BIOST 514/517
Biostatistics I / Applied Biostatistics I

Kathleen Kerr, Ph.D.
Associate Professor of Biostatistics
University of Washington

Lecture 15:

Inference for Medians and Variances;
Bootstrapping

November 27-December 2, 2013

Lecture Topics

- Inference for Medians
- Inference for Variances
- Bootstrapping

Medians and Variances

- We discuss medians and variances together because there are not convenient, reliable formulas to estimate confidence intervals for either one.
- Bootstrapping will help us make inference for these parameters.
- An important distinction:
  - Inference for medians would probably be quite widely used if more convenient
  - Inference for variances is less often of scientific interest

Median

- Scientific: The median may be the preferred summary of effect when it is important to show an effect across all subjects
  - The mean detects effects that occur in a small subset
- Statistical: The median tends to be more efficiently estimated than the mean when the data are distributed with heavy tails

Median

- "In theory," inference for the sample median can use asymptotic theory
- The sample median is asymptotically Normally distributed

$$X_m \sim N\left(mdn(X), \frac{1}{4n\left[f(mdn(X))\right]^2} \right)$$

- The formula for the standard error is difficult to use in practice
  - Depends on the distribution of the data

Median: Testing

- Testing for equality of medians can be done by creating a binary variable indicating whether an observation is above the pooled median for the two groups
- Under the null hypothesis that the median is the same in the two groups
  - The median would be the same in the pooled data
  - Any observation should have the same probability of being above the sample median, regardless of which group it comes from

Median: Testing

- In the PBC data, the median bilirubin across both groups (treated/untreated) is 1.35
- **`gen dibili=bilirubin>1.35`**
- **`gen tmt=2-treatment`**
- **`cs dibili tmt`**

```
                 | tmt                    |
                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+------------
           Cases |        83          73  |        156
        Noncases |        75          81  |        156
-----------------+------------------------+------------
           Total |       158         154  |        312
                 |                        |
            Risk |  .5253165     .474026  |         .5
                 | chi2(1) =     0.82  Pr>chi2 = 0.3650
```

Median: Testing

- The previous slide show exactly what STATA is doing when testing the median:
- **`median bilirubin, by(tmt)`**

```
Median test

   Greater |
  than the |         tmt
    median |         0          1 |     Total
-----------+----------------------+----------
        no |        81         75 |       156
       yes |        73         83 |       156
-----------+----------------------+----------
     Total |       154        158 |       312

          Pearson chi2(1) =   0.8206   Pr = 0.365
```

Median: Testing

- Note: The test gives a p-value, does not give a confidence interval
- Note: Previous comments hold about testing baseline variables based on group assignment in randomized interventional trials

Testing the Median: Wilcoxon?

- The Wilcoxon Rank Sum test is sometimes described as testing equality of medians
- This is, in general, incorrect.
  - Only true if you assume the shape of the distributions is exactly the same in the two groups,
    - There is only a “location shift”
  - In my opinion, this is an inappropriate assumption
    - Assuming something about distributions that is far more detailed than what you are trying to detect

STATA: Testing Quantiles

- The ‘**`median`**’ function is a command to test equality of medians across groups
- However, the methodology is general and can be applied to other quantiles
  - Just a few steps
  - Rarely of interest

Variance

- Most textbook examples of comparing variances are trying to decide which two-sample t-test to use
- As discussed, this is misguided. Use the test that does not assume the variances are equal in the two groups
- There are situations where comparing variances is important
  - Comparing different ways of measuring fine particulate air pollution. Machines can be calibrated to remove systematic bias, so the variability of measurement is the most important characteristic.
  - Quantitative genetics: many tests for heritability of a trait are based on comparing the variability of the trait is more or less genetically heterogeneous groups

Variance

- The sampling distribution of the variance is asymptotically Normal. However, it converges to a Normal distribution very slowly, so this result is only useful with very large sample sizes.
- If the data are Normally distributed, then s^2/σ^2 has a χ^2 distribution with n-1 degrees of freedom. In software, this result is the basis of inference for variances.
- Unfortunately, if the distribution isn't Normal…
  - Skewed
  - More outliers than Normal

