Assignment 3

● Graded

Student

James La Fontaine

Total Points

18 / 20 pts

Question 1

Question 1 — **1** / 1 pt

1.1 **1(a)i** — **1** / 1 pt

✔ **+ 1 pt** Any attempt Q1

+ 0 pts No attempt Q1

− 1 pt Questions weren't matched i.e. the following message comes up 'The student did not select pages for this question...'

− 1 pt Incredibly messy (e.g. blank pages, illegible writing, code and graphs handdrawn/handwritten, etc.)

Question 2

Question 2 **7** / 7 pts

2.1 2(a) **1** / 1 pt

Assuming that these follow a Bi(30,p) distribution, estimate p.

✔ **+ 1 pt** $\hat{p}$=0.3054167 with code or working e.g.
`p.hat <- sum(germinations * count) / (80 * 30)`

+ 0.5 pts Correct working/code but didn't give phat=0.3054167
OR correct answer but no code/working
OR reasonable attempt but some mistake e.g. data entry error
OR used fitdistr and got 0.3053711 (more accurate solution exists)

+ 0 pts Incorrect/no attempt

2.2 2(b) **3** / 3 pts

Design a set of classes suitable for carrying out a goodness-of-fit test for a binomial distribution.

✔ **+ 3 pts** Classes are: **(0–5, 6, 7, 8, 9, 10, 11, 12, 13–30)**
Accept equivalent groupings e.g. (0–5.5, 5.5–6.5, 6.5–7.5, ...) or writing ≤ 5 for first class, etc.
WITH SENSIBLE JUSTIFICATION

+ 2 pts Classes: **(0–6, 7, 8, 9, 10, 11, 12–30)** or equivalent. Student used **observed** rather than expected but 'correctly' merged tails

+ 2 pts Correct classes but justification unclear/missing

+ 2 pts Correct approach and some attempt made to merge tails, but final classes suboptimal

+ 1 pt Attempted to calculate E's but didn't merge tails at all
OR
Other reasonable attempt not stated above

+ 0 pts Incorrect/no attempt

2.3 2(c) **3** / 3 pts

✔ **+ 1 pt** Employed a chi-squared approach

✔ **+ 1 pt** Binned observed to match whatever classes chosen in (b) ((e.g. O=(5, 4, 10, 16, 9, 11, 13, 4, 8) or chi-squared=5.924))
Or used **observed** in (b) but correctly calculating expected/pi in matching classes

✔ **+ 1 pt** Correctly adjusted dof=k–2, where k=#classes
((e.g. updated p-value=1-pchisq(5.924, 7)=0.54865 OR 1-pchisq(5.0762,5)=0.40665))

+ 0 pts Incorrect/no attempt or no working (e.g. just wrote down the p-value)

Question 3

Question 3 4 / 6 pts

3.1 **3(a)** 2 / 2 pts

Find the cdf of the sample minimum

✔ **+ 2 pts** $F(x) = 1 - x^{-\theta}$

$Pr(X_{(1)} > x) = (1 - F(x))^n = x^{-\theta n}$

$F_1(x) = Pr(X_{(1)} < x) = 1 - x^{-\theta n}$

(other correct approaches exist)

+ 1 pt Some reasonable attempt but some mistake OR correct but no working given

+ 0 pts No attempt/incorrect

3.2 **3(b)** 1 / 2 pts

Find the p-quantile in terms of p and θ

+ 2 pts

By definition, $p = F(\pi_p) = 1 - \pi_p^{-\theta}$. Solving for π_p,

$$\pi_p^{-\theta} = 1 - p$$
$$\pi_p^{\theta} = \frac{1}{1-p}$$
$$\pi_p = \left(\frac{1}{1-p}\right)^{\frac{1}{\theta}}.$$

✔ **+ 1 pt** Some reasonable attempt but some mistake OR correct but no working given

+ 0 pts Incorrect/no attempt

3.3 **3(c)** 1 / 2 pts

Find the asymptotic variance of the sample median

+ 2 pts Correct

(2 marks) The median of X is $m = \pi_{0.5} = 2^{1/\theta}$. To find the asymptotic variance of $\hat{M}$, we first need to find $f(m)$,

$$f(m) = \theta\left(2^{1/\theta}\right)^{-(\theta+1)} = \theta\, 2^{-(1+1/\theta)}.$$

