Bootstrapping a Confidence Interval for a Difference in Means

Jennifer Thompson, MPH

MS Meeting, October 18, 2007

1 Background

This data is from a study of mechanically ventilated ICU patients which compared two methods of weaning patients off ventilation. The standard method involves interrupting ventilator support in what is known as a "spontaneous breathing trial" (SBT). The experimental method includes an SBT, but adds a "spontaneous awakening trial" (SAT), which interrupts sedatives. The main outcome is ventilator-free days, which are days alive and off the ventilator. To avoid mortality bias, ventilator-free days were calculated as follows:

- All patients who died within 28 days of enrollment have a value of 0.

- All patients who survived the study period have a value of 28 - number of days on vent.

Because of the high mortality rate in this population (critically ill adults), ventilator-free days are very heavily skewed, as seen on page 2.

As is typically done for skewed data, we used a nonparametric Wilcoxon test to determine whether there was a difference between the two groups in terms of time off the ventilator.

Table 1: Descriptive Statistics by Study arm

	N	SBT	SBT+SAT	Test Statistic
		$N = 168$	$N = 167$	
Ventilator-free days during first 28 days of study	332	0 8 24 (12±12)	0 20 26 (15±12)	$F_{1,330} = 6$, $P = 0.02$

$_a$ $_b$ $_c$ represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD.
N is the number of non–missing values.
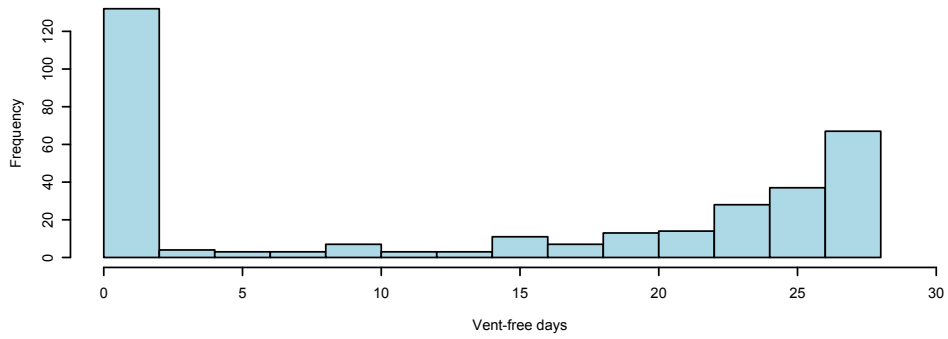Test used: Wilcoxon test

This test showed a significant difference in the distribution of ventilator-free days; the intervention group had significantly more days alive and off the vent than did the usual care group.
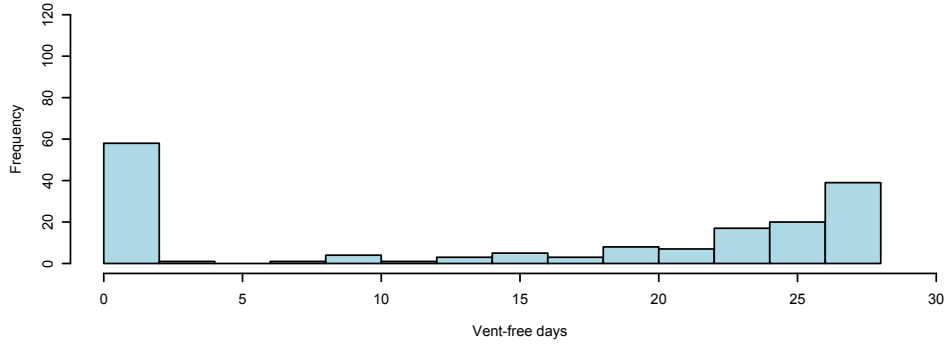
2 Problem

