STA 104 Applied Nonparametric Statistics

Chapter 4: One-Way Layout Problems: Nonparametric One-Way Analysis of Variance

---

Xiner Zhou

Department of Statistics, University of California, Davis

Table of contents

The Fligner-Wolfe Test for Treatments versus a Control

In this section, we discuss a test procedure specifically designed for the setting where one of the treatments corresponds to a control or baseline set of conditions and we are interested in assessing which, if any, of the treatments is better than the control.

Without loss of generality, we label the treatments so that the control corresponds to treatment 1. In this setting, the null hypothesis of interest is still the same, but now it corresponds to the statement that none of the treatments $2, \ldots, k$ is different from the control (treatment 1). This is usually expressed as

$$H_0 : [\tau_i = \tau_1, i = 2, \ldots, k] .$$

τ_1 : control group

$\tau_2 \cdots \tau_k$: treatment groups

$\Leftrightarrow$

$\tau_2 = \tau_1$

$\tau_3 = \tau_1$

$\vdots$

$\tau_k = \tau_1$

Hypothesis [3]

$k=3$: $\tau_2 \geq \tau_1$ $>$
$\tau_3 \geq \tau_1$ $>$

One-Sided Upper-Tail Test:

$$H_0 : [\tau_i = \tau_1, \text{ for } i = 2, \ldots, k]$$

$$H_1 : [\tau_i \geq \tau_1 \text{ ,, for } i = 2, \ldots, k, \text{ with at least one strict inequality}]$$

One-Sided Lower-Tail Test:

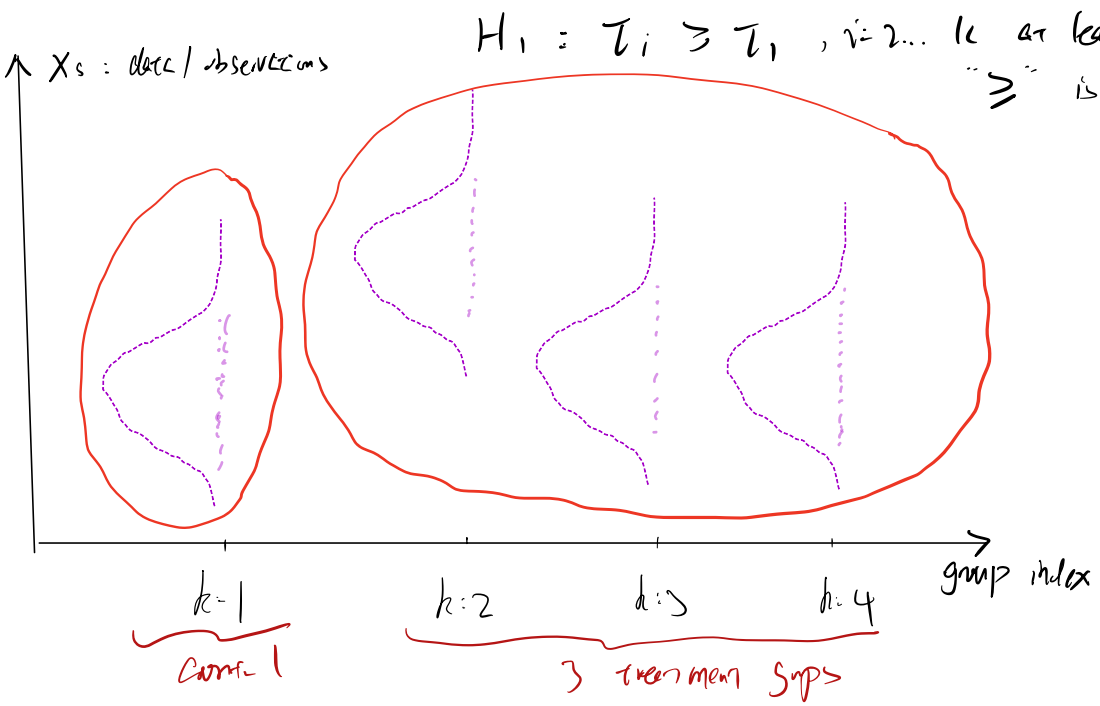
$$H_0 : [\tau_i = \tau_1, \text{ for } i = 2, \ldots, k]$$

$$H_1 : [\tau_i \leq \tau_1 \text{ ,, for } i = 2, \ldots, k, \text{ with at least one strict inequality}]$$

