

## Chapter 556

# Kruskal-Wallis Tests (Simulation)

---

## Introduction

This procedure analyzes the power and significance level of the Kruskal-Wallis Test which is used to test statistical hypotheses in a one-way experimental design. For each scenario that is set up, two simulation studies are run. One simulation estimates the significance level and the other estimates the power.

---

## Technical Details

*Computer simulation* allows us to estimate the power and significance level that is actually achieved by a test procedure in situations that are not mathematically tractable. Computer simulation was once limited to mainframe computers. But, in recent years, as computer speeds have increased, simulation studies can be completed on desktop and laptop computers in a reasonable period of time.

The steps to a simulation study are

1. Specify how the test is carried out. This includes indicating how the test statistic is calculated and how the significance level is specified.
2. Generate random samples from the distributions specified by the alternative hypothesis. Calculate the test statistics from the simulated data and determine if the null hypothesis is accepted or rejected. Tabulate the number of rejections and use this to calculate the test's power.
3. Generate random samples from the distributions specified by the null hypothesis. Calculate each test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. Tabulate the number of rejections and use this to calculate the test's significance level.
4. Repeat steps 2 and 3 several thousand times, tabulating the number of times the simulated data leads to a rejection of the null hypothesis. The power is the proportion of simulated samples in step 2 that lead to rejection. The significance level is the proportion of simulated samples in step 3 that lead to rejection.

---

## Generating Random Distributions

Two methods are available in **PASS** to simulate random samples. The first method generates the random variates directly, one value at a time. The second method generates a large pool (over 10,000) of random values and then draws the random numbers from this pool. This second method can cut the running time of the simulation by 70%.

The second method begins by generating a large pool of random numbers from the specified distributions. Each of these pools is evaluated to determine if its mean is within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean. This population is then sampled with replacement using the uniform distribution. We have found that

## Kruskal-Wallis Tests (Simulation)

this method works well as long as the size of the pool is at least the maximum of twice the number of simulated samples desired and 10,000.

## Kruskal-Wallis One-Way ANOVA on Ranks

This section summarizes the Kruskal-Wallis rank-sum test. See Conover (1999) and Gibbons and Chakraborti (2011) for more details.

### Assumptions

1. A simple random sample is available from each group. This implies that the individual subjects with the sample are independent.
2. There is mutual independence between samples.
3. The measurement scale is at least ordinal (the data can be ranked).
4. The  $g$  distributions are identical, or some groups tend to produce larger values than others.

### Test Statistic

Suppose that  $X_{kj}$  represents the (at least ordinal) data value of the  $j^{\text{th}}$  subject in the  $k^{\text{th}}$  group. Let  $r_{kj}$  represent the rank of  $X_{kj}$  when the  $N$  values are ranked without regard to their group. Let  $n_1, n_2, \dots, n_g$  denote the number of subjects in each group and let  $N$  denote the total sample size of all groups.

The formula for the Kruskal-Wallis test is

$$T = \frac{\frac{12}{N^2 + N} \sum_{k=1}^g \frac{R_k^2}{n_k} - 3(N + 1)}{\left[ 1 - \frac{\sum_{i=1}^G (t_i^3 - t_i)}{N^3 - N} \right]}$$

where  $R_k$  is the sum of the ranks of the data  $k^{\text{th}}$  group ( $R_k = \sum_{j=1}^{n_k} r_{kj}$ ),  $G$  is the number sets of tied ranks, and  $t_i$  is the number of tied values within set  $i$ .

### Distribution of T

If these assumptions are met, the distribution of  $T$  follows the *Chi-square* distribution with  $g-1$  degrees of freedom. These results can be used to form an approximate statistical test of the null hypothesis that the distributions of all groups are equal versus the alternative hypothesis that at least one group tends to produce larger values than the others.

## Example 1 – Power at Various Sample Sizes

For this first example we repeat Example 1 of the regular One-Way ANOVA procedure. This will allow you to compare the values obtained by simulation with the actual values obtained from the theoretical results.

An experiment is being designed to compare the means of four groups using an  $F$  test with a significance level of 0.05. Previous studies have shown that the standard deviation within a group is 18. Treatment means of 40, 10, 10, and 10 represent clinically important treatment differences. To better understand the relationship between power and sample size, the researcher wants to compute the power for group sample sizes of 4, 8, and 12. The group sample sizes are equal.

### Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab	
Solve For .....	<b>Power</b>
Alpha.....	<b>0.05</b>
Simulations .....	<b>5000</b>
Random Seed.....	<b>5756037</b> (for Reproducibility)
Number of Groups .....	<b>4</b>
Group Allocation .....	<b>Equal (n = n1 = n2 = ...)</b>
n (Group Size) .....	<b>4 8 12</b>
Group 1 Power Distribution .....	<b>Normal(M1 SD)</b>
Group 2 Power Distribution .....	<b>Normal(M2 SD)</b>
Group 3 Power Distribution .....	<b>Normal(M2 SD)</b>
Group 4 Power Distribution .....	<b>Normal(M2 SD)</b>
Parameter 1 Name .....	<b>M1</b>
Parameter 1 Value(s).....	<b>40</b>
Parameter 2 Name .....	<b>M2</b>
Parameter 2 Value(s).....	<b>10</b>
Parameter 6 Name .....	<b>SD</b>
Parameter 6 Value(s).....	<b>18</b>
Specify Alpha Distributions .....	<b>All equal to the group 1 distribution of the power simulation</b>

## Output

Click the Calculate button to perform the calculations and generate the following output.

### Simulation Summary and Numeric Results

#### Simulation Summary

Group	Simulation Distribution	
	Power (H1)	Alpha (H0)
1	Normal(M1 SD)	Normal(M1 SD)
2	Normal(M2 SD)	Normal(M1 SD)
3	Normal(M2 SD)	Normal(M1 SD)
4	Normal(M2 SD)	Normal(M1 SD)

Number of Groups	4
Random Number Pool Size	10000
Number of Simulations	5000
Random Seed	5756037 (User-Entered)
Run Time	1.61 seconds

#### Numeric Results

Solve For: [Power](#)

Scenario	Power	Sample Size		Alpha		Standard Deviation of $\mu$ 's   H1	Average of $\sigma$ 's   H1	M1	M2	SD
		Average Group n	Total N	Target	Actual					
		1	0.394	4	16					
2	0.854	8	32	0.05	0.044	13	17.9	40	10	18
3	0.979	12	48	0.05	0.046	13	18.0	40	10	18

- Scenario Identifies this line of the report. It will be used cross-reference this line in other reports.
- $\mu$  The mean of a group.
- $\sigma$  The standard deviation within a group.
- H0 The null hypothesis that all means are equal.
- H1 The alternative hypothesis that at least one mean is different from the others.
- Power The probability of rejecting H0 when it is false. This is the actual value calculated by the power simulation.
- n The average group sample size. The average of the individual group sample sizes.
- N The total sample size found by summing all group sample sizes.
- Target Alpha The planned probability of rejecting a true H0. This is the planned Type-1 error rate.
- Actual Alpha The alpha achieved by the test as calculated by the alpha simulation. Note that the alpha simulation is separate from the power simulation.
- SD of  $\mu$ 's|H1 The standard deviation of the group  $\mu$ 's used in the power simulation. This measures the magnitude of the difference among the  $\mu$ 's.
- Average of  $\sigma$ 's|H1 The mean of the group  $\sigma$ 's used in the power simulation. This measures the magnitude of the  $\sigma$ 's.
- M1, M2, SD, etc. The value(s) of the user-specified distribution parameters.

#### Summary Statements

A one-way design with 4 groups will be used to test whether there is a difference among the medians (or distributions) of the 4 groups. The comparison will be made using a Kruskal-Wallis chi-square test, with a target Type I error rate ( $\alpha$ ) of 0.05. Based on 5000 simulations of the null distributions: Normal(M1 SD); Normal(M1 SD); Normal(M1 SD); and Normal(M1 SD), and of the alternative distributions: Normal(M1 SD); Normal(M2 SD); Normal(M2 SD); and Normal(M2 SD) (where M1 = 40, M2 = 10, and SD = 18), with group sample sizes of 4, 4, 4, and 4 (for a total of 16 subjects), the power is 0.394. (Additional details: The standard deviation of the group means under the null hypothesis is 0, and the standard deviation of the group means under the alternative hypothesis is 13. The average of the within-group standard deviations, assuming the alternative hypothesis distributions, is 18. The actual Type I error rate ( $\alpha$ ), based on the null hypothesis distribution simulations, is 0.034.)

## Kruskal-Wallis Tests (Simulation)

.  
.  
.

**References**

---

Conover, W.J. 1999. Practical Nonparametric Statistics. 3rd Edition. John Wiley & Sons. New York.  
Devroye, Luc. 1986. Non-Uniform Random Variate Generation. Springer-Verlag. New York.  
Gibbons, J.D. and Chakraborti, S. 2011. Nonparametric Statistical Inference, 5th Edition. CRC Press.  
Milliken, G.A and Johnson, D.E. 1984. Analysis of Messy Data: Volume 1. Van Nostrand Reinhold. New York.  
Winer, B.J., Brown, D.R, and Michels, K.M. 1991. Statistical Principles in Experimental Design. 3rd Edition. McGraw-Hill. New York.

