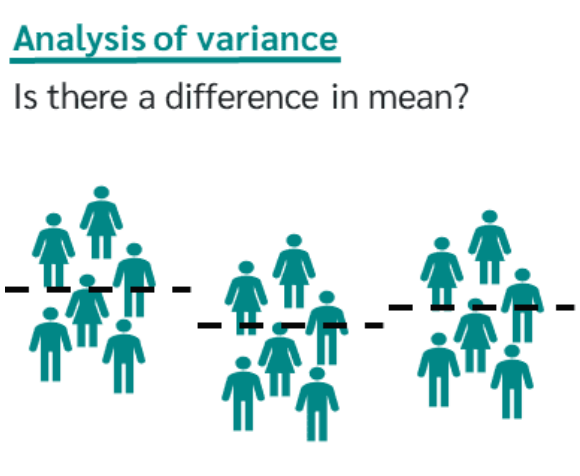
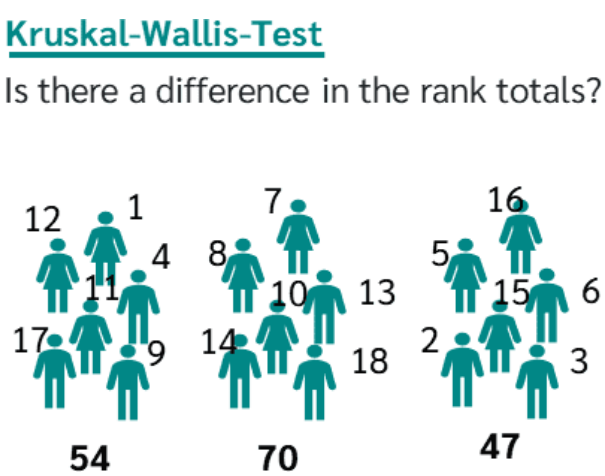
Comparing Centers of Several Independent Groups

EN5423 | Spring 2024

w10_several_independent_01.pdf
(Week 10)

Contents





Intro

• **Ex 1)** Concentrations of volatile organic compounds are measured in shallow ground waters across a ***multi-county area***. The wells sampled can be classified as being contained in one of ***seven land-use types***: undeveloped, agricultural, wetlands, low-density residential, high-density residential, commercial, and industrial/transportation. Do the concentrations of volatiles differ between these types of surface land-use, and if so, how?

• **Ex 2)** Alkalinity, pH, iron concentrations, and biological diversity are measured at low flow for small streams draining areas mined for coal. Each stream drains either **unmined land**, **land strip-mined and then abandoned**, or **land strip-mined and then reclaimed**. The streams also drain one of two rock units, a **sandstone** or a **limestone formation**. Do drainages from mined and unmined lands differ in quality? What effect has reclamation had? Are there differences in chemical or biological quality owing to rock type separate and distinct from the effects owing to mining impacts?

• **Ex 3) Three methods** for **field sampling** and **extraction of an organic chemical** are to be compared **at numerous wells**. Are there differences among concentrations produced by the three processes? These must be discerned above the well-to-well differences in concentration that contribute considerable noise to the data.

• The methods of this chapter, comparing centers of ***several independent groups***, can be used to answer questions such as those above. These methods are extensions of the ones introduced in week09 and week09-01; in this chapter ***more than two groups of data will be compared***.

• The **parametric technique** in this situation is analysis of variance (ANOVA). More robust nonparametric and permutation techniques are also presented for the frequent situations where data ***do not meet the assumptions of ANOVA***.

• First consider the effect of only ***one grouping variable***, also called a ***factor***. A factor is a categorical variable suspected of influencing the measured data, analogous to an explanatory variable in regression.

Example Scenario: The Effect of Soil Type on Plant Growth
Objective: Determine if different types of soil affect the growth rate of a specific plant species. Factor (Grouping Variable): **Type of Soil** Levels of the Factor: 1) Sandy Soil, 2) Clay Soil, 3) Loamy Soil

• The factor is made up of more than one level and each level is defined by a group of observations. Levels may be an ordered low-medium-high change in intensity or unordered categories such as different locations or times that represent a change in underlying influences.

• The factor consists of a set of *k* groups, with each data point belonging in one of the *k* groups.

