Appendix T

Standard on Ratio Studies

Approved April 2013

4.5 Sample Representativeness

In general, a ratio study is valid to the extent that the sample is sufficiently *representative* of the population.

The distribution of ratios in the population cannot be ascertained directly and appraisal accuracy can vary from property to property. By definition, a ratio study sample would be representative when the distribution of ratios of properties in the sample reflects the distribution of ratios of properties in the population. Representativeness is improved when the sample proportionately reflects major property characteristics present in the population of sold and unsold properties. As long as sold and unsold parcels are appraised in the same manner and the sample is otherwise representative, statistics calculated in a sales ratio study can be used to infer appraisal performance for unsold parcels.

However, if parcels that sell are selectively reappraised based on their sale prices and if such parcels are in the ratio study, uniformity inferences will not be accurate (appraisals appear more uniform than they are). In this situation, measures of appraisal level also will not be supportable unless similar unsold parcels are appraised by a model that produces the same overall percentage of market value (appraisal level) as on the parcels that sold (see Appendix E, "Sales Chasing Detection Techniques"). Assessing officials must incorporate a quality control program; including checks and audits of the data, to ensure that sold and unsold parcels are appraised at the same level.

Operationally, representativeness is improved when the following occur:

1. Appraisal procedures used to value the sample parcels are similar to procedures used to value the corresponding population
2. Accuracy of recorded property characteristics data for sold property does not differ substantially from that of unsold property,
3. Sample properties are not unduly concentrated in certain areas or types of property whose appraisal levels differ from the general level of appraisal in the population
4. Sales have been appropriately screened and validated (see Appendix A).

The first requirement generally is met unless sampled parcels are valued or updated differently from nonsampled parcels, or unless appraisals of sample parcels were done at a different time than appraisals of nonsampled parcels. For example, it is unlikely that the sample is representative of unsold parcels when the sample consists mostly of new construction, first-time sales of improved properties, condominium conversions, or newly platted lots.

The second requirement is met only if value-related property characteristics are updated uniformly for all property in a class as opposed to being updated only upon sale.

The third requirement relates to the extent to which appraisal performance for the sample reflects appraisal performance for the population.

The fourth requirement generally is met when the sales to be used in the sample are properly screened, adjusted if necessary, and validated.

4.6 Acquisition and Validation of Sales Data

Sales data are important in ratio studies and play a crucial role in any credible and efficient mass appraisal system. In some instances, it may be necessary to make adjustments to sales prices so they are more representative of the market. When there is more than one sale of the same property during a study period, only one of the transactions should be used in the ratio study. For guidelines on sales validation see Appendix A.

5. Ratio Study Statistics and Analyses

Once data have been properly collected, reviewed, assembled, and adjusted, outlier handling and statistical analysis can begin. This process involves the following steps.

1. A ratio should be calculated for each observation in the sample by dividing the appraised (or assessed) value by the sale price.
2. Graphs and exhibits can be developed that show the distribution of the ratios.
3. Exploratory data analysis, including outlier identification and screening, and tests of the hypotheses of normality may be conducted.
4. Ratio study statistics of both appraisal level and uniformity should be calculated.
5. Reliability measures should be calculated.

An example of a ratio study statistical analysis report is given in table 1-1.

5.1 Data Displays

Displays or exhibits that provide a profile or picture of ratio study data are useful for illustrating general patterns and trends, particularly to nonstatisticians. The particular form of the displays, as well as the data used (e.g., sales prices, sales ratios, and property characteristics) depends on the purposes of the particular display. Types of displays useful in ratio studies are arrays, frequency distributions, histograms, plots, and maps (Gloudemans 1999).