$\tau_2 \leq \tau_1$
$\tau_3 \leq \tau_1$ $<$

[3] We do not discussed a FW test designed for a two-sided alternative. The "natural" two-sided alternative for this treatment versus control setting corresponds to [either $\tau_i \geq \tau_1$ for all $i = 2, \ldots, k$ or $\tau_i \leq \tau_1$ for all $i = 2, \ldots, k$, with at least one strict inequality]. We feel that it is rather unlikely that we would find ourselves in such a setting where either all the treatments are better than the control or all the treatments are worse than the control, but we have no idea which of the two cases pertains.

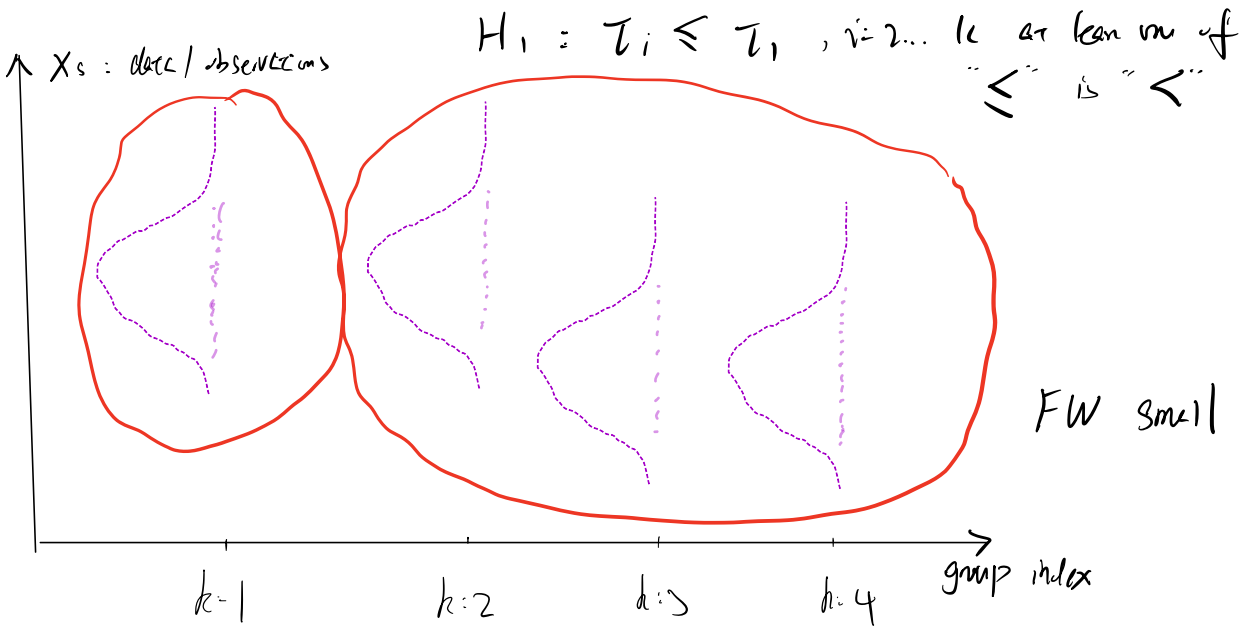
Suppose: $X \uparrow$ better

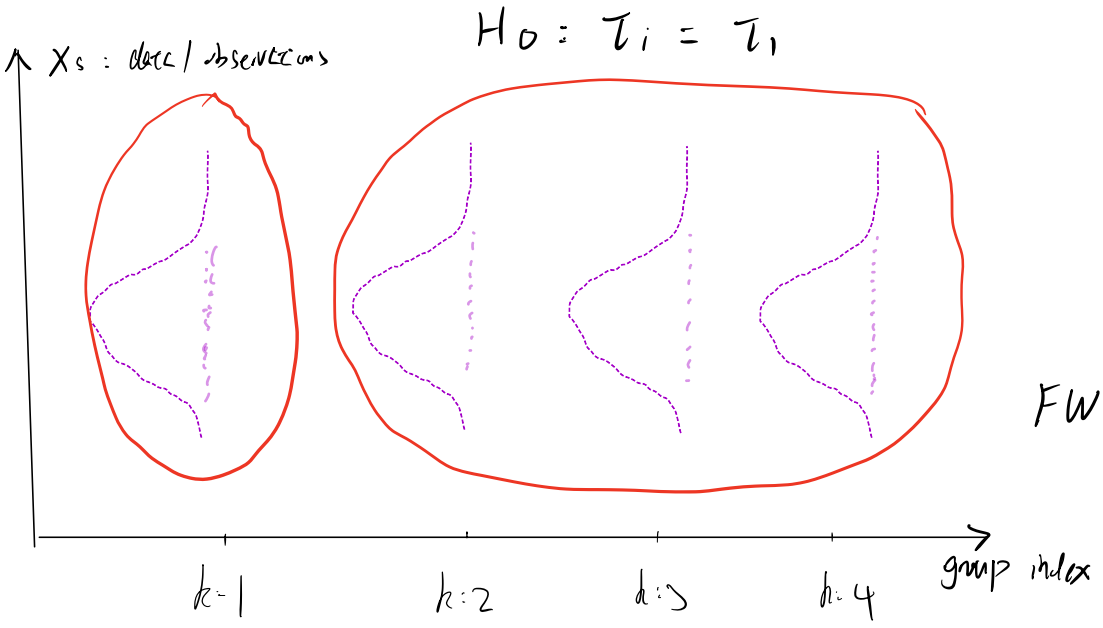


Step 1: jointly rank X_{ij}'s $\rightarrow R_{ij}$

Step 2: Sum up ranks associated with all treatment grps

$$FW = \sum_{i=2}^{K} \sum_{j=1}^{n_i} R_{ij} \quad \text{large}$$





FW not too large
nor too small

Did we see new test ? Yes & H_0

FW test is equivalent as one-sided Wilcoxon rank sum test for two-sample population.

$H_0: \Delta = \tau_y - \tau_x = \tau_2 - \tau_1 = 0$

$H_1: \underbrace{\Delta > 0}_{\tau_2 > \tau_1}$

$W = \sum_{i=1}^{n} R(Y_i)$

Motivation

First combine all N observations from the k samples and order them from least to greatest.

$\Rightarrow$ Letting R_{ij} denote the rank of X_{ij} in this joint ranking, the Fligner-Wolfe statistic FW is then the sum of these joint ranks for the non-control treatments,

$$\mathrm{FW} = \sum_{j=2}^{k} \sum_{i=1}^{n_j} R_{ij}$$

