Topic 7

STAT 621 Lecture Notes
The Kruskal-Wallis Test

In this chapter we consider generalizations of the location problem. Here interest is in comparing the location parameters, specifically the medians, of three or more populations. Objectives are similar to those in the normal-theory analysis of variance (ANOVA). Specifically the hypotheses we will be interested in testing are

$H_0: \theta_1 = \theta_2 = \cdots = \theta_k$ $\qquad$ θ_i = median pop'n i

$H_1: \theta_i \neq \theta_j$ some $i \neq j$

The Kruskal-Wallis test is a distribution-free procedure for performing this hypothesis test. There are many variations on this hypothesis test, as well as other types of procedures that make up a distribution-free approach to ANOVA. We (probably) won't cover these, but I'll give them a mention.

- Test ordered alternatives: $H_1: \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$ when strict inequality some $i \neq j$
- Compare treatment to control
- Multiple comparisons: m tests, P(at least 1 type I error) $\gg \alpha$ if m large
- Contrasts (linear combinations of θ_i)

Data: k treatments or distinct groups with n_j observations from the jth group, for a total of $N = \sum_{j=1}^{k} n_j$ observations, $X_{1j}, X_{2j}, \ldots, X_{n_j j}$ for $j = 1, \ldots, k$. In tabular form,

Treatment

1	2	$\cdots$	k
X_{11}	X_{12}		X_{1k}
X_{21}	X_{22}		X_{2k}
X_{31}	$\vdots$		$\vdots$
$\vdots$			
$X_{n_1 1}$	$X_{n_2 2}$		$X_{n_k k}$

Assumptions:

A1. Everything independent: $X_{ij} \perp X_{lm}$ $i \neq l, j \neq m$

A2. $X_{1j}, X_{2j} \ldots X_{n_j j}$ iid $\sim F_j$ continuous cdf, $j = 1 \ldots k$

A3. $F_j(t) = F_j(t - \tau_j)$ when F continuous cdf, median θ

$\Rightarrow \tau_j$ = shift in distribution (ie median) pop'n j
= treatment effect

Correspondence to the normal theory ANOVA model:

Anova: $Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$

μ = overall mean

τ_i = treatment effect pop'n i

$\varepsilon_{ij} \sim N(0, \sigma^2)$

$\Rightarrow Y_{ij} \sim N(\mu + \tau_j, \sigma^2)$ = shifted Normals

Here: could write

$Y_{ij} = \theta + \tau_i + \varepsilon_{ij}$

θ = overall median

$\Rightarrow Y_{ij}$ shifted F

just not assume normal

Null Hypothesis: For the testing procedures we will cover in this chapter, the null hypothesis will be

$$H_0 : \tau_1 = \ldots = \tau_k$$

What does this imply about the distributions of the k groups?

$F_1 = F_2 = \cdots = F_k = F$

all same

Hypotheses: Here we test our null hypothesis against the alternative that at least two of the treatment effects differ. Specifically,

$$H_1: \tau_i \neq \tau_j \quad \text{some } i \neq j$$

Test Statistic: Order the combined N observations from smallest to largest. Let r_{ij} denote the rank of X_{ij} and compute

$$R_j = \sum_{i=1}^{n_j} r_{ij} \qquad \text{and} \qquad R_{.j} = \frac{R_j}{n_j}$$

for $j = 1, \ldots, k$. What do these quantities represent?

R_j = sum ranks group j

$R_{\cdot j}$ = mean rank group j

The Kruskal-Wallis test statistic is defined as

$$H = \underbrace{\frac{12}{N(N+1)} \sum_{j=1}^{k} n_j \left(R_{.j} - \frac{N+1}{2} \right)^2} = \left(\frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} \right) - 3(N+1).$$

Let's examine this a bit. What would we expect under the null? The alternative?

Note: overall ave rank: $\frac{1}{N} \sum_{j=1}^{k} \sum_{i=1}^{n_j} r_{ij} = \frac{1}{N} \frac{N(N+1)}{2} = \frac{(N+1)}{2}$

and $R_{\cdot j}$ = within group ave.

$\Rightarrow \left(R_{\cdot j} - \frac{N+1}{2}\right)^2$ = sq. deviation btw. group mean & overall mean

so $\sum n_j \left(R_{\cdot j} - \frac{N+1}{2}\right)^2$ ~ similar to SSTreatment in Anova (but on ranks)

$\Rightarrow \frac{12}{N(N+1)} \sum n_j \left(R_{\cdot j} - \frac{N+1}{2}\right)^2$ ~ MS Treatment

$\underline{H \text{ large}}$: supports H_A

$\underline{H \text{ small}}$: consistent with H_0

Rejection Rule: How can we determine whether or not to reject the null here?

In general: Reject H_0 if H big (how big?)

-Need sampling Dist!

Critical values may be found with the function `cKW` in the `NSM3` package. Alternatively p-values are given with the `pKW` function.

Notes:

(1) Large Sample Approximation: For large sample sizes (in every group or treatment), the sampling distribution of H is approximately Chisquared$(k-1)$. The large-sample test rejects for unusually large H compared to this distribution. Reject for

$$H \geq \chi^2_{k-1,\alpha}$$

where $\chi^2_{k-1,\alpha}$ is the upper-tailed critical value for a Chisquared$(k-1)$ distribution.

The R command `qchisq(a, df, lower.tail=F)` will return the value in the χ^2_{df} distribution that has area `a` in its upper-tail (critical value). The command `pchisq(H, df, lower.tail=F)` will return the area to the right of the value `H` for a χ^2_{df} distribution (p-value).

