An Introduction to Distribution-free Statistical Methods

Douglas G. Bonett
University of California, Santa Cruz

5/4/2016

Overview

- Need for distribution-free statistical methods
- Limitations of distribution-free methods
- Distribution-free alternatives to popular hypothesis testing methods
- Distribution-free confidence intervals to accompany distribution-free tests
- Sample size requirements for distribution-free hypothesis testing and interval estimation
- R commands for distribution-free tests and new user-defined R functions for confidence intervals and sample size requirements

Need for Distribution-free Methods

The popular *t*-tests and ANOVAs assume that the response variable has an approximate normal distribution. These methods can give misleading results when the normality assumption is violated and the sample size is small (less than 20 per group).

Tests for normality are not useful in small samples because they often fail to detect serious departures from normality.

Studies that use *t*-tests and ANOVAs in small samples are susceptible to *replication failure* if the response variable is substantially nonnormal. This is disconcerting because normality tests usually fail to warn the researcher of nonnormality in small samples.

Need for Distribution-free Methods (*continued*)

Even if the sample size is not small, so that the traditional methods can be expected to perform properly under moderate non-normality, traditional methods involve tests and confidence intervals for parameters (e.g., means and standard deviations) that can be misleading in non-normal populations.

For example, tests and confidence intervals for means can gives misleading results because the mean is a misleading measure of centrality in skewed distributions. Distribution-free tests and confidence intervals for medians are useful when the response variable is skewed because the median is a more meaningful measure of centrality than the mean in these situations.

Some distribution-free tests are more powerful than *t*-tests and ANOVAs if the response variable is extremely non-normal.

Limitations of Distribution-free Methods

If the response variable is only moderately non-normal, distribution-free tests can be much less powerful than *t*-tests and ANOVAs, and distribution-free confidence intervals can be much less precise than the traditional confidence intervals.

Recommendations:

1) Use distribution-free methods in studies where the sample size must be small because of cost and other considerations, but the effect size is expected to be large and detectable in a small sample.

2) Use distribution-free methods in studies where the response variable is believed to have a moderately or highly non-normal distribution where the traditional parameters (e.g., mean, standard deviation, Pearson correlation) could be misleading descriptions of the population.

Distribution-free Alternative to One-sample *t*-test

The one-sample *t*-test is a test of $\mathrm{H}_0\colon \mu = h$ where μ is the population mean and h is the hypothesized value of the population mean. The p-value for this test tends to be too small if the sample size is small and response variable is skewed.

The *sign test* is a distribution-free alternative to the one-sample *t*-test. The sign test is a test of $\mathrm{H}_0\colon \theta = h$ where θ is the population *median* and h is the hypothesized value of the population median. The sign test is trustworthy for any sample size and for any response variable distribution shape.

Distribution-free Alternative (*continued*)

If the p-value for the sign test is less than α we reject H_0: $\theta = h$ and accept H_1: $\theta > h$ if the sample median is greater than h or accept H_2: $\theta < h$ if the sample median is less than h. The probability of making a directional error (i.e., accepting H_1 when H_2 is true or accepting H_2 when H_1 is true) is at most α /2.

The sign test is substantially less powerful than the one-sample *t*-test when the response variable is approximately normal, although it can be more powerful if the response variable is highly leptokurtic (i.e., more peaked with heavier tails than a normal distribution).

Distribution-free Confidence Interval for a Median

A confidence interval for a population mean can have a true coverage probability that is substantially less than the specified confidence level if the response variable is skewed and the sample size is small.

Even if the sample size is large, a distribution-free confidence interval for a population median, rather than a mean, could be more meaningful if the response variable is skewed.

Note: Using a degree-of-belief definition of probability, rather than a relative frequency definition, provides a very simple and useful interpretation of a computed confidence interval.

Example

A random sample of 15 UCSC students was obtained and each student was asked to keep a journal and record how much they spent during a 30-day period on food. The hypothetical results (in dollars) are:

385 497 376 405 520 790 480 530 345 371 468 501 586 583 460

A sign test was used to test the null hypothesis that the population median 30-day food expenditure for the population of UCSC student is equal to the national median amount of $570.

The results of the sign test indicate that we can reject the null hypothesis and accept the alternative hypothesis that the population median 30-day food expenditure is less than $570 ($p$ = .035). We are 95% confident the population median is between $385 and $530.

Distribution-free Alternative to Two-sample *t*-test

The *Mann-Whitney test* is a distribution-free alternative to the two-sample *t*-test.