• For example, the data could be ***calcium concentrations*** from wells in one of *k* aquifers, and the ***objective is to determine whether the calcium concentrations differ among the aquifers***.

• The various aquifers are the groups or levels. Within each group (aquifer) there are n_j observations (the sample size of each of the *j* groups is not necessarily the same). Observation y_{ij} is the i^{th} of n_j observations in group *j*, so that *i* =1, 2, ..., n_j for the j^{th} of *k* groups *j* =1, 2, ..., *k*. The total number of observations *N* is thus

$$N = \sum_{j=1}^{k} n_j \tag{1}$$

which simplifies to $N = k \cdot n$ when the sample size $n_j = n$ for all k groups (equal sample sizes per group, ***also called a balanced design***).

• When data within each of the groups ***are normally distributed*** and ***possess identical variances***, ***classical ANOVA can be used***.

• Analysis of variance, ANOVA, is a parametric test, determining whether all group means are equal. **ANOVA is analogous to a *t*-test between three or more groups of data and is restricted by the same assumptions as the *t*-test**.

• When data in each group ***do not have identical variance***, an adjustment similar to the one for the *t*-test will improve on classical ANOVA (Welch, 1951).

• When data in each group ***do not follow a normal distribution*** a ***permutation test*** can check differences between group means.

• When the objective is to determine ***whether some groups have higher or lower values than others*** and is ***not focused on the mean as a parameter***, nonparametric tests such as the Kruskal-Wallis (KW) and Brunner-Dette-Munk (BDM) tests will have more power than parametric ANOVA methods (table 1). When the null hypothesis is rejected, these tests ***do not state*** which group or groups differ from the others!

• We therefore discuss multiple comparison tests—tests ***for determining which groups differ from others***. ***These methods are then expanded to evaluating the effects of two factors simultaneously*** (table 2).

• These factorial methods determine ***whether neither, or one or both of two factors significantly affect the values of observed data***. Although higher numbers of factors can be evaluated, the design of those studies and the tests that follow are beyond the scope of this book.

• We finish the chapter by discussing repeated measures designs, the extension of the matched-pairs tests of week09-02 and week10-01 to situations where three or more related observations are taken on each subject or block (table 3).

Table 1. Hypothesis tests with ***one factor*** and their characteristics.
[ANOVA, analysis of variance; BDM, Brunner-Dette-Munk test; MCT, multiple comparison test. HA is the alternative hypothesis, the signal to be found if it is present]

	Objective of test (H_A)				
	Data from at least one group is frequently higher than the other groups		Mean of at least one group is higher than the mean of the other groups		
Test	Kruskal-Wallis test	BDM test	ANOVA	Welch's adjusted ANOVA	Permutation test on group means
Class of test	Nonparametric	Nonparametric	Parametric	Parametric	Permutation
Distributional assumption for group data	None	None	Normal distribution; equal variances	Normal distribution	Exchangeable
Multiple comparison test	Pairwise rank-sum tests or Dunn's test	Pairwise rank-sum tests or Dunn's test	Tukey's MCT	Tukey's MCT	Tukey's MCT

Table 2. Hypothesis tests with ***two factors*** and their characteristics.
[ANOVA, analysis of variance; BDM, Brunner-Dette-Munk test; MCT, multiple comparison test. HA is the alternative hypothesis, the signal to be found if it is present]

	Objective of test (H_A)		
	Data from at least one group is frequently higher than the other groups	Mean of at least one group is higher than the mean of the other groups	
Test	BDM two-factor test	Two-factor ANOVA	Two-factor permutation test
Class of test	Nonparametric	Parametric	Permutation
Distributional assumption for group data	None	Normal distribution; equal variances	Exchangeable
Multiple comparison test	Pairwise rank-sum tests	Two-factor Tukey's MCT	Two-factor Tukey's MCT

Table 3. Hypothesis tests for ***repeated measures*** and their characteristics.
[ANOVA, analysis of variance; BDM, Brunner-Dette-Munk test; MCT, multiple comparison test. HA is the alternative hypothesis, the signal to be found if it is present]

	Objective of test (H_A)		
	Data from at least one group is frequently higher than the other groups		Mean of at least one group is higher than the mean of the other groups
Test	Friedman test	Aligned-rank test	ANOVA without replication
Class of test	Nonparametric	Nonparametric	Parametric
Distributional assumption for group data	None	Symmetry	Normal distribution; equal variances
Multiple comparison test	Paired Friedman comparison tests	Tukey's MCT on aligned ranks	Pairwise paired *t*-tests

