MAST20005/MAST90058: Assignment 3 Solutions

1.
```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, 24.6, 23.3, 19.3, 24.4,
       24.4, 29.5, 27.6, 21.7, 25.4, 39.4)
```

Let $D = Y - X$.

(a) i. $H_0 \colon m_D = 0$ versus $H_1 \colon m_D \neq 0$.

```r
binom.test(sum(y > x), 17)
```

```
##
##  Exact binomial test
##
## data:  sum(y > x) and 17
## number of successes = 5, number of trials = 17, p-value = 0.1435
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.1031355 0.5595827
## sample estimates:
## probability of success
##              0.2941176
```

The p-value is 0.14, so we cannot reject H_0.

ii. $H_0 \colon m_D = 0$ versus $H_1 \colon m_D \neq 0$.

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

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

The p-value is 0.02, so we reject H_0.

iii. $H_0 \colon \mu_D = 0$ versus $H_1 \colon \mu_D \neq 0$.

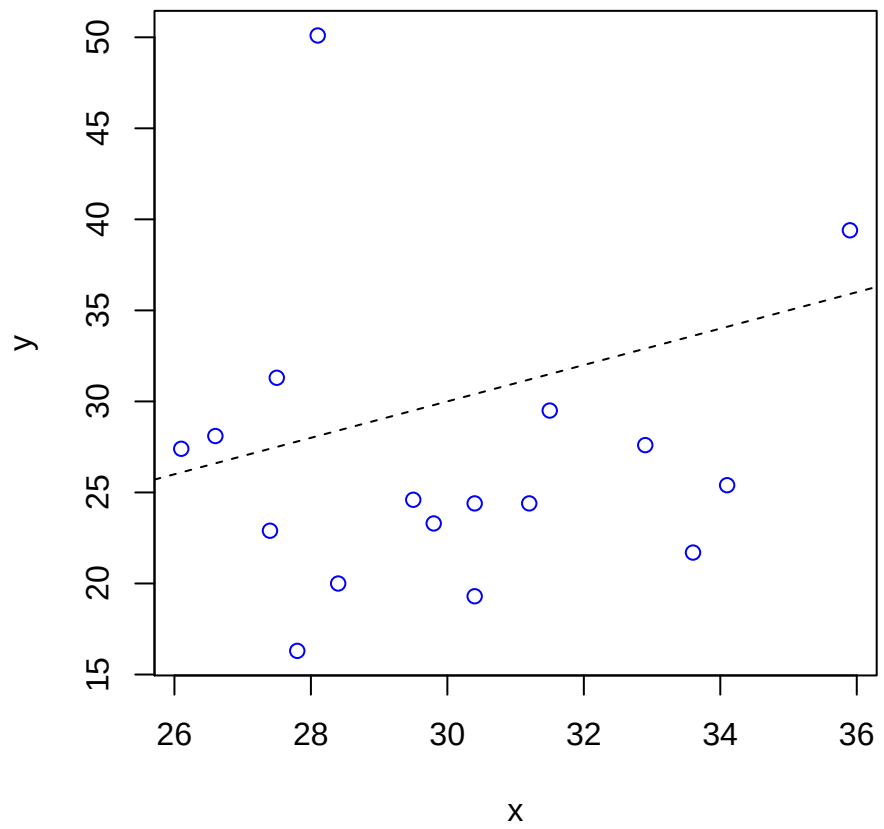
```r
t.test(x, y, paired = TRUE)
```

```
##
##  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 is 0.12, so we cannot reject H_0.

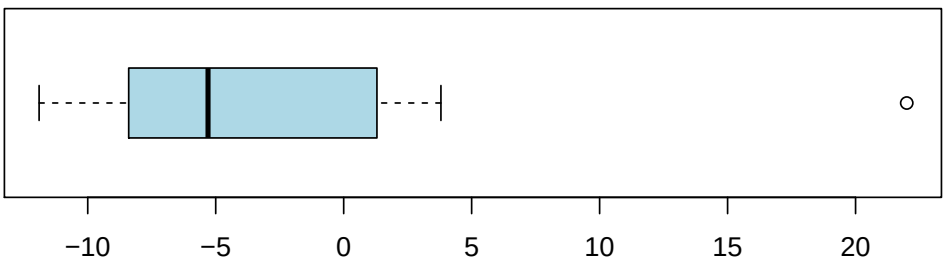
(b) Interestingly, the Wilcoxon test is the only one that leads to a rejection of H_0. The sign test doesn't reject, which is usually unsurprising since it has low power. However, the t-test, which usually has the highest power, also doesn't reject here while the Wilcoxon does. What's going on? Let's look at the data:

