

One Sample t Test

Nathaniel E. Helwig

University of Minnesota

1 How Beer Changed the World

William Sealy Gosset was a chemist and statistician who was employed as the Head Experimental Brewer at Guinness in the early 1900s. As a part of his job, Gosset was responsible for taking samples of beer and performing various analyses to ensure that the beer was of the proper composition and quality. While at Guinness, Gosset taught himself experimental design and statistical analysis (which were new and exciting fields at the time), and he even spent some time in 1906–1907 studying in Karl Pearson’s laboratory. Gosset’s work often involved collecting a small number of samples, and testing hypotheses about the mean of the population from which the samples were drawn. For example, in the brewery, Gosset may want to test whether the population mean amount of some ingredient (e.g., barley) was equal to the intended value. And, on the farm, Gosset may want to test how different farming techniques and growing conditions affect the barley yields. Through this work, Gosset noticed that the normal distribution was not adequate for testing hypotheses about the mean in small samples of data with an unknown variance.

As a reminder, if $X \sim N(\mu, \sigma^2)$, then $\bar{x} \sim N(\mu, \sigma^2/n)$ where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, which implies

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

i.e., the Z test statistic follows a standard normal distribution. However, in practice, the population variance σ^2 is often unknown, so the sample variance $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ must be used in its place. With the help of Pearson, Gosset discovered that when the

sample variance is used in place of the true population variance, the test statistic

$$T = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

where t_ν denotes Student's t distribution (Student, 1908) with ν degrees of freedom. Note that Gosset published this work under the pseudonym “Student” because Guinness did not allow employees to include their name or affiliation on any scientific publications.

2 Student's t Distribution

Student's t distribution is a family of distributions that depend on a single parameter: the degrees of freedom parameter ν . As discussed in the previous section, Student's t distribution arises when we standardize a normal variable using the population mean and the sample estimate of the variance. However, Student's t distribution can be a reasonable approximation to a standardized variable formed from a variety of different population distributions, i.e., the distribution of the test statistic $T = \frac{\bar{x} - \mu}{s/\sqrt{n}}$, can often be reasonably approximated by a t_{n-1} distribution even if the random variable X follows some non-normal distribution.

Student's t distribution has the follow properties:

- PDF: $f(t) = \frac{(1+t^2)^{-\frac{\nu+1}{2}}}{\sqrt{\nu} B(\frac{1}{2}, \frac{\nu}{2})}$ where $B(u, v) = \int_0^1 t^{u-1} (1-t)^{v-1} dt$ is the beta function
- CDF: $F(t) = 1 - \frac{1}{2} I_{\frac{\nu}{t^2+\nu}}(\frac{\nu}{2}, \frac{1}{2})$ where $I_x(u, v) = \frac{B(x; u, v)}{B(u, v)}$ is the regularized incomplete beta function and $B(x; u, v) = \int_0^x t^{u-1} (1-t)^{v-1} dt$ is the incomplete beta function
- Mean: $E(T) = 0$ if $\nu > 1$; otherwise undefined
- Variance: $\text{Var}(T) = \frac{\nu}{\nu-2}$ if $\nu > 2$, $\text{Var}(T) = \infty$ if $1 < \nu \leq 2$; otherwise undefined

Relations to other distributions:

- $t_\nu \rightarrow N(0, 1)$ as $\nu \rightarrow \infty$
- $T = Z\sqrt{\frac{\nu}{V}}$ where $Z \sim N(0, 1)$ and $V \sim \chi_\nu^2$ (assuming Z and V are independent)
- $T^2 \sim F(1, \nu)$

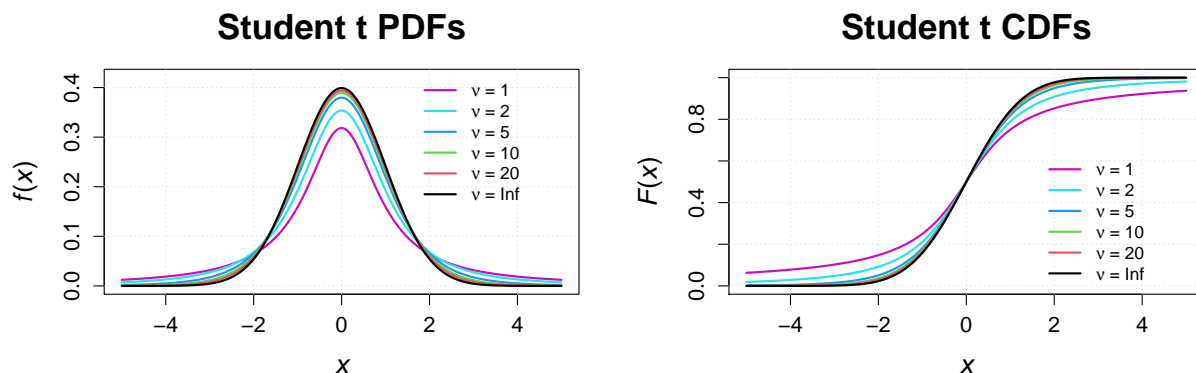


Figure 1: Student's t distribution PDFs and CDFs with various different degrees of freedom.