Using the asymptotic distribution of sample quantiles, we deduce that,

$$\mathrm{var}(\hat{M}) \to \frac{1}{4nf(m)^2} = \frac{1}{4n\theta^2\, 2^{-(2+2/\theta)}} = \frac{4^{1/\theta}}{n\theta^2}.$$

or $\frac{2^{2/\theta}}{n\theta^2}$ or $\frac{1}{n\theta^2 0.5^{2/\theta}}$

✔ **+ 1 pt** Some reasonable attempt but some mistake OR correct but no/incorrect/insufficient working given

+ 0 pts Incorrect/no attempt

Question 4

Question 4 **6** / 6 pts

4.1 4(a) **4** / 4 pts

- ✔ **+ 0.5 pts** $H_0 : \alpha_i = 0$ for all i
  H_1 : at least one of the α_i is non-zero
  (need both hypotheses ... ok to also state in words or use β instead of α). Correct variations involving μ also exist.
  Also acceptable: $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$

- ✔ **+ 0.5 pts** Assumptions (doesn't need to be complete):
  The model is $Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$ with i=1,...,4, j=1,...,5 and k=1,2 where ε_{ijk} denote independent errors such that $\varepsilon_{ijk} \sim N(0, \sigma^2)$. Further, $\sum \alpha_i = 0$ and $\sum \beta_j = 0$.

- ✔ **+ 1 pt** .Fitting a two-way anova e.g. using aov or anova(lm ...
  need to provide some output/code & needs to be two-way anova to get mark (award even if data entered incorrectly)

- ✔ **+ 0.5 pts** Answer mark: Test statistic: F=5.8632 (need to state)

- ✔ **+ 0.5 pts** Answer mark: p-value=0.002614 (need to state)

- ✔ **+ 1 pt** We have strong evidence that the power output is affected by the angle of elevation. (need STATEMENT, not just 'reject')
  OR wrong model (e.g. one-way anova) but made statement consistent with p-value they found.