Note:

1. **ANOVA (Analysis of Variance)**:
   - **Purpose**: ANOVA is used to determine if there are any statistically significant differences between the means of three or more independent (unrelated) groups.
   - **Outcome**: It tests the null hypothesis that all group means are equal. If the ANOVA is significant, it tells us that at least one group mean is different from the others, but it does not specify which groups differ.
2. **Adjustments for Variance**:
   - If the groups do not have identical variances, an adjustment like Welch's ANOVA is used, which does not assume equal variances across groups.
3. **Non-Normal Distribution**:
   - When data do not follow a normal distribution, nonparametric tests like permutation tests are recommended to check differences between group means.
4. **Nonparametric Tests (KW and BDM)**:
   - **Purpose**: These are used when the data do not meet the assumptions necessary for ANOVA (like normality and homogeneity of variances) or when the focus is not just on the mean but on differences in distribution patterns across groups.
   - **Outcome**: Kruskal-Wallis (KW) and Brunner-Dette-Munk (BDM) tests can indicate whether differences exist among the groups, but similar to ANOVA, they do not specify which specific groups are different from each other.
5. **Identification of Specific Group Differences**:
   - To identify which specific groups differ from each other following a significant ANOVA, KW, or BDM test, multiple comparison tests (like Tukey's HSD for ANOVA) are necessary. These tests compare each pair of groups to pinpoint where the differences lie.
6. **Factorial Methods**:
   - When evaluating the effects of two factors simultaneously, factorial ANOVA is used to see not only the individual effects of each factor but also whether there is an interaction effect between the factors.
7. **Extension to Repeated Measures (paired or related data)**:
   - For designs where the same subjects are measured under different conditions (repeated measures), the methodology extends to include tests that can handle the correlated nature of the data (e.g., repeated measures ANOVA).

In summary, ANOVA, KW, and BDM tests indicate whether there are differences among groups but require follow-up multiple comparison tests to determine which specific groups differ. These insights align well with the descriptions provided in your text.

Dealing with experiments involving more than three factors:

1. Factorial ANOVA (N-way ANOVA)

- **Description**: This extends the basic concept of ANOVA to more than two factors. N-way ANOVA can handle interactions among multiple factors simultaneously, allowing you to determine not only the main effects of each factor but also how these factors interact with each other.
- **Use**: Ideal when all factors are categorical, and you are interested in the impact on a continuous outcome variable.

2. Mixed-Effects Models

- **Description**: These models are useful when you have both fixed effects (factors of primary interest whose levels are fixed and generalizable) and random effects (factors that introduce variability into the data, but are not the primary interest, such as random batches or subjects).
- **Use**: Particularly valuable in complex designs where some factors might be nested or hierarchical, or when data are collected at different levels (such as repeated measures over time).

3. Multivariate Analysis of Variance (MANOVA)

- **Description**: MANOVA extends ANOVA to multiple dependent variables. It is used when you want to analyze the effect of one or more independent factors on multiple dependent variables simultaneously.
- **Use**: Useful when the dependent variables are expected to be correlated or when the impact of independent variables is to be assessed across several outcome measures.

4. Generalized Linear Models (GLMs) and Generalized Linear Mixed Models (GLMMs)

- **Description**: GLMs extend linear models to outcomes that are not normally distributed. GLMMs further extend GLMs by including both fixed and random effects.
- **Use**: Appropriate for handling various types of data distributions (e.g., binary, count data) and accounting for complex error structures and hierarchical data.

5. Multilevel Models (Hierarchical Models)

- **Description**: These are a form of regression models designed to handle data that are grouped or clustered.
- **Use**: Effective for data with multiple levels of analysis (such as students within schools, patients within hospitals), allowing for modeling of dependencies within data clusters.

6. Structural Equation Modeling (SEM)

- **Description**: SEM is a comprehensive statistical approach that includes multiple regression equations simultaneously and is used to assess relationships among observed and latent variables.

- **Use**: Suitable for complex models with latent variables, multiple pathways, and feedback loops, often used in social sciences and psychology.

