6 Introduction to formal statistical inference

Formal statistical inference uses probability theory to quantify the reliability of data-based conclusions. We want information on a population. We can use:

for example: true mean fill weight of food jars.
average number of cycles to failure of a kind of spring
true mean breaking strength of a wire rope

1. Point estimates:

   e.g. sample mean

   For example, measure breaking strength of 6 wire ropes as
   5, 3, 7, 3, 10, 1

   $$\text{estimate } \mu \approx \bar{x} = \frac{5+3+7+3+10+1}{6} = 4.83 \text{ tons.}$$

2. Interval estimates:

   μ is likely to be inside the interval $(4.83-2, 4.83+2) = (2.83, 6.83)$

   We are confident that the true mean breaking strength μ is somewhere in $(2.83, 6.83)$.

   But, <u>how confident</u>?

6.1 Large-sample confidence intervals for a mean

Many important engineering applications of statistics fit the following mold. Values for parameters of a data-generating process are unknown. Based on data, the goal is

1. identify an interval of values likely to contain an unknown paramet
2. quantify "how likely" the interval is to cover the correct value.

Definition 6.1. A *confidence interval* for a parameter (or function of one or more parameters) is a data-based interval of numbers thought likely to contain the parameter (or function of one or more parameters) possessing a stated probability-based confidence or reliability.

A confidence interval is a realization of a **random interval**, an interval on the real line with a random variable at one or both of the endpoints.

Example 6.1 (Instrumental drift)**.** Let Z be a measure of instrumental drift of a random voltmeter that comes out of a certain factory. Say $Z \sim N(0,1)$. Define a random interval:

$$(Z-2, Z+2)$$

What is the probability that -1 is inside the interval?

$$P(-1 \text{ in } (Z-2, Z+2)) = P(Z-2 < -1 < Z+2)$$

$$= P(Z-1 < 0 < Z+3)$$

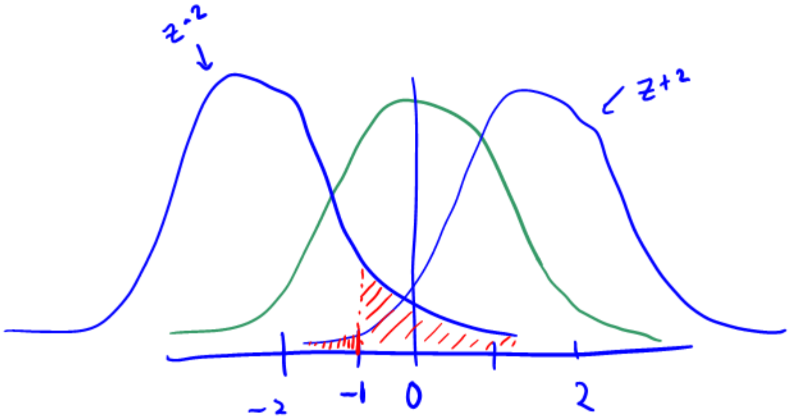
$$= P(-1 < -Z < 3) \overset{\text{multiply by } -1}{=} P(1 > Z > -3)$$

$$= P(-3 < Z < 1)$$

$$= P(Z \le 1) - P(Z \le -3)$$

$$= \Phi(1) - \Phi(-3)$$

$$= 0.84$$



Example 6.2 (More practice)**.** Calculate:

1. $P(2 \text{ in } (X-1, X+1))$, $X \sim N(2, 4)$ $= 2^2$

$$P(2 \in (X-1, X+1)) = P(X-1 < \overset{\downarrow}{2} < X+1)$$
$$= P(-1 < 2 - X < 1)$$
$$= P(-1 < X-2 < 1)$$
$$= P\left(-\tfrac{1}{2} < \frac{X-2}{2} < \tfrac{1}{2}\right)$$
$$= P(-0.5 < Z < 0.5)$$
$$= \Phi(0.5) - \Phi(-0.5)$$
$$= 0.6915 - 0.3085$$
$$= 0.383$$

