38. Apply the two-sided Wilcoxon rank sum test procedure from Section 4.1 to the salivation data in Table 5.7 by finding the appropriate P-value. Compare the conclusion indicated by this Wilcoxon rank sum procedure with that indicated by the Kolmogorov–Smirnov procedure in Example 5.4. Comment on your findings.

Table 5.7 Mean Drop Differences

Feedback group		No-Feedback group	
−.15	5	2.55	11
8.60	19	12.07	20
5.00	16	.46	7
3.71	14	.35	6
4.29	15	2.69	12
7.74	17	−.94	2
2.48	10	1.73	9
3.25	13	.73	8
−1.15	1	−.35	4
8.38	18	−.37	3

Source: F. C. Delse and B. W. Feather (1968).

Let X_i represents No-Feedback group. $i=1,2\cdots 10$

Y_j represents Feedback group. $j=1,2\cdots 10$

① Wilcoxon rank sum test.

$N = m+n = 10+10 = 20.$

$\hat{W} = \sum_{j=1}^{10} S_{Y_j} = 5+19+16+14+15+17+10+13+1+18 = 128$

Under the null: $E_0(\hat{W}) = (n+m+1)n/2 = 105.$

$Var_0(\hat{W}) = nm(m+n+1)/12 = 175.$

$W = \hat{W} - E_0(\hat{W}) / \sqrt{Var_0(\hat{W})} \xrightarrow{d} N(0,1)$

$\approx 1.74.$

Let $\alpha = 0.05.$ $\Phi(1-\frac{1}{2}\alpha) = 1.96.$ $1.74 < 1.96 \Rightarrow$ fail to reject the null. $\hat{p} \approx 0.082.$

② Kolmogorov–Smirnov Test.

From the textbook. $J = \frac{mn}{d} \max\{|F_{10}(Z_{(i)}) - G_{10}(Z_{(i)})|\}$

$= 6.$

$J^* = \frac{d}{(mnN)^{\frac{1}{2}}} J = 10/\sqrt{10\times10\times20} \cdot 6 \approx 1.34.$

Let $\alpha = 0.05$, $q^*_{0.05} = 1.358$, $J^* < q^*_{0.05}$

$\Rightarrow$ fail to reject the null. $\hat{p} \approx 0.0511.$

Finding: In both tests, the results agree with each other, that we fail to reject the null. But in terms of p-value, KS test tends to reject the null compared with Wilcoxon rank test.

39. Generate the exact null distribution of J (5.70) for the setting $m = 3$, $n = 3$. (See Comment 38.)

$m=3,\ n=3 \qquad N = m+n = 6. \qquad d = 3. \qquad \binom{6}{3} = 20.$

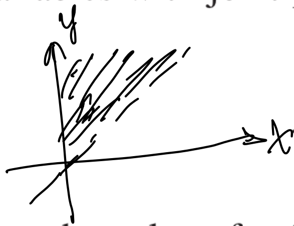
$$J = mn/d \cdot \max\{|F_3(Z_{(i)}) - G_3(Z_{(i)})|\}.$$

$$= 3 \max\{|F_3(Z_{(i)}) - G_3(Z_{(i)})|\}$$

Meshing	$(F_3(Z_{(1)} \cdots Z_{(6)}))$	$(G_3(Z_{(1)} \cdots Z_{(6)}))$	J
XXXYYY	$(\frac{1}{3}, \frac{2}{3}, 1, 1, 1, 1)$	$(0, 0, 0, \frac{1}{3}, \frac{2}{3}, 1)$	3
XXYXYY	$(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1, 1, 1)$	$(0, 0, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1)$	2
XXYYXY	$(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}, 1, 1)$	$(0, 0, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	2
XXYYYX	$(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	$(0, 0, \frac{1}{3}, \frac{2}{3}, 1, 1)$	2
XYXYYX	$(\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	$(0, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1, 1)$	1
XYYXYX	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	$(0, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1, 1)$	1
XYYYXX	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1)$	$(0, \frac{1}{3}, \frac{2}{3}, 1, 1, 1)$	2
YXYYXX	$(0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1)$	$(\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1, 1, 1)$	2
YYXYXX	$(0, 0, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1)$	$(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1, 1, 1)$	2.
YYYXXX	$(0, 0, 0, \frac{1}{3}, \frac{2}{3}, 1)$	$(\frac{1}{3}, \frac{2}{3}, 1, 1, 1, 1)$	3
XYXYXY	$(\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1, 1)$	$(0, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	1.
XYYXXY	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1, 1)$	$(0, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	1
YYXXXY	$(0, 0, \frac{1}{3}, \frac{2}{3}, 1, 1)$	$(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	2
YXYXYX	$(0, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	$(\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1, 1)$	1
YXXYYX	$(0, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1, 1)$	1
YXXYXY	$(0, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1, 1)$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	1
YXXXYY	$(0, \frac{1}{3}, \frac{2}{3}, 1, 1, 1)$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1)$	2
XYXXYY	$(\frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1, 1, 1)$	$(0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1)$	2
YXYXXY	$(0, \frac{1}{3}, \frac{1}{3}, \frac{2}{3}, 1, 1)$	$(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}, 1, 1)$	1
YYXXYX	$(0, 0, \frac{1}{3}, \frac{2}{3}, \frac{2}{3}, 1)$	$(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}, \frac{2}{3}, 1, 1)$	2.