If the response variable is approximately normal, the Mann-Whitney test is only slightly less powerful than the two-sample *t*-test. If the response variable is leptokurtic, the Mann-Whitney test can be more powerful than the two-sample *t*-test.

In a 2-group experiment, the Mann-Whitney test is a test of the null hypothesis $\mathrm{H_0}\colon \pi = .5$ where π is the probability that a person would score higher under Treatment 1 than Treatment 2. Note that $\mathrm{H_0}\colon \pi = .5$ implies $\mathrm{H_0}\colon \theta_1 = \theta_2$ if the shape of the response variable distribution, assuming no treatment effect, is the same in each condition.

The probability of a directional error (accepting $\mathrm{H_1}\colon \pi > .5$ when $\mathrm{H_2}\colon \pi < .5$ is true or accepting $\mathrm{H_2}\colon \pi < .5$ when $\mathrm{H_1}\colon \pi > .5$ is true) is at most α /2.

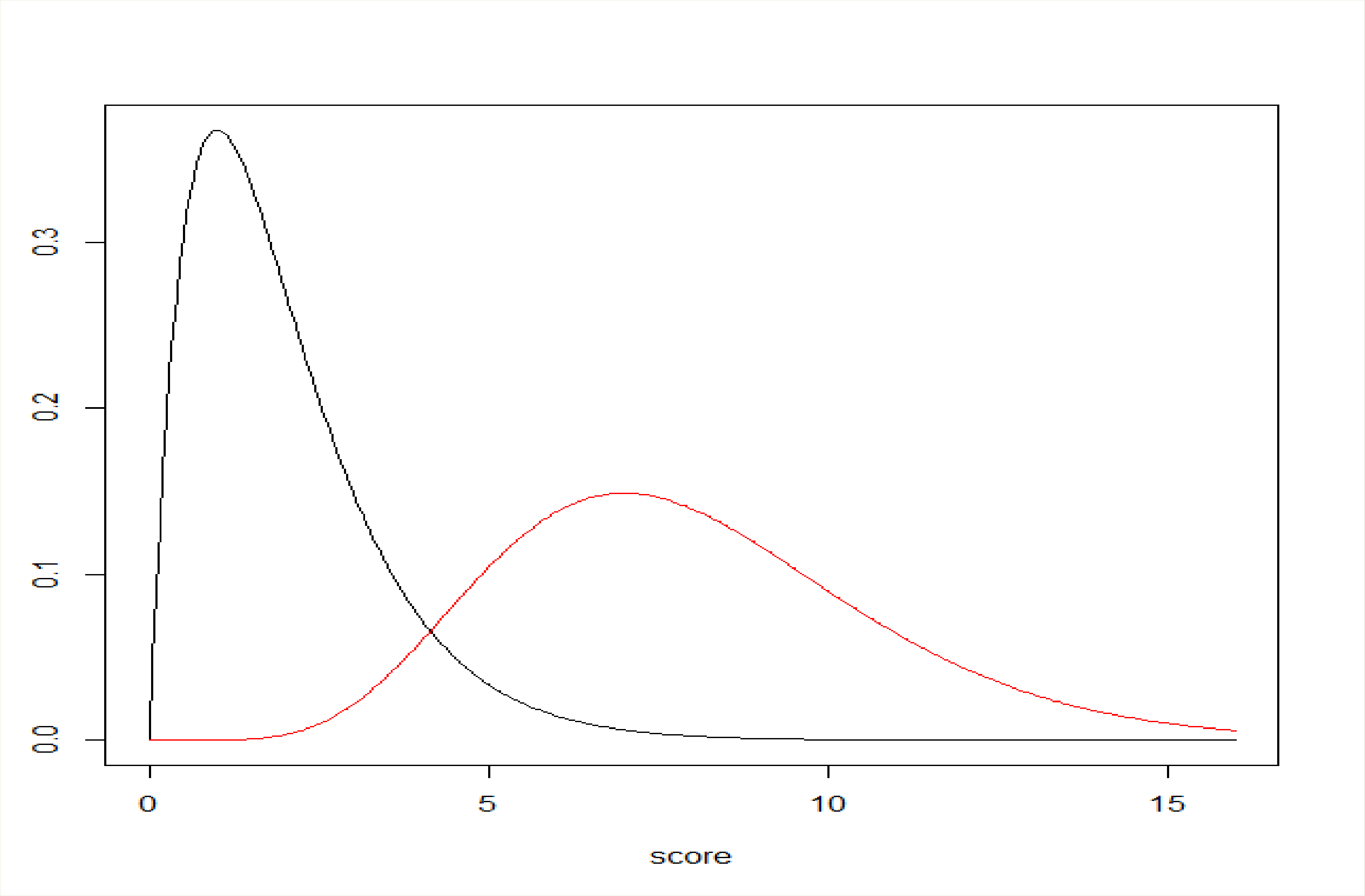
Distribution-free Confidence Interval for a Difference in Medians (2-group design)

Some computer packages that compute the Mann-Whitney test will also compute a confidence interval for a “shift parameter”. The shift parameter is equal to the difference in population medians <u>only</u> if the distribution of the response variable has the same shape in each treatment condition – this is a very unrealistic assumption.