---

These reports show the output for this run. We will annotate the Numeric Results report.

**Power**

This is the probability of rejecting a false null hypothesis. This value is estimated by the power simulation. The Power and Alpha Confidence Interval report displayed next will provide estimates of the precision of these power values.

**Average Group Size n**

This is the average of the group sample sizes.

**Total Sample Size N**

This is the total sample size of the study.

**Target Alpha**

The target value of alpha: the probability of rejecting a true null hypothesis. This is often called the significance level.

**Actual Alpha**

This is the value of alpha estimated by the alpha simulation. It should be compared with the Target Alpha. The Power and Alpha Confidence Interval report displayed next will provide estimates of the precision of these Actual Alpha values.

**SD of  $\mu$ 's | H1**

This is the standard deviation of the hypothesized means of the power (H1) simulation distributions,  $\sigma_m | H1$ . Under H0, this value is zero. So this value represents the magnitude of the difference among the means that is being tested.

Note that a popular definition of effect size is the ratio of  $\sigma_m | H1$  and  $\sigma | H1$ .

**Average of  $\sigma$ 's | H1**

This is the mean of the group standard deviations calculated from the power simulation distributions,  $\sigma | H1$ .

**M1**

These are the values entered for M1, the group 1 mean in the power simulation.

## Kruskal-Wallis Tests (Simulation)

**M2**

These are the values entered for M2, the groups 2, 3, and 4 means in the power simulation.

**SD**

These are the values entered for SD, the standard deviation.

**Power and Alpha Confidence Intervals from Simulations**

Power and Alpha Confidence Intervals from Simulations											
Scenario	Total Sample Size N	Power	95% Confidence Interval Limits for Power		Alpha		95% Confidence Interval Limits for Alpha		M1	M2	SD
			Lower	Upper	Target	Actual	Lower	Upper			
1	16	0.394	0.380	0.407	0.05	0.034	0.029	0.039	40	10	18
2	32	0.854	0.844	0.864	0.05	0.044	0.039	0.050	40	10	18
3	48	0.979	0.975	0.983	0.05	0.046	0.040	0.052	40	10	18
N			The total sample size found by summing all group sample sizes.								
Power			The probability of rejecting H0 when it is false. This is the actual value calculated by the power simulation.								
Lower and Upper Limits of a 95% C.I. for Power			The limits of an exact, 95% confidence interval for power based on the binomial distribution. They are calculated from the power simulation.								
Target Alpha			The desired probability of rejecting a true null hypothesis at which the tests were run.								
Actual Alpha			The alpha achieved by the test as calculated by the alpha simulation.								
Lower and Upper Limits of a 95% C.I. for Alpha			The limits of an exact, 95% confidence interval for alpha based on the binomial distribution. They are calculated from the alpha simulation.								
M1, M2, SD, etc.			The value(s) of the user-specified distribution parameters.								

**Total Sample Size N**

This is the total sample size of the study.

**95% Confidence Interval Limits for Power**

These are the limits of an exact, 95% confidence interval for power using the power simulation. The confidence interval is based on the binomial distribution. The width of this confidence interval is directly related to the number of simulations that were used.

**95% Confidence Interval Limits for Alpha**

These are the limits of an exact, 95% confidence interval for alpha using the alpha simulation. The confidence interval is based on the binomial distribution. The width of this confidence interval is directly related to the number of simulations that were used. Since the target alpha is 0.05, 0.05 should be within these limits.

## Kruskal-Wallis Tests (Simulation)

## Detailed Results Reports

## Details for Scenario 1 of the Numeric Results

Group	ni	Percent of N	$\mu$   H0	$\mu$   H1	$\sigma$   H0	$\sigma$   H1
1	4	25	40	40	18.2	17.9
2	4	25	40	10	18.0	18.2
3	4	25	40	10	18.2	17.7
4	4	25	40	10	17.8	18.1
<b>Group</b>	<b>N</b>		<b>SD of <math>\mu</math>'s   H0</b>	<b>SD of <math>\mu</math>'s   H1</b>	<b>Average of <math>\sigma</math>'s   H0</b>	<b>Average of <math>\sigma</math>'s   H1</b>
All	16	100	0	13	18.0	18.0

(More Reports Follow)

These reports show the details of each scenario.

**Scenario (in Title)**

This is the row number of the Numeric Results report about which this report gives the details.

**Group**

This is the number of the group shown on this line.

**ni**

This is the sample size of each group. This column is especially useful when the sample sizes are unequal.

