

## Activity #18b: Confidence Intervals - Practice & the Bootstrap Method

Scenario: When a shipment of coal is traded, the value of the shipment is determined by a number of its properties. An important example is the so-called gross calorific value, which characterizes the heat content and is a numerical value in megajoules per kilogram (MJ/kg). The International Organization of Standardization (ISO) issues standard procedures for the determination of these properties. For the gross calorific value, there is a method known as ISO 1928. When the procedure is carried out properly, resulting measurement errors are known to be approximately normal with a standard deviation of about 0.1 MJ/kg. Labs that operate according to standard procedures receive ISO certificates. In the following table, 23 ISO 1928 measurements are given for a shipment of Osterfeld coal coded 262DE27:

23.870	23.730	23.712	23.760	23.640	23.850	23.840	23.860	23.940	23.830	23.877	23.700
23.796	23.727	23.778	23.740	23.890	23.780	23.678	23.771	23.860	23.690	23.800	

Source: A.M.H. van der Veen and A.J.M. Broos. Interlaboratory study programme "ILS coal characterization" – reported data. Technical report, NMI Van Swinden Laboratorium B.V., The Netherlands, 1996.

If you were to calculate the average of these 23 values, you would find the average to be 23.788.

1) Calculate and interpret a 95% confidence interval for  $\mu$ :

2) Calculate a 99% confidence interval for  $\mu$ . Why is this interval wider than the 95% confidence interval?

Scenario: Suppose we receive another shipment of 258GB41 coal from Daw Mill, where there were actually some doubts whether the stated accuracy of the ISO 1928 method was attained. We, therefore, doubt whether  $\sigma = 0.1$  MJ/kg. We measure 22 ISO 1928 measurements for the shipment and find:

30.990	31.030	31.060	30.921	30.920	30.990	31.024	30.929	31.050	30.991	31.208
31.330	30.830	30.810	31.060	30.800	31.091	31.170	31.026	31.020	30.880	31.125

Source: A.M.H. van der Veen and A.J.M. Broos. Interlaboratory study programme "ILS coal characterization" – reported data. Technical report, NMI Van Swinden Laboratorium B.V., The Netherlands, 1996.

From these 22 values, you could calculate the average to be 31.012 and the standard deviation to be 0.1294.

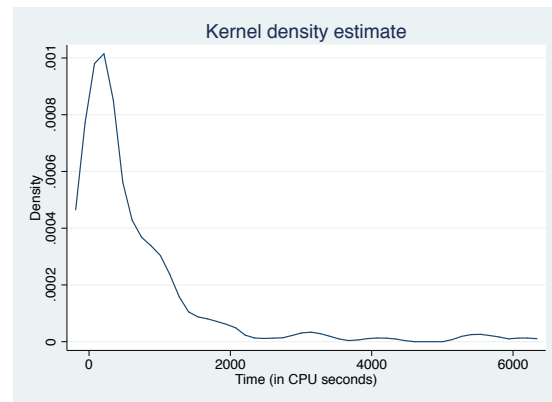
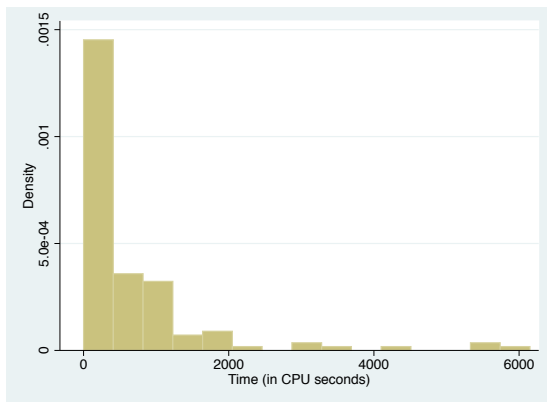
3) Calculate and interpret a 95% confidence interval for  $\mu$ . How does the width of this interval compare to the width of the interval you calculated in question #1?

- 4) Recall that these methods only apply if the normal distribution is a good model for our sample averages. So far, you've calculated confidence intervals for averages based on sample sizes of 23 and 22. Did the normal distribution apply to these scenarios? How do you know?

It is not uncommon that these methods to construct confidence intervals are used even when the normal distribution is *not* a good model for the data. In some cases, this isn't a big problem. For large datasets, the central limit ensures that this method provides confidence intervals with approximately correct confidence levels.

If you doubt the normality of data and you do *not* have a large sample, usually the best thing to do is use the *bootstrap method*.

Scenario: The following graphs display the observed lengths of time (in CPU seconds) between successive failures that occur during the execution of a certain computer command. In other words, these are times between successive computer crashes:



From these 136 observations, we can calculate a mean of 649.1397 and a standard deviation of 1036.2.

- 5) Let's ignore the fact that the distribution is not normal (or, let's assume with  $n=136$ , that the central limit theorem will allow us to assume the sample averages follow an approximately normal distribution). We could calculate a 90% confidence interval for  $\mu$  as:

$$\bar{X} \pm Z_{.05} \left( \frac{\sigma}{\sqrt{n}} \right) = 649.1397 \pm 1.645 \left( \frac{1036.2}{\sqrt{136}} \right) = (502.98, 795.30)$$

The width of this confidence interval is 292.32 and it is centered at 649.1397.

