The Student t Distribution

t Student distribution

Definition: The (Student) t distribution with degrees of freedom parameter ν is a continuous probability distribution with probability density

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2} \qquad \text{for all } t \in R.$$

The t distribution

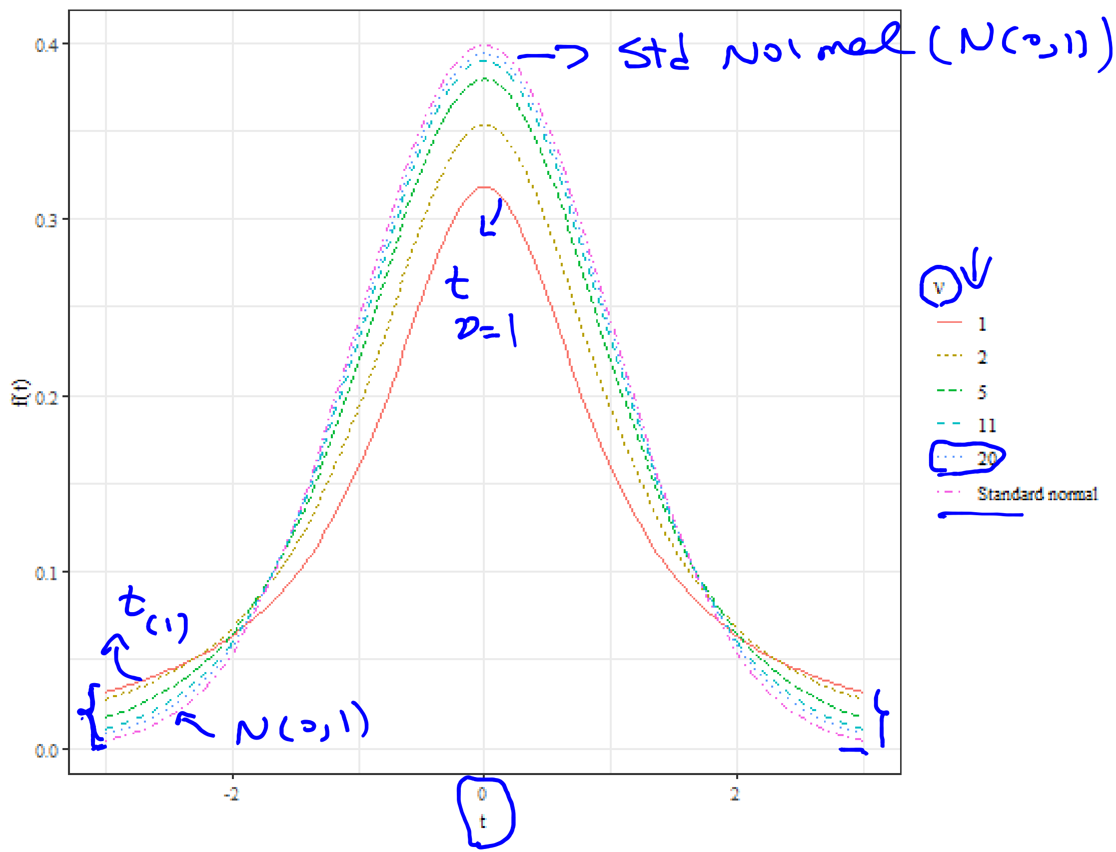
> - is bell-shaped and symmetric about 0
> - has fatter tails than the normal, but approaches the shape of the normal as $\nu \to \infty$.

(heavy tails)

$\Gamma(x) = (x!$

t Student distribution

We use the t table (Table B.4 in Vardeman and Jobe) to calculate quantiles.



t Student distribution

Example: Say $T \sim t_5$. Find c such that $P(T \leq c) = 0.9$.

ν

Table B.4
t Distribution Quantiles

ν	$Q(.9)$	$Q(.95)$	$Q(.975)$	$Q(.99)$	$Q(.995)$	$Q(.999)$	$Q(.9995)$
1	3.078	6.314	12.706	31.821	63.657	318.317	636.607
2	1.886	2.920	4.303	6.965	9.925	22.327	31.598
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869

So, $P(T \leq c) = 0.9$ holds true if $c = 1.476$ (by the table).

$P(T \leq 1.476) = 0.9$

- Quiz 3 on Thursday. (In-class)
  - Bring your Calculator!
- HW 8 Solution Posted.
- Sample quiz, topic outline, formula sheet for quiz 3 posted to "Exam Materials"

Small-sample Confidence Interval for μ when σ is unknown

Small-sample confidence intervals, σ unknown

If we can assume that $X_1, \ldots, X_n$ are iid with mean μ and variance σ^2, and are also normally distributed, if $n < 25$, we cannot use CLT.