$\Rightarrow$ When some of the τ_i 's are strictly greater than the control effect τ_1, we would expect the joint ranks for the observations from those treatments to be larger than the joint ranks for the control observations. The net result would be a larger value of FW. This suggests rejecting H_0 in favor of H_1 for large values of FW .

Derivation of null distribution using permutation

$m : X$
$n : Y$
$\Longleftrightarrow$
$m = n_1 : X =$ control grp
$n = \sum_{i=2}^{k} n_i : Y =$ all treatment grps

FW can be viewed as a two-sample Wilcoxon rank sum statistic computed for the $m = n_1$ control treatment observations (playing the role of the X 's in the two sample setting) and the $n = \sum_{j=2}^{k} n_j$ combined observations from treatments $2, \ldots, k$ (playing the role of the Y 's in the two-sample setting).

As a result, the null distribution of FW is the same as that of the Wilcoxon rank sum statistic with sample sizes m, n.

Thus, the critical value f_α is just the upper α th percentile w_α for the null distribution of the Wilcoxon rank sum statistic with sample sizes m, n.

Large sample approximation of null distribution

The statistic FW has the same probability distribution as the null distribution of the two-sample Wilcoxon rank sum statistic W with sample sizes m, n. Hence, it follows directly from the Large-Sample Approximation for two-sample Wilcoxon rank sum statistic

$$\mathrm{FW}^* = \frac{\mathrm{FW} - \frac{n(N+1)}{2}}{\sqrt{mn(N+1)/12}} \sim N(0,1)$$

has, as $\min(n_1, N^*)$ tends to infinity, an asymptotic $N(0,1)$ distribution when H_0 is true.