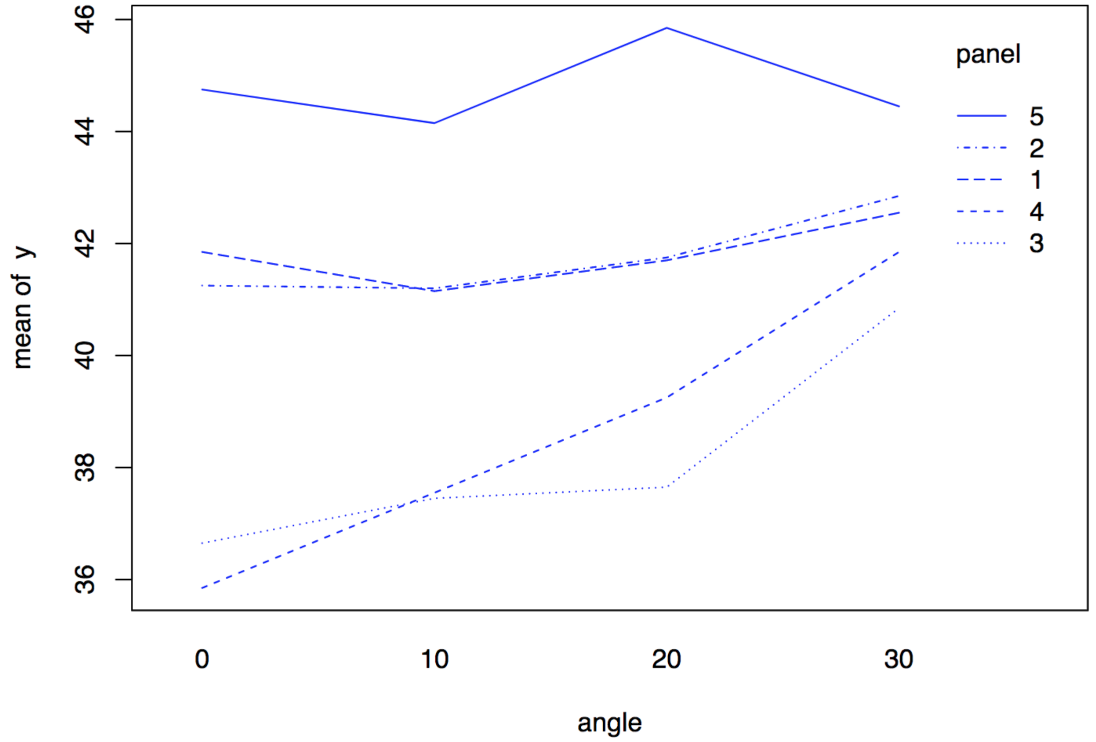
- **+ 0 pts** No attempt/incorrect

1 alternative hypothesis is not clear

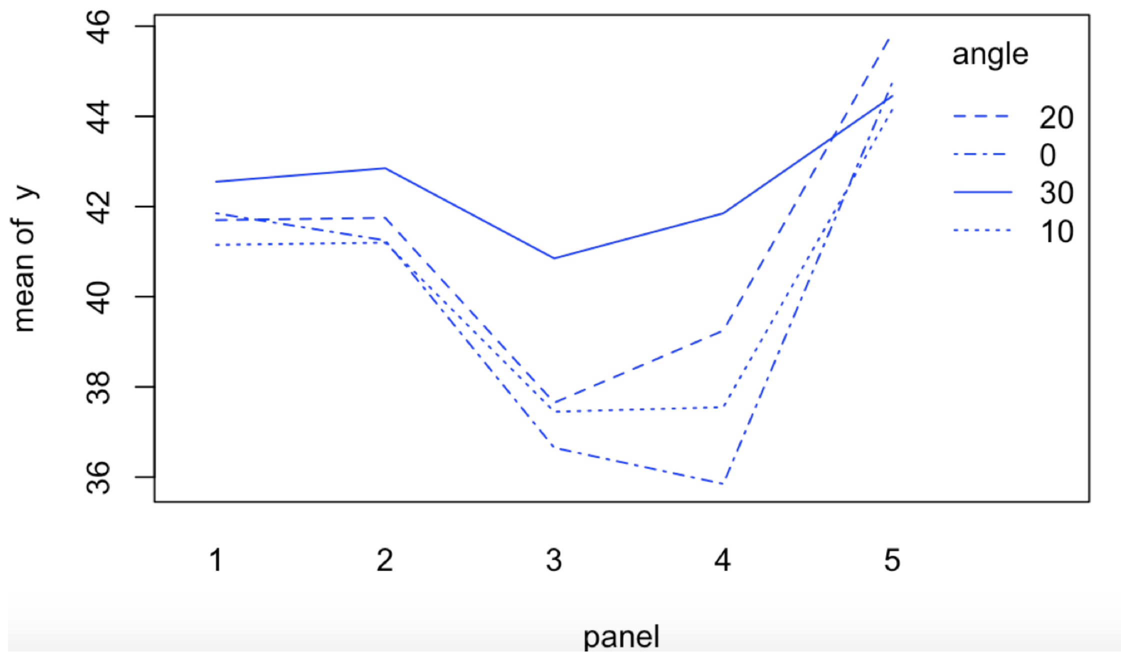
4.2 4(b) **2** / 2 pts

✔ **+ 1 pt** Fitting model with interaction

✔ **+ 1 pt** Interaction plot



OR



+ 0 pts Incorrect/no attempt

No questions assigned to the following page.

Assignment 3

Name: James La Fontaine

Student Number: 1079860

Tutorial Day and Time: Friday 2:15 PM – 4:15 PM

Tutor's Name: Haoyu Yang

Question assigned to the following page: 1.1

Question 1

1ai)

H_0: $m_X = m_Y$

H_1: $m_X \neq m_Y$

```
binom.test(12, 17)
```

```
##
##  Exact binomial test
##
## data:  12 and 17
## number of successes = 12, number of trials = 17, p-value = 0.1435
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.4404173 0.8968645
## sample estimates:
## probability of success
##              0.7058824
```

The p-value > 0.05 so we cannot reject the null hypothesis that $m_X = m_Y$ at the significance level of 5%.

1aii)

H_0: $m_X = m_Y$

H_1: $m_X \neq m_Y$

```
x = c(26.1, 26.6, 27.4, 27.5, 27.8, 28.1, 28.4, 29.5, 29.8, 30.4, 30.4, 31.2,
31.5, 32.9, 33.6, 34.1, 35.9)
y = c(27.4, 28.1, 22.9, 31.3, 16.3, 50.1, 20.0, 24.6, 23.3, 19.3, 24.4, 24.4,
29.5, 27.6, 21.7, 25.4, 39.4)

wilcox.test(x, y, paired = TRUE)
```

```
##
##  Wilcoxon signed rank exact test
##
## data:  x and y
## V = 124, p-value = 0.02322
## alternative hypothesis: true location shift is not equal to 0
```

The p-value < 0.05 so we reject the null hypothesis that $m_X = m_Y$ at the significance level of 5% and can conclude that there is sufficient evidence to show that the location of X and Y differ.