It is not easy to prove but,

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

We can then use $t_{n-1,1-\alpha/2}$ instead of $z_{1-\alpha/2}$ in the confidence intervals.

Note that the df (degree of freedom) for the t distribution is $n - 1$.

Large Sample Inference

Confidence Interval

CI for μ

CI for μ unknown σ^2

t Distribution

Small n unknown σ^2

Small-sample confidence intervals, σ unknown

- *Two-sided* $100(1-\alpha)\%$ confidence interval for μ

$$\left(\bar{x} - \boxed{t_{n-1,1-\alpha/2}}\frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,\,1-\alpha/2}\frac{s}{\sqrt{n}}\right)$$

- *One-sided* $100(1-\alpha)\%$ confidence interval for μ with a upper confidence bound

$$\left(-\infty\ ,\ \bar{x} + \boxed{t_{n-1,\,1-\alpha}}\frac{s}{\sqrt{n}}\right)$$

- *One-sided* $100(1-\alpha)\%$ confidence interval for μ with a lower confidence bound

$$\left(\bar{x} - \boxed{t_{n-1,\,1-\alpha}}\frac{s}{\sqrt{n}}\ ,\ +\infty\right)$$

Example: [Concrete beams]

10 concrete beams were each measured for flexural strength (MPa). Assuming the flexural strengths are iid ~~normal~~, calculate and interpret a two-sided 99\% CI for the flexural strength of the beams.

$\boxed{\alpha = 0.01}$

[1] 8.2 8.7 7.8 9.7 7.4 7.8 7.7 11.6 11.3 11.8

$$\begin{cases} n = 10 \\ \alpha = 0.01 \end{cases}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}(8.2 + 8.7 + \cdots + 11.8) = 9.2$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{1}{10-1}\sum_{i=1}^{10}(x_i - 9.2)^2$$

$$= \frac{1}{9}\left[(8.2-9.2)^2 + (8.7-9.2)^2 + \cdots\right.$$

$+ (11.8 - 9.2)^2 \Big] = \cdots$

$$S = \sqrt{S^2} = 1.76$$

$n = 10$

$\alpha = 0.01$

$\bar{x} = 9.2$

$S = 1.76$

$\Rightarrow$ two-sided 99% CI:

$$\left(\bar{x} - t_{(n-1,\,1-\alpha/2)} \frac{s}{\sqrt{n}},\ \bar{x} + t_{(n-1,\,1-\alpha/2)}\, s/\sqrt{n}\right)$$

$$= \left(9.2 - t_{9,\,0.995} \frac{1.76}{\sqrt{10}},\ 9.2 + t_{9,\,0.995} \frac{1.76}{\sqrt{10}}\right)$$

$$= \left(9.2 - 3.25 \times \frac{1.76}{\sqrt{10}},\ 9.2 + 3.25 \frac{1.76}{\sqrt{10}}\right)$$

$$= (7.393,\ 11.007)$$

we are 99% confident that the true mean of flexural strength of this kind of beam is between 7.393 and 11.007 MPa.

Large Sample Inference

Confidence Interval

CI for μ

CI for μ unknown σ^2

t Distribution

Small n unknown σ^2

Example: [Concrete beams]

Is the true mean flexural strength below the minimum requirement of 11 MPa? Find out with the appropriate 95% CI.

Find a one-sided CI (upper CI)

$n=10$

$\alpha=0.05$

$$\left(-\infty,\ \bar{x}+t_{(n-1,1-\alpha)}\frac{s}{\sqrt{n}}\right)$$

$$=\left(-\infty,\ 9.2+t_{9,1-0.05}\frac{1.76}{\sqrt{10}}\right)$$

$$=\left(-\infty,\ 9.2+t_{9,0.95}\cdot\frac{1.76}{\sqrt{10}}\right)$$

$$=\left(-\infty,\ 9.2+1.833\times\frac{1.76}{\sqrt{10}}\right)$$

$$=\left(-\infty,\ 10.22\right)$$

Interpretation:

we're 95% confident that the true mean flexural strength is below 10.22

$\Rightarrow$ Since this < 11, it meets the requirement.