Graphic displays can be used to

- indicate whether a sample is sufficiently representative of the properties in a stratum
- indicate the degree of nonnormality in the distribution of ratios
- depict the overall level of appraisal

Table 1-1. Example of Ratio Study Statistical Analysis Data Analyzed

Rank of ratio of observation	Appraised value ($)	Sale Price ($)	Ratio (AV/SP)
1	48,000	138,000	0.348
2	28,800	59,250	0.486
3	78,400	157,500	0.498
4	39,840	74,400	0.535
5	68,160	114,900	0.593
6	94,400	159,000	0.594
7	67,200	111,900	0.601
8	56,960	93,000	0.612
9	87,200	138,720	0.629
10	38,240	59,700	0.641
11	96,320	146,400	0.658
12	67,680	99,000	0.684
13	32,960	47,400	0.695
14	50,560	70,500	0.717
15	61,360	78,000	0.787
16	47,360	60,000	0.789
17	58,080	69,000	0.842
18	47,040	55,500	0.848
19	136,000	154,500	0.880
20	103,200	109,500	0.942
21	59,040	60,000	0.984
22	168,000	168,000	1.000
23	128,000	124,500	1.028
24	132,000	127,500	1.035
25	160,000	150,000	1.067
26	160,000	141,000	1.135
27	200,000	171,900	1.163
28	184,000	157,500	1.168
29	160,000	129,600	1.235
30	157,200	126,000	1.248
31	99,200	77,700	1.277
32	200,000	153,000	1.307
33	64,000	48,750	1.313
34	192,000	144,000	1.333
35	190,400	141,000	1.350
36	65,440	48,000	1.363

Note: Due to rounding, totals may not add to match those on following table, which reports results of statistical analysis of above data.

Results of statistical analysis

Statistic	Result
Number of observations in sample	36
Total appraised value	$3,627,040
Total sale price	$3,964,620
Average appraised value	$100,751
Average sale price	$110,128
Mean ratio	0.900
Median ratio	0.864
Weighted mean ratio	0.915
Coefficient of dispersion (COD)	29.8%
Price-related differential (PRD)	0.98
Price-related bias (PRB) coefficient (t-value)	.232 (3.01)
95% median two-tailed confidence interval	(0.684, 1.067)
95% weighted mean two-tailed confidence interval	(0.806, 1.024)
Normal distribution of ratios (0.05 level of significance)	Reject—D'Agostino, Pearson K^2, and Shapiro-Wilk W
Date of analysis	9/99/9999
Category or class being analyzed	Residential

- depict the degree of uniformity
- depict the degree of value bias (regressivity or progressivity)
- compare the level of appraisal or degree of uniformity among strata
- detect outlier ratios
- identify specific opportunities to improve mass appraisal performance
- track performance measures over time

5.2 Outlier Ratios

Outlier ratios are very low or high ratios as compared with other ratios in the sample. The validity of ratio study statistics used to make inferences about population parameters could be compromised by the presence of outliers that distort the statistics computed from the sample. One extreme outlier can have a controlling influence over some statistical measures. However, some statistical measures, such as the median ratio, are resistant to the influence of outliers and trimming would not be required. Although the coefficient of dispersion (COD) is affected by extreme ratios, it is affected to a lesser extent than the coefficient of variation (COV) and the mean. The weighted mean and price-related differential (PRD) are sensitive to sales with high prices even if the ratios on higher priced sales do not appear unusual relative to other sales. Regression analysis, sometimes used in assessment ratio analyses (e.g., when ratios are regressed on sales prices or property characteristics, such as lot size or living area), is also affected by outliers: both ratio outliers and outliers based on the comparison characteristics (an excellent treatment of the assumptions made in regression and deviations from can be found in Cook, R.D. and Weisberg, S. 1982).

Outlier ratios can result from any of the following:

1. an erroneous sale price
2. a nonmarket sale
3. unusual market variability
4. a mismatch between the property sold and the property appraised
5. an error in the appraisal of an individual parcel
6. an error in the appraisal of a subgroup of parcels
7. any of a variety of transcription or data handling errors

In preparing any ratio study, outliers should be

1. identified
2. scrutinized to validate the information and correct errors
3. trimmed if necessary to improve sample representativeness

indicates, for example, that assessment ratios fall by 4.5% when values double and increase by 4.5% when values are halved. Like all regression coefficients, the statistical reliability of the PRB can be gauged by noting its t-value and related significance level, and by computing confidence intervals. In table 1-4 the PRB is -0.035 and is not statistically significant.