7. Design of Experiments (DoE) with Response Surface Methodology (RSM)

- **Description**: DoE involves systematic methods for planning experiments so that the data obtained can be analyzed to yield valid and objective conclusions. RSM uses quantitative data from appropriate experiments to solve multivariate equations simultaneously.
- **Use**: Ideal for optimizing processes or products involving several variables and where responses are influenced by multiple variables simultaneously.

1 The Kruskal-Wallis Test (One Factor)

1.1 Null and Alternate Hypotheses for the Kruskal-Wallis Test

The Kruskal-Wallis (KW) test objectives are stated by the null and alternate hypotheses:

H_0: All groups of data have identical distributions.
H_A: At least one group differs in its distribution.

For the tests of this chapter, the alternate hypothesis H_A is always two-sided; no prior direction of difference is hypothesized, but only whether group differences exist or not.

1.2 Assumptions of the Kruskal-Wallis Test

• For the general objective of determining whether all groups are similar in value, or alternatively that one or more groups more ***frequently have higher or lower values than the other groups***, ***no assumptions are required about the shape of the distributions***.

• They may be normal, lognormal, or anything else. If the alternate hypothesis is true, they may have different distributional shapes.

• This difference ***is not attributed solely to a difference in median***, though that is one possibility.

• The test can determine differences where, for example, one group is a control group with only background concentrations, whereas the others combine background concentrations with higher concentrations owing to contamination.

• For example, 35 percent of the data in one group may have concentrations indicative of contamination and yet group medians remain similar. The KW test can see this type of change in the upper 35 percent as dissimilar to the control group.

• The test is sometimes stated with a more specific objective—as a test for difference in medians. This objective requires that all other characteristics of the data distributions, such as spread or skewness, be identical—though not necessarily on the original scale. This parallels

the rank-sum test. As a specific test for difference in medians, the Kruskal-Wallis null and alternate hypotheses are

H_0: The medians of the groups are identical.
H_A: At least one group median differs from the others.

• As with the rank-sum test, the KW test statistic and *p*-value computed for data that are transformed using any monotonic transformation give identical test statistics and *p*-values to those using data on the original scale.

• Thus, there is little incentive to search for transformations (to normality or otherwise) as the test is applicable in many situations.

1.3 Computation of the Exact Kruskal-Wallis Test

• The exact method determines a *p*-value by computing all possible test statistics when the observed data are rearranged, calculating the probability of obtaining the original test statistic or those more extreme.

• It is needed only for quite small sample sizes—three groups with $n_j \leq 5$, or with four or more groups of size $n_j \leq 4$ (Lehmann, 1975). ***Otherwise, the large sample approximations are very close to their exact values***. To compute the test, all *N* observations from all groups are jointly ranked from 1 to *N*, smallest to largest. These ranks R_{ij} are used to compute the average rank $\bar{R}_j$ for each of the *j* groups, where n_j is the number of observations in the j^{th} group:

$$\bar{R}_j = \frac{\sum_{i=1}^{n_j} R_{ij}}{n_j} \tag{2}$$

Compare $\bar{R}_j$ to the overall average rank $R_{ij} = (N+1)/2$, squaring and weighting by sample size, to form the test statistic *K*:

$$K = \frac{12}{N(N+1)} \sum_{j=1}^{k} n_j \left[\bar{R}_j - \frac{N+1}{2}\right]^2 \tag{3}$$

• When the null hypothesis is true, the average rank for each group should be similar to one another and to the overall average rank of (N + 1) / 2.

• When the alternative hypothesis is true, the average rank for some of the groups will differ from others, some higher than (N + 1) / 2 and some lower.

• The test statistic *K* will equal 0 if all groups have identical average ranks and will be positive if average group ranks differ.

• The null hypothesis is rejected when *K* is sufficiently large. Conover (1999) provided tables of exact *p*-values for *K* for small sample sizes.

• An example computation of K is shown in table 4. In past years, K would have been compared to the 0.95 quantile of the chi-squared distribution with $k-1$ degrees of freedom, which for these data would be 7.815 (3 degrees of freedom).

• Because K in table 4 does not exceed the 7.815, the null hypothesis is not rejected at an α of 0.05.

• The provided Python script will compute the proportion of the chi-square distribution that equals or exceeds the test statistic value of 2.66. That proportion is the p-value, here 0.44. Because 0.44 is higher than α, the null hypothesis is not rejected and the values in each group are not considered to be different.