Large Sample Inference

Confidence Interval

CI for μ

CI for μ unknown σ^2

t Distribution

Small n unknown σ^2

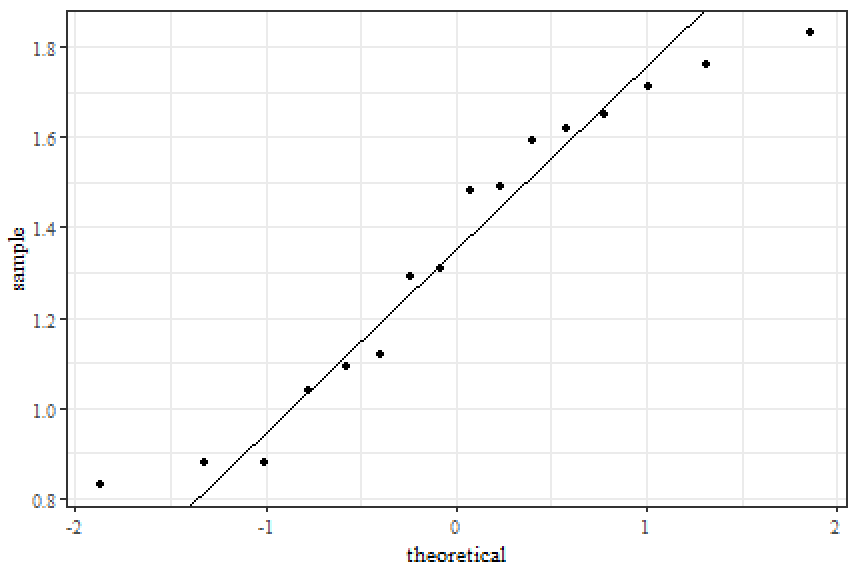
Example: [Paint thickness]

Consider the following sample of observations on coating thickness for low-viscosity paint. (mm)

```
 [1] 0.83 0.88 0.88 1.04 1.09 1.12 1.29
 [8] 1.31 1.48 1.49 1.59 1.62 1.65 1.71
[15] 1.76 1.83
```

$n=16$

A normal QQ plot shows that they are close enough to normally distributed.



Large Sample Inference

Confidence Interval

CI for μ

CI for μ unknown σ^2

t Distribution

Small n unknown σ^2

Example: [Paint thickness]

Calculate and interpret a two-sided 90% confidence interval for the true mean thickness.

① $n = 16$, $100(1-\alpha)\% = 90\% \Rightarrow \alpha = 0.1$

② $\bar{x} = \frac{1}{16}(0.83 + \cdots + 1.83) = 1.35$ mm

③ $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

$= \sqrt{\frac{1}{16-1}\left[(0.83-1.35)^2 + \cdots (1.83-1.35)^2\right]}$

$= 0.34$ $\boxed{mm}$

$$90\% \ CI : \left(\bar{x} - t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}\right)$$

$$: \left(1.35 - t_{(15, 1-\frac{0.1}{2})} \frac{0.34}{\sqrt{16}}, 1.35 + t_{(15, 1-\frac{0.1}{2})} \frac{0.34}{\sqrt{16}}\right)$$

$$: (1.35 - 1.75 \times 0.085, 1.35 + 1.75 \times 0.085)$$

$$= (1.201, 1.499)$$

- we are 90% confident that the true mean tickness of viscosity falls between 1.201 and 1.499.

Let's Wrap Up

Common Assumptions and Common Statements

Suppose that $X_1, X_2, \ldots, X_n$ are random variables whose values will be determined based on the results of random events.

> **Large Sample Size**, **Known Variance**
> Assuming:
>
> - $E(X_i) = \mu$,
> - $n \geq \cancel{30}$, 25
> - $Var(X_i) = \sigma^2$ is known
>
> Then by CLT,
>
> $$* \quad \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \dot{\sim} N(0,1)$$

$100(1-\alpha)\%$ Confidence interval for μ:

$$\bar{x} \pm z_{1-\alpha/2} \sqrt{\frac{\sigma^2}{n}}$$

Common Assumptions and Common Statements

Large Sample Size, **Unknown Variance**
Assuming:

- $E(X_i) = \mu$,
- $n \geq \cancel{30}$, 25
- $Var(X_i)$ is unknown, but sample variance $S^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$ can be calculated

Then by CLT and convergence of sample variance

$$\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim N(0, 1)$$

`

$100 \cdot (1 - \alpha)\%$-Confidence interval for μ:

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

Common Assumptions and Common Statements

Small Sample Size, **Unknown Variance**

Assuming:

- $E(X_i) = \mu$,
- $n < \cancel{30}$, 25
- $Var(X_i)$ is unknown, but sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ can be calculated

Then by CLT and convergence of sample variance

$$\frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t_{n-1}$$

`

$100 \cdot (1 - \alpha)\%$-Confidence interval for μ:

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

Common Assumptions and Common Statements

With the last set of assumptions, we can conclude that $\frac{\bar{X}-\mu}{\sqrt{S^2/n}}$ follows a "t-distribution with $n-1$ degrees of freedom"

The t-distribution looks a lot like a standard normal distribution and we use it the same way:

- It is symmetric
- It is centered at 0
- Important quantiles are collected together in tables for reference

*

It only has one parameter, the degrees of freedom. In this class, the degrees of freedom are related to the number of parameters being tested

n

* degrees of freedom = (# of observations) - (# of $\underbrace{\text{parameters}}_{1}$)