However, confidence intervals are a better descriptor of what's happening in the data than p-values alone. The question is, in this case, what measure to use and how to calculate the CI. Because of the skewed nature of the data, a typical calculation for the difference in means and CI is not the way to go. The median, normally the choice for skewed data, doesn't fit this extremely skewed distribution because of the extreme number of zeroes. The Hodges-Lehmann estimator, an option in many cases with a skewed distribution, also balks at extreme numbers of zeroes. After lots of back and forth with Frank, we wound up with this solution: use the difference in means as a descriptive statistic, because that fits the data better than a median, is the clinical norm and is more intuitive for physicians reading about the study. But rather than using the typical CI calculation, we decided to bootstrap a 95% confidence interval.
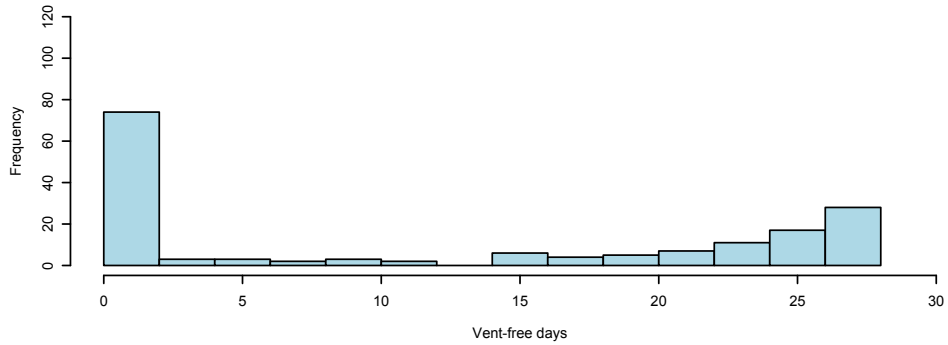
Figure 1: Histograms of Ventilator-Free Days

Histogram of Ventilator-Free Days, All Patients

Histogram of Ventilator-Free Days, Intervention Group

Histogram of Ventilator-Free Days, Control Group

3   Method and Code

The code for this procedure can be found on the Twiki under **BootstrapMeansSoftware**. We use the `smean.cl.boot()` function in the `Hmisc` library and write a function to return an object with the difference in means and 95% CI:

```
bootdif <- function(y, g) {
  ## Ensure treatment group is a factor
  g <- as.factor(g)

  ## Use the smean.cl.boot function to bootstrap means for
   # variable y for each treatment group (A and B); this code
   # uses 2000 samples, but can easily be changed
  a <- attr(smean.cl.boot(y[g==levels(g)[1]],
        B=2000, reps=TRUE), 'reps')
  b <- attr(smean.cl.boot(y[g==levels(g)[2]],
        B=2000, reps=TRUE), 'reps')

  ## Calculate the observed mean difference between groups
  meandif <- diff(tapply(y, g, mean, na.rm=TRUE))

  ## Calculate the 2.5 and 97.5 percentiles of the differences
   # in bootstrapped means
   # (can easily be changed for 90% CI, 99% CI, etc.)
  a.b <- quantile(b-a, c(.025,.975))

  ## Prepare object to return
  res <- c(meandif, a.b)
  names(res) <- c('Mean','.025','.975')

  res
}
```

The grouping variable is passed to `g` and the analysis variable to `y`. The function then uses `smean.cl.boot()` to bootstrap means for each treatment group. The `diff()` and `tapply()` functions are used to calculate the difference in observed means between treatment groups. The `quantile()` function calculates the 2.5th and 97.5th percentiles for the differences in bootstrapped means. The observed difference and percentiles are returned in the object `res`.