Unacceptable vertical inequities should be addressed through reappraisal or other corrective actions. In some cases, additional stratification can help isolate the problem. Measures of level computed for value strata should not be compared as a way of determining vertical inequity because of a boundary effect that is most pronounced in the highest and lowest strata (Schultz 1996).

5.7 Tests of Hypotheses

An appropriate test should be used whenever the purpose of a ratio study is implicitly or explicitly to test a hypothesis. A hypothesis is essentially a tentative answer to a question, such as, Are residential and commercial properties appraised at equal percentages of market value? A test is a statistical means of deciding whether the answer "yes" to such a question can be rejected at a given level of confidence. In this case, if the test leads to the conclusion that residential and commercial properties are not appraised at equal percentages of market value, some sort of corrective action on the part of assessing officials is clearly indicated.

Tests are available to determine whether the

- level of appraisal of a stratum fails to meet an established standard
- meaningful differences exist in the level of appraisal between two or more strata
- high-value properties are appraised at a different percentage of market value than low-value properties

Appropriate tests are listed in table 1-2 and discussed in Gloudemans (1999), *Property Appraisal and Assessment Administration* (IAAO 1990), and *Improving Real Property Assessment* (IAAO 1978, 137–54).

5.8 The Normal Distribution

Many conventional statistical methods assume the sample data conform to the shape of a bell curve, known as the normal (or Gaussian) distribution. Performance measures based on the mean or standard deviation can be misleading if the study sample does not meet the assumption of normality. As a first step in the analysis, the distribution of sample ratios should be examined to reveal the shape of the data and uncover any unusual features. Although ratio study samples typically do not conform to the normal distribution, graphical techniques and numerical tests can be used to explore the data thoroughly. Traditional choices are the binomial, chi-square, and Lilliefors tests. Newer and more powerful procedures are the Shapiro-Wilk W, the D'Agostino-Pearson K^2, and the Anderson-Darling A^2 tests (D'Agostino and Stephens 1986).

5.9 Parametric and Distribution-Free (Nonparametric) Statistics

For every problem that might be solved by using statistics, there is usually more than one measure or test. These measures and tests can be divided into two broad categories: parametric and distribution-free (nonparametric). Parametric statistics assume the population data conform to a known family of probability distributions (such as the normal distribution). When the mean, weighted mean, and standard deviation are used in this context, they tend to be more meaningful. Distribution-free statistics make less restrictive assumptions and do not require knowledge about the shape of the underlying population distribution. Given similar distribution of ratios in the underlying populations, distribution free tests, such as the Mann-Whitney test, can determine the likelihood that the level of assessment

Table 1-2. Tests of Hypotheses

Null Hypothesis	Nonparametric Test	Parametric Test
1. Ratios are normally distributed.	Shapiro-Wilk W test D'Agostino-Pearson K^2 test Anderson-Darling A^2 test Lillifores Test	N/A
2. The level of appraisal meets legal requirements.	Binomial test	t-test
3. Two property groups are appraised at equal percentages of market value.	Mann-Whitney test	t-test
4. Three or more property groups are appraised at equal percentages of market value.	Kruskal-Wallis test	Analysis of Variance
5. Low- or high-value properties are appraised at equal percentages of market value.	Spearman Rank test	PRB, correlation or regression analysis
6. Sold and unsold parcels are treated equally.	Mann-Whitney test	t-test

to determine whether it can be reasonably concluded that appraisal level differs from the established goal in a particular instance. Additionally, when uniformity measures show considerable variation between ratios, level measurements may be less meaningful.

9.1.1 Purpose of Level-of-Appraisal Standard

Jurisdictions that follow the IAAO recommendation of annual revaluations (*Standard on Property Tax Policy* [IAAO 2010] and *Standard on Mass Appraisal of Real Property* [IAAO 2013]) and comply with USPAP standard rules should be able to develop mass appraisal models that maintain an overall ratio level of 100 percent (or very near thereto). However, the local assessor may be compelled to follow reappraisal cycles defined by a legal authority or public policy that can extend beyond one year. During extended cycles the influence of inflation or deflation can shift the overall ratio.

