Resampling Statistics

These statistics may be appropriate in circumstances where you are uncomfortable with the normality assumption common to parametric inferential statistics, and perhaps in some other circumstances as well. David Howell was of the opinion that resampling statistics will replace the traditional nonparametric statistics, and perhaps the traditional parametric statistics, in time.

David Howell's Visual Basic Resampling package can be downloaded at https://www.uvm.edu/~dhowell/StatPages/Resampling/ResamplingPackage.zip. Prior to his death in October of 2018, Dave was developing R code to do resampling statistics, see Howell’s Resampling Statistics page. Releases of Windows more recent than Windows 7 have problems running Dave's program.

Bootstrapping

With this approach, one constructs a sampling distribution by repeatedly sampling, with replacement, from the actual sample of data at hand. This is much like what we did back in PSYC 6430 when we employed Monte Carlo methods to construct sampling distributions of the mean, variance, standard deviation, $z$, and $t$, but then we sampled from mathematically defined populations (such as the standard normal population).

Confidence Interval for a Median

Consider the example presented displayed on Figure 18.1 of David Howell's Statistical Methods for Psychology, 8th edition. The data at hand are 20 scores on a memory task. We wish to construct an 95% confidence interval for the median. We chose the median because the distribution appears to be distinctly skewed. Here is how we construct the confidence interval:

- We assume that the population is distributed exactly as is the sample.
- We randomly draw one score from the sample. We record it, replace it, and draw another. We repeat this process 20 times (20 because there are 20 scores in the sample). We compute the median of the obtained resample and record it.
- We repeat this process, obtaining a second sample of 20 scores and computing and recording a second median.
- We continue until we have obtained a large number (10,000 or more) of resample medians. We obtain the probability distribution of these medians and treat it like a sampling distribution.
- From the obtained sampling distribution, we find the .025 and the .975 percentiles. These define the confidence limits.

The sampling distribution obtained by Howell appears on page 662 of his textbook and indicates a 95% CI of [6, 10].
I used Dave's software to analyze the data found at http://core.ecu.edu/psyc/wuenschk/StatData/MedianBoot.dat

I booted the program Resampling.exe.

I clicked Help, Program Description. Read the little help file there.

I clicked Analysis, Bootstrapping, Bootstrapped Median.

I clicked File, Open, MedianBoot.dat, Open.

I set NReps to 10,000, CI to 95%, clicked Run.

Here are the results

Confidence Interval for Pearson $r$

Consider the data on misanthropy, idealism, and attitude about animals, which we analyzed back in PSYC 6430. We found a significant correlation between misanthropy and attitude about animals for nonidealists but not for idealists. Let us now put a confidence interval on the correlation we obtained with the nonidealists. Here is how we construct the confidence interval:

- We construct a bootstrap sample of 91 pairs of scores. That is, we sample, with replacement, from our actual sample, 91 pairs of scores. We compute Pearson $r$ on that sample of 91 pairs of scores and then record the value of $r$.
- We repeat this process many times, each time recording the value of $r$. 
• From the resulting distribution of values of $r$, we obtain the .025 and .975 percentiles. This is the confidence interval for our sample $r$

I used Howell's resampling program to construct this confidence interval:

• The data are at http://core.ecu.edu/psyc/wuenschk/StatData/CorrReSample.dat
• Click Analysis, Bootstrapping, Bootstrapped Correlation.
• Click File, Open, CorrReSample.dat, Open.
• Set NReps to 10,000, CI to 95%, click Run.

Here is the output:

[Image of the output from the Bootstrapping Correlations program]

You know that the traditional independent samples $t$ test is equivalent to a test of the null hypothesis that the point-biserial $r$ is zero in the population. Accordingly, it might well make sense to use this correlation program for a bootstrapping test of the difference in means between two independent samples -- just code group membership with numbers (like 1 and 2) and run the program. If the confidence interval for the point biserial $r$ does not include zero, then the two groups differ significantly.
Permutation/Randomization Tests

With this approach one takes the data at hand, randomly assigns scores to groups (without replacement), and then computes, on the obtained sample(s), the relevant statistic. This procedure is repeated many times, obtaining a sampling distribution of the statistic of interest.

Two Independent Samples

Consider the data on page 667 of Howell's textbook. We have a sample of 49 scores in the success group, and 18 in the fail group. Here is how we conduct a permutation/randomization test:

- We take all $49 + 18 = 67$ scores and throw them into one pot. After mixing them, we select one score and assign it to the success group. Without replacing it, we select a second score and assign it to the success group. We continue this until we have assigned 49 scores to the success group. The remaining 17 scores are assigned to the fail group.
- We compute the medians of the resulting groups and then the difference between those medians. We record the difference in medians.
- We put the 67 scores back in the pot and repeat the process, obtaining and recording a second difference in medians for the resulting sample of 49 scores assigned to one group, 17 to the other group.
- We repeat this procedure a large number of times. The resulting set of differences in medians is our sampling distribution.
- From the sampling distribution, we determine what proportion of the differences had absolute values as large or larger than that observed in our actual data. If that proportion is less than some criterion (typically .05), we reject the null hypothesis of no difference in population medians.
- From Figure 18.6 we see that $p = .0498$.

I used Howell's software to conduct this test:

- The data are at: [http://core.ecu.edu/psyc/wuenschk/StatData/2IndReSamples.dat](http://core.ecu.edu/psyc/wuenschk/StatData/2IndReSamples.dat)
- Click Analysis, Randomization Tests, Compare Medians of Two Samples.
- Click File, Open, 2IndReSamples.dat, Open.
- Set NReps to 10,000, CI to 95%, click Run.
- What is reported here as a confidence interval is NOT what you usually think of as a confidence interval for the difference in medians, but is rather the nonrejection region for the test of our null hypothesis of no difference in medians. If the difference in medians for our two samples falls outside of that nonrejection interval, then we reject the null hypothesis of no difference in medians.