4 Example

4.1 Using the Function and Printing Within Text

So what does this actually look like? We start with a data frame that includes patient ID, treatment group, and ventilator-free days.

```
> head(subset(abc.analyze, select = c(id, group.pp, vent.free.death.28)))

    id group.pp vent.free.death.28
1 1001  SBT+SAT          25.949306
2 1002      SBT          26.052083
3 1003  SBT+SAT           0.000000
4 1004      SBT           0.000000
5 1005  SBT+SAT          26.893056
6 1006      SBT           1.158333
```

The grouping and analysis variables are passed to the `bootdif()` function we defined earlier.

```
> g <- abc.analyze$group.pp
> y <- abc.analyze$vent.free.death.28
> ci.vfds <- bootdif(y, g)
> ci.vfds

     Mean       .025       .975
3.1206996 0.5569981 5.5915029
```

To put these results in a print-friendly form, I did the following:

```
> meandiff.vfds <- round(ci.vfds[1], 1)
> lcl.vfds <- round(ci.vfds[2], 1)
> ucl.vfds <- round(ci.vfds[3], 1)
```

"Bootstrapping with 2000 samples was used to calculate nonparametric 95% confidence intervals for the difference in mean VFDs, resulting in a difference in means of 3.1 and CI of (0.6, 5.6)."

4.2 Printing Results From Multiple Variables

Another option is putting the numbers in a data frame and printing the data frame, which is especially helpful when you have more than one variable for which you're bootstrapping CIs. For example, if I was looking at vent-free days during the first 14 days of the study and during the first 28 days, I would do the following in addition to what has already been calculated:

```
> g <- abc.analyze$group.pp
> y <- abc.analyze$vent.free.death.14
> ci.vfds.14 <- bootdif(y, g)
> meandiff.vfds.14 <- round(ci.vfds.14[1], 1)
> lcl.vfds.14 <- round(ci.vfds.14[2], 1)
> ucl.vfds.14 <- round(ci.vfds.14[3], 1)
> boot.cis <- data.frame(outcome = c("Ventilator-Free Days During First 28 Study Days",
+      "Ventilator-Free Days During First 14 Study Days"), numbers = c(paste(meandiff.vfds,
+      " (", lcl.vfds, ", ", ucl.vfds, ")", sep = ""), paste(meandiff.vfds.14,
+      " (", lcl.vfds.14, ", ", ucl.vfds.14, ")", sep = "")))

> latex(boot.cis, file = "", rowname = NULL, where = "!htbp", col.just = c("l",
+      "c"), caption = "Difference in Means (Intervention - Usual Care) and\nBootstrapped 95\\% CIs",
+      colheads = c("Outcome", "Difference in Means (CI)"))
```

Table 2: Difference in Means (Intervention - Usual Care) and Bootstrapped 95% CIs

Outcome	Difference in Means (CI)
Ventilator-Free Days During First 28 Study Days	3.1 (0.6, 5.6)
Ventilator-Free Days During First 14 Study Days	1.4 (0.2, 2.6)

5 Other Notes

5.1 Three or More Groups

The **BootstrapMeansSoftware** twiki page also includes code to look at means for three groups and their pairwise differences.

5.2 Bootstrapped CI and Wilcoxon p-Value

Because the bootstrapped CI is not related to the Wilcoxon test we originally performed, you may get results that seem very different. For example, our p-value was 0.02, but our CI for the difference in means is (0.6, 5.6). It's important to be able to clearly articulate the difference to investigators.

5.3 Please Note:

If you include this bootstrapping code within a complete analysis file, the values will change slightly each time you run the program, since you are re-bootstrapping each time. This caused a little bit of confusion for our investigators. Although I haven't tried it, one suggestion I have is to run the bootstrapping separately and save the results in a separate file or workspace, then read those results into your main analysis file.

References

[1] R. Ihaka and R. Gentleman. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5:299-314, 1996. Software available from www.r-project.org.

[2] B. Efron and R.J. Tibshirani. An Introduction to the Bootstrap. New York, N.Y.: Chapman & Hall/CRC; 1993.

[3] Internal twiki site with code: http://biostat.mc.vanderbilt.edu/BootstrapMeansSoftware

[4] Hmisc documentation: http://biostat.mc.vanderbilt.edu/s/Hmisc/html/00Index.html