The purpose of a performance standard that allows reasonable variation from 100 percent of market value is to recognize uncontrollable sampling error and the limiting conditions that may constrain the degree of accuracy that is possible and cost-effective within an assessment jurisdiction. Further, the effect of performance standards on local assessors must be considered in light of public policy and resources available.

9.1.2 Confidence Intervals in Conjunction with Performance Standards

The purpose of confidence intervals and similar statistical tests is to determine whether it can be reasonably concluded that the appraisal level differs from the established performance standard in a particular instance. A conclusion of noncompliance requires a high degree of confidence; thus, a 90 percent (two-tailed) or 95 percent (one-tailed) confidence level should be used, except for small or highly variable samples. The demonstration ratio study report in table 1-4 presents 95% two-tailed confidence interval estimates for the mean, median, and weighted mean ratio.

9.2 Appraisal Uniformity

Assuming the existence of an adequate and sufficiently representative sample, if the uniformity of appraisal is unacceptable, model recalibration and/or reappraisal should be undertaken. It is important to recognize that the COD is a point estimate and, especially for small samples, should not be accepted as proof of assessment uniformity problems. Proof can be provided by recognized statistical tests, including bootstrap confidence intervals.

In unusually homogeneous strata, low CODs can be anticipated. In all other cases, CODs less than 5 percent should be considered suspect and possibly indicative of nonrepresentative samples or selective reappraisal of selling parcels.

9.2.1 Uniformity among Strata

Although the goal is to achieve an overall level of appraisal equal to 100 percent of the legal requirement, ensuring uniformity in appraisal levels among strata also is important. The level of appraisal of each stratum (class, neighborhood, age group, market areas, and the like) should be within 5 percent of the overall level of appraisal of the jurisdiction. For example, if the overall level of appraisal of the jurisdiction is 1.00, but the appraisal

Table 1-4. Demonstration Ratio Study Report

Rank	Parcel #	Appraised value	Sale price*	Ratio	Statistic	Result
1	9	$87,200	138,720	0.629	Number (n)	17
2	10	38,240	59,700	0.641	Total appraised value	$1,455,330
3	11	96,320	146,400	0.658	Total sale price	$1,718,220
4	12	68,610	99,000	0.693	Avg appraised value	$85,608
5	13	32,960	47,400	0.695	Avg sale price	$101,072
6	14	50,560	70,500	0.717		
7	15	61,360	78,000	0.787	Mean ratio	0.827
8	16	47,360	60,000	0.789	Median ratio	0.820
9	17	56,580	69,000	0.820	Weighted mean ratio	0.847
10	18	47,040	55,500	0.848		
11	19	136,000	154,500	0.880	Coefficient of dispersion	14.5
12	20	98,000	109,500	0.895	Price-related differential	0.98
13	21	56,000	60,000	0.933	PRB	−0.035
14	22	159,100	168,000	0.947	PRB coefficient (*t*-value)	0.135 (2.4)
15	23	128,000	124,500	1.028		
16	24	132,000	127,500	1.035	95% conf. int. mean (two-tailed)	0.754 to 0.901
17	25	160,000	150,000	1.067	95% conf. int. median (two-tailed)	0.695 to 0.933
					95% conf. int. wtd. mean (two-tailed)	0.759 to 0.935

Date: 0/0/00. *No outlier trimming*

** or adjusted sale price*

International Journal of Endocrinology & Metabolism
KOWSAR
Home | For Authors | Submit Online | Current Issue | Search

Int J Endocrinol Metab. 2012 Spring; 10(2): 486–489. PMCID: PMC3693611
Published online 2012 Apr 20. doi: 10.5812/ijem.3505

Normality Tests for Statistical Analysis: A Guide for Non-Statisticians

Asghar Ghasemi[1] and Saleh Zahediasl[1,*]

[1]Endocrine Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran
*Corresponding author: Saleh Zahediasl, Endocrine Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, P.O. Box: 19395-4763, Tehran, IR Iran. Tel.: +98-2122409309, Fax: +98-2122402463, E-mail: zahedi@endocrine.ac.ir.