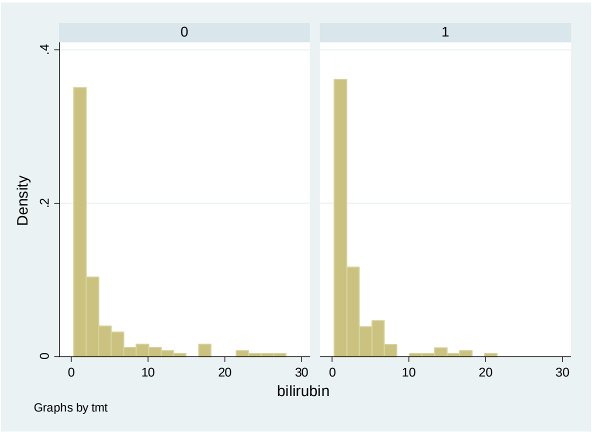
… this result does not hold, and inference is anti-conservative

  - Confidence intervals too narrow
  - "5% test" could have type I error rate 20-30%

Variance: STATA

- Two-sample tests for equality of variances are based on comparing the ratio of the sample variances to the ratio of two χ^2 distributions, which has an F distribution
- STATA does not provide confidence intervals

```
sdtest bilirubin, by(tmt)
Variance ratio test
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     154    3.648701    .4256316    5.281949    2.807828    4.489575
       1 |     158    2.873418    .2886962    3.628855    2.303188    3.443647
---------+--------------------------------------------------------------------
combined |     312     3.25609    .2564786    4.530315    2.751437    3.760742
------------------------------------------------------------------------------
    ratio = sd(0) / sd(1)                                         f =   2.1186
Ho: ratio = 1                                  degrees of freedom = 153, 157
    Ha: ratio < 1               Ha: ratio != 1                 Ha: ratio > 1
  Pr(F < f) = 1.0000         2*Pr(F > f) = 0.0000           Pr(F > f) = 0.0000
```



Other Measures of Spread?

- We might consider other measures of spread, e.g. the IQR, but there are no simple formulas for their distributions

BOOTSTRAPPING

Sampling Distributions

- The concept of a Sampling Distribution underlies inferential statistics
  - The distribution of a statistic across conceptual replications of a study
- In practice, we do not see sampling distributions
  - We only have our 1 study
  - Statistical theory tells us about the sampling distribution
- E.g., In Hypothesis Testing, statistical theory tells us about the sampling distribution of the test statistic when the null hypothesis is true

Bootstrapping

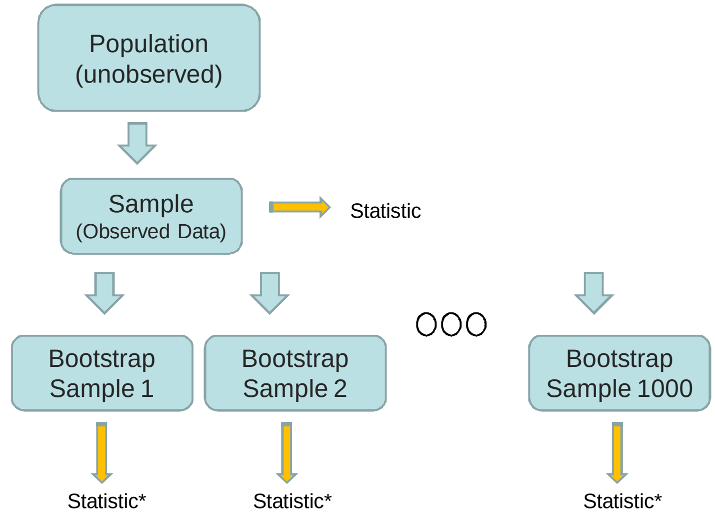
- Perform *in silico* replication of the study.
- If the sample size is large enough, the sample "stands in" for the population
- Your data are an independent sample from the population.
- Take independent samples from your data to mimic replications of your study.

Basic Strategy

- Pretend that the sample is the population.
- Sample randomly (and with replacement) from the sample to generate pseudosamples.
  - Each psuedosample uses same sample size.
  - Each observation equally likely to be sampled at each "draw" when making a pseudosample

Why sample with replacement?