Instead of reporting a confidence interval for a shift parameter, it is usually better to report a confidence interval for a difference in population medians $(\theta_1 - \theta_2)$ that does not assume equal response variable distribution shapes.

However, the confidence interval for a difference in medians that does not assume equal distribution shapes will usually be wider that the confidence interval for a shift parameter.

Example where Median Difference > Mean Difference



Distribution-free Confidence Interval for a Ratio of Medians (2-group design)

A confidence interval for a ratio of medians (θ_1/θ_2) is also available that assumes ratio scale measurements but does not assume equal response variable distribution shapes.

If the response variable has been measured on a ratio scale, a confidence interval for a ratio of medians might be more meaningful than a confidence interval for a difference in medians. For example, suppose reaction time was measured in two groups (0.1% blood alcohol vs. 0.02% or less) a 95% confidence interval for a ratio of medians was [1.25, 1.50]. We could say that the median reaction time would be 1.25 to 1.50 <u>times greater</u> if everyone in the population had a 0.1% blood alcohol level rather than a 0.02% or less blood alcohol level.

Distribution-free Confidence Interval for a 2-group Standardized Effect Size

Standardized effect sizes can be used to describe the magnitude of an effect in the absence of any information about the metric of the response variable.

In a 2-group design, Cohen's d is a popular standardized measure of effect size. However, if the response variable is not approximately normal, Cohen's d is difficult to interpret and the confidence interval for the population value can have a true coverage probability that is much less than the specified confidence level.

The Mann-Whitney parameter (π) is a value from -1 to 1 and is another type of standardized effect size. A confidence interval for the Mann-Whitney parameter does not require any assumptions about the shapes of the response variable distributions in the two treatment conditions.

Example

A random sample of 8 male frequent marijuana users and a random sample of 10 male non-users were obtained from a University of Colorado research participant pool. Amygdala activity levels of all participants were obtained while participants listened to an audio recording with high emotional content. The hypothetical activity scores are:

Users:	14.6	5.1	8.1	22.7	6.4	4.4	19.0	3.2		
Non-users:	9.4	10.3	58.3	66.0	31.0	46.2	12.0	19.0	71.0	129.0

