.mean)^2 )/MSE
)
test = pwr.anova.test(k = 5, f = effect.size, sig.level = 0.05, power = 0.99)
n = ceiling(test$n)

# I-c
test = pwr.anova.test(k = 5, f = effect.size, sig.level = 0.01, power = 0.99)
n = ceiling(test$n)

# II-a
give.me.CI = function(ybar,ni,ci,MSE,multiplier){
  if (sum(ci) != 0 & sum(ci !=0 ) != 1){
    return("Error - you did not input a valid contrast")
  } else if (length(ci) != length(ni)){
    return("Error - not enough contrasts given")
  } else {
    estimate = sum(ybar*ci)
    SE = sqrt(MSE*sum(ci^2/ni))
    CI = estimate + c(-1,1)*multiplier*SE
    result = c(estimate,CI)
    names(result) = c("Estimate","Lower Bound","Upper Bound")
    return(result)
  }
}

t.value = qt(1-0.05/2, sum(group.nis) - length(group.nis))
# c(breast, bronchus, colon, ovary, stomach)
ci = c(0, -1, 0, 0, 1)
the.ci = give.me.CI(group.means,group.nis,ci,MSE,t.value)
lower <- round(the.ci[["Lower Bound"]], 4)
upper <- round(the.ci[["Upper Bound"]], 4)

# II-c
# c(breast, bronchus, colon, ovary, stomach)
ci = c(0, -1/2, 1, 0, -1/2)
```

```r
the.ci = give.me.CI(group.means,group.nis,ci,MSE,t.value)
lower <- round(the.ci[["Lower Bound"]], 4)
upper <- round(the.ci[["Upper Bound"]], 4)

# III-a
green <- read.csv("Green.csv")
the.model = lm(weight ~ group, data = green)
anova.table = anova(the.model)
the.p = round(anova.table["group", "Pr(>F)"], 4)

# III-b
group.means = by(green$weight,green$group,mean)
gammai = group.means - mean(group.means)

# IV-a
group.nis = by(green$weight,green$group,length)
the.model = lm(weight ~ group, data = green)
anova.table = anova(the.model)
MSE = anova.table[2,3]
alpha = 0.10
the.power = round(give.me.power(group.means,group.nis,MSE,alpha), 4)

# IV-b
overall.mean = sum(group.means*group.nis)/sum(group.nis)
effect.size = sqrt(
  sum( group.nis/sum(group.nis) *(group.means -overall.mean)^2 )/MSE
)
test = pwr.anova.test(k = 3, f = effect.size, sig.level = 0.05, power = 0.99)
n = ceiling(test$n)

# IV-d
t.value = qt(1-0.05/2, sum(group.nis) - length(group.nis))
# c(ctrl, trt1, trt2)
ci = c(1/2, 1/2, -1)
the.ci = give.me.CI(group.means,group.nis,ci,MSE,t.value)
lower <- round(the.ci[["Lower Bound"]], 4)
upper <- round(the.ci[["Upper Bound"]], 4)
```