**Percent of N**

This is the percentage of the total sample that is allocated to each group.

 **$\mu$  | H0 and  $\mu$  | H1**

These are the means that were used in the alpha and power simulations, respectively.

 **$\sigma$  | H0 and  $\sigma$  | H1**

These are the standard deviations that were obtained by the alpha and power simulations, respectively. Note that they often are not exactly equal to what was specified because of the error introduced by simulation.

**SD of  $\mu$ 's | H0 or H1**

These are the standard deviations of the means that were obtained by the alpha and power simulations, respectively. Under H0, this value should be near zero. The H0 value lets you determine if your alpha simulation was correctly specified. The H1 value represents the magnitude of the effect size (when divided by an appropriate measure of the standard deviation).

Kruskal-Wallis Tests (Simulation)

Average of  $\sigma$ 's | H0 or H1

These are the average of the individual group standard deviations that were obtained by the alpha and power simulations, respectively.

Dropout-Inflated Sample Size Report

Dropout-Inflated Sample Size					
Average Group Sample Size n	Group	Dropout Rate	Sample Size ni	Dropout-Inflated Enrollment Sample Size ni'	Expected Number of Dropouts Di
4	1 - 4	20%	4	5	1
	Total		16	20	4
8	1 - 4	20%	8	10	2
	Total		32	40	8
12	1 - 4	20%	12	15	3
	Total		48	60	12

n	The average group sample size.
Group	Lists the group numbers.
Dropout Rate	The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
ni	The evaluable sample size for each group at which power is computed (as entered by the user). If ni subjects are evaluated out of the ni' subjects that are enrolled in the study, the design will achieve the stated power.
ni'	The number of subjects that should be enrolled in each group in order to obtain ni evaluable subjects, based on the assumed dropout rate. ni' is calculated by inflating ni using the formula ni' = ni / (1 - DR), with ni' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lohknygina, Y. (2018) pages 32-33.)
Di	The expected number of dropouts in each group. Di = ni' - ni.

**Dropout Summary Statements**

---

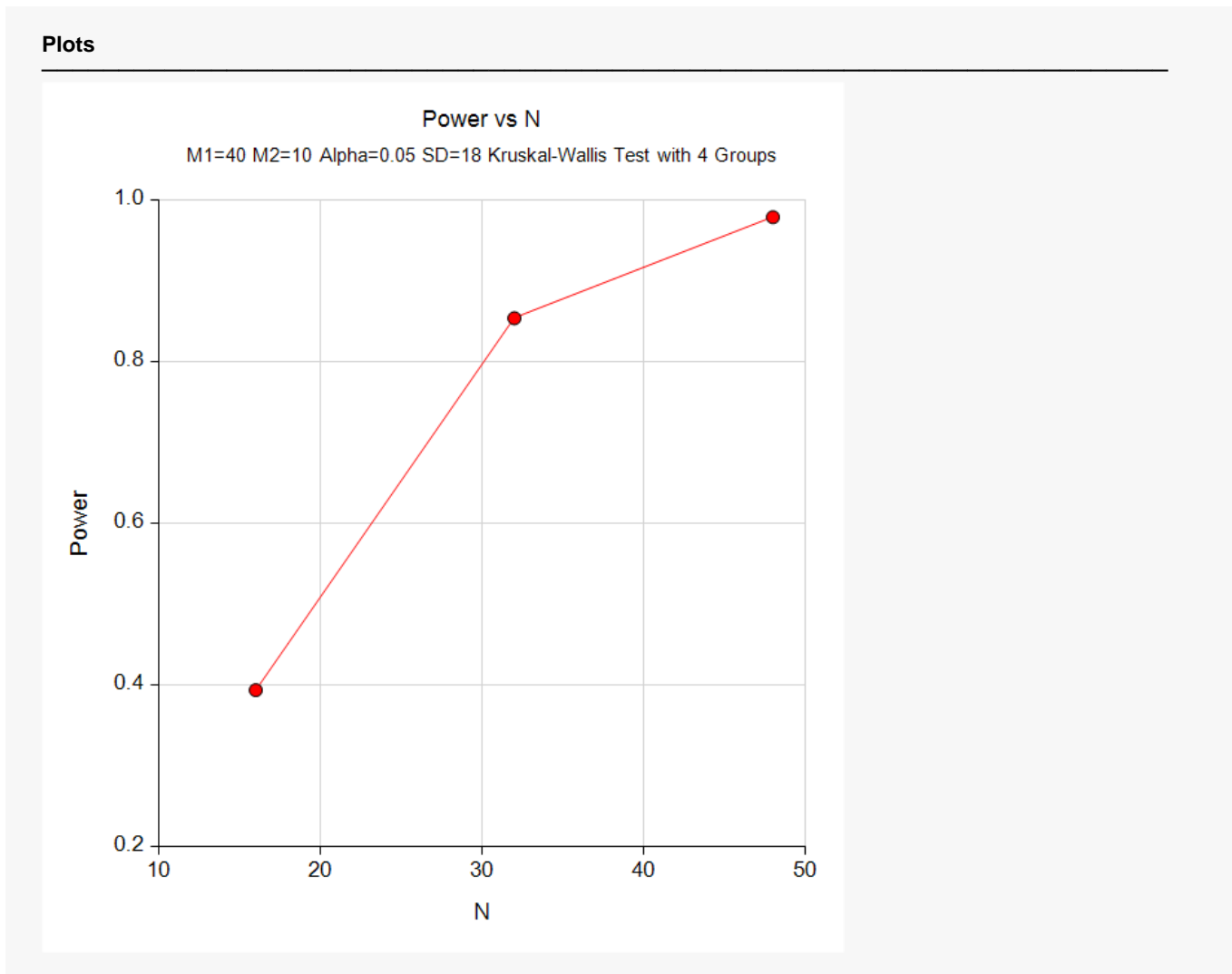
Anticipating a 20% dropout rate, group sizes of 5, 5, 5, and 5 subjects should be enrolled to obtain final group sample sizes of 4, 4, 4, and 4 subjects.