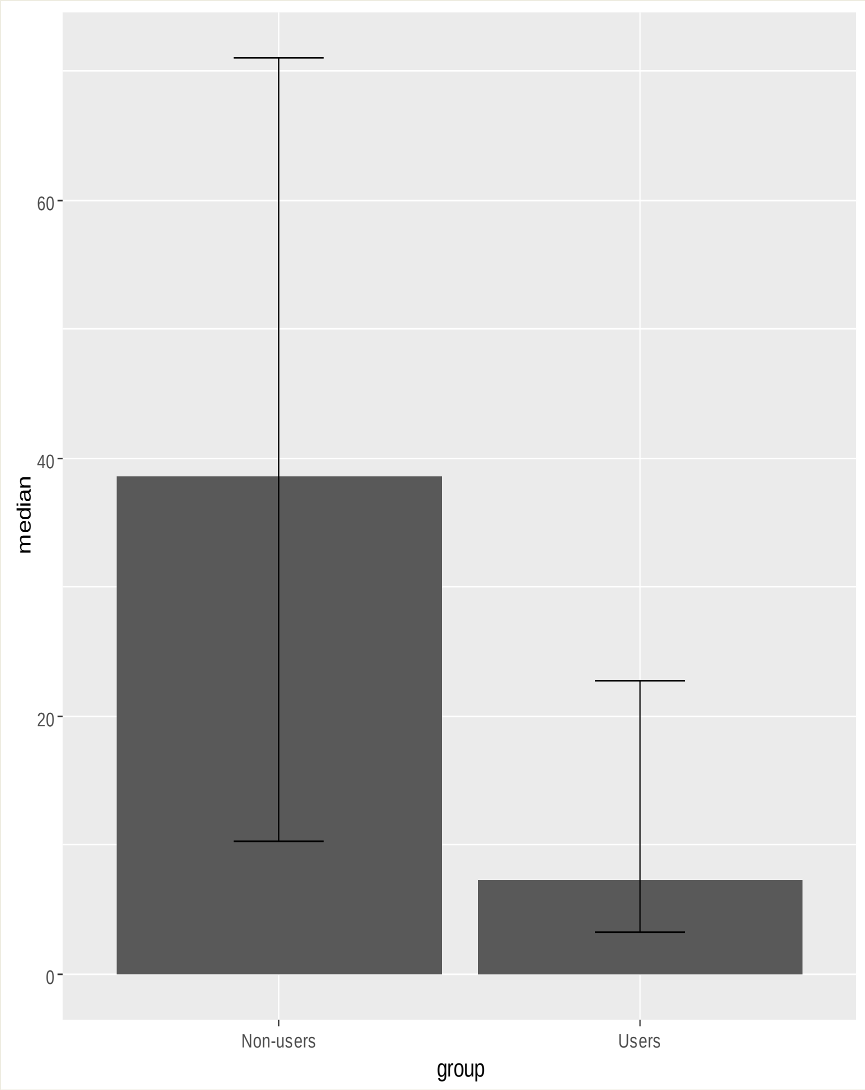
Example

The p-value for the Mann-Whitney test is .006 and we can reject $\mathrm{H_0}$: $\pi =$ .5. The confidence interval for π (given below) suggests that we can accept the alternative hypothesis π > .5.

A 95% confidence interval for $\theta_2 - \theta_1$ is [4.29, 58.4]. We can be 95% confident that the median amygdala activity score in the population of marijuana non-users is 4.29 to 58.4 greater than the median amygdala activity score in the population of marijuana users. In the absence of information regarding the psychological meaning of activity scores, a confidence interval for a standard measure of effect size (π) will be easier to interpret.

The 95% confidence interval for π is [.629, 1.00] which indicates that the probability of a randomly selected marijuana non-user having a higher amygdala activity score in response to the emotional video than a randomly selected marijuana user is .629 to 1.00.

Median Bar Chart with Confidence Interval Lines



Distribution-free Alternative to One-way ANOVA
(between-subjects design)

Like the two-sample *t*-test, the one-way ANOVA should not be used in small samples when the response variable is skewed. Even in large samples, a comparison of medians would be more interesting than a comparison of means if the response variable is skewed.

The *Krusal-Wallis test* is a distribution-free alternative to the one-way ANOVA. If the shape of the response variable distribution in the absence of any treatment effect is assumed to be the same for all treatments, then the Kruskal-Wallis test is a test of $\mathrm{H}_0\!: \theta_1 = \theta_2 = \ldots = \theta_k$.

Jonckheere Test for an Ordered Between-subjects Factor

If the levels of the between-subjects factor are ordered, the *Jonckheere test* is a test of the null hypothesis

$$\mathrm{H}_0\colon \theta_1 = \theta_2 = \dots = \theta_k$$

against the alternative hypothesis

$$\mathrm{H}_1\colon \theta_1 \le \theta_2 \le \dots \le \theta_k$$

Unlike the Kruskal-Wallis test where the alternative hypothesis simply states that there is at least one pair of population medians that are not identical, accepting H_1 using the Jonckheere test provides useful information about the order of the treatment effects.

Between-subject Pairwise Comparisons

The Kruskal-Wallis test has the same intrinsic limitation as the one-way ANOVA. Specifically, rejecting H_0 does not provide any information about how the population medians are ordered or by how much they differ. The Jonckheere test does provide order information.

For an unordered factor, Mann-Whitney tests for all pairwise comparisons and confidence intervals for all pairwise differences or ratios of medians would provide useful information.

For an ordered factor, Mann-Whitney tests and confidence intervals could be limited to the adjacent pairs.

A Bonferroni adjusted alpha level should be used when computing pairwise tests and confidence intervals ($\alpha^* = \alpha/[k(k-1)]$ for all pairwise comparisons and $\alpha^* = \alpha/(k-1)$ for adjacent comparisons).