subtract X

multiply by -1

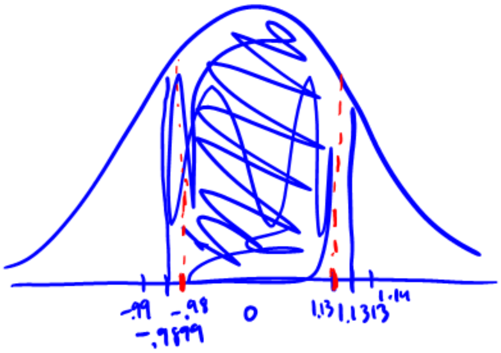
divide by σ

recognize standardized

$Z \sim N(0,1)$

2. $P(6.6 \text{ in } (X-2, X+1))$, $X \sim N(7, 2)$

$$P(6.6 \in (X-2, X+1)) = P(X-2 < 6.6 < X+1)$$
$$= P(-2 < 6.6 - X < 1)$$
$$= P(-1 < X - 6.6 < 2)$$
$$= P(-1.4 < X - 7 < 1.6)$$
$$= P\left(-\frac{1.4}{\sqrt{2}} < \frac{X-7}{\sqrt{2}} < \frac{1.6}{\sqrt{2}}\right)$$
$$= P(-0.9899 < Z < 1.1313) \qquad Z \sim N(0,1)$$



"at least"

$$\approx P(-.98 < Z < 1.13)$$
$$= \Phi(1.13) - \Phi(-.98)$$
$$= 0.8708 - 0.1635 = 0.7073$$

Example 6.3 (Abstract random intervals)**.** Let's say $X_1, X_2, \ldots, X_n$ are iid with $n \geq 25$, mean μ, variance σ^2. We can find a random interval that provides a lower bound for μ with $1-\alpha$ probability:

Want a s.t. $P(\mu \in (a, \infty)) = 1 - \alpha$

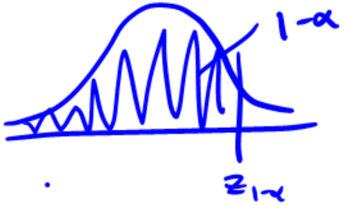
We know $\bar{X} \,\dot{\sim}\, N\left(\mu, \frac{\sigma^2}{\sqrt{n}}\right)$ by CLT

$$\Rightarrow \boxed{\frac{\bar{X} - \mu}{\sigma^2/\sqrt{n}}} \,\dot{\sim}\, N(0,1) \quad \text{by standardization}$$

new notation

Let $\boxed{z_{1-\alpha}}$ denote the $1-\alpha$ quantile of $N(0,1)$

$Z \sim N(0,1)$, $P(Z \leq z_{1-\alpha}) = 1 - \alpha$



$$\Rightarrow P\left(\frac{\bar{X} - \mu}{\sigma^2/\sqrt{n}} \leq z_{1-\alpha}\right) \approx 1 - \alpha$$

$$P\left(\bar{X} - \mu \leq z_{1-\alpha} \frac{\sigma^2}{\sqrt{n}}\right) \approx 1 - \alpha$$

$$P\left(\bar{X} - z_{1-\alpha} \frac{\sigma^2}{\sqrt{n}} \leq \mu\right) \approx 1 - \alpha$$

$$P\left(\mu \in \left(\underset{\uparrow \text{ random variable}}{\boxed{\bar{X}}} - z_{1-\alpha} \frac{\sigma^2}{\sqrt{n}}, \infty\right)\right) \approx 1 - \alpha$$

Calculate:

1. $P(\mu \in (-\infty, \overline{X} + z_{1-\alpha}\frac{\sigma}{\sqrt{n}})), X \sim N(\mu, \sigma^2)$

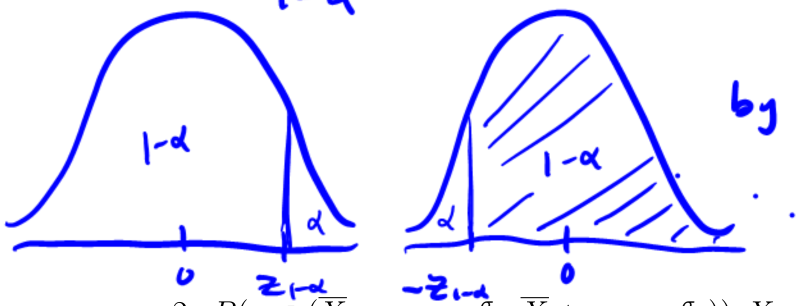
$$= P\left(\mu < \overline{X} + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(-z_{1-\alpha}\frac{\sigma}{\sqrt{n}} < \overline{X} - \mu\right)$$

$$= P\left(-z_{1-\alpha} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)$$

$$\approx P\left(-z_{1-\alpha} < Z\right) \quad Z \sim N(0,1) \text{ by CLT if } n \geq 25$$

$$= 1 - \alpha$$



by symmetry

2. $P(\mu \in (\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})), X \sim N(\mu, \sigma^2)$

$$= P\left(\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu - \overline{X} < z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(-z_{1-\alpha/2} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right)$$

