Homework 5

20307100013 蒋翌坤

Solution for Problem 1

1.1

To get Kendall's τ, we use the following test statistic: $T = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} U_{ij}V_{ij}}{\left[\left(\sum_{i=1}^{n}\sum_{j=1}^{n} U_{ij}^2\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{n} V_{ij}^2\right)\right]^{1/2}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} U_{ij}V_{ij}}{n(n-1)}$, where $U_{ij} = sgn\left(X_j - X_i\right)$, $V_{ij} = sgn\left(Y_j - Y_i\right)$. $Z = \frac{3\sqrt{n(n-1)}}{\sqrt{2(2n+5)}} T \sim N(0,1)$

We can calculate Z statistic and compare it with the critical value $z_{\alpha/2} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$. If $|Z| \leq z_{\alpha/2}$, we cannot reject the null hypothesis that X and Y are independent.

From 1,000 simulations, 93.8% of the simulations reach the conclusion that the null hypothesis cannot be rejected. Therefore, we cannot reject the null hypothesis that X and Y are independent using Kendall's τ.

1.2

To get Spearman's correlation, we use the following test statistic: $R = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\left[\left(\sum_{i=1}^{n}(R_i - R)^2\right)\left(\sum_{i=1}^{n}(S_i - \bar{S})^2\right)\right]^{1/2}} = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)}$, where $R_i = rank(X_i)$, $S_i = rank(Y_i)$, $D_i = R_i - S_i$. $t = \frac{R\sqrt{n-2}}{1-R^2} \sim t_{n-2}$

We can calculate t statistic and compare it with the critical value $t_{\alpha/2} = t_{n-2}^{-1}\left(1 - \frac{\alpha}{2}\right)$. If $|t| \leq t_{\alpha/2}$, we cannot reject the null hypothesis that X and Y are independent.

From 1,000 simulations, 93.6% of the simulations reach the conclusion that the null hypothesis cannot be rejected. Therefore, we cannot reject the null hypothesis that X and Y are independent using Spearman's correlation.

Solution for Problem 2

We can use the same method from Problem 2 to test the null hypothesis using Kendall's τ and Spearman's correlation.

From 1,000 simulations, 94.2% of the simulations reach the conclusion that the null hypothesis cannot be rejected using Kendall's τ; 93.9% of the simulations reach the conclusion that the null hypothesis cannot be rejected using Spearman's correlation. Therefore, we cannot reject the null hypothesis that X and Y are independent using Kendall's τ nor Spearman's correlation.

Kendall's τ is more powerful in this setting because X and Y are actually independent and Kendall's τ get more accurate results from simulations then Spearman's correlation.

Appendix

The following provides R code I used to complete this homework. You can also find the R code and output at https://thisiskunmeng.github.io/nonparametric/hw5.html

R code for Solution for Problem 1

```r
#' rank function
rank = function(x){
    n = length(x)
    rank = c()
    for (i in 1:n){
        rank = c(rank, sum(x <= x[i]))
    }
    return(rank)
#' Kendall's test
#' @return TRUE if the null hypothesis is not rejected
kendall <- function(x, y, a) {
    n <- length(x)
    u_ij <- c()
    v_ij <- c()
    for (i in 1:n) {
        for (j in 1:n) {
            u_ij <- c(u_ij, sign(x[j] - x[i]))
            v_ij <- c(v_ij, sign(y[j] - y[i]))
        }
    }
    t <- sum(u_ij * v_ij) / (n * (n - 1))
    z <- t * 3 * sqrt(n * (n - 1) / (2 * (2 * n + 5)))
    z_a <- qnorm(1 - a / 2)
    return(abs(z) <= z_a)
}
#' Spearman's test
#' @return TRUE if the null hypothesis is not rejected
spearman <- function(x, y, a) {
    n <- length(x)
    x_rank <- rank(x)
    y_rank <- rank(y)
    d <- x_rank - y_rank
    r <- 1 - 6 * sum(d^2) / (n * (n^2 - 1))
    t <- r * sqrt((n - 2) / (1 - r^2))
    t_a <- qt(1 - a / 2, n - 2)
    return(abs(t) <= t_a)
}
simulate_once <- function(seed) {
    set.seed(seed)
    n_pair <- 50
    x <- runif(n_pair, min = -1, max = 1)
    y <- rnorm(n_pair, mean = sqrt(1 - x^2), sd = 1)
    a <- 0.05
    return(c(kendall(x, y, a), spearman(x, y, a)))
}
n_simulate <- 1000
kendall_result <- c()
spearman_result <- c()
for (i in 1:n_simulate) {
    result <- simulate_once(seed = i)
```

```r
    kendall_result <- c(kendall_result, result[1])
    spearman_result <- c(spearman_result, result[2])
}
print(paste("Kendall's test: ", sum(kendall_result) / n_simulate))
print(paste("Spearman's test: ", sum(spearman_result) / n_simulate))
# "Kendall's test:  0.938", "Spearman's test:  0.936"
```

R code for Solution for Problem 2

```r
simulate_once <- function(seed) {
    set.seed(seed)
    n_pair <- 10
    x <- runif(n_pair, min = 1, max = 2)
    y <- rnorm(n_pair, mean = x / 2, sd = 1)
    a <- 0.05
    return(c(kendall(x, y, a), spearman(x, y, a)))
}
n_simulate <- 1000
kendall_result <- c()
spearman_result <- c()
for (i in 1:n_simulate) {
    result <- simulate_once(seed = i)
    kendall_result <- c(kendall_result, result[1])
    spearman_result <- c(spearman_result, result[2])
}
print(paste("Kendall's test: ", sum(kendall_result) / n_simulate))
print(paste("Spearman's test: ", sum(spearman_result) / n_simulate))
# "Kendall's test:  0.942", "Spearman's test:  0.939"
```