Sandra Villamar
MATH 185

HW #2

1. Let $Y_{(1)}, \dots, Y_{(m)}$ be order statistics for y sample. For fixed j, the # of X_i w/ $X_i < Y_{(j)}$ is:

$$(\text{rank}(Y_{(j)})) - (\# Y_k : Y_k \leq Y_{(j)}) = R_j - j$$

So we have:

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} I_{i,j} = \sum_{j=1}^{m} \sum_{i=1}^{n} I_{i,j} = \sum_{j=1}^{m} (R_j - j) = \sum_{j=1}^{m} R_j - \sum_{j=1}^{m} j = W - \frac{1}{2}(m)(m+1)$$

$$\Longrightarrow \boxed{W = U + \frac{1}{2}(m)(m+1)}$$

2. ① $E(I_{i,j}) = P(X_i < Y_j) = \boxed{\frac{1}{2}}$

↳ - since X and Y are i.i.d., $1 = P(X_i < Y_j) + P(Y_j < X_i) = 2P(X_i < Y_j)$

- same distribution ⇒ equally likely to be lower than the other

② $E(I_{i,j} I_{i,l}) = P(X_i < Y_j \text{ AND } X_i < Y_l) = P(X_i = \min\{X_i, Y_j, Y_l\}) = \boxed{\frac{1}{3}}$

↳ - since i.i.d. means X_i, Y_j, Y_l all have equal chance to be the minimum.

③ $E(I_{i,j} I_{k,l}) = E(I_{i,j}) E(I_{k,l}) = (\frac{1}{2})(\frac{1}{2}) = \boxed{\frac{1}{4}}$

↳ - we can apply ① since $I_{i,j}$ and $I_{k,l}$ are independent events

```r
#PART 3
unexp <- c(8,11,12,14,20,43,111)
exp <- c(35,56,83,92,128,150,176,208)
expNew <- exp - 25
#HT
wilcox.test(unexp, expNew, alternative = "less", exact = TRUE)
#CI
wilcox.test(unexp, exp, exact = TRUE, conf.int = T)


#PART 4
pH <- c(7.02,7.34,7.28,7.09,7.45,7.40,7.32)
wilcox.test(pH, exact = TRUE, conf.int = T)
```

3. (See R code above!!)

The results from the HT were:

```
        Wilcoxon rank sum test

data:  unexp and expNew
W = 11, p-value = 0.02704
alternative hypothesis: true location shift is less than 0
```

Hence, since the **p-value = .02704 < alpha = .05, we reject the null hypothesis** that $\Delta = 25$, and move towards the alternative hypothesis that $\Delta > 25$.

The results from the CI were:

```
        Wilcoxon rank sum test

data:  unexp and exp
W = 5, p-value = 0.005905
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 -156  -24
sample estimates:
difference in location
                   -79
```

But since R computes the difference by: unexposed – exposed, we must flip the CI for our case, where we have defined Δ = exposed – unexposed. Hence, the CI for Δ at alpha = 5% is **[24, 156]**.

4. (See R code above!!)

The 95% signed-rank CI for the population median is: **[7.09, 7.40]**.

The results from the test are shown below:

```
        Wilcoxon signed rank test

data:  pH
V = 28, p-value = 0.01563
alternative hypothesis: true location is not equal to 0
95 percent confidence interval:
 7.09 7.40
sample estimates:
(pseudo)median
          7.29
```