2. The normal and t-distributions

The normal distribution is simply a distribution with a certain shape. It is "normal" because many things have this same shape. The normal distribution is the “bell-shaped distribution” that describes how so many natural, machine-made, or human performance outcomes are distributed. If you ever took a class when you were "graded on a bell curve", the instructor was fitting the class' grades into a normal distribution—not a bad practice if the class is large and the tests are objective, since human performance in such situations is normally distributed. This chapter will discuss the normal distribution and then move onto a common sampling distribution, the t-distribution. The t-distribution can be formed by taking many samples (strictly, all possible samples) of the same size from a normal population. For each sample, the same statistic, called the t-statistic, which we will learn more about later, is calculated. The relative frequency distribution of these t-statistics is the t-distribution. It turns out that t-statistics can be computed a number of different ways on samples drawn in a number of different situations and still have the same relative frequency distribution. This makes the t-distribution useful for making many different inferences, so it is one of the most important links between samples and populations used by statisticians. In between discussing the normal and t-distributions, we will discuss the central limit theorem. The t-distribution and the central limit theorem give us knowledge about the relationship between sample means and population means that allows us to make inferences about the population mean.

The way the t-distribution is used to make inferences about populations from samples is the model for many of the inferences that statisticians make. Since you will be learning to make inferences like a statistician, try to understand the general model of inference making as well as the specific cases presented. Briefly, the general model of inference-making is to use statisticians' knowledge of a sampling distribution like the t-distribution as a guide to the probable limits of where the sample lies relative to the population. Remember that the sample you are using to make an inference about the population is only one of many possible samples from the population. The samples will vary, some being highly representative of the population, most being fairly representative, and a few not being very representative at all. By assuming that the sample is at least fairly representative of the population, the sampling distribution can be used as a link between the sample and the population so you can make an inference about some characteristic of the population.

These ideas will be developed more later on. The immediate goal of this chapter is to introduce you to the normal distribution, the central limit theorem, and the t-distribution.

Normal things

Normal distributions are bell-shaped and symmetric. The mean, median, and mode are equal. Most of the members of a normally distributed population have values close to the mean—in a normal population 96 per cent of the members (much better than Chebyshev's 75 per cent), are within 2 σ of the mean.

2. The normal and t-distributions

Statisticians have found that many things are normally distributed. In nature, the weights, lengths, and thicknesses of all sorts of plants and animals are normally distributed. In manufacturing, the diameter, weight, strength, and many other characteristics of man- or machine-made items are normally distributed. In human performance, scores on objective tests, the outcomes of many athletic exercises, and college student grade point averages are normally distributed. The normal distribution really is a normal occurrence.

If you are a skeptic, you are wondering how can GPAs and the exact diameter of holes drilled by some machine have the same distribution—they are not even measured with the same units. In order to see that so many things have the same normal shape, all must be measured in the same units (or have the units eliminated)—they must all be "standardized." Statisticians standardize many measures by using the STANDARD deviation. All normal distributions have the same shape because they all have the same relative frequency distribution *when the values for their members are measured in standard deviations above or below the mean.*

Using the United States customary system of measurement, if the weight of pet cats is normally distributed with a mean of 10.8 pounds and a standard deviation of 2.3 pounds and the daily sales at The First Brew Expresso Cafe are normally distributed with μ=\$341.46 and σ=\$53.21, then the same proportion of pet cats weigh between 8.5 pounds (μ-1σ) and 10.8 pounds (μ) as the proportion of daily First Brew sales which lie between $\mu - 1\sigma$ (\$288.25) and μ (\$341.46). Any normally distributed population will have the same proportion of its members between the mean and one standard deviation below the mean. Converting the values of the members of a normal population so that each is now expressed in terms of standard deviations from the mean makes the populations all the same. This process is known as "standardization" and it makes all normal populations have the same location and shape.