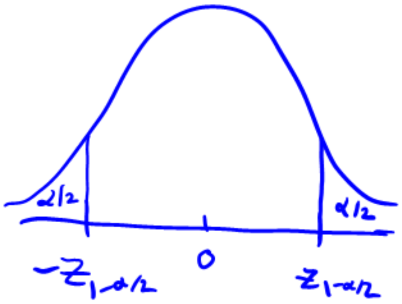
standardization $\frac{X - EX}{\sqrt{\operatorname{var} X}}$, mean 0 variance 1

$$\approx P\left(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}\right) \quad Z \sim N(0,1) \text{ by CLT for } n \geq 25$$

$$= \Phi\left(z_{1-\alpha/2}\right) - \Phi\left(-z_{1-\alpha/2}\right)$$

$$= 1 - \frac{\alpha}{2} - \frac{\alpha}{2}$$

$$= 1 - \alpha$$



6.1.1 A Large-n confidence interval for μ involving σ

A $\underline{1-\alpha}$ **confidence interval** for an unknown parameter is the realization of a random interval that contains that parameter with probability $\underbrace{1-\alpha}_{\text{called "confidence level"}}$.

For random variables $X_1, X_2, \ldots, \overset{n\geq 25}{X_n}$ iid with $\mathrm{E}(X_1) = \mu$, $\mathrm{Var}(X_1) = \sigma^2$, a $1-\alpha$ confidence interval for μ is

depends on data

$$(\overline{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \boxed{\overline{x}} + z_{1-\alpha/2}\frac{\sigma \leftarrow}{\sqrt{n}})$$

which is a **realization** from the random interval (has random variables as endpoints)

$$(\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}).$$

- Two-sided $1-\alpha$ confidence interval for μ

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

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

- One-sided $1-\alpha$ confidence interval for μ with a upper confidence bound

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

- One-sided $1-\alpha$ confidence interval for μ with a lower confidence bound

$$\left(\bar{x} - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}, \infty\right)$$

Example 6.4 (Fill weight of jars)**.** Suppose a manufacturer fills jars of food using a stable filling process with a known standard deviation of $\sigma = 1.6$g. We take a sample of $n = 47$ jars and measure the sample mean weight $\overline{x} = 138.2$g. A two-sided 90% confidence interval ($\alpha = 0.1$) for the true mean weight μ is:

$$\left(\bar{x} - z_{1-0.1/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{1-0.1/2}\frac{\sigma}{\sqrt{n}}\right)$$

$$= \left(138.2 - z_{.95}\frac{1.6}{\sqrt{47}},\ 138.2 + z_{.95}\frac{1.6}{\sqrt{47}}\right)$$

$$= (138.2 - 1.64(0.23),\ 138.2 + 1.64(0.23))$$

$$= (137.82,\ 138.58)$$

Could also write as 138.2 ± 0.38 g

Interpretation:

→ We are 90% confident that the true mean fill is between 137.82 and 138.58g.

If we took 100 more samples of 47 jars each, roughly 90 of those samples would produce confidence intervals containing the true mean fill weight.

What if we just want to be sure that the true mean fill weight is high enough?

We could use a one-sided 90% CI with a lower bound.

$$\left(\bar{x} - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}, \infty\right)$$

$$= \left(138.2 - z_{.9}\frac{1.6}{\sqrt{47}}, \infty\right)$$

$$= (138.2 - 1.28(0.23), \infty)$$

$$= (137.91, \infty)$$

We are 90% confident that the true mean fill weight is above 137.91.

Example 6.5 (Hard disk failures)**.** F. Willett, in the article "The Case of the Derailed Disk Drives?" (*Mechanical Engineering*, 1988), discusses a study done to isolate the cause of link code A failure in a model of Winchester hard disk drive. For each disk, the investigator measured the breakaway torque (in. oz.) required to loosen the drive's interrupter flag on the stepper motor shaft. Breakaway torques for 26 disk drives were recorded, with a sample mean of 11.5 in. oz. Suppose you know the true standard deviation of the breakaway torques is 5.1 in. oz. Calculate and interpret:

1. A two-sided 90% confidence interval for the true mean breakaway torque of the relevant type of Winchester drive.

2. An analogous two-sided 95% confidence interval.

Example 6.6 (Width of a CI)**.** If you want to estimate the breakaway torque with a 2-sided, 95% confidence interval with ±2.0 in. oz. of precision, what sample size would you need?

