



MAKING SENSE OF DATA

Intermediate series

THE T-DISTRIBUTION

Copyright © 2012 by City of Bradford MDC

Sampling distributions
Confidence intervals
Hypothesis tests
The t-distribution

Many different distributions exist in statistics, and one of the most commonly used distributions is the t -distribution. This topic covers the basic characteristics of the t -distribution, how to use the t -table to find probabilities, and how it's used to solve problems in its most well-known settings — confidence intervals and hypothesis tests.

Basics of the t -distribution

The normal distribution is the well-known bell-shaped distribution whose mean is μ and whose standard deviation is σ . The t -distribution can be thought of as related to the normal distribution — it looks similar to a normal distribution in that it has a basic bell shape with an area of 1 under it, but is shorter and flatter than a normal distribution. Like the standard normal (Z) distribution, it is centered at 0, but its standard deviation is proportionally larger compared to the Z -distribution.

As with normal distributions, there is a family of different t -distributions. Each t -distribution is distinguished by what statisticians call *degrees of freedom*, which are related to the sample size of the data set. If your sample size is n , the degrees of freedom for the corresponding t -distribution is $n - 1$. For example, if your sample size is 10, you use a t -distribution with $10 - 1$ or 9 degrees of freedom, denoted t_9 .

Smaller sample sizes have flatter t -distributions than larger sample sizes. The larger the sample size is, and the larger the degrees of freedom, the more the t -distribution looks like a standard normal distribution (Z -distribution); and the point where they become very similar is about the point where the sample size is 30. (This result is due to the Central Limit Theorem.)

Figure 1 shows different t -distributions for different sample sizes, and how they compare to the Z -distribution.

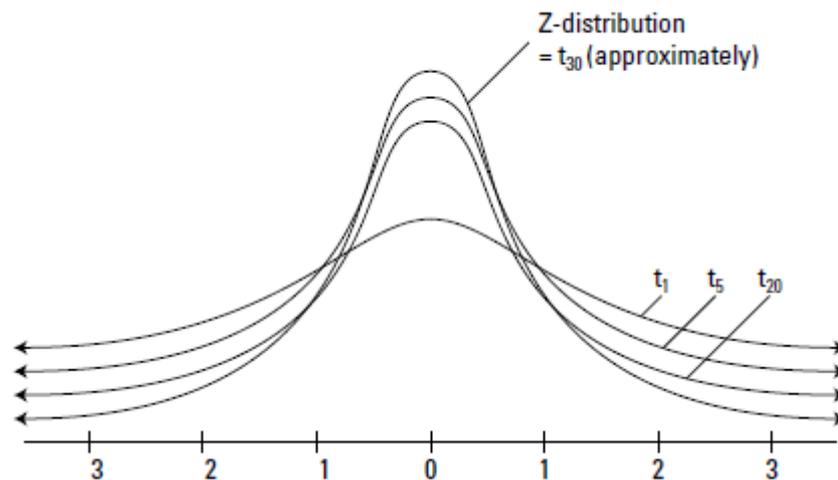


Figure 1: t -distributions for different sample sizes

Understanding the t -table

Each t -distribution has its own shape and its own set of probabilities, so one size doesn't fit all. To help with this, statisticians have come up with one abbreviated table that you can use to mark off certain points of interest on several different t -distributions whose degrees of freedom range from 1 to 30 (see "The t -table" at the end of this topic).

If you look at the column headings in the t -table, you see selected values from 0.40 to 0.0005. These numbers represent right tail probabilities (the probability of being larger than a certain value). The numbers moving down any given column represent the values on each t -distribution having those right tail probabilities. For example, under the "0.05" column, the first number is 6.3138. This represents the number on the t_1 distribution (one degree of freedom) whose probability to the right equals 0.05. Further down that column in row 15 you see 1.7531. This is the number on the t_{15} distribution whose probability to the right is 0.05.

Note that as the degrees of freedom of the t -distribution increase (as you move down any given column); the t -values get smaller. The last row of the t -table corresponds to the values on the Z -distribution. The degrees of freedom for the last row are listed as "infinity" (∞) — the t -distribution approaches a Z -distribution as n gets infinitely large.

t-distributions and ...

Hypothesis tests

The most common use by far of the *t*-distribution is in hypothesis testing — in particular the case where you do a hypothesis test for one population mean. You use a *t*-distribution when you do not know the standard deviation of the population (σ) and you have to use the standard deviation of the sample (s) to estimate it. Typically in these situations you have a small sample size n , but not always.

Finding critical values

If you don't know the population standard deviation and are using the sample standard deviation instead, you pay a penalty: a *t*-distribution with more variability and bigger tails. When you do a hypothesis test, your “cut-off point” for rejecting the null hypothesis H_0 is further out than it would have been if you had more data and could use the *Z*-distribution.