Example

21 randomly selected college students were randomly assigned to three groups. All participants viewed the same 4-second video of a moving car hitting a stopped car and were later asked to estimate the speed (in MPH) of the moving car when the moving car "*bumped* into the stopped car" (group 1), "*crashed* into the stopped car" (group 2) or "*smashed* into the stopped car" (group 3). The hypothetical MPH estimates are:

Bumped:	17	8	10	12	16	14	13
Crashed:	12	22	20	12	19	18	20
Smashed:	25	26	22	32	30	22	27

Jonckheere p-value: <.001 [reject H_0 and accept H_1: $\theta_1 \leq \theta_2 \leq \theta_3$]

95% CI for $\theta_{bump} - \theta_{crash}$: [-11.5, -0.6]

95% CI for $\theta_{crash} - \theta_{smash}$: [-12.7, -1.3]

Distribution-free Alternative to Paired-samples *t*-test

In a paired-samples design, each participant is assigned to scores, y_1 and y_2. Recall that the paired-sample *t*-test is a test of $\mathrm{H}_0\colon \mu_d = 0$ where μ_d is the population mean of the difference scores $(y_1 - y_2)$. The *Wilcoxon signed-rank sum test* is a distribution-free alternative to the paired-samples *t*-test. The Wilcoxon signed-rank test is a test of $\mathrm{H}_0\colon \theta_d = 0$ where θ_d is the population median of the difference scores.

A confidence interval for the population median difference (θ_d) or the population median of ratios (y_1/y_2), assuming ratio scale measurements, is an informative supplement to the Wilcoxon signed-rank test.

Note: If the distribution of difference scores is symmetric (and it can be shown that it will be symmetric in a within-subjects experiment under the null hypothesis of no treatment effect), then $\mathrm{H}_0\colon \theta_d = 0$ implies $\mathrm{H}_0\colon \mu_d = 0$.

Distribution-free Confidence Interval for a Paired-samples Standardized Effect Size

A confidence interval for a median difference might be difficult to interpret if the metric of the response variable is not well understood. In these situations, a confidence interval for the population proportion of people who would have a y_1 score that is greater than their y_2 score is a useful standardized measure of effect size in paired-samples designs.

Example

The reaction times under two conditions (upright vs. 60^0 rotated objects) were recorded for a random sample of eight college students. The hypothetical data (in ms) are:

0^0	621	589	604	543	588	647	639	590
60^0	704	690	741	625	724	736	780	625

The Wilcoxon signed-rank test rejects H_0: $\theta_d = 0$ (p = .008), and a 95% confidence interval for the median difference ($60^0 - 0^0$) is [35, 141]. We can be 95% confident that the population median difference is between 35 and 141 ms. Also, we can be 95% confident that between 62.2% and 100% of the population would have slower reaction times to the 60^0 rotated objects than the upright objects.

Distribution-free Alternative to One-way ANOVA
(within-subjects design)

Like the paired-samples *t*-test, the one-way within-subjects ANOVA should not be used in small samples when the response variable is skewed. Even in large samples, a comparison of medians would be more interesting than a comparison of means if the response variable is skewed.

The *Friedman test* is a distribution-free alternative to the within-subjects one-way ANOVA. If the shape of the response variable distribution in the absence of any treatment effect is assumed to be the same for all treatments, then the Friedman test is a test of $\mathrm{H}_0\colon \theta_1 = \theta_2 = \ldots = \theta_k$.

Page Test for an Ordered Within-subjects Factor

If the levels of the within-subjects factor are ordered, the *Page test* is a test of the null hypothesis

$$\mathrm{H}_0\colon \theta_1 = \theta_2 = \ \dots = \theta_k$$

against the alternative hypothesis

$$\mathrm{H}_1\colon \theta_1 \leq \theta_2 \leq \ \dots \ \leq \theta_k$$

Unlike the Friedman test where the alternative hypothesis simply states that there is at least one pair of population medians that are not identical, accepting H_1 using the Page test provides useful information about the order of the treatment effects.

Example

Six randomly select college students viewed three photos of the same person and were asked to rate on a 1-20 scale how angry the person was when the photo was take. The person in the photo was a professional actor who was instructed to look "mildly annoyed", "moderately angry", and "extremely angry". The hypothetical ratings are:

mild:	4	2	1	6	7	4
moderate:	5	4	7	8	12	15
extreme:	9	14	10	12	16	17

Page p-value: <.01 [reject H_1 and accept H_1: $\theta_1 \leq \theta_2 \leq \theta_3$]

95% CI for $\theta_{mild\ -mod}$: [-11, -1]

95% CI for $\theta_{mod\ -ext}$: [-10, -2]

Within-subject Pairwise Comparisons

The Friedman test has the same intrinsic limitation as the one-way within-subjects ANOVA. Specifically, rejecting H_0 does not provide any information about how the population medians are ordered or by how much they differ. The Page test does provide order information.