This report shows the sample sizes adjusted for dropout. In this example, dropout is assumed to be 20%. You can change the dropout rate on the Reports tab.



## Kruskal-Wallis Tests (Simulation)

## Plots Section



This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the power of increasing the sample size.

## Example 2 – Validation using Fleiss (1986)

Fleiss (1986) page 374 presents an example of determining an appropriate sample size when using an F-test in an experiment with 4 groups; means of 9.775, 12, 12, and 14.225; standard deviation of 3; alpha of 0.05, and beta of 0.20. He finds a sample size of 11 per group when using the F-Test. We anticipate that a slightly larger value will be needed for the Kruskal-Wallis Test.

For reproducibility, we'll use a random seed of 5550568.

### Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

#### Design Tab

Solve For .....	<b>Sample Size</b>
Power.....	<b>0.80</b>
Alpha.....	<b>0.05</b>
Simulations .....	<b>5000</b>
Random Seed.....	<b>5550568</b> (for Reproducibility)
Number of Groups .....	<b>4</b>
Group Allocation .....	<b>Equal (n1 = n2 = ...)</b>
Group 1 Power Distribution.....	<b>Normal(M1 SD)</b>
Group 2 Power Distribution.....	<b>Normal(M2 SD)</b>
Group 3 Power Distribution.....	<b>Normal(M3 SD)</b>
Group 4 Power Distribution.....	<b>Normal(M4 SD)</b>
Parameter 1 Name .....	<b>M1</b>
Parameter 1 Value(s).....	<b>9.775</b>
Parameter 2 Name .....	<b>M2</b>
Parameter 2 Value(s).....	<b>12</b>
Parameter 3 Name .....	<b>M3</b>
Parameter 3 Value(s).....	<b>12</b>
Parameter 4 Name .....	<b>M4</b>
Parameter 4 Value(s).....	<b>14.225</b>
Parameter 6 Name .....	<b>SD</b>
Parameter 6 Value(s).....	<b>3</b>
Specify Alpha Distributions .....	<b>All equal to the group 1 distribution of the power simulation</b>

## Output

Click the Calculate button to perform the calculations and generate the following output.

Simulation Summary													
Simulation Distribution													
Group	Power (H1)		Alpha (H0)										
1	Normal(M1 SD)		Normal(M1 SD)										
2	Normal(M2 SD)		Normal(M1 SD)										
3	Normal(M3 SD)		Normal(M1 SD)										
4	Normal(M4 SD)		Normal(M1 SD)										
Number of Groups		4											
Random Number Pool Size		10000											
Number of Simulations		5000											
Random Seed		5550568 (User-Entered)											
Run Time		3.35 seconds											
Numeric Results													
Solve For: <a href="#">Sample Size</a>													
Sample Size													
Scenario	Power		Average Group n		Alpha		Standard Deviation of $\mu$ 's   H1	Average of $\sigma$ 's   H1	M1	M2	M3	M4	SD
	Target	Actual	Total N	Target	Actual								
1	0.8	0.805	12	48	0.05	0.046	1.6	3	9.8	12	12	14.2	3

Note that **PASS** has found the group sample size to be 12, which is only slightly larger than 11, as expected.