For example, suppose you have a two-tailed hypothesis test for one population mean where $\alpha = 0.05$ and you have a sample size (n) of 100. If you are doing a two-sided hypothesis test for one population mean, you can use $Z = \pm 1.96$ as your critical value to determine whether to reject H_0 . But if n is, say, 8, the critical value for this same test would be $t_7 = \pm 2.365$; see the *t*-table, row 7, column “0.025” to obtain this number. (Remember a two-tailed hypothesis test with significance level 5% has 2.5% = 0.025 in each tail area.) This means you have to submit more evidence to reject H_0 if you only have 8 items of data than if you have 100 items. In other words, your goal line to score and find a “statistically significant result” is set at 2.365 for the *t*-distribution versus 1.96 for the *Z*-distribution.

Finding p -values

Recall from the “Hypothesis tests” topic that a p -value is the probability of obtaining a result beyond your test statistic on the appropriate sampling distribution. In terms of p -values, the same test statistic has a larger p -value on a t -distribution than on the Z -distribution. A test statistic far out on the leaner Z -distribution has little area beyond it. But that same test statistic out on the bigger t -distribution has more area beyond it, and that’s exactly what the p -value represents.

Suppose your sample size n is 10, your test statistic (referred to as the t -value) is 2.5, and your alternative hypothesis, H_1 , is the greater-than alternative. Because the sample size is 10, you use the t -distribution with $10 - 1 = 9$ degrees of freedom to calculate your p -value. This means you look at the row in the t -table that has a 9 in the degrees of freedom (df) column. Your test statistic (2.5) falls between two values: 2.262 (the “0.025” column) and 2.821 (the “0.01” column). The p -value is between $0.025 = 2.5\%$ and $0.01 = 1\%$. You don’t know exactly what the p -value is, but because 1% and 2.5% are both less than the typical cut-off of 5% , you can reject H_0 .

For a less-than alternative hypothesis, your test statistic would be a negative number (to the left of 0 on the t -distribution). In this case, you want to find the percentage below, or to the left of, your test statistic to get your p -value. Yet negative test statistics don’t appear on the t -table. However, the percentage to the left (below) a negative t -value is the same as the percentage to the right (above) the positive t -value, due to the distribution being symmetric. So, to find the p -value for your negative test statistic, look up the positive version of your test statistic in the t -table, and find the corresponding right tail probability.

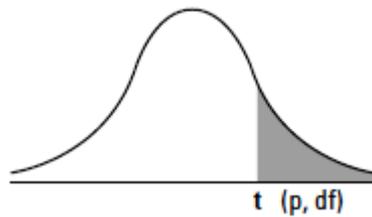
If the alternative hypothesis (H_1) has the not-equal-to alternative, you double the percentage that you get to obtain your p -value. That’s because the p -value in this case represents the chance of being beyond your test statistic in either the positive or negative direction.

Confidence intervals

The t -distribution is used in a similar way with confidence intervals. If your data have a normal distribution and either 1) the sample size n is small; or 2) you don't know the population standard deviation, σ , and you must use the sample standard deviation, s , to substitute for it, you use a value from a t -distribution with $n - 1$ degrees of freedom instead of a z -value in your formulas for the confidence intervals for the population mean.

For example, to make a 95% confidence interval for μ where $n = 9$ you add and subtract 2.306 times the standard error of \bar{X} when you use the t -distribution versus adding and subtracting 1.96 times the standard error when you use a z -value.

The *t*-table



df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.3249	1.0000	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192
2	0.2887	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991
3	0.2767	0.7649	1.6377	2.3534	3.1825	4.5407	5.8409	12.9240
4	0.2707	0.7407	1.5332	2.1318	2.7765	3.7470	4.6041	8.6103
5	0.2672	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688
6	0.2648	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588
7	0.2632	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079
8	0.2619	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413
9	0.2610	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809
10	0.2602	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693	4.5869
11	0.2596	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058	4.4370
12	0.2590	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545	4.318
13	0.2586	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123	4.2208
14	0.2582	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768	4.1405
15	0.2579	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467	4.0728
16	0.2576	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208	4.0150
17	0.2573	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982	3.9651
18	0.2571	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784	3.9216
19	0.2569	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609	3.8834
20	0.2567	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453	3.8495
21	0.2566	0.6864	1.3232	1.7207	2.0796	2.5177	2.8314	3.8193
22	0.2564	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188	3.7921
23	0.2563	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073	3.7676
24	0.2562	0.6849	1.3178	1.7109	2.0639	2.4922	2.7969	3.7454
25	0.2561	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874	3.7251
26	0.2560	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787	3.7066
27	0.2559	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707	3.6896
28	0.2558	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633	3.6739
29	0.2557	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564	3.6594
30	0.2556	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500	3.6460
∞	0.2533	0.6745	1.2816	1.6449	1.9600	2.3264	2.5758	3.2905