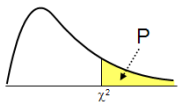
Table 4. Kruskal-Wallis test statistic computation for fecal coliform counts (Lin and Evans, 1980).

	Ranks R_{ij}						$\bar{R}_j$
Summer	6	12	15	18	21	24	16
Fall	5	8.5	11	14	19.5	22	13.3
Winter	2	4	8.5	13	16	19.5	10.5
Spring	1	3	7	10	17	23	10.2

$$\bar{R}_{ij} = \frac{16 + 13.3 + 10.5 + 10.2}{4} = \frac{24+1}{2} = 12.5$$

$$K = \frac{12}{24(25)}(6(16-12.5)^2 + 6(13.3-12.5)^2 + 6(10.5-12.5)^2 + 6(10.2-12.5)^2) = 2.66$$

$$x^2_{0.95,(3)} = 7.815$$



	P										
DF	0.995	0.975	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.790
18	6.265	8.231	22.760	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.900	27.204	30.144	32.852	33.687	36.191	38.582	41.610	43.820
20	7.434	9.591	25.038	28.412	31.410	34.170	35.020	37.566	39.997	43.072	45.315
21	8.034	10.283	26.171	29.615	32.671	35.479	36.343	38.932	41.401	44.522	46.797
22	8.643	10.982	27.301	30.813	33.924	36.781	37.659	40.289	42.796	45.962	48.268
23	9.260	11.689	28.429	32.007	35.172	38.076	38.968	41.638	44.181	47.391	49.728
24	9.886	12.401	29.553	33.196	36.415	39.364	40.270	42.980	45.559	48.812	51.179
25	10.520	13.120	30.675	34.382	37.652	40.646	41.566	44.314	46.928	50.223	52.620
26	11.160	13.844	31.795	35.563	38.885	41.923	42.856	45.642	48.290	51.627	54.052
27	11.808	14.573	32.912	36.741	40.113	43.195	44.140	46.963	49.645	53.023	55.476
28	12.461	15.308	34.027	37.916	41.337	44.461	45.419	48.278	50.993	54.411	56.892
29	13.121	16.047	35.139	39.087	42.557	45.722	46.693	49.588	52.336	55.792	58.301
30	13.787	16.791	36.250	40.256	43.773	46.979	47.962	50.892	53.672	57.167	59.703
31	14.458	17.539	37.359	41.422	44.985	48.232	49.226	52.191	55.003	58.536	61.098

$$x^2_{0.44,(3)} = 2.66\ (p = 0.44)$$

Characteristics of the Chi-Squared Distribution:

The chi-squared distribution is a widely used theoretical probability distribution in statistics, particularly useful for hypothesis testing and constructing confidence intervals. It is especially important in scenarios involving categorical data analyzed using tests such as the chi-squared test for independence in contingency tables and the chi-squared test for goodness of fit.

- **Skewness**: The chi-squared distribution is skewed right, meaning it has a long right tail. The distribution becomes more symmetric as the degrees of freedom increase.

- **Degrees of Freedom**: This is a critical parameter that shapes the distribution. The degrees of freedom (df) are typically associated with the number of independent variables contributing to a sum of squares calculation. In the context of chi-squared tests, the degrees of freedom often relate to the number of categories or groups minus any constraints required by the data or parameters estimated.

- **Origin**: The distribution is strictly positive, with values ranging from zero to infinity. It only takes on non-negative values because it is defined as the sum of the squares of a number of standard normal distributions.

- **Applications**:

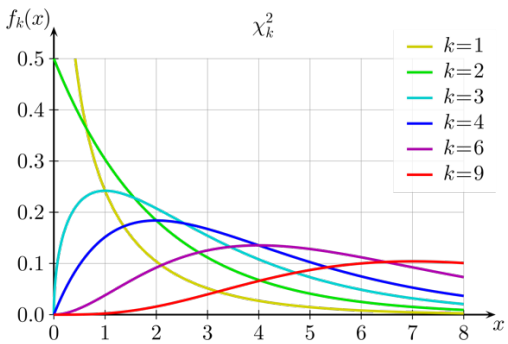
  - **Goodness of Fit**: Used to determine how well a theoretical distribution fits an observed distribution. This is often used in the context of fitting observed data to theoretical models.
  - **Test of Independence**: In contingency tables, where you want to assess whether two categorical variables are independent of each other.
  - **Variance Estimates**: It helps in estimating the variance of normally distributed samples when the underlying population variance is unknown.