With an unordered within-subjects factor, Wilcoxon tests for all pairwise comparisons and confidence interval for all pairwise medians of differences provide useful information.

With an ordered within-subjects factor, Wilcoxon tests and confidence intervals for medians of differences could be limited to the adjacent pairs.

A Bonferroni adjusted alpha level should be used when computing pairwise tests and confidence intervals ($\alpha^* = \alpha/[k(k-1)]$ for all pairwise comparisons and $\alpha^* = \alpha/(k-1)$ for adjacent pairs).

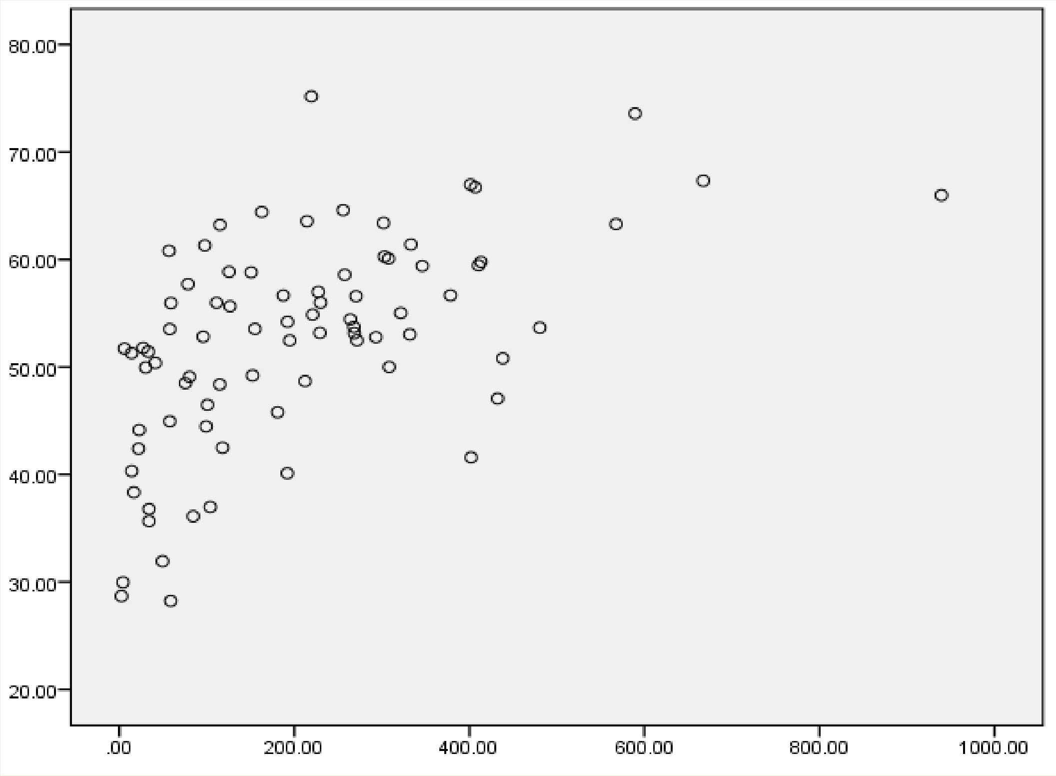
Spearman Correlation as an Alternative to the Pearson Correlation

A Pearson correlation describes the magnitude of a *linear* relation between two quantitative variables x and y. The confidence interval for a Pearson correlation assumes bivariate normality and will have a coverage probability less than the stated confidence level if x or y is leptokurtic regardless of sample size.

The *Spearman correlation* describes the strength of a *monotonic* relation between x and y, and a confidence interval for the population Spearman correlation does not assume bivariate normality. The Spearman correlation can be used as a standardized measure of effect size to accompany a Jonckheere test in single-factor study where the factor is ordered.

The Spearman correlation is less influenced by extreme x or y scores than the Pearson correlation. If x and y have a nonlinear but monotonic relation, the Spearman correlation can be substantially larger than the Pearson correlation.

Example of a Monotonic Relation



Pearson correlation: .56

Spearman correlation: .59

Distribution-free Confidence Interval for Difference and Average of Two Spearman Correlations

A Spearman correlation between y and x can be compared in two different populations (e.g., men and women, adolescents and adults, minority and nonminority, etc.).

A confidence interval for the difference in population Spearman correlations provide information about the direction of the difference and the magnitude of the difference.

A confidence interval for a difference in Spearman correlations also can be used to provide replication evidence – the Spearman correlation in the published study can be compared with the Spearman correlation in the replication study. If the results replicate, a confidence interval for the average Spearman correlation can be substantially narrower than the confidence intervals for either of the separate correlations.

