1 Задача 2.4

Формальная постановка задачи: Заданы выборки $\{X_i\}$, $\{Y_i\}$, $N = 14$ мощность выборок.
Нулевая гипотеза:$\mathbf{P}(X_i > Y_i) = 0.5$
Альтернатива: $\mathbf{P}(X_i > Y_i) \neq 0.5$

Используем критерий знаков $T = \sum_{i=1}^{N} [X_i > Y_i]$
Получаем, что $T \sim Bi(n, p = \frac{1}{2})$ - Распределение в условиях истинности нулевой гипотезы. $p-value = \frac{C_n^k}{2^n}$, где $k = T$. Таким образом, если $p-value \in [\frac{\alpha}{2}; \frac{1-\alpha}{2}]$ нулевая гипотеза принимается

Другой подход: предположим, что значения времени восстановления температуры - нормальная случайная величина. Формальная постановка задачи: Заданы выборки $\{X_i^{(1)}\}$, $\{X_i^{(2)}\}$, $N = 14$ мощность выборок. $X^{(k)} \sim N(\mu_k, \sigma_k)$
Нулевая гипотеза:$\mu_1 = \mu_2$
Альтернатива: $\mu_1 \neq \mu_2$
Статистика $-T\left(\mathbf{x}_1^{(n)}, \mathbf{x}_2^{(n)}\right) = \frac{\mathbf{x}_1 - \mathbf{X}_2}{s}\sqrt{n_1}$ где $S = \sqrt{\frac{1}{m-1}\sum_{i=1}^{n}\left(D_i - \bar{D}\right)^2}$, $D_i = X_{1i} - X_{2i}$
$T\left(\mathbf{X}_1^{(n)}, \mathbf{X}_2^{(n)}\right) \sim St(n-1)$ - Нулевое распределение. Таким образом, получаем что нулевая гипотеза отклоняется если $T\left(\mathbf{x}_1^{(n)}, \mathbf{x}_2^{(n)}\right) \notin \left[St(13)_{\frac{\alpha}{2}}, St(13)_{\frac{1-\alpha}{2}}\right]$

2 Задача 4.3

Недостаток предложенного подхода заключается в отсутствии поправок на множественную проверку гипотез.
Воспользуемся методом Холма, так как характер зависимости между статистиками неизвестен. Так как критерий Холма - наиболее мощный, контролирующий FWER.
По определению, FWER - групповая вероятность ошибки первого рода (familywise error rate).

$$FWER = \mathbf{P}(\text{число ошибок первого рода} > 0)$$

Метод Холма обеспечивает контроль над FWER на уровне α. Он гарантирует, что величина FWER никогда не будет больше чем α.
Из за применения метода Холма количество ошибок второго рода может увеличиться, в следствие чего мощность может снизиться.
Главный недостаток метода - мы не учитывали характер взаимосвязи статистик.
Новое условие дает означает, что имеющиеся у нас статистики независимы. Желание правительства означает необходимость минимизировать вероятность ошибки второго рода, поэтому необходимо контролировать величину FDR. FDR - ожидаемая доля ложных отклонений гипотез (false discovery rate). Воспользуемся методом Бенджамини-Хохберга. Метод обеспечивает контроль над FDR на уровне α при условии, что статистики независимы. Тогда мы получим, что мощность метода возросла,и соответственно количество ошибок второго рода меньше. Мощность возросла, однако это достигается за счет увеличения числа ошибок первого рода