Received 2011 Nov 21; Revised 2012 Jan 21; Accepted 2012 Jan 28.

Abstract

Go to:

Statistical errors are common in scientific literature and about 50% of the published articles have at least one error. The assumption of normality needs to be checked for many statistical procedures, namely parametric tests, because their validity depends on it. The aim of this commentary is to overview checking for normality in statistical analysis using SPSS.

Keywords: Normality, Statistical Analysis

1. Background

Go to:

Statistical errors are common in scientific literature, and about 50% of the published articles have at least one error (1). Many of the statistical procedures including correlation, regression, t tests, and analysis of variance, namely parametric tests, are based on the assumption that the data follows a normal distribution or a Gaussian distribution (after Johann Karl Gauss, 1777–1855); that is, it is assumed that the populations from which the samples are taken are normally distributed (2-5). The assumption of normality is especially critical when constructing reference intervals for variables (6). Normality and other assumptions should be taken seriously, for when these assumptions do not hold, it is impossible to draw accurate and reliable conclusions about reality (2, 7).

With large enough sample sizes (> 30 or 40), the violation of the normality assumption should not cause major problems (4); this implies that we can use parametric procedures even when the data are not normally distributed (8). If we have samples consisting of hundreds of observations, we can ignore the distribution of the data (3). According to the central limit theorem, (a) if the sample data are approximately normal then the sampling distribution too will be normal; (b) in large samples (> 30 or 40), the sampling distribution tends to be normal, regardless of the shape of the data (2, 8); and (c) means of random samples from any distribution will themselves have normal distribution (3). Although true normality is considered to be a myth (8), we can look for normality visually by using normal plots (2, 3) or by significance tests, that is, comparing the sample distribution to a normal one (2, 3). It is important to ascertain whether data show a serious deviation from normality (8). The purpose of this report is to overview the procedures for checking normality in statistical analysis using SPSS.

2. Visual Methods

Go to:

Visual inspection of the distribution may be used for assessing normality, although this approach is usually unreliable and does not guarantee that the distribution is normal (2, 3, 7). However, when data are presented visually, readers of an article can judge the distribution assumption by themselves (9). The frequency distribution (histogram), stem-and-leaf plot, boxplot, P-P plot (probability-probability plot), and Q-Q plot (quantile-quantile plot) are used for checking normality visually (2). The frequency distribution that plots the observed values against their frequency, provides both a visual judgment about whether the distribution is bell shaped and insights about gaps in the data and outliers outlying values (10). The stem-and-leaf plot is a method similar to the histogram, although it retains information about the actual data values (8). The P-P plot plots the cumulative probability of a variable against the cumulative probability of a particular distribution (e.g., normal distribution). After data are ranked and sorted, the corresponding z-score is calculated for each rank as follows: $z = x - \bar{x} / s$. This is the expected value that the score should have in a normal distribution. The scores are then themselves converted to z-scores. The actual z-scores are plotted against the expected z-scores. If the data are normally distributed, the result would be a straight diagonal line (2). A Q-Q plot is very similar to the P-P plot except that it plots the quantiles (values that split a data set into equal portions) of the data set instead of every individual score in the data. Moreover, the Q-Q plots are easier to interpret in case of large sample sizes (2). The boxplot shows the median as a horizontal line inside the box and the interquartile range (range between the 25 th to 75 th percentiles) as the length of the box. The whiskers (line extending from the top and bottom of the box) represent the minimum and maximum values when they are within 1.5 times the interquartile range from either end of the box (10). Scores greater than 1.5 times the interquartile range are out of the boxplot and are considered as outliers, and those greater than 3 times the interquartile range are extreme outliers. A boxplot that is symmetric with the median line at approximately the center of the box and with symmetric whiskers that are slightly longer than the subsections of the center box suggests that the data may have come from a normal distribution (8).