Examples

1) Suppose the estimated Spearman correlation between "loneliness" and "self-confidence" was .46 in one sample of 250 in-state college students and .24 another sample of 200 out-of-students college students. The 95% confidence interval for the difference in population Spearman correlations is [.05, .39]. This results suggest that the correlation between loneliness and self-confidence is .05 to .39 greater for in-state students than out-of-state students.

2) Suppose a published study reported a Spearman correlation of .65, and a replication study obtained a Spearman correlation of .58. Both studies used a sample size of 100. The 95% confidence for the difference is [-.12, .26] which includes 0 and hence provides evidence of "effect size replication". The confidence intervals in both studies had lower limits greater than 0 which also provides evidence of "directional replication". The 95% confidence interval for the average Spearman correlation is [.57, .66].

Distribution-free Alternative to Test and Confidence Interval for Simple Linear Regression Slope

The simple linear regression model $y_i = \beta_0 + \beta_1 x_i + e_i$ assumes that the prediction errors (e_i) are normally distributed and are equally variable across the levels of x. Hypothesis testing and confidence interval results for the population slope (β_1) can be misleading if these assumptions are violated.

The *Theil-Sen test* and confidence interval for β_1 are distribution-free alternatives to the traditional test and confidence interval. Even if all assumptions of the linear regression model could be satisfied, the Theil-Sen test of $\mathrm{H}_0\!: \beta_1 = 0$ is only slightly less powerful than the traditional test and the Theil-Sen confidence interval for β_1 is only slightly less precise than the traditional confidence interval. If the assumptions are violated, the Theil-Sen test can be more powerful and the Theil-Sen confidence interval can be more precise than the traditional methods.

Example

Suppose a sample of 15 LGBT college students were given tests to measure anxiety and post-traumatic stress. The hypothetical scores are:

Anxiety:	21	4	9	12	35	18	10	22	24	1	6	8	13	16	19
PTSD:	67	28	30	28	52	40	25	37	44	10	14	20	28	40	51

The p-value for the Theil-Sen test of β_1 = 0 is .0007, and the p-value for the traditional test of β_1 = 0 is .0002. The Theil-Sen confidence interval for β_1 is [0.42, 0.66], and the traditional confidence interval is [0.27, .68].

In this example, the traditional results are misleading because the prediction errors are heteroscedastic and slightly skewed.

Variability Assessment

Social scientists tend to focus on group means when analyzing their data and reporting the results. This can lead to overly simplified generalizations (stereotypes) about the effects of different types of treatments or different sub-populations of people.

An analysis of the variability of the response variable can provide important insights regarding individual differences.

In an experiment, current practice recommends a particular treatment if the treatment produces the best mean response. However, if that treatment also had the lowest variability in responses, such a finding would provide a more compelling argument in favor of that treatment. Also, if one treatment has considerably greater variability than another treatment, there could be theoretically interesting interaction effects worthy of further investigation.

Variability Assessment (*continued*)

Given the benefits of statistically analyzing variability in additional to centrality, why don’t researchers test hypotheses or report confidence intervals for measures of variability such as the standard deviation?

Classical methods for analyzing standard deviations assume that the response variable has a normal distribution, and these methods can give very misleading results if the assumption has been violated *even in large samples*.

Also, the standard deviation is most easily interpreted in normal distributions because it is the point of inflection on the normal curve.

Variability Assessment (*continued*)

An alternative measure of variability is the *mean absolute devi*ation (MAD) from the median. This measure of variability has a simple interpretation for any shape distribution. Furthermore, tests and confidence intervals for the population MAD (τ) have recently been developed that do not assume normality of the response variable.

True distribution-free methods perform properly for any type of distribution and for any sample size. In comparison, these new statistical methods for analyzing MADs are not distribution-free in the usual sense but they are asymptotically (large sample) distribution free (ADF). However, they exhibit distribution-free properties in moderate sized samples.

Variability Assessment (*continued*)

Available methods:

- confidence interval for τ
- confidence interval for τ_1/τ_2 (2-group design)
- confidence interval for τ_1/τ_2 (paired-samples design)
- confidence interval for a coefficient of dispersion (COD = $\tau\ /\theta$)
- confidence interval for $(\tau_1/\theta_1)/(\tau_2/\theta_2)$ (2-group design)

Example 1

A random sample of 20 3rd grade students who qualified for ESL training were randomly assigned to receive traditional ESL training or a new ELS training program. An English proficiency test was given to all 20 students at the end of the academic year. The hypothetical tests scores are:

New:	41	46	49	44	51	42	48	44	43	50
Traditional:	30	37	14	33	46	25	38	35	32	7

95% CI for $\theta_1 - \theta_2$: [1.70, 23.3]

95% CI for τ_1/τ_2: [0.17, 0.82]

The new ESL program produces a higher median <u>and</u> less variability in English proficiency scores than the traditional ESL program.