Procedure

Permutation

To test

One-Sided Upper-Tail Test:

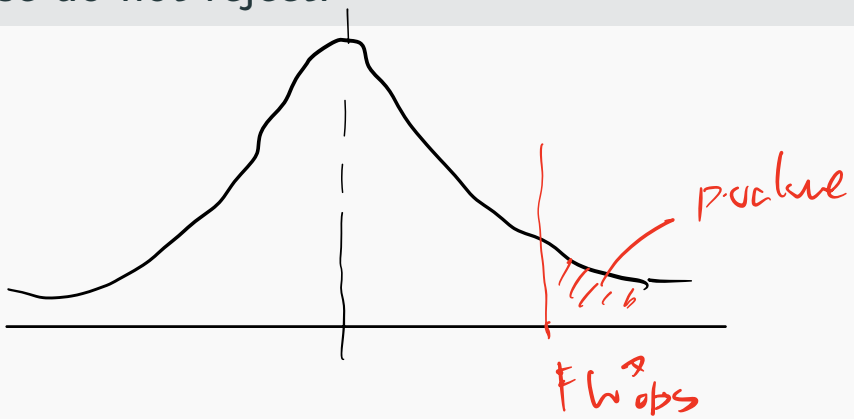
$$H_0 : [\tau_i = \tau_1, \text{ for } i = 2, \ldots, k]$$

$$H_1 : [\tau_i \geq \tau_1 \text{ , , for } i = 2, \ldots, k, \text{ with at least one strict inequality}]$$

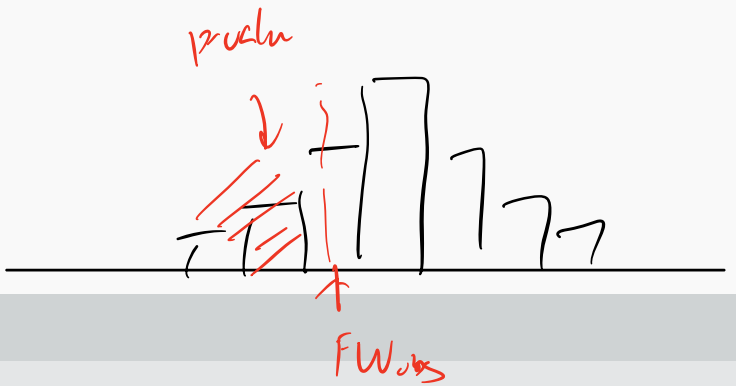
at the α level of significance, Reject H_0 if $\mathrm{FW} \geq f_\alpha$; otherwise do not reject.

Large-sample approximation

Reject H_0 if $\mathrm{FW}^* \geq z_\alpha$; otherwise do not reject.



Procedure



Permutation

To test

One-Sided Lower-Tail Test:

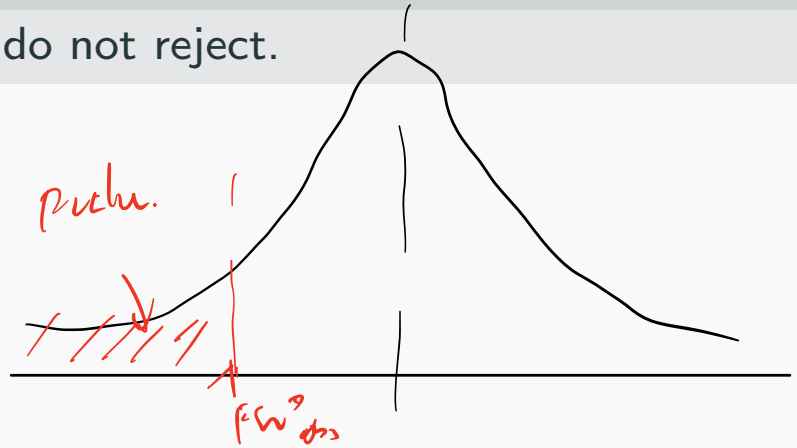
$$H_0 : [\tau_i = \tau_1, \text{ for } i = 2, \ldots, k]$$

$$H_1 : [\tau_i \leq \tau_1 \text{ ,, for } i = 2, \ldots, k, \text{ with at least one strict inequality}]$$

at the α level of significance, Reject H_0 if $\mathrm{FW} \leq f_{1-\alpha}$; otherwise do not reject.

