Assignment 2 - ANSWER BOOKLET

COMP616 / STAT604

Julia Calma 22167597

INSTRUCTIONS: Use this file to write your answer to Assignment 2. You will need to knit it to PDF and then submit this PDF to Canvas.

Insert as many R code chunks as needed, using:

```
# copy and paste
```

If mathematical notation/ formulae are needed, please copy-paste and modify those in the .Rmd files used in the labs.

ANSWERS FOR QUESTION 1

1(a)

```
weight <- c(19.0, 17.2, 16.7, 17.0, 17.4, 12.4, 10.1, 14.6, 13.4, 14.4)
height <- c(101.0, 113.2, 102.9, 80.4, 96.9, 108.7, 97.2, 98.4, 104.0, 81.9)
#Spearman correlation
spearman<-cor.test(weight,height,method="spearman")
#Kendall Tau
kendall<-cor.test(weight,height,method="kendall")

print(spearman)
```

```
##
##  Spearman's rank correlation rho
##
## data:  weight and height
## S = 180, p-value = 0.8114
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##         rho
## -0.09090909
```

```
print(kendall)
```

```
##
##  Kendall's rank correlation tau
##
```

```
## data:  weight and height
## T = 21, p-value = 0.8618
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##         tau
## -0.06666667
```

1(b)

$$p = 2 * \sum_{k=0}^{2} \binom{12}{k} (0.5)^{12} = 0.03857421875$$

```
# Define the parameters
n <- 12
x <- 10
T <- min(x, n - x)  # Smaller of number of positives and negatives

# Calculate the p-value for the two-sided test
p_value <- 2 * pbinom(T, n, 0.5)

# Output the result
p_value
```

```
## [1] 0.03857422
```

Conclusion: We see that the p-value (0.03) is lower than the significance level,0.5. This means that we reject our null hypothesis,therefore the median of a population is not 55.

1(c)

Run test would be suitable to determine if the sequence of the colours of the balls is random or not as it its main purpose is testing randomness and we only have two colours.

$H_0 : The\ sequence\ of\ colors\ is\ random$

$H_1 : The\ sequence\ of\ colors\ is\ not\ random$

Max of R = 2×min(n1,n2)+1 Max of R = 2×min(4,6)+1 Max of R = 2x4+1

Max of R = 9 Min of R = 2

1(d)

$H_0 : M_{locA} = M_{locB}$

$H_1 : M_{locA} > M_{locB}$

```
locA <- c(6,5,4,3,7,8)
locB <- c(3,2,7,4,6,4)

wilcox.test(locA, locB, alternative="greater", exact=TRUE,correct=FALSE)
```

```
## Warning in wilcox.test.default(locA, locB, alternative = "greater", exact =
## TRUE, : cannot compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test
##
## data:  locA and locB
## W = 24.5, p-value = 0.146
## alternative hypothesis: true location shift is greater than 0
```

Conclusion: We reject the null hypothesis as p-values is greater than significance level. Therefore, we conclude that the UV index location A is higher than that at location B.

1(e)

The probability we are looking for is $P(C|Pos)$, which can be expressed as

$$P(C|Pos) = \frac{P(Pos|C)P(C)}{P(Pos)}.$$

From the information given, we have

- $P(C) = 0.10$
- $P(NoC) = 0.90$
- $P(Pos|C) = 0.90$
- $P(Neg|NoC) = 0.95$
- $P(Pos|NoC) = 0.05$
- $P(Neg|C) = 0.10$

To find $P(Pos)$, we use the identity

$$\begin{aligned} P(Pos) &= P(Pos|C)P(C) + P(Pos|NoC)P(NoC) \\ &= 0.90 \times 0.10 + 0.05 \times 0.90 \\ &= 0.945 \end{aligned}$$

Thus,

$$P(C|Pos) = \frac{P(Pos|C)P(C)}{P(Pos)} = \frac{0.90 \times 0.10}{0.945} = 0.0952.$$

9.52%

ANSWERs FOR QUESTION 2

2(a)

For the purpose of determining if there is a difference in the median mathematical skills among the staff from the above four departments, I will choose the Kruskall Wallis Test. This is because it is used to compare the medians of two or more samples using the ranks of the data.

2(b)

Let M denote the median. The hypotheses can be written as:

H_0: $M_{\text{artdes}} = M_{\text{educ}} = M_{\text{midwife}} = M_{\text{law}}$

H_1: not all medians are equal.

2(c)

```
results <- c(41, 52, 32, 44, 45, 49, 43, 60, 55, 53, 48, 26, 27, 46, 51, 17, 30, 23, 39, 37, 13, 29)
dep <- c(rep("artdes",6),rep("educ",5),rep("midwife",5),rep("law",6))
print(rbind(dep,results))
```

```
##         [,1]     [,2]     [,3]     [,4]     [,5]     [,6]     [,7]   [,8]
## dep     "artdes" "artdes" "artdes" "artdes" "artdes" "artdes" "educ" "educ"
## results "41"     "52"     "32"     "44"     "45"     "49"     "43"   "60"
##         [,9]   [,10]  [,11]  [,12]     [,13]     [,14]     [,15]     [,16]
## dep     "educ" "educ" "educ" "midwife" "midwife" "midwife" "midwife" "midwife"
## results "55"   "53"   "48"   "26"      "27"      "46"      "51"      "17"
##         [,17] [,18] [,19] [,20] [,21] [,22]
## dep     "law" "law" "law" "law" "law" "law"
## results "30"  "23"  "39"  "37"  "13"  "29"
```