Example 2

A random sample of 7th grade girls and boys where asked to keep track of the number of hurtful text messages, posts, or emails they received during a 30-day period. With frequency count data, the group with the larger mean or median is expected to also have greater variability. The researcher wants to know how the relative variability (τ/θ) in cyberbully events compares for boys and girls.

Suppose the 95% CI for $(\tau_{girl}/\theta_{girl})/(\tau_{boy}/\theta_{boy})$ was [1.25, 2.80].

This result suggests that there is greater relative variability in cyberbully events for girls than boys. Additional research could attempt to explain the greater variability experienced by girls.

Sample Size Requirements

Sample size planning is an important aspect in the design of most studies. If the sample size is too small, hypothesis tests will lack power and confidence intervals can be uselessly wide.

In the behavioral sciences, sample sizes that are too large are wasteful of valuable human resources and could prevent other researchers from obtaining their required samples.

Funding agencies usually require a sample size justification, and some journals now require authors to show that the sample size is large enough to have adequate power or precision in an effort to reduce the number of studies that might fail to replicate.

Sample Size Requirements (*continued*)

Sample size formulas are available that can be used to approximate the sample size needed to test a hypothesis (e.g., sign test, Mann-Whitney test, Wilcoxon test) with desired power and a specified alpha level.

Other sample size formulas are available that can be used to estimate a parameter (e.g., Spearman correlation, Mann-Whitney parameters, proportion greater than median, median, median difference, linear contrast of medians, MAD, ratio of MADs) with desired confidence and confidence interval width.

Sample Size Requirements for Medians

Sample size planning for a median, median difference, or linear contrast of medians confidence interval is complicated by the fact that the sampling variability of a median depends on the shape of the distribution.

The sample size formulas require the researcher to specify a “shape” parameter.

This shape parameter can be set to 1, 1.3, or 1.57 if the response variable is expected to be extremely skewed, mildly skewed, or approximately normal, respectively.

Sample Size Requirements for MADs

Sample size planning for a MAD or MAD ratio confidence interval is complicated by the fact that the sampling variability of a MAD depends on the shape of the distribution.

The sample size formulas require the researcher to specify two shape parameters (`KUR` and `SKW`).

Set `KUR` to 1.3 for a flat distribution, 1.57 for a normal distribution, and 2.0 for long-tailed distribution (larger values of `KUR` give larger sample size requirements).

Set `SKW` to 0 for a symmetric distribution, 0.25 for a mildly skewed distribution, and 0.5 for an extremely skewed distribution (larger values of `SKW` give larger sample size requirements).

Examples

To obtain a 95% confidence interval for a population Spearman correlation that has a width of .3, assuming the correlation is about .5

```
sizeCISpear(.05, .5, .3)
[1] 109
```

a sample size of 109 is required.

To obtain a 95% confidence interval for a difference in population medians (2-group design) that has a width of 8, assuming the response variable has a variance of about 144 with a shape parameter of 1 (highly slewed)

```
sizeCImedian2(.05, 144, 1.0, 8)
[1] 72
```

a sample size of 72 per group is required.

Examples (*continued*)

To obtain a Mann-Whitney test that has power = .9 for alpha = .05 and assuming a .7 probability of scoring higher under treatment 1 than treatment 2,

```
sizePOWmann(.05, .7, .9)
[1] 44
```

a sample size of 44 per group is required.

To obtain a Sign test for a within-subjects design that has power = .9 for alpha = .05 and assuming a .75 probability of scoring higher under treatment 1 than treatment 2

```
sizePOWsignWS(.05, .75, .9)
1] 32
```

a sample size of 32 is required.

R Commands

A Word file that contains R commands and functions for many distribution-free methods can be downloaded from the CSASS website.

This file contains the commands and examples to compute the distribution-free hypothesis tests (Sign, Mann-Whitney, Kruskal-Wallis, Jonckheere, Friedman, Page, Wilcoxon, Theil-Sen), confidence intervals (median, median difference, median ratio, linear contrast of medians, Mann-Whitney parameter, MAD, ratio of MADs, COD, and ratio of CODs) and many sample size problems.

Thank you.

Questions or comments？