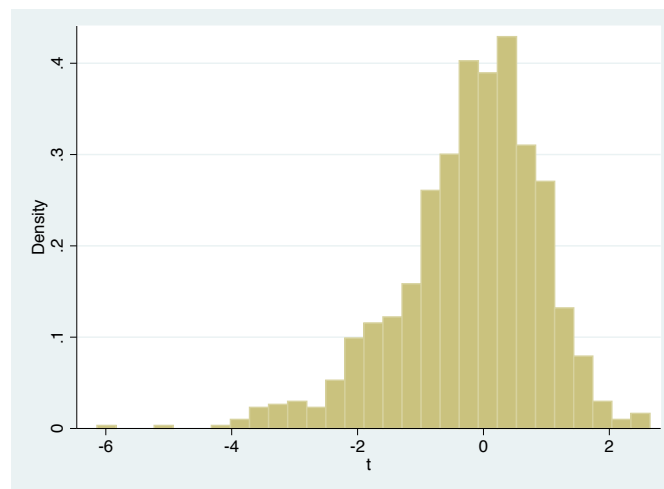
6) Rather than ignoring the non-normality of the distribution (or relying on the central limit theorem), we could construct a 90% *bootstrap-t confidence interval*. Here are the steps we would follow:

1. Let's treat our sample of 136 observations as if it is the entire population of interest. Let's randomly sample  $n=136$  observations **with replacement** from our population of size  $n=136$ . Since we're sampling with replacement, the same observation could be chosen more than once in our *bootstrap sample*.
2. In fact, let's have a computer repeat this process 1,000 times. So, as a result, we'll have 1,000 random bootstrap samples of size  $n=136$ .
3. For each bootstrap sample, let's calculate the sample mean.
4. Let's also calculate a t-statistic for each bootstrap sample:  $t^* = \frac{\bar{X}^* - \bar{X}}{s^*/\sqrt{n}}$   
where \* represents something we've calculated from our bootstrap sample.
5. Now that we have 1,000  $t^*$  statistics, we can find the 5th and 95th percentiles by
  - Ordering our  $t^*$  statistics from smallest (1st) to largest (1000th)
  - Our 5th percentile is the 50th observation
  - Our 95th percentile is the 951st observation
6. We use the 5th and 95th percentiles as the critical values in our confidence interval formula (instead of the t or z statistic)

I used a computer program called *Stata* to run this bootstrap procedure. I entered in the following syntax:

```
> bootstrap t=(r(mean)-649.1397)/(r(sd)/(136^.5)), rep(1000): summarize Time
```

This syntax caused Stata to randomly select 136 observations (with replacement) 1,000 times and record the t-statistic for each sample. I was left with 1,000 t-statistics, which are displayed in the following histogram:



I then ordered the t-statistics from smallest-to-largest and found the 50th and 951st values to be -2.332 and 1.348. With these t-values, I can then calculate:

$$\bar{X} \pm t \left( \frac{\sigma}{\sqrt{n}} \right) = 649.1397 \pm t^* \left( \frac{1036.2}{\sqrt{136}} \right)$$

$$649.1397 - 1.348 \left( \frac{1036.2}{\sqrt{136}} \right) = 529.37$$

$$649.1397 - (-2.332) \left( \frac{1036.2}{\sqrt{136}} \right) = 856.35$$

So, our bootstrap confidence interval is (529.37, 856.35).

The following table summarizes the confidence intervals we have calculated for this computer crash data:

Method	Normality Assumption (standard confidence interval formula)	Bootstrap t Method
Lower bound	502.98	529.37
Upper bound	795.30	856.35
Width of interval	292.32	326.98
Center of interval	649.14	692.86

Comparing the confidence interval methods, we can see that the bootstrap method produced a wider interval. It also produced an interval that is not centered around the sample mean (649.1397), but skewed to the right side. This is one of the features of the bootstrap: if the distribution from which the data originate is skewed, this is reflected in the bootstrap confidence interval. In some sense, then, the bootstrap adapts to the shape of the distribution from which it was calculated, and in this way, it leads to more accurate confidence statements than we get from using the method that assumes normality.

7) Looking at the bootstrap t-values again, I see the 25th value is -2.7865 and the 976th value is 1.65218. Using these values, construct a bootstrap confidence interval for  $\mu$ . What is the corresponding confidence level for this interval? How does it compare to an interval (of the same confidence) using the formulas that assume normality?

8) In January of 2002, the Euro was introduced. Soon after, stories started to circulate that some of the Euro coins would not be fair coins, because the "national side" of some coins would be too heavy or too light (see, for example, the New Scientist of January 4, 2002. A Belgian 1 Euro was tossed 250 times which resulted in 140 heads and 110 tails. Construct a 95% confidence interval for the probability of getting heads with this coin.