This standardization process is accomplished by computing a "z-score" for every member of the normal population. The z-score is found by:

$$z = (x - \mu)/\sigma$$

This converts the original value, in its original units, into a standardized value in units of "standard deviations from the mean." Look at the formula. The numerator is simply the difference between the value of this member of the population, x, and the mean of the population μ. It can be measured in centimeters, or points, or whatever. The denominator is the standard deviation of the population, σ, and it is also measured in centimeters, or points, or whatever. If the numerator is 15cm and the standard deviation is 10cm, then the z will be 1.5. This particular member of the population, one with a diameter 15cm greater than the mean diameter of the population, has a z-value of 1.5 because its value is 1.5 standard deviations greater than the mean. Because the mean of the x's is μ, the mean of the z-scores is zero.

We could convert the value of every member of *any* normal population into a z-score. If we did that for any normal population and arranged those z-scores into a relative frequency distribution, they would all be the same. Each and every one of those standardized normal distributions would have a mean of zero and the same shape. There are many tables which show what proportion of any normal population will have a z-score less than a certain value. Because the standard normal distribution is symmetric with a mean of zero, the same proportion of the population that is less than some positive z is also greater than the same negative z. Some values from a "standard normal" table appear below:

Proportion below	.75	.90	.95	.975	.99	.995

z-score	0.674	1.282	1.645	1.960	2.326	2.576

John McGrath has asked Kevin Schmidt "How much does a pair of size 11 mens dress socks usually weigh?" Kevin asks the people in quality control what they know about the weight of these socks and is told that the mean weight is 4.25 ounces with a standard deviation of .021 ounces. Kevin decides that Mr. McGrath probably wants more than the mean weight, and decides to give his boss the range of weights within which 95% of size 11 men's dress socks falls. Kevin sees that leaving 2.5% (.025) in the left tail and 2.5% (.025) in the right tail will leave 95% (.95) in the middle. He assumes that sock weights are normally distributed, a reasonable assumption for a machine-made product, and consulting a standard normal table, sees that .975 of the members of any normal population have a z-score less than 1.96 and that .975 have a z-score greater than -1.96, so .95 have a z-score between ±1.96..

Now that he knows that 95% of the socks will have a weight with a z-score between ±1.96, Kevin can translate those z's into ounces. By solving the equation for both +1.96 and -1.96, he will find the boundaries of the interval within which 95% of the weights of the socks fall:

$$1.96 = (x - 4.25)/.021$$

solving for x, Kevin finds that the upper limit is 4.29 ounces. He then solves for z=-1.96:

$$- 1.96 = (x - 4.25)/ .021$$

and finds that the lower limit is 4.21 ounces. He can now go to John McGrath and tell him: "95% of size 11 mens' dress socks weigh between 4.21 and 4.29 ounces."

The central limit theorem

If this was a statistics course for math majors, you would probably have to prove this theorem. Because this text is designed for business and other non-math students, you will only have to learn to understand what the theorem says and why it is important. To understand what it says, it helps to understand why it works. Here is an explanation of why it works.

The theorem is about sampling distributions and the relationship between the location and shape of a population and the location and shape of a sampling distribution generated from that population. Specifically, the central limit theorem explains the relationship between a population and the distribution of sample means found by taking all of the possible samples of a certain size from the original population, finding the mean of each sample, and arranging them into a distribution.

The sampling distribution of means is an easy concept. Assume that you have a population of x's. You take a sample of n of those x's and find the mean of that sample, giving you one $\bar{x}$. Then take another sample of the same size, n, and find its $\bar{x}$ ...Do this over and over until you have chosen all possible samples of size n. You will have generated a new population, a population of $\bar{x}$'s. Arrange this population into a distribution, and you have the sampling distribution of means. You could find the sampling distribution of medians, or variances, or some other sample statistic by collecting all of the possible samples of some size, n, finding the median, variance, or other statistic about each sample, and arranging them into a distribution.