```
plot(x, y, col = "blue")
abline(0, 1, lty = 2)  # reference line (y = x)
```



The above tests only look at differences, so we can look directly at those:

```
boxplot(y - x, horizontal = TRUE, col = "lightblue")
```



The bulk of the data indicate a difference in the two groups, with $y < x$ generally. However, one of the observed pairs is an outlier giving a very different result in the reverse direction. This had a substantial effect on the t-test, which is why it did not reject H_0, while the Wilcoxon test was less sensitive to it and was still able to reject. This outlier should be investigated further to see if it is an error in the data or if it can be deemed unrepresentative of the question being studied. If so, then it would be appropriate to omit the outlier and re-run the test, which would result in a clear reject of the null. If not, and we do indeed expect to see such outliers, then we would likely consider the t-test inappropriate here because the distribution of differences is highly skewed and therefore not well represented by the t-test assumptions, and therefore we would prefer the Wilcoxon test. Either way, we would reject the null and conclude that the data show good evidence in support of the conclusion that X and Y differ.

(c) Note that since we are doing paired tests, we only need to know the distribution of the differences.

```
B <- 10000  # number of simulations
outcome1 <- numeric(B)
outcome2 <- numeric(B)
outcome3 <- numeric(B)
for (i in 1:B) {
    d <- rnorm(17, 3, 5)  # generate data (of the differences)
    outcome1[i] <- binom.test(sum(d > 0), 17)$p.value < 0.05
    outcome2[i] <- wilcox.test(d)$p.value < 0.05
    outcome3[i] <- t.test(d)$p.value < 0.05
}
# Point estimates.
c(mean(outcome1), mean(outcome2), mean(outcome3))

## [1] 0.4848 0.6010 0.6339

# 95% confidence intervals.
prop.test(sum(outcome1), B)$conf.int

## [1] 0.4749625 0.4946493
## attr(,"conf.level")
## [1] 0.95

prop.test(sum(outcome2), B)$conf.int

## [1] 0.591315 0.610607
## attr(,"conf.level")
## [1] 0.95

prop.test(sum(outcome3), B)$conf.int

## [1] 0.6243581 0.6433385
## attr(,"conf.level")
## [1] 0.95
```

Note: As an extension exercise, prove that the following calculations give the exact power for the sign test and t-test respectively.

```
1 - pbinom(12, 17, pnorm(0.6)) +
    pbinom( 4, 17, pnorm(0.6))

## [1] 0.4818206

1 - pt(qt(0.975, 16), 16, sqrt(17) * 3 / 5) +
    pt(qt(0.025, 16), 16, sqrt(17) * 3 / 5)

## [1] 0.6419633
```

2.
```
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)
```

(a) Each observation is a count across 30 experiments. By assuming a binomial distribution, we are modelling each experiment as a Bernoulli trial with a common success parameter, p. Therefore, we can simply merge all of the data together and consider them as a single set of 80×30 Bernoulli trials; or, equivalently, the sum of the successes across all experiments follows a $\text{Bi}(80 \times 30, p)$ distribution. The MLE of p is therefore the sample proportion across all experiments.