No questions assigned to the following page.

1aiii)

$H_0: \mu_X = \mu_Y$

$H_1: \mu_X \neq \mu_Y$

```r
x = c(26.1, 26.6, 27.4, 27.5, 27.8, 28.1, 28.4, 29.5, 29.8, 30.4, 30.4, 31.2,
31.5, 32.9, 33.6, 34.1, 35.9)
y = c(27.4, 28.1, 22.9, 31.3, 16.3, 50.1, 20.0, 24.6, 23.3, 19.3, 24.4, 24.4,
29.5, 27.6, 21.7, 25.4, 39.4)

t.test(x, y, paired = TRUE)

##
##  Paired t-test
##
## data:  x and y
## t = 1.6402, df = 16, p-value = 0.1205
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.954790  7.484202
## sample estimates:
## mean of the differences
##                3.264706
```

The p-value > 0.05 so we cannot reject the null hypothesis that $\mu_X = \mu_Y$ at the significance level of 5%.

1b)

The sign test has a larger type II error rate / smaller power and so it is plausible that the null hypothesis has incorrectly not been rejected in this test. For the t-test we have made the assumption that the differences between X and Y are normally distributed and since we have a small sample size of only 17, it is plausible that these differences do not follow a normal distribution. It would be more appropriate to give more consideration to the outcome of the Wilcoxon signed-rank test in this case, which simply assumes that the differences between X and Y are continuous and follow a symmetrical distribution, which is a reasonable assumption under the null hypothesis. Therefore, there is mild evidence that X and Y differ in location, however, further testing with a larger sample would be required to make stronger conclusions.

No questions assigned to the following page.

1c)

```
B = 20000
n = 17
numRejectionsSign = 0
numRejectionsWilcoxon = 0
numRejectionsT = 0

for (i in 1:B) {
  numSuccesses = 0
  sampleDifference = rnorm(n, 3, 5)
  for (number in sampleDifference) {
    if (sign(number) == 1) {
      numSuccesses = numSuccesses + 1
    }
  }
  if (binom.test(numSuccesses, n)$p.value < 0.05) {
    numRejectionsSign = numRejectionsSign + 1
  }
  if (wilcox.test(sampleDifference)$p.value < 0.05) {
    numRejectionsWilcoxon = numRejectionsWilcoxon + 1
  }
  if (t.test(sampleDifference)$p.value < 0.05) {
    numRejectionsT = numRejectionsT + 1
  }
}

powerSign = numRejectionsSign / B
powerWilcoxon = numRejectionsWilcoxon / B
powerT = numRejectionsT / B

cat("Simulated power of sign test: ", powerSign, "\n")

Simulated power of sign test:  0.4859

cat("Simulated power of Wilcoxon test: ", powerWilcoxon, "\n")

Simulated power of Wilcoxon test:  0.60815

cat("Simulated power of t-test: ", powerT, "\n")

Simulated power of t-test:  0.64395
```

Question assigned to the following page: 2.1

Question 2

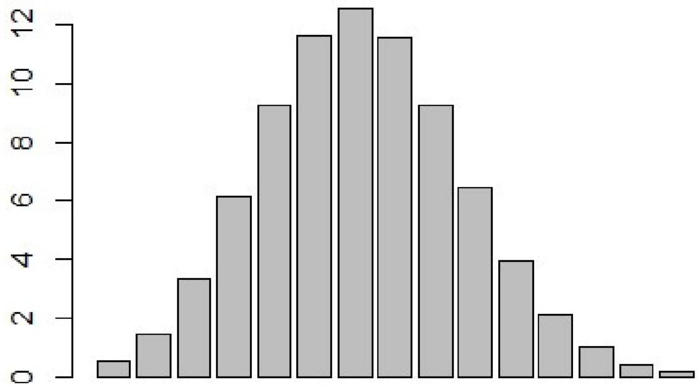
2a)

```
germinations = c(3, 4, 5, 6, 7, 8 ,9, 10, 11, 12, 13, 17)
count = c(1, 2, 2, 4, 10, 16, 9, 11, 13, 4, 7, 1)

experiments = data.frame(germinations, count)

data = rep(experiments$germinations, experiments$count)

p1 = prop.test(sum(data), 80*30)

p1

##
##  1-sample proportions test with continuity correction
##
## data:  sum(data) out of 80 * 30, null probability 0.5
## X-squared = 362.7, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.2871084 0.3243539
## sample estimates:
##         p
## 0.3054167

prop.estimate = as.numeric(p1$estimate)
```