Calculation: Mathematically, if $Z_1, Z_2, \ldots, Z_k$ are k independent standard normal random variables, then the sum of their squares,

$$Q = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

is distributed as a chi-squared distribution with k degrees of freedom, denoted as

$$Q \sim \chi^2(k)$$



Squared Standard Normal Variables: The chi-squared distribution is the distribution of a sum of the squares of a number of independent standard normal random variables. In the context of these tests, when adjustments are made for the number of observations and groups, the squared terms in the calculation of the test statistic *K* follow this logic.

- **Degrees of Freedom**: The degrees of freedom for the chi-squared test statistic in this context are $k-1$ (where k is the number of groups), reflecting the number of independent comparisons that can be made between the groups.

- **Large Sample Approximation**: The chi-squared approximation is used because it provides a reliable measure of significance for large samples. For smaller samples, exact calculations might be needed, as noted by Conover.

1.4 The Large-sample Approximation for the Kruskal-Wallis Test

• The distribution of *K* when the null hypothesis is true can be approximated quite well at small sample sizes by a chi-square distribution with $k-1$ degrees of freedom.

• The degrees of freedom is a measure of the number of independent pieces of information used to construct the test statistic. If all data are divided by their overall mean to standardize the dataset, then when any $k-1$ average group ranks are known, the final (k^{th}) average rank can be computed from the others.

$$\bar{R}_k = \frac{N}{n_k} \cdot \left(1 - \sum_{j=1}^{k-1} \frac{n_j}{N} \bar{R}_j \right) \qquad \text{Eq (2)}$$

• Therefore, there are actually only $k-1$ independent pieces of information as represented by $k-1$ average group ranks.

• From these and the overall average rank, the k^{th} average rank is fixed. This is the constraint represented by the degrees of freedom. The null hypothesis is rejected when the approximate *p*-value is less than α. The provided Python script computes the large-sample approximation results.

Example 1:

Knopman (1990) reported the specific capacity (discharge per unit time per unit drawdown) of wells within the Piedmont and Valley and Ridge Provinces of Pennsylvania.

Two hundred measurements from four rock types were selected from the report—see the provided dataset.

Boxplots for the four rock types are shown in figure 1. The fact that the boxes are flattened as a result of high outliers is important—the outliers may strongly affect the means and parametric tests, just as they affect your ability to see differences on the figure. Based on the outliers alone it appears that the variance differs among the groups, and all but the siliciclastic group are clearly non-normal.

The null hypothesis H_0 for the KW test on these data is that each of the four rock types has the same distribution (set of percentiles) for specific capacity. The alternate hypothesis H_A is that the distributions are not all the same, with at least one shifted higher than another (a two-sided test).