The central limit theorem is about the sampling distribution of means. It links the sampling distribution of $\bar{x}$'s with the original distribution of x's. It tells us that:

2. The normal and t-distributions

(1) The mean of the sample means equals the mean of the original population, $\mu_{\bar{x}} = \mu$. This is what makes $\bar{x}$ an unbiased estimator of μ.

(2) The distribution of $\bar{x}$'s will be bell-shaped, no matter what the shape of the original distribution of x's.

This makes sense when you stop and think about it. It means that only a small portion of the samples have means that are far from the population mean. For a sample to have a mean that is far from μ_x , almost all of its members have to be from the right tail of the distribution of x's, or almost all have to be from the left tail. There are many more samples with most of their members from the middle of the distribution, or with some members from the right tail and some from the left tail, and all of those samples will have an $\bar{x}$ close to μ_x .

(3a) The larger the samples, the closer the sampling distribution will be to normal, and

(3b) if the distribution of x's is normal, so is the distribution of $\bar{x}$'s.

These come from the same basic reasoning as 2), but would require a formal proof since "normal distribution" is a mathematical concept. It is not too hard to see that larger samples will generate a "more-bell-shaped" distribution of sample means than smaller samples, and that is what makes 3a) work.

(4) The variance of the $\bar{x}$'s is equal to the variance of the x's divided by the sample size, or:

$$\sigma^2_{\bar{x}} = \sigma^2 / n$$

therefore the standard deviation of the sampling distribution is:

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}$$

While it is a difficult to see why this exact formula holds without going through a formal proof, the basic idea that larger samples yield sampling distributions with smaller standard deviations can be understood intuitively. If $\sigma_{\bar{x}} = \sigma_{\bar{x}} / \sqrt{n}$ then $\sigma_{\bar{x}} < \sigma_A$. Furthermore, when the sample size, n, rises, $\sigma^2_{\bar{x}}$ gets smaller. This is because it becomes more unusual to get a sample with an $\bar{x}$ that is far from μ as n gets larger. The standard deviation of the sampling distribution includes an $(\bar{x} - \mu)$ for each, but remember that there are not many $\bar{x}$'s that are as far from μ as there are x's that are far from μ, and as n grows there are fewer and fewer samples with an $\bar{x}$ far from μ. This means that there are not many $(\bar{x} - \mu)$ that are as large as quite a few (x -μ) are. By the time you square everything, the average $(\bar{x} - \mu)^2$ is going to be much smaller that the average $(x - \mu)^2$, so, $\sigma_{\bar{x}}$ is going to be smaller than σ_x . If the mean volume of soft drink in a population of 12 ounce cans is 12.05 ounces with a variance of .04 (and a standard deviation of .2), then the sampling distribution of means of samples of 9 cans will have a mean of 12.05 ounces and a variance of .04/9=.0044 (and a standard deviation of .2/3=.0667).

You can follow this same line of reasoning once again, and see that as the sample size gets larger, the variance and standard deviation of the sampling distribution will get smaller. Just remember that as sample size grows, samples with an $\bar{x}$ that is far from μ get rarer and rarer, so that the average $(\bar{x} - \mu)^2$ will get smaller. The average $(\bar{x} - \mu)^2$ is the variance. If larger samples of soft drink bottles are taken, say samples of 16, even fewer of the samples will have means that are very far from the mean of 12.05 ounces. The variance of the sampling distribution when n=16 will therefore be smaller. According to what you have just learned, the variance will be only .04/16=.0025 (and the standard deviation will be .2/4=.05). The formula matches what logically is happening; as the samples get bigger, the probability of getting a sample with a mean that is far away from the population mean

gets smaller, so the sampling distribution of means gets narrower and the variance (and standard deviation) get smaller. In the formula, you divide the population variance by the sample size to get the sampling distribution variance. Since bigger samples means dividing by a bigger number, the variance falls as sample size rises. If you are using the sample mean as to infer the population mean, using a bigger sample will increase the probability that your inference is very close to correct because more of the sample means are very close to the population mean.. There is obviously a trade-off here. The reason you wanted to use statistics in the first place was to avoid having to go to the bother and expense of collecting lots of data, but if you collect more data, your statistics will probably be more accurate.