6.1.2 A generally applicable large-n confidence interval for μ

Although the equations for a $1-\alpha$ confidence interval is mathematically correct, it is severely limited in its usefulness because

If $n \geq 25$ and σ is *unknown*, $Z = \frac{\overline{X}-\mu}{s/\sqrt{n}}$, where

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

is still **approximately standard normally distributed**. So, you can replace σ in the confidence interval formula with the sample standard deviation, s.

- Two-sided $1-\alpha$ confidence interval for μ

- One-sided $1-\alpha$ confidence interval for μ with a upper confidence bound

- One-sided $1-\alpha$ confidence interval for μ with a lower confidence bound

Example 6.7. Suppose you are a manufacturer of construction equipment. You make 0.0125 inch wire rope and need to determine how much weight it can hold before breaking so that you can label it clearly. Here are breaking strengths, in kg, for 41 sample wires:

```
 [1] 100.37  96.31  72.57  88.02 105.89 107.80  75.84  92.73  67.47  94.87
[11] 122.04 115.12  95.24 119.75 114.83 101.79  80.90  96.10 118.51 109.66
[21]  88.07  56.29  86.50  57.62  74.70  92.53  86.25  82.56  97.96  94.92
[31]  62.00  93.00  98.44 119.37 103.70  72.40  71.29 107.24  64.82  93.51
[41]  86.97
```

The sample mean breaking strength is 91.85 kg and the sample standard deviation is 17.6 kg. Using the appropriate 95% confidence interval, try to determine whether the breaking strengths meet the requirement of at least 85 kg.

6.2 Small-sample confidence intervals for a mean

The most important practical limitation on the use of the methods of the previous sections is

That restriction comes from the fact that without it,

So, if one mechanically uses the large-n interval formula $\overline{x} \pm z\frac{s}{\sqrt{n}}$ with a small sample,

If it is sensible to model the observations as iid normal random variables, then we can arrive at inference methods for small-n sample means.

6.2.1 The Student t distribution

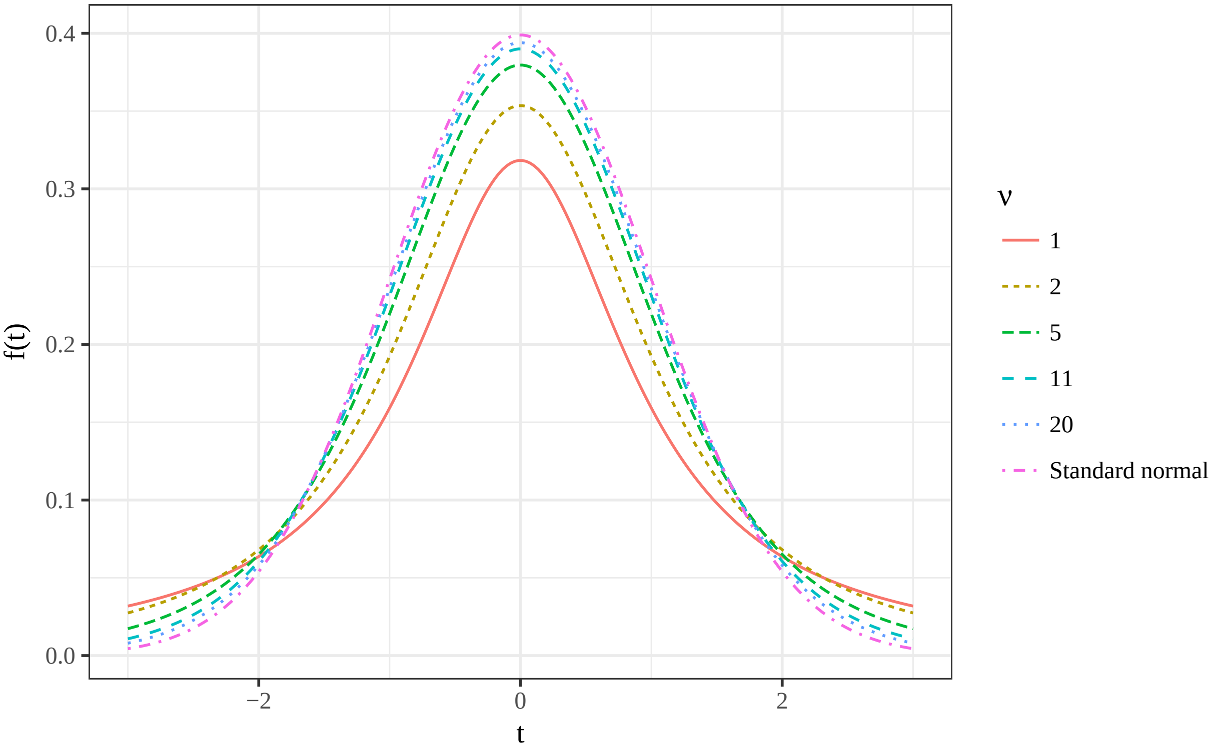
Definition 6.2. The *(Student) t distribution with degrees of freedom parameter ν* is a continuous probability distribution with probability density

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

The t distribution

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

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



Example 6.8 (t quantiles)**.** Say $T \sim t_5$. Find c such that $P(T \leq c) = 0.9$.