3. Normality Tests

Go to:

The normality tests are supplementary to the graphical assessment of normality (8). The main tests for the assessment of normality are Kolmogorov-Smirnov (K-S) test (7), Lilliefors corrected K-S test (7, 10), Shapiro-Wilk test (7, 10), Anderson-Darling test (7), Cramer-von Mises test (7), D'Agostino skewness test (7), Anscombe-Glynn kurtosis test (7), D'Agostino-Pearson omnibus test (7), and the Jarque-Bera test (7). Among these, K-S is a much used test (11) and the K-S and Shapiro-Wilk tests can be conducted in the SPSS Explore procedure (Analyze → Descriptive Statistics → Explore → Plots → Normality plots with tests) (8).

The tests mentioned above compare the scores in the sample to a normally distributed set of scores with the same mean and standard deviation; the null hypothesis is that "sample distribution is normal." If the test is significant, the distribution is non-normal. For small sample sizes, normality tests have little power to reject the null hypothesis and therefore small samples most often pass normality tests (7). For large sample sizes, significant results would be derived even in the case of a small deviation from normality (2, 7), although this small deviation will not affect the results of a parametric test (7). The K-S test is an empirical distribution function (EDF) in which the theoretical cumulative distribution function of the test distribution is contrasted with the EDF of the data (7). A limitation of the K-S test is its high sensitivity to extreme values; the Lilliefors correction renders this test less conservative (10). It has been reported that the K-S test has low power and it should not be seriously considered for testing normality (11). Moreover, it is not recommended when parameters are estimated from the data, regardless of sample size (12).

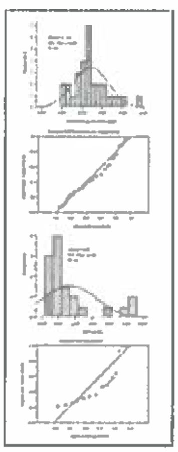
The Shapiro-Wilk test is based on the correlation between the data and the corresponding normal scores (10) and provides better power than the K-S test even after the Lilliefors correction (12). Power is the most frequent measure of the value of a test for normality—the ability to detect whether a sample comes from a non-normal distribution (11). Some researchers recommend the Shapiro-Wilk test as the best choice for testing the normality of data (11).

4. Testing Normality Using SPSS

Go to:

We consider two examples from previously published data: serum magnesium levels in 12–16 year old girls (with normal distribution, n = 30) (13) and serum thyroid stimulating hormone (TSH) levels in adult control subjects (with non-normal distribution, n = 24) (14). SPSS provides the K-S (with Lilliefors correction) and the Shapiro-Wilk normality tests and recommends these tests only for a sample size of less than 50 (8).

In Figure, both frequency distributions and P-P plots show that serum magnesium data follow a normal distribution while serum TSH levels do not. Results of K-S with Lilliefors correction and Shapiro-Wilk normality tests for serum magnesium and TSH levels are shown in Table. It is clear that for serum magnesium concentrations, both tests have a p-value greater than 0.05, which indicates normal distribution of data, while for serum TSH concentrations, data are not normally distributed as both p values are less than 0.05. Lack of symmetry (skewness) and pointiness (kurtosis) are two main ways in which a distribution can deviate from normal. The values for these parameters should be zero in a normal distribution. These values can be converted to a z-score as follows:



Figure

Histograms (Left) and P-P Plots (Right) for Serum Magnesium and TSH Levels

Table

Skewness, kurtosis, and Normality Tests for Serum Magnesium and TSH Levels Provided by SPSS

$Z_{Skewness} = Skewness\text{-}0 / SE_{Skewness}$ and $Z_{Kurtosis} = Kurtosis\text{-}0 / SE_{Kurtosis}$.

An absolute value of the score greater than 1.96 or lesser than -1.96 is significant at $P < 0.05$, while greater than 2.58 or lesser than -2.58 is significant at $P < 0.01$, and greater than 3.29 or lesser than -3.29 is significant at $P < 0.001$. In small samples, values greater or lesser than 1.96 are sufficient to establish normality of the data. However, in large samples (200 or more) with small standard errors, this criterion should be changed to ± 2.58 and in very large samples no criterion should be applied (that is, significance tests of skewness and kurtosis should not be used) (2). Results presented in Table indicate that parametric statistics should be used for serum magnesium data and non-parametric statistics should be used for serum TSH data.

