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

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.

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,

Assumptions:

A1.

A2.

A3.

Correspondence to the normal theory ANOVA model:

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?

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

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?

The Kruskal-Wallis test statistic is defined as

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

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

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

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

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

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
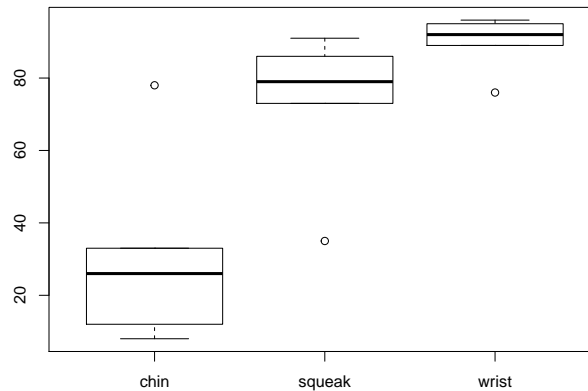
Compute the H statistic

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

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

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

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

   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
```

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
```

<u>Null Distribution of H:</u> 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?

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?

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

6

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

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

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!