Statistics continued

Tech Talent South (original author)
Ai Yukino (PDF rewrite)

Confidence Intervals

A confidence interval is a range of values where we believe the true value lies.

We commonly use a 95% or 99% confidence interval.

We calculate this using the mean of our data and the standard deviation. We can use this to find a range that the true mean most likely lies in.

Still, it may not lie in this range. What we are saying is 95% of experiments will include the true mean, and 5% won't. This means there is a 5% chance our confidence interval does not include the true meann.

Confidence Intervals

There are many ways to calculate a confidence interval, but we are going to create ours first from a z-score, so we'll need to calculate one. A z score tells you how many standard deviations from the mean your score is.

To get this score, we subtract the mean from the test score and divide by the standard deviation.

$$z = \frac{(x - \mu)}{\sigma}$$

If our test score is 100, our mean is 80, and the standard deviation is 30, the z score would be:

$$z = \frac{100 - 80}{30}$$
$$z = \frac{20}{30}$$
$$z = 0.67$$

Confidence Intervals

When multiple samples are involved, we will need to account for the stand error.

$$z = \frac{x - \mu}{\sigma / n}$$

This will tell you how many standard errors there are between the sample mean and the population mean.

Confidence Intervals

Example problem: In general, the mean height of women is 65" with a standard deviation of 3.5". What is the z score that represents finding a random sample of 50 women with a mean height of 70", assuming the heights are normally distributed?

$$z = \frac{x - \mu}{\sigma}$$
$$z = \frac{70 - 65}{{}^{3.5}\!/\!{\sqrt{50}}} = \frac{5}{0.495} = 10.1$$

What does this z-score mean then? What is it a measurement of?

Confidence Intervals

Recall that our z-scores reflect a measurement in standard deviations from the mean. Typically, when making confidence intervals, we are aiming for a 95% confidence interval, which represents a z-score of around 1.96. This represents the score you would get with a conventional significance level of $\alpha = 0.05$, and the resultant interval

Confidence Intervals

Great! We have a z score. Let's get back to calculating the confidence interval.

The formula is:

$$X \pm z \cdot ({}^{\sigma}/_{\sqrt{n}})$$

X - the mean around which we will center our confidence interval

z - the z score

s - the standard deviation

n - the number of observations

Confidence Intervals

Let's try an example. We have an apple orchard, and want to estimate the true mean of their size.

We get a mean weight of 5.7 oz with a standard deviation of 2.3.

Using this data, try to calculate the 95% confidence interval yourself.

(Remember, for a 95% confidence interval, the z score is 1.96.)

Confidence Intervals

$$x = 5.7$$
$$z = 1.96 \text{ oz}$$
$$\sigma = 2.3$$
$$n = 50$$

$$5.7 \pm 1.96 \cdot ({}^{2.3}/\sqrt{50})$$
$$5.7 \pm 1.96 \cdot 0.325$$
$$5.7 \pm 0.638$$

The true mean of the weight of the apples is likely to lie between 5.062 and 6.338oz.

Confidence Intervals

Again, this only means that we expect 95% of other samples to fall into this range. 5% will still outside the range.

Confidence Intervals

Z-intervals aren't the only option out there! In our previous example, we didn't use the population's data, just what we had from our sample.

A t-interval can be more useful when there is an unknown population mean. We rely on our sample standard deviation to calculate the margin of error.

Confidence Intervals

The formula for a t interval is as follows:

$$x \pm t \cdot ({}^{s}/_{\sqrt{n}})$$

where t is a critical value from the t-distribution, s is the sample standard deviation, and n is the sample size.

Remember: We only use a t-interval when we have a population standard deviation that is UNKNOWN and the original population is normal for the sample, OR when sample size is greater than or equal to 30.

Confidence Intervals

But what is a critical value?

To get the value of t, we need to first calculate degrees of freedom. This is equal to our sample size minus 1.

$$df = n - 1$$

We then use this number in a t-table.

Confidence Intervals

Let's try a similar example with a t interval.

We get a sample of 20 random apples.

We get a mean weight of 5.2 oz with a standard deviation of 3.5.

Using this data, try to calculate the 95% confidence interval yourself.

Confidence Intervals

$$x = 5.2$$
$$n = 20$$
$$s = 3.5$$
$$df = 20 - 1$$
$$t = 2.09302$$

$$5.2 \pm 2.09302 \cdot ({}^{3.5}/\sqrt{20})$$
$$5.2 \pm 0.7826237921$$

Our confidence interval is between 4.42 and 5.98