The probability density function (PDF) for Student's t distribution looks much like a normal distribution, in the sense that it is bell shaped and symmetric around the mean. Furthermore, the PDF for Student's t distribution approaches the PDF for the standard normal distribution as the degrees of freedom parameter ν get large. For small ν , the Student's t PDF has “fatter tails”, which implies that extreme observations are more likely to occur when generating data from a t_ν distribution (compared to a $N(0, 1)$ distribution).

3 Degrees of Freedom Detour

Definition. In statistics, the degrees of freedom associated with some quantity is the number of values that are free to vary.

Suppose that we have an independent sample of data x_1, \dots, x_n where $x_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$. The vector of sample observations $(x_1, \dots, x_n)^\top$ has n degrees of freedom, and note that

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{bmatrix} + \begin{bmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix}$$

where the first vector $(\bar{x}, \dots, \bar{x})^\top$ has a single degree of freedom (because it is the same number replicated n times), and the second vector $(x_1 - \bar{x}, \dots, x_n - \bar{x})^\top$ has $n - 1$ degrees of freedom (because the deviation scores have one constraint, i.e., that $\sum_{i=1}^n (x_i - \bar{x}) = 0$).

4 Student's Test of Statistical Significance

Suppose that we have collected an independent and identically distributed (iid) sample of observations x_1, \dots, x_n , where each observation is assumed to follow a normal distribution with mean μ and variance σ^2 , i.e., $x_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$. Furthermore, suppose that we want to test a null hypothesis about the mean. As a reminder, we could test the following hypotheses:

- $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ (exact H_0 with two-sided H_1)
- $H_0 : \mu \geq \mu_0$ versus $H_1 : \mu < \mu_0$ (inexact H_0 with less than H_1)
- $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$ (inexact H_0 with greater than H_1)

where $\mu_0 \in \mathbb{R}$ is the null hypothesized value of the mean.

As a test statistic, we will use Student's t statistic

$$T_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ is the sample mean and $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ is the sample variance. Assuming that the null hypothesis is true, we just need to compare T_0 to the quantiles of a t_{n-1} distribution to calculate the p-value. For the directional (one-sided) tests, the p-values are

- $H_0 : \mu \geq \mu_0$ versus $H_1 : \mu < \mu_0$, the p-value is defined as $p = P(T < T_0)$
- $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$, the p-value is defined as $p = P(T > T_0)$

where T is a random variable that follows Student's t distribution with $n - 1$ degrees of freedom, i.e., $T \sim t_{n-1}$. For the two-sided test, the p-value is defined as

$$p = P(T < -|T_0|) + P(T > |T_0|) = 2P(T > |T_0|)$$

where $T \sim t_{n-1}$. Note that we can simply double the probability that T is greater than the absolute value of T_0 because the t_{n-1} distribution is symmetric around zero. Although the two-sided alternative hypothesis is the default in many statistical softwares, the one-sided alternative is often more appropriate for applications—because the researcher often has an idea of the direction of the effect being tested.

Example 1. A store sells “16-ounce” boxes of Captain Crisp cereal. A random sample of 9 boxes was taken and weighed. The results were

15.5, 16.2, 16.1, 15.8, 15.6, 16.0, 15.8, 15.9, 16.2

Suppose that we want to test the null hypothesis $H_0 : \mu \geq \mu_0$ versus the alternative hypothesis $H_1 : \mu < \mu_0$. The sample mean and variance are

$$\bar{x} = \frac{1}{9}(15.5 + 16.2 + 16.1 + 15.8 + 15.6 + 16.0 + 15.8 + 15.9 + 16.2) = 15.9$$

$$s^2 = \frac{1}{8} \left(\sum_{i=1}^9 x_i^2 - 9\bar{x}^2 \right) = \frac{1}{8} (2275.79 - 9(15.9)^2) = 0.0625$$

which implies that the observed T test statistic and p-value have the form

$$T_0 = \frac{15.9 - 16}{\sqrt{0.0625/9}} = -1.2 \quad \text{and} \quad P(T < -1.2) = 0.1322336$$

where $T \sim t_8$. To confirm this result in R, we can use the `t.test` function:

```
> x <- c(15.5, 16.2, 16.1, 15.8, 15.6, 16.0, 15.8, 15.9, 16.2)
> t.test(x, mu = 16, alternative = "less")
```

One Sample t-test

```
data: x
t = -1.2, df = 8, p-value = 0.1322
alternative hypothesis: true mean is less than 16
95 percent confidence interval:
 -Inf 16.05496
sample estimates:
mean of x
 15.9
```