- Theoretical reason:
- Practical reason:



Bootstrap Standard Errors

- From a large number of pseudosamples, we can estimate the sampling distribution of a wide variety of statistics
- The statistic is calculated on each pseudosample
- Across lots of pseudosamples (100s or 1000s), we have a distribution of the statistic. This can be taken as an estimate of the sampling distribution of the statistic.
- The SE of a statistic is just the SD of its sampling distribution.
  – Thus the SD of statistics* estimates SE(statistic)

Inference with Bootstrapped SE

- Providing that we know the statistic is approximately Normally distributed

$$100(1-\alpha)\% \text{ confidence interval is } (\theta_L, \theta_U)$$

$$\theta_L = \hat{\theta} - z_{1-\alpha/2}\ \hat{se}(\hat{\theta})$$

$$\theta_U = \hat{\theta} + z_{1-\alpha/2}\ \hat{se}(\hat{\theta})$$

Hypothesis tests based on

$$Z = \frac{\hat{\theta} - \theta_0}{\hat{se}(\hat{\theta})} \sim N(0,1)$$

Ex: SE of Sample Median

- Bootstrapped estimates of the standard error for sample median

```
                      Data          Median
Original sample:  {1, 5, 8, 3, 7}    5

Bootstrap 1     :  {1, 7, 1, 3, 7}    3
Bootstrap 2     :  {7, 3, 8, 8, 3}    7
Bootstrap 3     :  {7, 3, 8, 8, 3}    7
Bootstrap 4     :  {3, 5, 5, 1, 5}    5
Bootstrap 5     :  {1, 1, 5, 1, 8}    1
     etc.
```

1000 Bootstrapped Samples

- Descriptive statistics for the sample medians from 1000 bootstrapped samples

n	1000
Mean	4.964
Standard Deviation	1.914
Median	5
Minimum, Maximum	1, 8
25th, 75th %ile	3, 7

Inference for Sample Median

- From the above bootstrapped samples:
- Estimated SE sample median is 1.914
  - The standard deviation of the sample medians across the 1000 pseudosamples
- A 95% asymptotic (with n=5?) confidence interval (using the 0.975 quantile of the standard normal distribution) is thus

  5 +/- 1.96 * 1.914 = 1.25, 8.75

Bootstrapped Standard Errors

- There are some instances when bootstrapping does not work
  - For instance, no sample of continuous data is ever adequate to bootstrap the sampling distribution of the minimum or maximum
    - We can never mimic the chance to have observed more extreme values than were in our sample
- But as a general rule, bootstrapping behaves remarkably well for measures of location and variability

STATA: SE for Sample Median

- Stata can find bootstrapped standard errors and confidence intervals
- The **`summarize`** command computes the median; the manual tells me it is saved as r(p50)
- **`bootstrap r(p50), reps(1000): summarize bilirubin, detail`**
  - STATA takes a few seconds to run this command, as opposed to "instantaneously"

STATA: SE for Sample Median

```
Bootstrap results                        Number of obs      =       312
                                         Replications       =      1000

  command:  summarize bilirubin, detail
    _bs_1:  r(p50)

------------------------------------------------------------------------------
      |   Observed   Bootstrap                         Normal-based
      |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------+-----------------------------------------------------------------------
_bs_1 |       1.35   .1371632     9.84   0.000     1.081165    1.618835
------------------------------------------------------------------------------
```

Bootstrap Confidence Intervals

- "Normal confidence intervals" uses the bootstrap to estimate the standard error of the statistic, than uses +/- 1.96 standard errors for the confidence interval
- "Percentile confidence intervals" use the 2.5^{th} and 97.5^{th} percentiles of the bootstrap distribution as the confidence interval
- "BC" method tries to adjust the percentile limits for asymmetry of the distribution
  - BC=Bias-corrected
- When they all agree they are all likely to be reasonable reliable
  - STATA post-estimation command **`estat`**

STATA: SE for Sample Median

```
estat bootstrap, all

Bootstrap results                        Number of obs      =       312
                                         Replications       =      1000

      command:  summarize bilirubin, detail
        _bs_1:  r(p50)

--------------------------------------------------------------------------
      |Observed               Bootstrap
      |   Coef.       Bias    Std. Err.  [95% Conf. Interval]
------+-------------------------------------------------------------------
_bs_1 |    1.35     .03725   .13716324    1.081165   1.618835   (N)
      |                                        1.2        1.8   (P)
      |                                        1.2        1.8  (BC)
--------------------------------------------------------------------------
(N)    normal confidence interval
(P)    percentile confidence interval
(BC)   bias-corrected confidence interval
```