Back to statistics - remember the pebble mass distribution?

The probability of occurrence of specific values in a sample often takes on that bell-shaped Gaussian-like curve, as illustrated by the pebble mass data.
The Gaussian (normal) distribution of pebble masses looked a bit different from the probability distribution we derived directly from the sample, but it provided a close approximation of the sample probabilities.

You are in a good position now to answer someone’s questions about what the probability is that a pebble mass, strike, dip, porosity, etc. is likely to fall in a certain range of values. These probabilities can all be easily derived either directly from the statistics of your sample or inferred more generally from the corresponding Gaussian distribution as just discussed.
The pebble mass data represents just one of a nearly infinite number of possible samples that could be drawn from the parent population of pebble masses.

We obtained one estimate of the population mean and this estimate is almost certainly incorrect. Think of that big beach.

What might additional pebble mass samples look like?

These samples were drawn at random from a parent population having mean 350.18 and variance of 2273 (i.e. standard deviation of ~48).
Note that the sample means differ from the population mean.

The distribution of 35 means calculated from 35 samples drawn at random from a parent population with assumed mean of 350.18 and variance of 2273 (s = 47.676).

The mean of the above distribution of means is 350.45.
Their variance is 21.51 (i.e. standard deviation of 4.64).
The statistics of the distribution of means tells us something different from the statistics of the individual samples.

The statistics of the distribution of means gives us information about the variability we can anticipate in the mean value of 100 (or any number for that matter) - specimen samples.

Just as with the individual pebble masses observed in the sample, probabilities can also be associated with the possibility of drawing a sample with a certain mean and standard deviation.

This is how it works -

You go out to your beach and take a bucket full of pebbles in one area and then go to another part of the beach and collect another bucket full of pebbles.

You have two samples, and each has their own mean and standard deviation.

You ask the question - Is the mean determined for the one sample different from that determined for the second sample?

To answer this question you use probabilities determined from the distribution of means, NOT from those of an individual sample.

The means of the samples may differ by only 20 grams. If you look at the range of individual masses which is around 225 grams, you might conclude that these two samples are not really different.
However, you are dealing with means derived from individual samples each consisting of 100 specimens. The distribution of means is VERY different from the distribution of specimens. The range of possible means is much much smaller.

Thus, when trying to estimate the possibility that two means come from the same parent population, you need to examine probabilities based on the standard deviation of the means and NOT those of the specimens.

In the class example just presented we derived the mean and standard deviation of 35 samples drawn at random from a parent population having a standard deviation of 47.7.

Recall that the standard deviation of means was only 4.64 and that this is just about 1/10th the standard deviation of the sample.
This is the standard deviation of the sample means from the estimate of the true mean. This standard deviation is referred to as the standard error.

The standard error, $s_e$, is estimated from the standard deviation of the sample as:

$$s_e = \frac{s}{\sqrt{N}}$$

What is a significant difference?
To estimate the likelihood that a sample having a specific calculated mean and standard deviation comes from a parent population with given mean and standard deviation, one has to define some limiting probabilities.

There is some probability, for example, that you could draw a sample whose mean might be 10 standard deviations from the parent mean. It’s really small, but still possible.

What chance of being wrong will you accept?
This decision about how different the mean has to be in order to be considered statistically different is actually somewhat arbitrary.

In most cases we are willing to accept a one in 20 chance of being wrong or a one in 100 chance of being wrong.

The chance we are willing to take is related to the “confidence limit” we choose.
The confidence limits used most often are 95% or 99%. The 95% confidence limit gives us a one in 20 chance of being wrong. The confidence limit of 99% gives us a 1 in 100 chance of being wrong.

The risk that we take is referred to as the alpha level.

If our confidence limit is 95% our alpha level is 5% or 0.05.

If our confidence limit is 99% - α is 1% or 0.01

Whatever your bias may be - whatever your desired result - you can’t go wrong in your presentation if you clearly state the confidence limit you use.