5 Confidence Interval for the Mean

To form a $100(1 - \alpha)\%$ confidence interval for the mean, note that

$$1 - \alpha = P\left(t_{\alpha/2} < \frac{\bar{x} - \mu}{s/\sqrt{n}} < t_{1-\alpha/2}\right)$$

where t_α denotes the α -th quantile of the t_{n-1} distribution. Rearranging the terms inside the probability statement reveals that

$$\begin{aligned} 1 - \alpha &= P\left(t_{\alpha/2}s/\sqrt{n} < \bar{x} - \mu < t_{1-\alpha/2}s/\sqrt{n}\right) \\ &= P\left(t_{\alpha/2}s/\sqrt{n} - \bar{x} < -\mu < t_{1-\alpha/2}s/\sqrt{n} - \bar{x}\right) \\ &= P\left(\bar{x} - t_{\alpha/2}s/\sqrt{n} > \mu > \bar{x} - t_{1-\alpha/2}s/\sqrt{n}\right) \end{aligned}$$

which implies that the $100(1 - \alpha)\%$ confidence interval is given by

$$\left[\bar{x} - t_{1-\alpha/2}\frac{s}{\sqrt{n}}, \quad \bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}}\right]$$

and given that $-t_{\alpha/2} = t_{1-\alpha/2}$ the confidence interval can be written as

$$\bar{x} \pm t_{1-\alpha/2}\frac{s}{\sqrt{n}}$$

Note that the above confidence interval makes sense to use if the alternative hypothesis is two-sided. For directional (one-sided) tests, it would make more sense to use a confidence bound, which places all of the uncertainty in the direction relevant to the alternative:

- For $H_1 : \mu < \mu_0$, use an upper confidence bound: $[-\infty, \bar{x} + t_{1-\alpha}s/\sqrt{n}]$
- For $H_1 : \mu > \mu_0$, use a lower confidence bound: $[\bar{x} - t_{1-\alpha}s/\sqrt{n}, \infty]$

Example 2. For the previous example, the 95% one-sided upper confidence bound is

$$[-\infty, 16.05496]$$

which was output from the `t.test` function in the previous example.

6 Nonparametric Alternatives

Suppose that we want to test a null hypothesis about μ (or form a confidence interval for μ), but we are not certain that the assumption of normality is reasonable. Instead of comparing the observed test statistic T_0 to a Student's t_{n-1} distribution, we could compare the observed T_0 to a null distribution that is obtained in a nonparametric fashion. Such a test is referred to as a nonparametric randomization test, which is also known as a permutation test (see Helwig, 2019). Nonparametric tests involve randomly permuting aspects of the observed data under the assumption that the null hypothesis is true, which makes it possible to obtain a (conditional) null distribution of the test statistic without assuming a particular data generating distribution for the data.

Example 3. Revisiting the previous example, we could use the `np.loc.test` function in the `npctest` R package (Helwig, 2020) to conduct a nonparametric permutation test:

```
> library(npctest)
> np.loc.test(x, mu = 16, alternative = "less")

Nonparametric Location Test (One Sample t-test)
alternative hypothesis: true mean is less than 16
t = -1.2,  p-value = 0.1602
sample estimate:
mean
15.9
```

Note that the observed test statistic of $T_0 = -1.2$ is the same observed test statistic that we used for Student's t test. In this case, the p-value is an “exact” p-value, given that there are only $2^9 = 512$ elements of the null distribution (see the help files of the `np.loc.test` or Helwig (2019) for details). The p-value for the nonparametric test is similar to, but a bit larger than, the p-value that was obtained from Student's t test. In both cases, we would fail to reject the null hypothesis using any standard significance level. So, using both Student's t test and a nonparametric t test, we come to the same conclusion: we have no evidence to suggest that the Caption Crisp cereal boxes weigh less than what they claim.

If we wanted to form a confidence interval for the mean, we could use the `np.boot` function to compute a nonparametric confidence interval:

```
> set.seed(0)
> np.boot(x, mean)
```

```
Nonparametric Bootstrap of Univariate Statistic
using R = 9999 bootstrap replicates
```

```
t0: 15.9
SE: 0.0786
Bias: 4e-04
```

```
BCa Confidence Intervals:
```

```
      lower  upper
90% 15.7556 16.0111
95% 15.7333 16.0333
99% 15.6889 16.0778
```

Note that the `np.boot` function produces two-sided confidence intervals, so we would want to use the upper bound from the two-sided 90% confidence interval to form a 95% upper bound, which is given by $[-\infty, 16.0111]$.

References

- Helwig, N. E. (2019). Statistical nonparametric mapping: Multivariate permutation tests for location, correlation, and regression problems in neuroimaging. *WIREs Computational Statistics* 2, e1457.
- Helwig, N. E. (2020). *nptest: Nonparametric Tests*. R package version 1.0-2.
- Student (1908). The probable error of a mean. *Biometrika* 6(1), 1–25.