Here is the output:
Two Correlated Samples

Consider the data on page 665 of Howell's textbook. We have paired data from 19 subjects, and 19 signed difference scores. Here is how we conduct a permutation/randomization test:

- We conduct a one-sample $t$ test of the null that the mean difference score is zero. For our data, $t = 3.121$. If we refer this to the distribution of Student's $t$, we obtain $p = .006$, but we are not going to use Student's $t$. We are going to construct our own distribution of $t$.
- We take the 19 difference scores, strip them of their signs, and then randomly assign signs to them.
- We compute and record a one-sample $t$ for the resulting distribution of difference scores.
- We repeat this process a large number of times.
- From the resulting sampling distribution under the null hypothesis of no difference in population means (that is, a mean difference score of 0), we compute our $p$ value by finding what proportion of the resampled $t$ values differ from 0 by at least as much as did the $t$ from our actual samples.

I used Howell's software to analyze these data:

- The data are located at [http://core.ecu.edu/psyc/wuenschk/StatData/2CorrReSamples.dat](http://core.ecu.edu/psyc/wuenschk/StatData/2CorrReSamples.dat)
• Click Analysis, Randomization Tests, Two Paired Samples.
• Click File, Open, 2CorrReSamples.dat, Open.
• Set NReps to 10,000, click Run.
• Record the obtained \( t \) and the \( p \) value. Later you will share this information with the rest of the class.
• Note that this test compares means, not medians. The program computes \( t \) for the original paired sample and for each of the resamples. The \( p \) which is reported is not based on Student's \( t \), but rather on the location of the obtained \( t \) in the distribution of \( t \) values for the resamples.

Here is the output:

![Paired T Test Output](image)

**Closing Comments**

Dr. Howell's program is designed as a teaching tool rather than a research tool. If you wish to conduct resampling statistics for research purposes, you might want to get a commercial package -- unless you are as frugal as am I.
Other Software

John Grosberg offers a giftware program he has written, Statistics101. It executes the Resampling Stats language of Julian Simon and Peter Bruce. I cannot recommend this software for reasons I shall not elaborate here. Google “R resampling statistics.” You will get lots of hits. Replace R with SAS. Replace SAS with SPSS.

SPSS now will do bootstrapping. Here I used the same data used above to construct a confidence interval for Pearson r.

Analyze, Correlate Bivariate. Bootstrap

![Correlation screenshot]

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>X Pearson Correlation</td>
<td>1</td>
<td>.364**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>Bootstrap Bias</td>
<td>0</td>
<td>.001</td>
</tr>
<tr>
<td>Std. Error</td>
<td>0</td>
<td>.105</td>
</tr>
<tr>
<td>95% Confidence Interval Lower</td>
<td>1</td>
<td>.151</td>
</tr>
<tr>
<td>Upper</td>
<td>1</td>
<td>.560</td>
</tr>
</tbody>
</table>

The observed correlation is .364. The 95% confidence interval runs from .151 to .560. These results are similar to those obtained with Howell's software.
Now to compare two related samples using the same data used above with Howell’s software. Analyze, Nonparametric tests, Legacy Dialogs, 2 related samples, Exact, Monte Carlo.

![Image of software interface]

Ranks

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean Rank</th>
<th>Sum of Ranks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post - Pre</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative Ranks</td>
<td>4</td>
<td>4.63</td>
<td>18.50</td>
</tr>
<tr>
<td>Positive Ranks</td>
<td>15</td>
<td>11.43</td>
<td>171.50</td>
</tr>
<tr>
<td>Ties</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Post < Pre  
b. Post > Pre  
c. Post = Pre

Test Statistics\(^{a,c}\)

<table>
<thead>
<tr>
<th></th>
<th>Post - Pre</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Z</strong></td>
<td>-3.079(^b)</td>
</tr>
<tr>
<td>Asymp. Sig. (2-tailed)</td>
<td>.002</td>
</tr>
<tr>
<td>Monte Carlo Sig. (2-tailed)</td>
<td>.000</td>
</tr>
</tbody>
</table>

95% Confidence Interval

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Bound</td>
<td>.000</td>
</tr>
<tr>
<td>Upper Bound</td>
<td>.001</td>
</tr>
</tbody>
</table>

a. Wilcoxon Signed Ranks Test  
b. Based on negative ranks.  
c. Based on 10000 sampled tables with starting seed 2000000.

The confidence interval here is for the \(p\) value. Since it excludes .01, our test is significant beyond .01.

Another approach would be simply to bootstrap the confidence interval for the post minus pre difference scores.
### Descriptive Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Bias</th>
<th>Std. Error</th>
<th>95% Confidence Interval</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diff</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Minimum</td>
<td>-3.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum</td>
<td>35.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>6.8053</td>
<td>.0051</td>
<td>2.1601</td>
<td>3.0368</td>
<td>11.4892</td>
<td></td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>9.50359</td>
<td>-.57911</td>
<td>2.43017</td>
<td>4.18875</td>
<td>13.20427</td>
<td></td>
</tr>
<tr>
<td>Valid N (listwise)</td>
<td>N</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>

a. Unless otherwise noted, bootstrap results are based on 10000 bootstrap samples

[SPSS Exact Tests](#) – freebie, until the URL disappears.