2(d)

Sum of ranks for the staff from the Midwifery department: 2+4+5+15+18=44

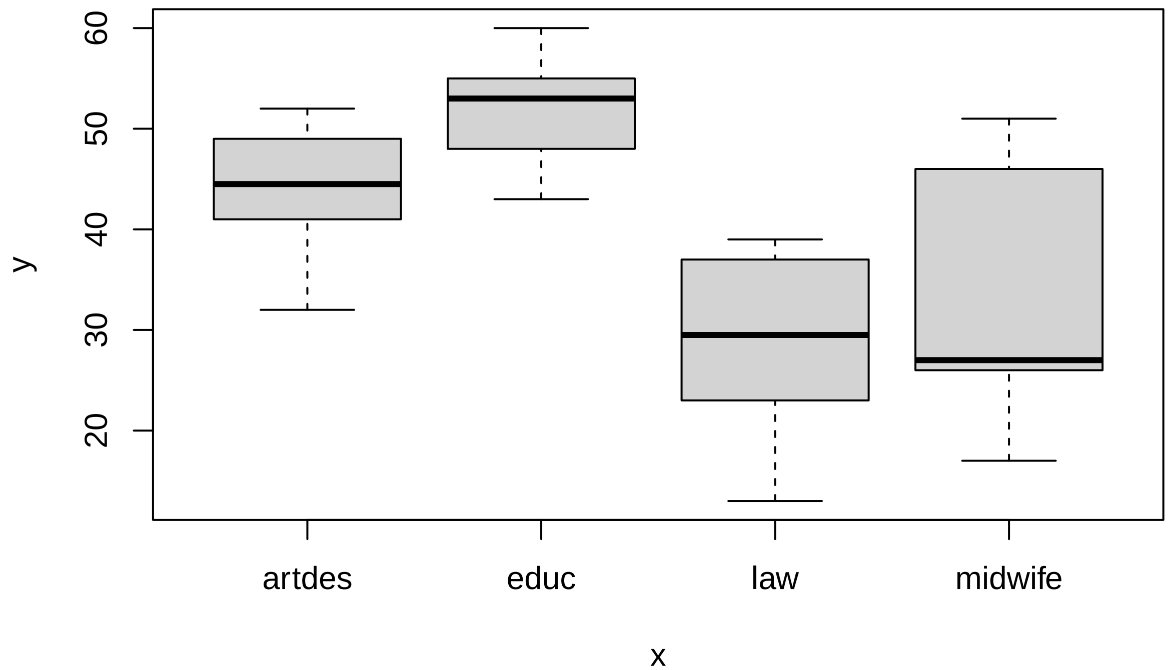
2(e)

```
kruskal.test(results,dep)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  results and dep
## Kruskal-Wallis chi-squared = 11.16, df = 3, p-value = 0.01089
```

Conclusion: We will reject the null hypothesis as p-value(0.01) is less than significance level (0.05), therefore we conclude that not all medians are equal. ## 2(f)

```r
plot(as.factor(dep),results)
```



From the boxplot, we can determine that the department that has the staff with the best math skills is education. Based on the medians,education is followed by art&design, law, and midwifery.

ANSWERS FOR QUESTION 3

3(a)

For the purpose of determining if there re any associations between the size of abalones and the areas, I will choose the Chi-sq Test of Independence. This is because it is used to identify if tqo qualitative factors are independent.

3(b)

H_0: the size of abalones and the areas are not associated

H_1: the size of abalones and the areas are associated

3(c)

```
dat <- as.table(rbind(c(150,57,14),c(96,45,38),c(52,25,23)))
dimnames(dat) <- list(Size = c("Large", "Medium", "Small"),
                       Areas = c("Area A","Area B","Area C"))
print(dat)
```

```
##         Areas
## Size      Area A Area B Area C
##   Large     150     57     14
##   Medium     96     45     38
##   Small      52     25     23
```

3(d)

```
prop.table(dat,1)
```

```
##         Areas
## Size          Area A     Area B     Area C
##   Large  0.67873303 0.25791855 0.06334842
##   Medium 0.53631285 0.25139665 0.21229050
##   Small  0.52000000 0.25000000 0.23000000
```

From the table, it is quite clear that the proportion of Area A and B decreases with size of the abalones. With such an apparent trend, it is expected that the null hypothesis will be rejected.

3(e)

Expected number of large abalones in Area C=(221*75)/500= 33.15 Expected number of medium abalones in Area C=(179*75)/500= 26.85

3(f)

```r
chisq.test(dat)
```

```
##
##  Pearson's Chi-squared test
##
## data:  dat
## X-squared = 24.561, df = 4, p-value = 6.164e-05
```

Conclusion:The p-value (0.00006164) is less than the significance level (0.05), therefore our null hypothesis is rejected. There is evidence suggesting that size of abalones and areas are associated.

END OF ASSIGNMENT 2