$J = k$	1	2	3
$P\{J = k\}$	$\frac{2}{5}$	$\frac{1}{2}$	$\frac{1}{10}$

4. Let (X_1, Y_1) and (X_2, Y_2) be independent and identically distributed continuous bivariate random variables with joint probability density function



$$f_{X,Y}(x,y) = \begin{cases} e^{-y}, & 0 < x < y < \infty, \\ 0, & \text{elsewhere.} \end{cases} \qquad e^{-y} \cdot \mathbb{1}_{\{0<x<y\}}$$

Calculate the value of τ for this bivariate distribution.

$$\tau = P\{(X_2 - X_1)(Y_2 - Y_1) > 0\} - P\{(X_2 - X_1)(Y_2 - Y_1) < 0\}$$
$$= 2P\{(X_2 - X_1)(Y_2 - Y_1) > 0\} - 1$$

Note that $P\{(X_2 - X_1)(Y_2 - Y_1) > 0\} = P\{X_2 > X_1, Y_2 > Y_1\} + P\{X_1 > X_2, Y_1 > Y_2\}$

Since $(X_1, Y_1) \perp (X_2, Y_2)$,

$$P\{X_2 > X_1, Y_2 > Y_1\} = P\{X_1 > X_2, Y_1 > Y_2\}$$

$$P\{(X_2 - X_1)(Y_2 - Y_1) > 0\} = 2P\{X_2 > X_1, Y_2 > Y_1\}.$$

Let $F_{XY}(x,y) = \int_{x,y} f_{XY}(u,v)\, d(u,v)$

$$= \int_0^x \int_u^y e^{-v}\, dv\, du$$

$$= 1 - e^{-x} - x e^{-y}.$$

$$P\{X_2 > X_1, Y_2 > Y_1\} = \int_x^\infty \int_0^\infty F_{XY}(x,y) f_{XY}(x,y)\, dx\, dy$$

$$= \int_0^\infty \int_x^\infty (1 - e^{-x} - x e^{-y}) e^{-y}\, dy\, dx$$

$$= \int_0^\infty e^{-x} - e^{-2x} - \frac{x}{2} e^{-2x}\, dx$$

$$= 1 - \frac{1}{2} - \frac{1}{8} = \frac{3}{8}$$

$$\tau = 4P\{X_2 > X_1, Y_2 > Y_1\} - 1 = \frac{1}{2}$$

15. For the case of $n = 5$ untied bivariate (X, Y) observations, obtain the form of the exact null (H_0) distribution of K. (See Comment 6.)

$n=5$ $\quad K^*$: $1,\ \frac{4}{5},\ \frac{3}{5},\ \frac{2}{5},\ \frac{1}{5},\ 0,\ -\frac{1}{5},\ -\frac{2}{5},\ -\frac{3}{5},\ -\frac{4}{5},\ -1$

$n! = 120$ $\quad P\{K^*=k\}$: $\frac{1}{120}$ $\frac{1}{30}$ $\frac{3}{40}$ $\frac{1}{8}$ $\frac{1}{6}$ $\frac{11}{60}$ $\frac{1}{6}$ $\frac{1}{8}$ $\frac{3}{40}$ $\frac{1}{30}$ $\frac{1}{120}$

19. Gerstein (1965) studied the long-term pollution of Lake Michigan and its effect on the water supply for the city of Chicago. One of the measurements considered by Gerstein was the annual number of "odor periods" over the period of years 1950–1964. Table 8.8 contains this information for Lake Michigan for each of these 15 years.

Test the hypothesis that the degree of pollution (as measured by the number of odor periods) had not changed with time against the alternative that there was a general increasing trend in the pollution of Lake Michigan over the period 1950–1964. (See Comment 14.)

Table 8.8 Annual Number of Odor Periods for Lake Michigan for the Period 1950–1964

Year	Number of odor periods
1950	10
1951	20
1952	17
1953	16
1954	12
1955	15
1956	13
1957	18
1958	17
1959	19
1960	21
1961	23
1962	23
1963	28
1964	28

Source: H. H. Gerstein (1965).

Let X_i represents Number of odor period.
Y_i represent year. $i = 1, 2, \cdots 15$. $n = 15$

H_0: X_i and Y_i is uncorrelated.
Construct Kendall's statistics.
There in totall 15 observations. $\binom{15}{2} = 105$ pairs.

```
#Lake Michigan example. Kendall test
```{r}
# Example data: Two vectors of 15 observations each
x <- c(10, 20, 17, 16, 12, 15, 13, 18, 17, 19, 21, 23, 23, 28, 28)
y <- c(1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964)

# Calculate Kendall's tau
kendall_tau <- cor(x, y, method = "kendall")

# Print the result
print(kendall_tau)
```

[1] 0.6570738
```

$\hat{\tau} \simeq 0.657$.

Using large sample approximation:

Under H_0: $\quad \tau^* = \dfrac{\hat{\tau} - E_0(\hat{\tau})}{\sqrt{Var_0(\hat{\tau})}} \sim N(0,1)$

$E_0(\hat{\tau}) = 0 \qquad \sqrt{Var_0(\hat{\tau})} = \sqrt{2(2n+5)/9n(n-1)}$
$\simeq 0.1925$.

$\alpha = 0.05$ $\qquad \tau^* \simeq 3.41 > 1.96 = \Phi(1-\frac{1}{2}\alpha)$

$\Rightarrow$ reject the null

Pollution of Lake Michigan is correlated with time