(2) Ties: In the case of ties, use the modified test statistic

g = # tied values

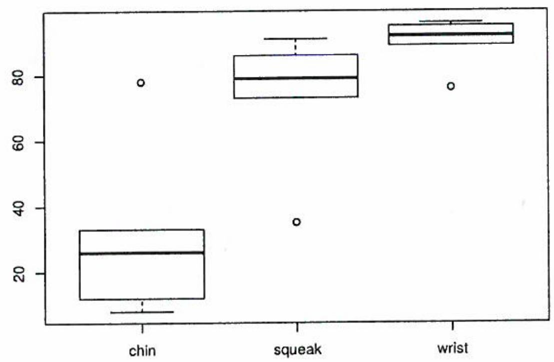
t_j = # groups in jth value

$$H' = \frac{H}{1 - \{\sum_{j=1}^{g}(t_j^3 - t_j)/(N^3 - N)\}}.$$

Use the same rejection rule for the case of no ties. Note that the Kruskal-Wallis test based on H' will have an *approximate* level α.

Example: (From *Nonparametric Statistical Inference*, Dickerson et al.) Hundreds of patents have been registered at the U.S. Patents office that claim to prevent snoring. A study was performed to compare the effectiveness of three of these patented devices: the clothes squeaker (sewn into the back of night clothes), the wrist tie (binds wrists to the side of the bed) and the chin strap (keeps the mouth shut). Fifteen men who snore were randomly divided into three groups, each assigned a device. Sleep was monitored for one night by a machine that measures snoring on a 100 point scale. Is there evidence that the devices differ in their (median) effectiveness? Do the Kruskal-Wallis test at $\alpha = .05$.

First a plot



$\frac{N+1}{2} = \frac{16}{2} = 8$

Compute the H statistic

```
squeak
   [1] 73   79   86   91   35
```

6 9 10 12 5

$R_1 = 6+9+10+12+5 = \underline{42}$ $\qquad R_{\cdot 1} = \frac{42}{5} = \underline{8.4}$

```
wrist
   [1] 96   92   89   95   76
```

15 13 11 14 7

$R_2 = 15+13+11+14+7 = \underline{60}$ $\qquad R_{\cdot 2} = \frac{60}{5} = \underline{12}$

```
chin
   [1] 12   26   33    8   78
```

2 3 4 1 8

$R_3 = 2+3+4+1+8 = \underline{18}$ $\qquad R_{\cdot 3} = \frac{18}{5} = \underline{3.6}$

$$H = \frac{12}{15(16)}\left[5(\overset{8.4}{\cancel{12}} - 8)^2 + 5(12-8)^2 + 5(3.6-8)^2\right] = \boxed{8.88}$$

```
# let's look up a (approximate) critical value
library(NSM3)
cKW(.05, c(5,5,5))
```

α n1 n2 n3

```
   Monte Carlo Approximation (with 10000 Iterations) used:
      Group sizes: 5 5 5
      For the given alpha=0.05, the upper cutoff value is Kruskal-Wallis H=5.78,
      with true alpha level=0.049
```

```
# or large-sample critical value
qchisq(.049, 2, lower.tail=F)
   [1] 6.03187
```

Reject H_0: at least one device has different median

You can also use one of two R functions to report the test statistic and p-value. These are the previously mentioned pKW as well as the function kruskal.test in base R. They both operate on an argument that provides the data for each group as the components of a list object. Although other syntax options are available.

```
pKW(x=list(squeak, wrist, chin))

   Group sizes: 5 5 5
   Kruskal-Wallis H Statistic: 8.88
   Monte Carlo (Using 10000 Iterations) upper-tail probability: 0.0029
```

```
kruskal.test(list(squeak,wrist,chin))

   Kruskal-Wallis rank sum test
   data:  list(squeak, wrist, chin)
   Kruskal-Wallis chi-squared = 8.88, df = 2, p-value = 0.0118
```

χ^2 approx

Null Distribution of H: Let's sketch out an approach to finding the sampling distribution of H under the null hypothesis. At least to see what's involved.

First recall we have k groups with sample sizes $n_1, \ldots, n_k$, for a total sample size of N. In general, how many different ranking arrangements are there? If the null hypothesis is true, what is the probability of any one of these arrangements?

N integers: pick

n_1 for group 1

n_2 for group 2

⋮

n_k for group k

ways: multinomial $= \frac{N!}{n_1! n_2! \cdots n_k!}$

Under H_0: each way equally likely:

Prob $= 1 \Big/ \frac{N!}{n_1! \cdots n_k!}$

Consider a simple example. Let k=3, $n_1 = 2$, $n_2 = 3$ and $n_3 = 2$. Suppose there are no ties. How many ways are there to rank the 2 observations in group1?

$N=7$ choose 2 ranks for group 1: $\binom{7}{2} = 21$

Now given the ranks for group 1, how many ways are there to rank the 3 observations in group 2?

5 left: $\binom{5}{3} = 10$

Then given the ranks for groups 1 and 2, how many ways are there to rank the remaining 2 observations?

2 left: $\binom{2}{2} = 1$

So how many total ranking arrangements are there? Does this agree with the general formula above?

$(21)(10)(1) = 210$ $\qquad \frac{7!}{2!\,3!\,2!} = 210$ ✓

Here are a few examples of various rankings and the resulting H value. Enumerating all of the 210 arrangements and finding their probabilities under the null hypothesis yields the sampling distribution of H. Just for this one example.

Group 1	Group 2	Group 3	h	$P(H = h)$
1, 2	3, 4, 5	6, 7	5.3571	1/210
1, 3	2, 4, 5	6, 7	4.4643	1/210
1, 4	2, 3, 5	6, 7	3.9286	1/210
.				
.				
.				
etc.				

Thank goodness for software!