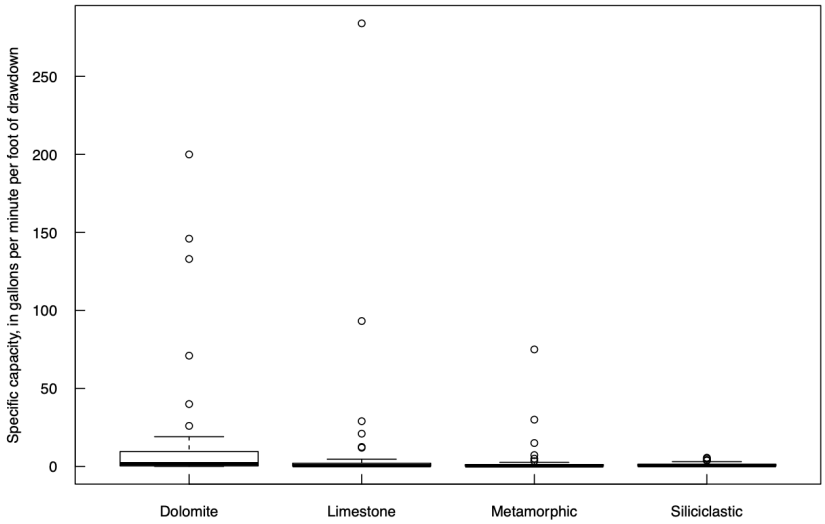


Figure 1. Boxplots of specific capacity of wells in four rock types (Knopman, 1990).

```
import pyreadr
import pandas as pd
import scipy.stats as stats

# Path to the RDA file
file_path = './specapic.rda'

# Reading the RDA file
result = pyreadr.read_r(file_path)

data = result['specapic']

groups = data.groupby('rock')['spcap'].apply(list)
statistic, p_value = stats.kruskal(*groups)

# Print the results
print(f"Kruskal-Wallis chi-squared = {statistic:.3f}, p-value =
{p_value:.5f}")

# To display degrees of freedom, it's the number of groups minus one
df = len(groups) - 1
print(f"Degrees of freedom = {df}")
```

```
Kruskal-Wallis chi-squared = 11.544, p-value = 0.00912
Degrees of freedom = 3
```

• The small *p*-value leads us to reject the null hypothesis of similarity of group percentiles. At least one group appears to differ in its frequency of high versus low values.

• A graph that visualizes what the Kruskal-Wallis test is testing for is the quantile plot. A quantile plot for natural logarithms of the four groups of specific capacity data is shown in figure 2.

• The dolomite group stands apart and to the right of the other three groups throughout most of its distribution, illustrating the Kruskal-Wallis conclusion of difference. Moving to the right at $y = 0.5$, three groups have similar medians but ***the dolomite group median is higher***.

• An experienced analyst can look for differences in variability and skewness by looking at the slope and shapes of each group's line.

• Boxplots are more accessible to nontechnical audiences, but quantile plots provide a great deal of detail while still illustrating the main points, especially for technical audiences.

• Alternative nonparametric tests, none of which have significant advantages over Kruskal-Wallis, include computing Welch's ANOVA on the ranks of data (Cribbie and others, 2007); the normal-scores test, of which there are two varieties (Conover, 1999); and a one-factor version of the BDM test (Brunner and others, 1997).

• The Cribbie and others (2007) procedure is similar to a *t*-test on ranks and is only an approximate nonparametric test. The normal-scores tests perform similarly to Kruskal-Wallis with a slight advantage over KW when data are normally distributed—water resources data rarely are. We discuss the two-factor version of the BDM test in section 6 as a nonparametric alternative to ANOVA.

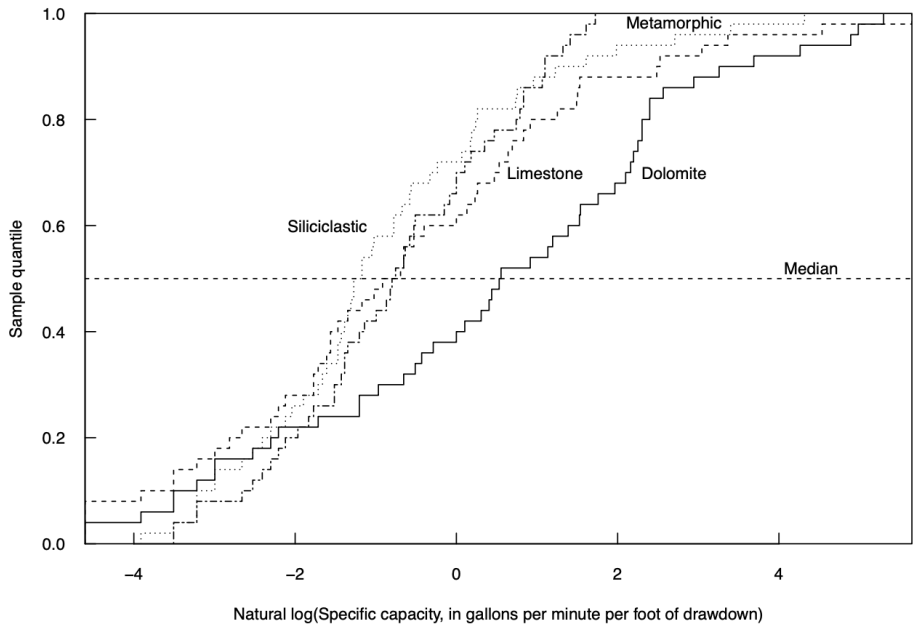


Figure 2. Quantile plots of the natural log of specific capacity for the four rock types from Knopman (1990). Three rock types have similar medians, but the dolomite group median is higher.