5. Conclusions

Go to:

According to the available literature, assessing the normality assumption should be taken into account for using parametric statistical tests. It seems that the most popular test for normality, that is, the K-S test, should no longer be used owing to its low power. It is preferable that normality be assessed both visually and through normality tests, of which the Shapiro-Wilk test, provided by the SPSS software, is highly recommended. The normality assumption also needs to be considered for validation of data presented in the literature as it shows whether correct statistical tests have been used.

Acknowledgments

Go to:

The authors thank Ms. N. Shiva for critical editing of the manuscript for English grammar and syntax and Dr. F. Hosseinpanah for statistical comments.

Footnotes

Go to:

Implication for health policy/practice/research/medical education: Data presented in this article could help for the selection of appropriate statistical analyses based on the distribution of data.

Please cite this paper as: Ghasemi A, Zahediasl S. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. Int J Endocrinol Metab. 2012;10(2):486-9. DOI: 10.5812/ijem.3505

Financial Disclosure: None declared.

Funding/Support: None declared.

References

Go to:

1. Curran-Everett D, Benos DJ. Guidelines for reporting statistics in journals published by the American Physiological Society. Am J Physiol Endocrinol Metab. 2004;287(2):E189–91. doi: 10.1152/ajpendo.00213.2004. [PubMed] [Cross Ref]

2. Field A. Discovering statistics using SPSS. 3 ed. London: SAGE publications Ltd; 2009. p. 822.

3. Altman DG, Bland JM. Statistics notes: the normal distribution. Bmj. 1995;310(6975):298. [PMC free article] [PubMed]

4. Pallant J. SPSS survival manual, a step by step guide to data analysis using SPSS for windows. 3 ed. Sydney: McGraw Hill; 2007. pp. 179–200.

5. Driscoll P, Lecky F, Crosby M. An introduction to everyday statistics-1. J Accid Emerg Med. 2000;17(3):205–11. [PMC free article] [PubMed]

6. Royston P. Estimating departure from normality. Stat Med. 1991;10(8):1283–93. [PubMed]

7. Oztuna D, Elhan AH, Tuccar E. Investigation of four different normality tests in terms of type 1 error rate and power under different distributions. Turkish Journal of Medical Sciences. 2006;36(3):171–6.

8. Elliott AC, Woodward WA. Statistical analysis quick reference guidebook with SPSS examples. 1st ed. London: Sage Publications; 2007.

9. Altman DG, Bland JM. Detecting skewness from summary information. Bmj. 1996;313(7066):1200. [PMC free article] [PubMed]

10. Peat J, Barton B. Medical Statistics: A guide to data analysis and critical appraisal. Blackwell Publishing; 2005.

11. Thode HJ. Testing for normality. New York: Marcel Dekker; 2002.

12. Steinskog DJ. A cautionary note on the use of the Kolmogorov-Smirnov test for normality. American Meteor Soc. 2007;135:1151–7.

13. Ghasemi A, Syedmoradi L, Zahediasl S, Azizi F. Pediatric reference values for serum magnesium levels in Iranian subjects. Scand J Clin Lab Invest. 2010;70(6):415–20. doi: 10.3109/00365513.2010.504280. [PubMed] [Cross Ref]

14. Zahediasl S, Brojeni NK, Ghasemi A, Faraji F, Hedayati M, Azizi F. Alterations in osmotic fragility of the red blood cells in hypo and hyperthyroid patients. J Endocrinol Invest. 2009;32(1):28–32. [PubMed]

IBM Knowledge Center

Search Content Products Help

SPSS Data Collection > SPSS Data Collection 6.0.1 >

prev next

Table of contents

IBM SPSS Data Collection V6.0.1 documentation

Version 6.0.1

Welcome to the IBM SPSS Data Collection V6.0.1 documentation, where you can find information about how to install, maintain, and use the IBM SPSS Data Collection applications.

Getting started

- Product overview
- What's new
- Samples
- IBM SPSS Data Collection Desktop installation guide
- IBM SPSS Data Collection Server installation guide
- Release notes

Common tasks

- Author tutorials
- Professional tutorials
- Data Entry and Interviewer tutorials
- Paper tutorials
- Interviewer Server tutorials

Troubleshooting and support

More information

English

SPSS Data Collection

Select

Select a specific version or edition of SPSS Data Collection documentation from the drop-down menu.