<table>
<thead>
<tr>
<th>Number of standard deviations</th>
<th>Area</th>
<th>Number of standard deviations</th>
<th>Area</th>
<th>Number of standard deviations</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.000</td>
<td>1.0</td>
<td>0.750</td>
<td>2.1</td>
<td>0.909</td>
</tr>
<tr>
<td>0.1</td>
<td>0.080</td>
<td>1.1</td>
<td>0.730</td>
<td>2.2</td>
<td>0.900</td>
</tr>
<tr>
<td>0.2</td>
<td>0.159</td>
<td>1.2</td>
<td>0.710</td>
<td>2.3</td>
<td>0.890</td>
</tr>
<tr>
<td>0.3</td>
<td>0.236</td>
<td>1.3</td>
<td>0.690</td>
<td>2.4</td>
<td>0.879</td>
</tr>
<tr>
<td>0.4</td>
<td>0.311</td>
<td>1.4</td>
<td>0.670</td>
<td>2.5</td>
<td>0.867</td>
</tr>
<tr>
<td>0.5</td>
<td>0.383</td>
<td>1.5</td>
<td>0.650</td>
<td>2.6</td>
<td>0.854</td>
</tr>
<tr>
<td>0.6</td>
<td>0.451</td>
<td>1.6</td>
<td>0.630</td>
<td>2.7</td>
<td>0.840</td>
</tr>
<tr>
<td>0.7</td>
<td>0.516</td>
<td>1.7</td>
<td>0.610</td>
<td>2.8</td>
<td>0.825</td>
</tr>
<tr>
<td>0.8</td>
<td>0.576</td>
<td>1.8</td>
<td>0.590</td>
<td>2.9</td>
<td>0.810</td>
</tr>
<tr>
<td>0.9</td>
<td>0.632</td>
<td>1.9</td>
<td>0.570</td>
<td>3.0</td>
<td>0.795</td>
</tr>
<tr>
<td>1.0</td>
<td>0.683</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the above table of probabilities (areas under the normal distribution curve), you can see that the 95% confidence limit extends out to 1.96 standard deviations from the mean.
The standard deviation to use when you are comparing means (are they the same or different?) is the standard error, $s_e$.

Assuming that our standard error is 4.8 grams, then $1.96s_e$ corresponds to ± 9.41 grams.

Notice that the 5% probability of being wrong is equally divided into 2.5% of the area greater than 9.41 grams from the mean and less than 9.41 grams from the mean.

You probably remember the discussions of one- and two-tailed tests.

The 95% probability is a two-tailed probability.
So if your interest is only to make the general statement that a particular mean lies outside $\pm 1.96$ standard deviations from the assumed population mean, your test is a two-tailed test.

and your $\alpha$ is 0.05

If you wish to be more specific in your conclusion and say that the estimate is significantly greater or less than the population mean, then your test is a one-tailed test.

then your $\alpha$ is 0.025

i.e. the probability of error in your one-tailed test is 2.5% rather than 5.

$\alpha$ is 0.025 rather than 0.05
Using our example dataset, we have assumed that the parent population has a mean of 350.18 grams, thus all means greater than 359.6 grams or less than 340.8 grams are considered to come from a different parent population - at the 95% confidence level.

<table>
<thead>
<tr>
<th>Mean Variance Mean Standard</th>
<th>Variance</th>
<th>Standard</th>
<th>95% C. L.</th>
</tr>
</thead>
<tbody>
<tr>
<td>348.84 2827.5 53.17 338.29 - 359.39</td>
<td>350.6 2192.59 46.82 341.31 - 359.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>356.43 2124.63 46.09 347.28 - 365.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>354.5 1977.63 44.47 345.68 - 363.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>348.42 2611.3 51.1 338.28 - 358.56</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that the samples we drew at random from the parent population have means which lie inside this range and are therefore not statistically different from the parent population.

It is worth noting that - we could very easily have obtained a different sample having a different mean and standard deviation. Remember that we designed our statistical test assuming that the SAMPLE mean and standard deviation correspond to those of the population.

<table>
<thead>
<tr>
<th>Mean Variance Mean Standard</th>
<th>Variance</th>
<th>Standard</th>
<th>95% C. L.</th>
</tr>
</thead>
<tbody>
<tr>
<td>348.84 2827.5 53.17 338.29 - 359.39</td>
<td>350.6 2192.59 46.82 341.31 - 359.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>356.43 2124.63 46.09 347.28 - 365.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>354.5 1977.63 44.47 345.68 - 363.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>348.42 2611.3 51.1 338.28 - 358.56</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This would give us different confidence limits and slightly different answers. Even so, the method provides a fairly objective quantitative basis for assessing statistical differences between samples.
Remember the t-test?

Tests for significance can be improved if we account for the fact that estimates of the mean derived from small samples are inherently sloppy estimates.

The t-test acknowledges this sloppiness and compensates for it by making the criterion for significant-difference more stringent when the sample size is smaller.

The z-test and t-test yield similar results for relatively large samples - larger than 100 or so.

The 95% confidence limit for example, diverges considerably from 1.96 s for smaller sample size. The effect of sample size (N) is expressed in terms of degrees of freedom which is N-1.