Question assigned to the following page: 2.2

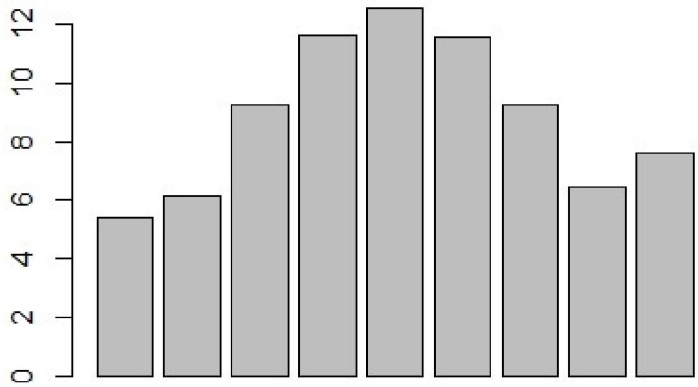
2b)

```
barplot(dbinom(3:17, 30, prop.estimate) * 80)
```




```
X1 <- cut(data, breaks = c(0, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5, 11.5, 12.5, Inf)
)
T1 <- table(X1)
grouped.data <- as.numeric(T1)

p <- c(pbinom(5, 30, prop.estimate), dbinom(6:12, 30, prop.estimate), 1 - pbi
nom(12, 30, prop.estimate))

barplot(p * 80)
```

Questions assigned to the following page: 2.2 and 2.3



2c)

```
chi1 = chisq.test(x=grouped.data, p = p)

chi1

##
##  Chi-squared test for given probabilities
##
## data:  grouped.data
## X-squared = 5.924, df = 8, p-value = 0.6557

X.squared = unname(chi1$statistic)

# recalculate the p-value using the correct degrees of freedom

1 - pchisq(unname(X.squared), length(grouped.data) - 2)

## [1] 0.5486503
```

The p-value 0.5487 > 0.05 and so there is insufficient evidence to conclude that there is a difference between a Binomial distribution and the distribution of the number of germinations of seeds of the tested plant.

Questions assigned to the following page: 3.1, 3.2, and 3.3

Question 3

a) $F_X(x) = \int_1^x \theta x^{-(\theta+1)}\,dx, \quad x \geq 1,\ \theta > 0$

$$= \theta \int_1^x x^{-\theta-1}\,dx = \theta\left[-\frac{x^{-\theta}}{\theta}\right]_1^x$$

$$= \theta\left(-\frac{x^{-\theta}}{\theta} + \frac{1}{\theta}\right) = 1 - x^{-\theta}$$

$$F_{X_{(1)}}(x) = P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x)$$

$$= 1 - [P(X > x)]^n = 1 - [1 - F(x)]^n$$

$$= 1 - (1 - (1 - x^{-\theta}))^n = 1 - (1 - 1 + x^{-\theta})^n$$

$$= 1 - (x^{-\theta})^n = 1 - x^{-n\theta}, \quad x \geq 1,\ \theta > 0$$

$$0 \quad \text{otherwise}$$

b) $\pi_p = F^{-1}(p)$

$$x = 1 - y^{-n\theta}$$

$$y^{-n\theta} = 1 - x$$

$$y = (1-x)^{-\frac{1}{n\theta}} = \frac{1}{(1-x)^{\frac{1}{n\theta}}} = F^{-1}(x)$$

$$\Rightarrow F^{-1}(p) = \frac{1}{(1-p)^{\frac{1}{n\theta}}}, \quad 0 \leq p < 1, \quad \theta > 0$$

c) Asymptotic variance of $\hat{M} \approx \frac{1}{4nf(m)^2}$

$$\approx \frac{1}{4n(\theta m^{-(\theta+1)})^2} \qquad m = \pi_{0.5} = F^{-1}(0.5)$$

$$\approx \frac{1}{4n\theta^2 m^{-2(\theta+1)}} \qquad = \frac{1}{0.5^{\frac{1}{n\theta}}} = 2^{\frac{1}{n\theta}}$$

$$\cancel{4n\theta^2 2^{-\frac{2(\theta+1)}{n\theta}}} \approx 4n\theta^2 2^{-\frac{2(\theta+1)}{n\theta}}$$