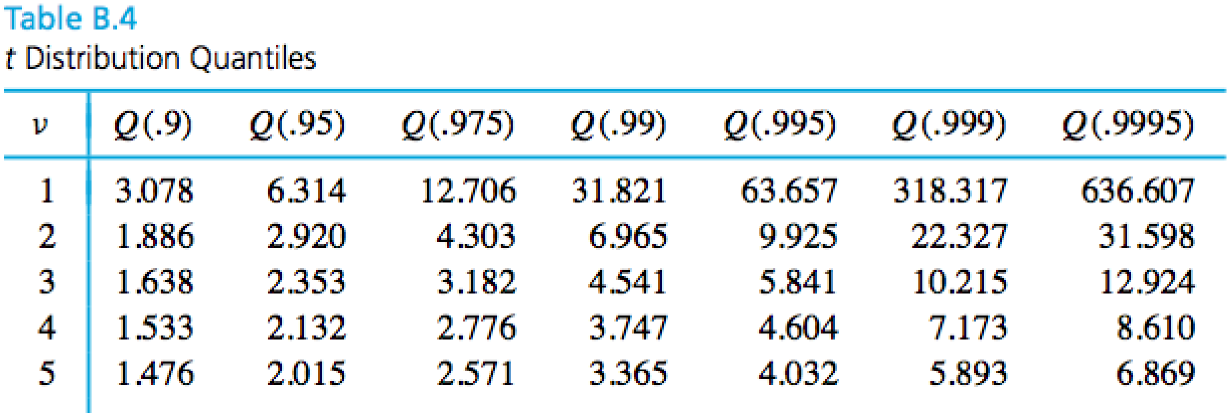


Table B.4

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

Figure 1: Student's t distribution quantiles.

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

- Two-sided $1-\alpha$ confidence interval for μ

- One-sided $1-\alpha$ confidence interval for μ with a upper confidence bound

- One-sided $1-\alpha$ confidence interval for μ with a lower confidence bound

Example 6.9 (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.

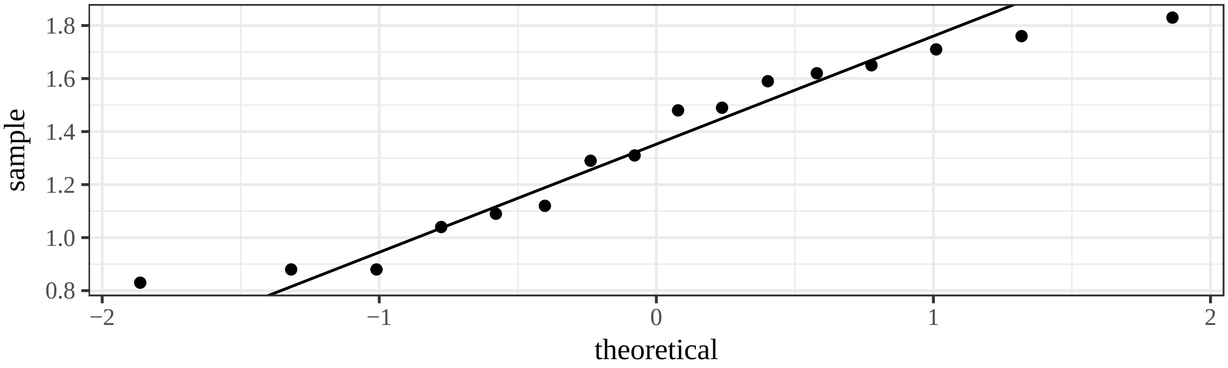
```
[1] 8.2 8.7 7.8 9.7 7.4 7.8 7.7 11.6 11.3 11.8
```

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

Example 6.10 (Paint thickness)**.** Consider the following sample of observations on coating thickness for low-viscosity paint.

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

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



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

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
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.849
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

This table was generated using MINITAB.