Learn more

- Online catalog of software products: General information about software products
- Explore IBM Systems: General information about Systems products
- IBM Cloud Computing: Materials about the promise of cloud
- Redbooks: Technical publications by experts about hundreds of subjects
- developerWorks: A community site with a vast amount of information for IT professionals

Find support

- IBM Support Portal
- IBM Software Support
- IBM Client Success Portal: Support for cloud solutions
- IBM Passport Advantage: Support for licensed software

Stay current

Significance value

The significance value, or p value, is the probability that a result occurred by chance. The significance value is compared to a predetermined cutoff (the significance level) to determine whether a test is statistically significant. If the significance value is less than the significance level (by default, 0.05), the test is judged to be statistically significant.

The significance value does not indicate whether a result is practically significant. Effect size is another measure from a statistical test. It helps determine the practical significance. IBM® Watson Analytics™ uses both the significance value and the effect size to determine whether a result is important enough to display.

Parent topic:
Statistical terms

Setting the Significance Levels

Version 6.0.1

By default, the column proportions, column mean, net difference test, and paired preference tests are run at the 5% significance level. However, you can optionally run a test at another significance level, such as the 10% or 1% significance level.

You can also run the test at two significance levels on the same table. In the resulting table, the IDs of columns that are significant at the higher level appear in upper case, and those that are significant at the lower level appear in lower case.

You select this option using the *SigLevel* statistics property. For example:

```
TableDoc.Tables.MyTable.Statistics[0].Properties["SigLevel"] = 1
```

The Statistics and Statistic objects implement the mrScriptBasic dynamic property expansion feature. This means that an alternative way of writing this would be:

```
TableDoc.Tables.MyTable.Statistics.ColumnProportions.SigLevel = 1
```

See the topic Dynamic Property Expansion for more information.

To run the test at two significance levels, use the *SigLevelLow* statistics property to display the lower significance level. For example:

```
TableDoc.Tables.MyTable.Statistics.ColumnProportions.SigLevel = 1
TableDoc.Tables.MyTable.Statistics.ColumnProportions.SigLevelLow = 5
```

In the resulting table, the IDs of columns that are significant at the higher level appear in upper case, and those that are significant at the lower level appear in lower case.

Note: If you are using two levels of significance, ensure that the value of the *SigLevelLow* property is greater than that of the *SigLevel* property, as it represents a higher probability that the results are due to chance, and therefore a lower level of significance.

319

Minitab® 17 Support

Search Minitab 17 Support

What value should I use for significance level?

Learn more about Minitab 17

Before you do a hypothesis test, you must choose a significance level for the test. Use the significance level to judge whether the test results are statistically significant. The significance level also determines the probability of error that is inherent in the test.

If the probability that an event occurs is less than α, the usual interpretation is that the event did not occur by chance. Formally, α is the maximum acceptable level of risk for rejecting a true null hypothesis (Type I error) and is expressed as a probability ranging between 0 and 1. The smaller the significance level, the less likely you are to make a Type I error, and the more likely you are to make a Type II error. Therefore, you should choose an alpha that balances these opposing risks of error based on their practical consequences in your specific situation.

Usually, a significance level (denoted as α or alpha) of 0.05 works well. A significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference.

When to choose a larger alpha

Choose a larger alpha, such as 0.10, to be more certain that you will not miss detecting a difference that might exist.

For example, an engine manufacturer wants to compare the stability of new ball bearings with the current ones. If the new ball bearings are less stable, customers could have disastrous consequences. Therefore, they choose an α of 0.1 to be more certain that they will detect any possible difference in the stability.

When to choose a smaller alpha

Choose a smaller alpha, such as 0.01, to be more certain that you will only detect a difference that really does exist.

For example, a pharmaceutical company wants to be very certain before making an advertising claim that its new product significantly reduces symptoms. The company chooses an α of 0.001 to be sure that any significant difference in symptoms that they detect actually does exist.