Large-sample approximation

Reject H_0 if $\mathrm{FW}^* \leq -z_\alpha$; otherwise do not reject.

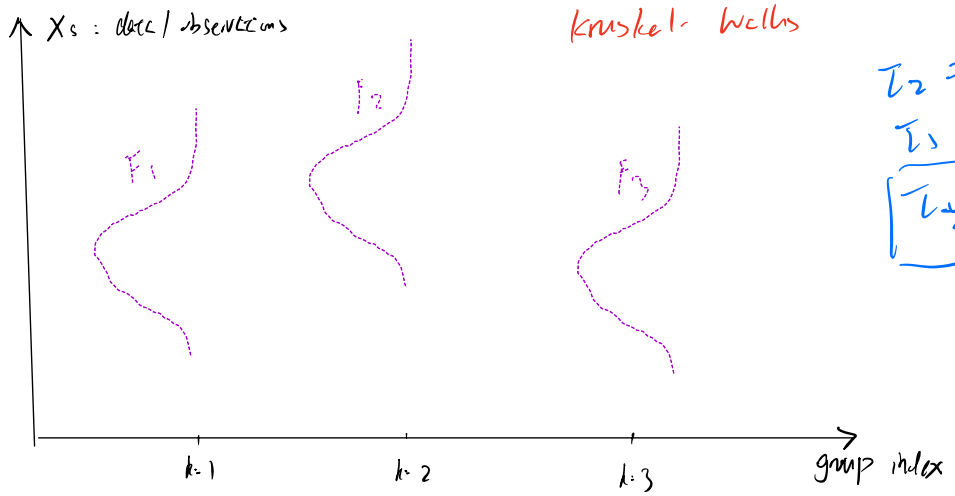


Remarks

The test deal with very restricted alternatives where all the treatments are either at least as good as the control or all the treatments are no better than the control, respectively. They are not appropriate tests when the possibility exists that some of the treatments might be better and some might be worse than the control. For such mixed alternatives, one would need to use the general alternatives Kruskal-Wallis test.

In many settings where we are interested in comparing a number of treatments with a control, we will have additional a priori information regarding the relative magnitude of the treatment effects. In a drug development, for instance, increasing dosage levels may be compared with a zero-dose control. If the treatment effects are not identical to that of the control, then it is often reasonable to assume that the higher the dose of the drug applied, the better (say, higher) will be the resulting effect on a patient, corresponding to monotonically ordered treatment effects.

However, it may also be the case that a subject might potentially succumb to toxic effects at high doses, thereby actually decreasing the associated treatment effects. Such a setting would correspond to an ordering in the treatment effects that is monotonically increasing up to a point, followed by a monotonic decrease; that is, an umbrella pattern on the treatment. This is a research topic very important for pharmaceutical industry – how to find the best dosage?

