Normal Distributions, z Values, and Statistics

Calculus II

College of the Atlantic

Winter 2023

The Normal Distribution

1. The normal distribution, with mean μ and standard deviation σ is:

$$p(x) \;=\; \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \; . \tag{1}$$

2. As with all probability distribution functions, to get a probability, it must be integrated.

3. The normal distribution does not have an anti-derivative that can be expressed in terms of elementary functions. Instead, it is evaluated numerically.

4. The normal distribution can be converted to a *standard normal* distribution:

$$p(z) \;=\; \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \; . \tag{2}$$

   A standard normal distribution is a normal distribution that has a mean of zero and a standard deviation of one.

5. Given x that is distributed normally with mean μ and standard deviation σ, transform x to z via:

$$z = \frac{x-\mu}{\sigma} \; . \tag{3}$$

   Then z is distributed according to a standard normal distribution, and the cumulative density function can be looked up in a z table.

The Central Limit Theorem and Its Implications

1. The sum of a bunch of random variables[1] is normally distributed, even if the individual random variables are *not* normally distributed. This statement is exact only in the limit that the sum has an infinitely large number of variables. But in practice the distribution is very well approximated by a normal distribution for a relatively small number of variables. (The t distribution is an approximation of the normal distribution for smaller samples.)

2. Suppose you take a bunch of measurements and calculate the mean. This is known as the *sample mean*, because it is the mean of a sample.

[1] There's a bit more mathematical fine print. The sum of the third moments of the variables must be finite, and the cube root of this sum, divided by the square root of the mean squared error, must tend to zero.

3. The sample mean is itself a random variable. How is this random variable distributed? I.e., if you calculated many different sample means from the same population, how much variance would you expect?

4. The central limit theorem says that the sample mean is normally distributed, at least in the limit that you take a reasonably large number of measurements when you form your sample.

5. What is the standard deviation of the sample mean? It turns out that this is given by $\sigma/\sqrt{n}$, where σ is the standard deviation of the population from which you are sampling. Note again that the distribution of the population need not be normal.

The Structure of Statistics, or at least Hypothesis Testing

1. Choose a null hypothesis. This is usually a statement about the distribution of a variable. We then test to see if the evidence suggests that the null hypothesis can be rejected.

2. Take your data and calculate a statistic. A statistic is just a fancy word for a function of your data. Usually this is the sample mean, but it also could be the sample standard deviation, or something more exotic like χ^2.

3. Determine the distribution of your statistic, assuming that the null hypothesis is true.

4. Reject (or not) the null hypothesis depending on how unlikely the observed statistic value is, assuming the null hypothesis is true.

5. The p-value is how likely it is that the null hypothesis generates a statistic value equal to or more extreme than the observed value.

Note: The wikipedia entries "Z-tests" and "hypothesis testing" are quite good.