Example: One sample t-test (one-tailed)

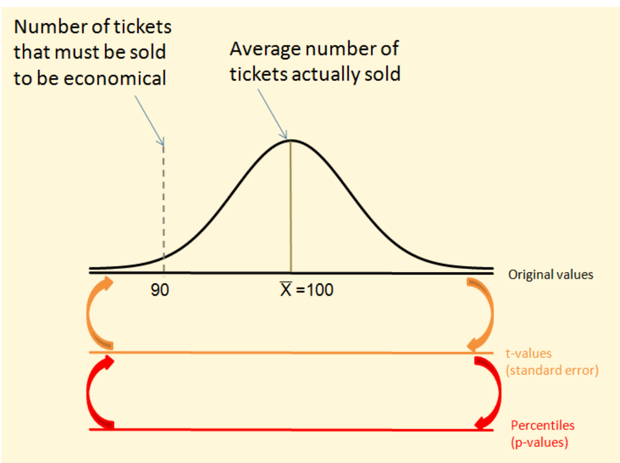
RenR 480/711

Thursday, November 9, 2017

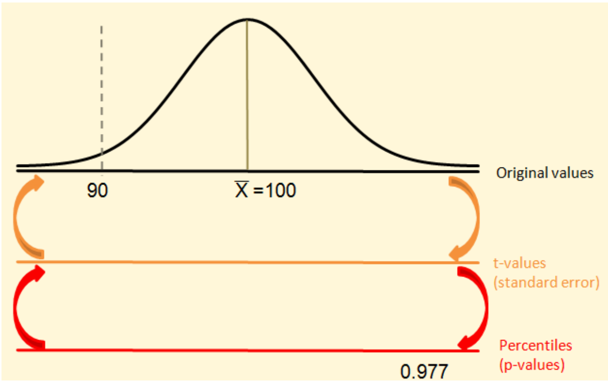
Many airlines sell more tickets than are available for a given journey in the anticipation of people not showing up on the day. Over 50 years in the industry, one airline determined that a popular journey needs to have at least 90 people on a flight to be economical. Over the years, they find a mean of 100 tickets sold and a standard error of 10. They need to be 97.7% confident that, most of the time, they will not have less than 90 tickets sold. How many tickets should they aim to sell?

First, recognize what type of test this is: There is one sample (one mean) of concern. The hypothesis is also directional: they need to be confident that it is more than 90 tickets sold (i.e., you only care about one side of the distribution; not the other). Therefore, you can assume one-tailed test. Thus, it is a one sample, one tailed t-test.

Now let's think in terms of the sampling distribution – it might help to draw this out:

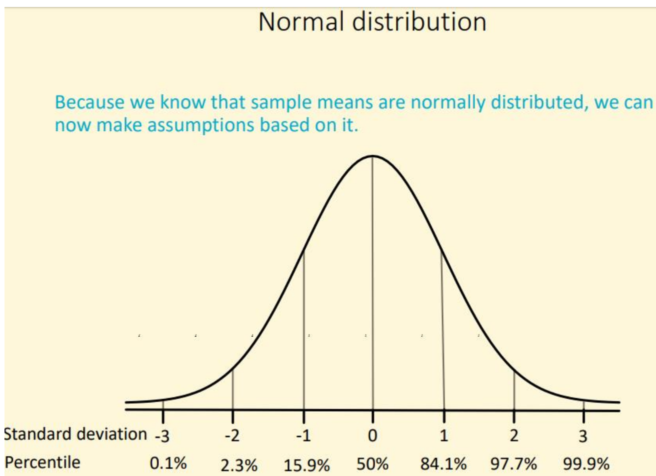


Now we have to figure out what our percentiles are. We want to be sure that we never have too little, so we know it must be greater than 90, therefore, we are looking at the percentiles on the right side of the distribution (we don't need to split up the 2.5% into both tails – we can keep it all on the right side). Therefore, being 97.7% sure might look like this:



Now we need to convert from the percentile scale up to the critical t-value (and then eventually to the original value that we need to answer the question).

To do so, we need to think about the distribution a little bit. We know from class that a normal distribution looks like this:

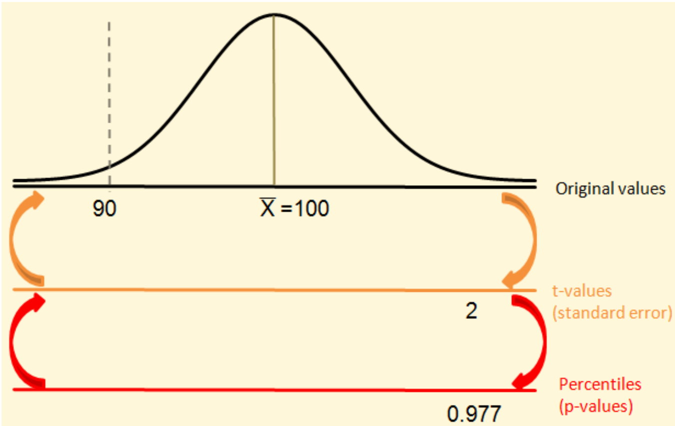


You can see that, for a normal distribution, 97.7% of the data falls to the left of +2 standard deviations from the mean. This is good to "just know", but we can verify that in R with qnorm(0.977). (q*norm* refers to normal distribution.) Due to the bell curve's symmetry, a percentile of 2.3 is -2 standard deviations from the mean.

On a normal distribution, a percentile of 97.5 is associated with 1.96. (A percentile of 2.5 is associated with -1.96.) This value is also "good to know", but can be verified in R with qnorm(0.975).

However, we are using a t-distribution, also known as a sampling distribution. The shape of the sampling distribution changes with degrees of freedom. In our case, sample size is relatively large (50), so a degree of freedom of 49 (n-1) is probably getting close to normal, and is likely close to about 2. We can verify that in R with the qt(0.977, 49). (q*t* refers to the t-distribution.) In R, we get about 2.04.

Although we can use qt in R, we can also make some approximations knowing 97.7% is close to 2 with a large sample size, so we can fill that in on our scale:



To go up to the “original values” scale, we need to remember the formula to convert between the t-values to the original values:

$$value = t\text{-}value * SEM + \bar{x}$$

In our case, therefore:

t-value is ~2 (The “~” symbol means “approximately”)
SEM is known to be 10
Mean is known to be 100

Thus:

Value = 2 * 10 + 100
Value = 20 + 100
Value = 120

We now know that the airline company needs to sell 120 tickets to be 97.7% sure that a flight will not have less than 90 people sitting on that airplane.

Let’s say the airplane only carries 110 people, though. The airline has now over-booked the flight to avoid the risk of having too few people on the flight for the risk of inconveniencing a few people. (That helps explain why sometimes you might see people bumped from a flight.)