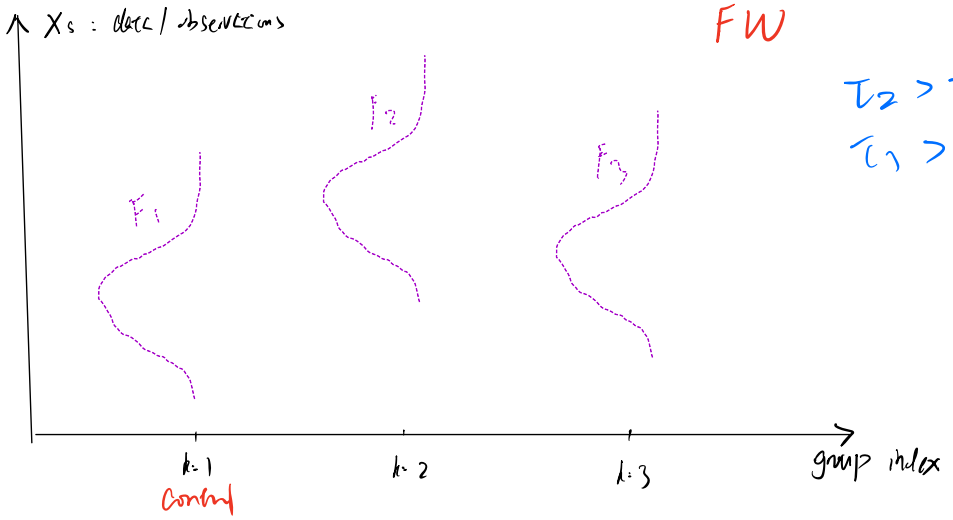


Kruskal-Wallis

$\tau_2 \neq \tau_1$

$\tau_3 \neq \tau_1$

$\boxed{\tau_2 \neq \tau_3}$

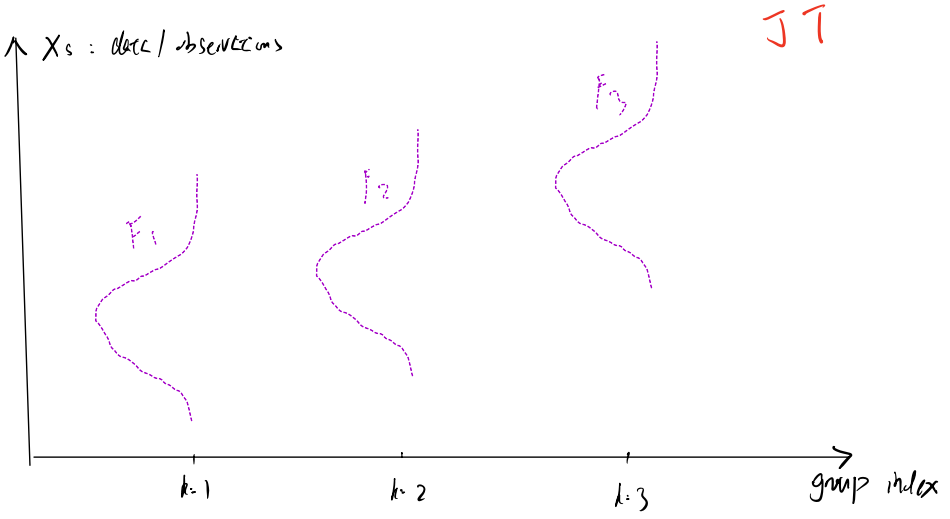


Control

FW

$\tau_2 > \tau_1$

$\tau_3 > \tau_1$



Control

JT

Example:Motivational Effect of Knowledge of Performance

For Hundal's (1969) study to assess the motivational effects of knowledge of performance, the no information category clearly serves as a control population, and it is very natural to ask if additional performance information of either type (rough or accurate) leads to improved performance as measured by an increase in the number of pieces processed.

Control (no information)	Group B (rough information)	Group C (accurate information)
39.5	37.5	48
35	40	40.5
38	47	45
42.5	44	43
44.5	41.5	46
41	42	50

$H_0: \tau_2 = \tau_1 , \tau_3 = \tau_1$

$H_1: \tau_2 \geq \tau_1 , \tau_3 \geq \tau_1$ at least one of $\geq$ is ">"

```
> sum(rank(c(39.5,35,38,42.5,44.5,41,37.5,40,47,44,41.5,42,48,40.5,45,43,46,50)
[1] 133
```

$$FW = 133$$

$$\mathrm{FW}^* = \frac{\mathrm{FW} - \frac{n(N+1)}{2}}{\sqrt{mn(N+1)/12}} = \frac{133 - 114}{\{112.12\}^{1/2}} = 1.79437$$

```
> pnorm(1.79437,lower.tail = F)
[1] 0.03637707
```

Thus, we have sufficient evidence from the Fligner-Wolfe treatments-versus-control test that additional performance knowledge (either rough or accurate) leads to an increase in the number of pieces produced.

check R functions for Wilcoxon rank sum test.