The t-distribution

The central limit theorem tells us about the relationship between the sampling distribution of means and the original population. Notice that if we want to know the variance of the sampling distribution we need to know the variance of the original population. You do not need to know the variance of the sampling distribution to make a point estimate of the mean, but other, more elaborate, estimation techniques require that you either know or estimate the variance of the population. If you reflect for a moment, you will realize that it would be strange to know the variance of the population when you do not know the mean. Since you need to know the population mean to calculate the population variance and standard deviation, the only time when you would know the population variance without the population mean are examples and problems in textbooks. The usual case occurs when you have to estimate both the population variance and mean. Statisticians have figured out how to handle these cases by using the sample variance as an estimate of the population variance (and being able to use that to estimate the variance of the sampling distribution). Remember that s^2 is an unbiased estimator of σ^2. Remember, too, that the variance of the sampling distribution of means is related to the variance of the original population according to the equation:

$$\sigma^2_{\bar{x}} = \sigma^2 / n$$

so, the estimated standard deviation of a sampling distribution of means is:

$$\text{estimated } \sigma_{\bar{x}} = s / \sqrt{n}$$

Following this thought, statisticians found that if they took samples of a constant size from a normal population, computed a statistic called a "t-score" for each sample, and put those into a relative frequency distribution, the distribution would be the same for samples of the same size drawn from any normal population. The shape of this sampling distribution of t's varies somewhat as sample size varies, but for any n it's always the same. For example, for samples of 5, 90% of the samples have t-scores between -1.943 and +1.943, while for samples of 15, 90% have t-scores between ± 1.761. The bigger the samples, the narrower the range of scores that covers any particular proportion of the samples. That t-score is computed by the formula:

$$t = (\bar{x} - \mu) / (s/\sqrt{n})$$

By comparing the formula for the t-score with the formula for the z-score, you will be able to see that the t is just an estimated z. Since there is one t-score for each sample, the t is just another sampling distribution. It turns out that there are other things that can be computed from a sample that have the same distribution as this t. Notice that we've used the sample standard deviation, s, in computing each t-score. Since we've used s, we've used up one degree of freedom. Because there are other useful sampling distributions that have this same shape, but use up various numbers of degrees of freedom, it is the usual practice to refer to the t-distribution not as the distribution

for a particular sample size, but as the distribution for a particular number of degrees of freedom. There are published tables showing the shapes of the t-distributions, and they are arranged by degrees of freedom so that they can be used in all situations.

Looking at the formula, you can see that the mean t-score will be zero since the mean $\bar{x}$ equals μ. Each t-distribution is symmetric, with half of the t-scores being positive and half negative because we know from the central limit theorem that the sampling distribution of means is normal, and therefore symmetric, when the original population is normal.

An excerpt from a typical t-table is printed below. Note that there is one line each for various degrees of freedom. Across the top are the proportions of the distributions that will be left out in the tail--the amount shaded in the picture. The body of the table shows which t-score divides the bulk of the distribution of t's for that df from the area shaded in the tail, which t-score leaves that proportion of t's to its right. For example, if you chose all of the possible samples with 9 df, and found the t-score for each, .025 (2 1/2 %) of those samples would have t-scores greater than 2.262, and .975 would have t-scores less than 2.262.

df	prob = .10	prob. = .05	prob. = .025	prob. = .01	prob. = .005
1	3.078	6.314	12.70	13.81	63.65
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
20	1.325	1.725	2.086	2.528	2.845
30	1.310	1.697	2.046	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
Infinity	1.282	1.645	1.960	2.326	2.58

Exhibit 3: A sampling of a student's t-table. The table shows the probability of exceeding the value in the body. With 5 df, there is a .05 probability that a sample will have a t-score > 2.015.