Question 4

4a)

```
Angle = c(rep(seq(0,30,10), each=10))

Panel = c(rep(rep(1:5, each = 2), 4))

Power = c(42.3, 41.4, 42.2, 40.3, 37.6, 35.7, 36.8, 34.9, 45.8, 43.7, 42.1, 4
0.2, 42.1,  40.3, 38.4, 36.5, 38.0, 37.1, 45.2, 43.1, 42.6, 40.8, 42.7, 40.8,
38.6, 36.7, 40.2, 38.3, 46.9, 44.8, 43.6, 41.5, 43.8, 41.9, 41.9, 39.8, 42.9,
40.8, 45.4, 43.5)

data = data.frame(Angle, Panel, Power)

model1 = lm(Power ~ factor(Angle) + factor(Panel), data = data)

anova(model1)

## Analysis of Variance Table
##
## Response: Power
##                Df  Sum Sq Mean Sq F value    Pr(>F)
## factor(Angle)  3  36.890  12.297  5.8632  0.002614 **
## factor(Panel)  4 235.522  58.880 28.0748 4.602e-10 ***
## Residuals     32  67.113   2.097
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let α = the effect of angle elevation on power output.

Let β = the effect of panel type on power output.

$\mu_{ij} = \mu + \alpha_i + \beta_j$

H_{0A}: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$

H_{1A}: $\overline{H}_{0A}$

F value = 5.8632

p-value = 0.002614

Assumptions:

- There is no statistical interaction between the factors and thus factor effects are additive.
- We have random samples drawn independently of each other from the different populations, each having a normal distribution.

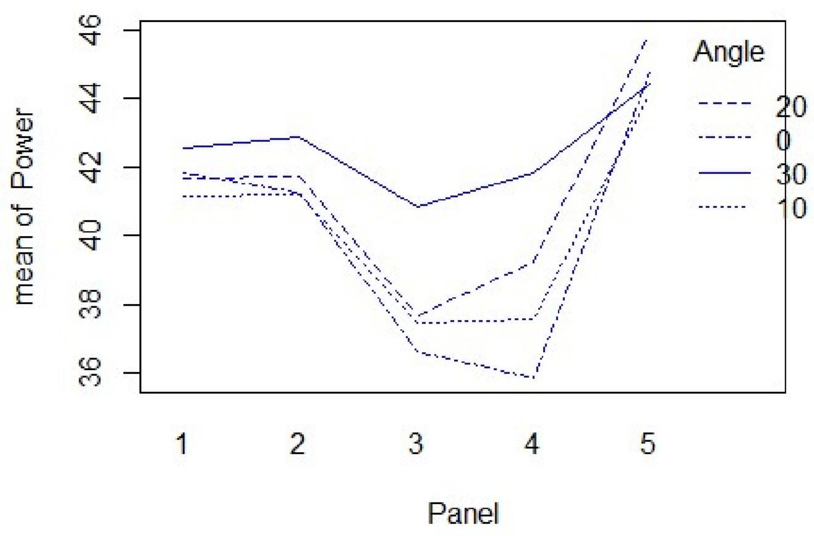
Questions assigned to the following page: 4.1 and 4.2

- All populations have the same variance, σ^2.

0.002614 < 0.05. Therefore, there is sufficient evidence to conclude at the 5% level of significance that the mean power output of solar panels varies between the different angles of elevation and thus that the angle of elevation influences mean power output.

4b)

```
model2 = lm(Power ~ factor(Angle) * factor(Panel), data = data)

anova(model2)

## Analysis of Variance Table
##
## Response: Power
##                             Df  Sum Sq Mean Sq F value    Pr(>F)
## factor(Angle)                3  36.890  12.297  6.9610  0.002163 **
## factor(Panel)                4 235.522  58.880 33.3317 1.383e-08 ***
## factor(Angle):factor(Panel) 12  31.782   2.649  1.4993  0.204458
## Residuals                   20  35.330   1.767
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

with(data, interaction.plot(Panel, Angle, Power, col = "blue"))
```



0.2045 > 0.05. Therefore, there is insufficient evidence to conclude at the the 5% level of significance that there is interaction between panel type and the angle of elevation.