```
p.hat <- sum(germinations * count) / (80 * 30)
p.hat

## [1] 0.3054167
```

(b) For the theoretical distribution we will use a binomial with p set to the MLE from above. The observations in the tails will have very low probability and will need to be merged together. To start with, let's check the expected counts if we don't merge any of the data in the table:

```
classProb1 <- c(pbinom(3,    30, p.hat),
                dbinom(4:13, 30, p.hat),
                pbinom(  13, 30, p.hat, lower.tail = FALSE))
names(classProb1) <- c("0-3", 4:13, "14-30")
80 * classProb1  # expected counts

##        0-3          4          5          6          7          8
##  0.6334039  1.4632685  3.3457675  6.1298929  9.2413297 11.6826318
##          9         10         11         12         13      14-30
## 12.5570966 11.5951642  9.2700609  6.4539162  3.9293512  3.6981166
```

On the lower tail we need to merge the observations from 0 to 5. On the upper tail we need to merge the 13 into the 14–30 class.

```
classProb2 <- c(pbinom(5,    30, p.hat),
                dbinom(6:12, 30, p.hat),
                pbinom(  12, 30, p.hat, lower.tail = FALSE))
names(classProb2) <- c("0-5", 6:12, "13-30")
```

```r
80 * classProb2  # expected counts

##       0-5         6         7         8         9        10        11
##  5.442440  6.129893  9.241330 11.682632 12.557097 11.595164  9.270061
##        12     13-30
##  6.453916  7.627468
```

This looks good now.

(c) Merging the data to correspond to the desired set of classes:

```r
count2 <- c(sum(count[1:3]), count[4:10], sum(count[11:12]))
names(count2) <- c("0-5", germinations[4:10], "13-30")
count2

##   0-5     6     7     8     9    10    11    12 13-30
##     5     4    10    16     9    11    13     4     8
```

Running the test in R:

```r
chisq.test(count2, p = classProb2)

##
##  Chi-squared test for given probabilities
##
## data:  count2
## X-squared = 5.924, df = 8, p-value = 0.6557
```

However, the default degrees of freedom is incorrect since it doesn't take into account the parameter we estimated. Re-calculating the p-value with the correct degrees of freedom:

```r
1 - pchisq(5.924, 7)

## [1] 0.5486511
```

The p-value is greater than our desired significance level (0.05) so we cannot reject the null. A binomial distribution fits the data reasonably well.

3. The cdf of X is $F(x) = \int_1^x \theta y^{-(\theta+1)} dy = \left[-y^{-\theta}\right]_1^x = 1 - x^{-\theta}$.

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

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

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

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

(c) 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,

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

4.
```r
# Set up the data.
y <- c(42.3, 41.4, 42.2, 40.3, 37.6, 35.7, 36.8, 34.9, 45.8, 43.7,
       42.1, 40.2, 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)
angle <- factor(rep(c(0, 10, 20, 30), each = 5 * 2))
panel <- factor(rep(1:5, times = 4, each = 2))

# Quick check that the factors are structured correctly.
angle

##  [1] 0  0  0  0  0  0  0  0  0  0  10 10 10 10 10 10 10 10 10 10 20 20 20
## [24] 20 20 20 20 20 20 20 30 30 30 30 30 30 30 30 30 30
## Levels: 0 10 20 30

panel

##  [1] 1 1 2 2 3 3 4 4 5 5 1 1 2 2 3 3 4 4 5 5 1 1 2 2 3 3 4 4 5 5 1 1 2 2 3
## [36] 3 4 4 5 5
## Levels: 1 2 3 4 5
```

(a)
```r
# Two-way ANOVA.
anova(lm(y ~ angle + panel))

## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## angle      3  36.890  12.297  5.8632  0.002614 **
## 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
```

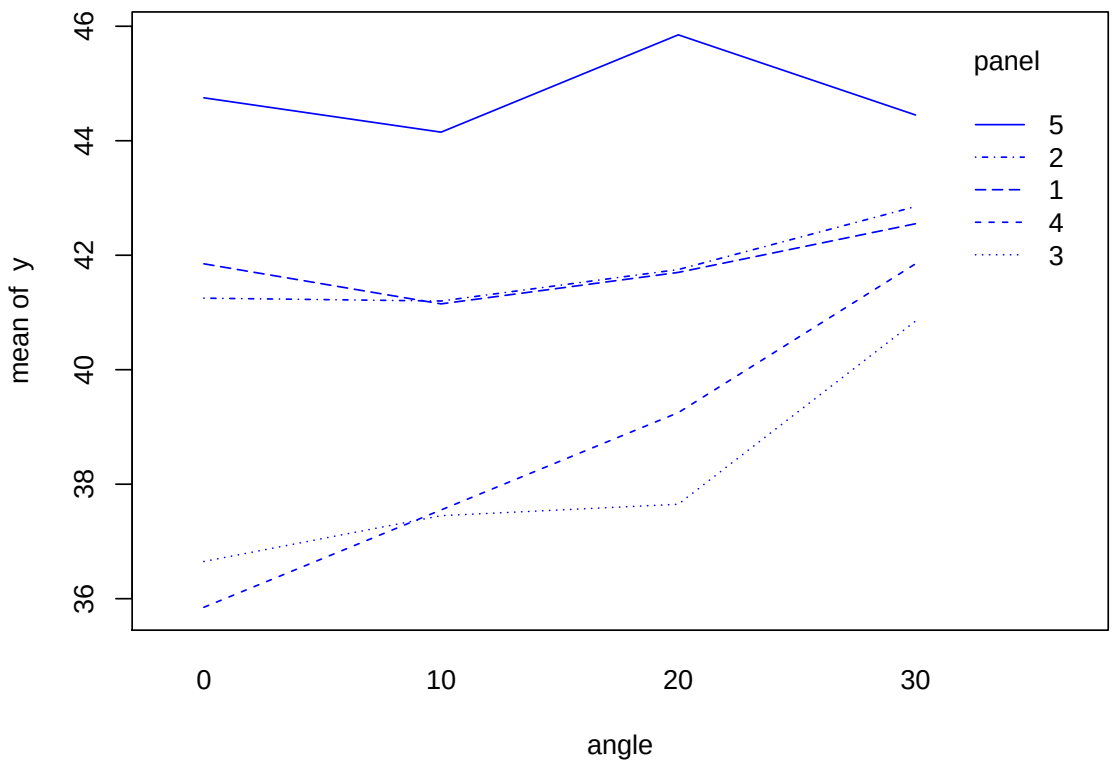
The outcome variable is modelled as $Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$, with $i = 1, \ldots, 4$ and $j = 1, \ldots, 5$ and $k = 1, 2$, where ε_{ijk} denote independent errors such that $\varepsilon_{ijk} \sim \mathrm{N}(0, \sigma^2)$. Further, we assume that $\sum_i \alpha_i = 0$ and $\sum_j \beta_j = 0$. The null hypothesis of interest is $H_0\colon \alpha_i = 0$ for all i, while the alternative hypothesis is that at least one of the α_i is non-zero.

The observed test statistic is $F = \frac{SS(\text{angle})/3}{SS(E)/32} = 5.8632$, which will follow an $F_{3,32}$ distribution under the null. We obtain a p-value of 0.003, which means we can reject H_0 at a 5% significance level. We have strong evidence that the power output is affected by the different angles tested. (Note that we also have very strong evidence that the power output differs between the different types of panels.)

(b) Since we have multiple observations for each factor combination, it **is** possible to test for interaction. Let's do that:

```
anova(lm(y ~ angle * panel))

## Analysis of Variance Table
##
## Response: y
##             Df  Sum Sq Mean Sq F value    Pr(>F)
## angle        3  36.890  12.297  6.9610  0.002163 **
## panel        4 235.522  58.880 33.3317 1.383e-08 ***
## angle: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
```

We obtain a p-value of 0.2, on the basis of which we have insufficient evidence to rule out the lack of any interaction (i.e. it is plausible that the interaction is negligible).

```
par(mar = c(4, 4, 1, 1))  # tighter margins
interaction.plot(angle, panel, y, col = "blue")
```