Since the t-distributions are symmetric, if 2 1/2% (.025) of the t's with 9df are greater than 2.262, then 2 1/2% are less than -2.262. The middle 95% (.95) of the t's, when there are 9df, are between -2.262 and +2.262. The

middle .90 of t=scores when there are 14df are between ±1.761, because -1.761 leaves .05 in the left tail and +1.761 leaves .05 in the right tail. The t-distribution gets closer and closer to the normal distribution as the number of degrees of freedom rises. As a result, the last line in the t-table, for infinity df, can also be used to find the z-scores that leave different proportions of the sample in the tail.

What could Kevin have done if he had been asked "about how much does a pair of size 11 socks weigh?" and he could not easily find good data on the population? Since he knows statistics, he could take a sample and make an inference about the population mean. Because the distribution of weights of socks is the result of a manufacturing process, it is almost certainly normal. The characteristics of almost every manufactured product are normally distributed. In a manufacturing process, even one that is precise and well-controlled, each individual piece varies slightly as the temperature varies some, the strength of the power varies as other machines are turned on and off, the consistency of the raw material varies slightly, and dozens of other forces that affect the final outcome vary slightly. Most of the socks, or bolts, or whatever is being manufactured, will be very close to the mean weight,or size, with just as many a little heavier or larger as there are that are a little lighter or smaller. Even though the process is supposed to be producing a population of "identical" items, there will be some variation among them. This is what causes so many populations to be normally distributed. Because the distribution of weights is normal, he can use the t-table to find the shape of the distribution of sample t-scores. Because he can use the t-table to tell him about the shape of the distribution of sample t-scores, he can make a good inference about the mean weight of a pair of socks. This is how he could make that inference:

STEP 1. Take a sample of n, say 15, pairs size 11 socks and carefully weigh each pair.

STEP 2. Find $\bar{x}$ and s for his sample.

STEP 3 (where the tricky part starts). Look at the t-table, and find the t-scores that leave some proportion, say .95, of sample t's with n-1df in the middle.

STEP 4 (the heart of the tricky part). Assume that his sample has a t-score that is in the middle part of the distribution of t-scores.

STEP 5 (the arithmetic). Take his $\bar{x}$, s, n, and t's from the t-table, and set up two equations, one for each of his two table t-values. When he solves each of these equations for m, he will find a interval that he is 95% sure (a statistician would say "with .95 confidence") contains the population mean.

Kevin decides this is the way he will go to answer the question. His sample contains pairs of socks with weights of :

4.36, 4.32, 4.29, 4.41, 4.45, 4.50, 4.36, 4.35, 4.33, 4.30, 4.39, 4.41, 4.43, 4.28, 4.46 oz.

He finds his sample mean, $\bar{x}$ = 4.376 ounces, and his sample standard deviation (remembering to use the sample formula), s = .067 ounces. The t-table tells him that .95 of sample t's with 14df are between ±2.145. He solves these two equations for μ:

$$+2.145 = (4.376 - \mu)/(.067/\sqrt{14}) \quad \text{and} \quad -2.145 = (4.376 - \mu)/(.067/\sqrt{14})$$

finding μ= 4.366 ounces and μ= 4.386. With these results, Kevin can report that he is "95 per cent sure that the mean weight of a pair of size 11 socks is between 4.366 and 4.386 ounces". Notice that this is different from when he knew more about the population in the previous example.

Summary

A lot of material has been covered in this chapter, and not much of it has been easy. We are getting into real statistics now, and it will require care on your part if you are going to keep making sense of statistics.

The chapter outline is simple:

- Many things are distributed the same way, at least once we've standardized the members' values into z-scores.
- The central limit theorem gives users of statistics a lot of useful information about how the sampling distribution of is related to the original population of x's.
- The t-distribution lets us do many of the things the central limit theorem permits, even when the variance of the population, s_x , is not known.

We will soon see that statisticians have learned about other sampling distributions and how they can be used to make inferences about populations from samples. It is through these known sampling distributions that most statistics is done. It is these known sampling distributions that give us the link between the sample we have and the population that we want to make an inference about.