21 Student's t distribution

21.1 $N(\mu, 1)$

.

The following data comes from sampling 10 independent random variables, all $N(\mu, 1)$, where μ needs to be estimated.

```
-0.68 1.69 1.49 1.88 2.75 1.58 2.20 2.32 1.28 1.69
```

$$n = 10, \quad \overline{X} = 1.62, \quad \sigma^2 = 1$$

The sample average $\sim N(\mu, 1/10)$. The 'Z-score' is

$$\sqrt{10}(\overline{X} - \mu)$$

We use this information to estimate μ.

$$\sqrt{10}(1.62 - \mu)$$

If we take the interval $[-1.96, 1.96]$, and $Z \sim N(0, 1)$, Z will be in this interval with 95% probability.

Where does 1.96 come from? From the tables, the probability $Z \leqslant 1.96$ is 97.5%, so the probability $Z \geqslant 1.96$ = probability $Z \leqslant -1.96 = 2.5\%$. $[-1.96, 1.96]$ is a symmetric interval with probability 95%, leaving 5% in the 'tails.'

$$-1.96 \leqslant \sqrt{10}(\overline{X} - \mu) \leqslant 1.96$$

with 95% probability. Now

That means that $\sqrt{10}(\overline{X} - \mu)$ will belong to the interval 19 times out of 20. We put this differently: knowing $\overline{X}$ but not μ, this will happen 19 times in 20, and

$$\overline{X} - \frac{1.96}{\sqrt{10}} \leqslant \mu \leqslant \overline{X} + \frac{1.96}{\sqrt{10}}$$

with 95% *confidence.*

$$1.00 \leqslant \mu \leqslant 2.24$$

with 95% confidence. This is 95% ***confidence interval*** for μ.

21.2 Confidence intervals for μ, variance unknown

'Student' was the pseudonym of William Sealey Gosset, who was a statistician working in Guinness's Brewery about a century ago. See Wikipedia for some background. His work involved testing fields of barley, etcetera; small samples, unlike much-repeatable laboratory measurements.

Here his methods are applied to estimating the mean of a normal distribution, *without knowing σ.* Except for that difference, confidence intervals can be constructed as previously (with σ known).

As usual, given $X_1, \ldots, X_n$,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$\overline{X}$ is close to μ and S^2 is close to σ^2.

For the previous example,

```
-0.68 1.69 1.49 1.88 2.75 1.58 2.20 2.32 1.28 1.69
```

$$n = 10, \quad \overline{X} = 1.62, \quad S^2 = .845, \quad S = 0.92$$

$$t = \sqrt{n}\frac{\overline{X} - \mu}{S} \tag{21.1}$$

is approximately $N(0,1)$: this is the Central Limit Theorem. Student's t-distribution is the *true* distribution of this quantity for given n.

As with the estimate of S, there is a small amount of cancellation, and the number of 'degrees of freedom' in the above random variable (21.1) is

$$n - 1.$$

(21.2) Proposition *When sample average and sample standard deviation are taken from a sampling of n iid random variables, all distributed $N(\mu, \sigma^2)$, The t-statistic (defined in Equation 21.1) follows*

Student's t-distribution *with* $n-1$ *degrees of freedom.*

The P% confidence interval for μ, based on Student's t-distribution, is

$$\left[\overline{X} \mp \alpha\frac{S}{\sqrt{n}}\right]$$

where α depends on the percentage level P and on the number of degrees of freedom, $n-1$. ∎

In this example, the t-statistic is

$$\sqrt{10}\left(\frac{1.62 - \mu}{0.92}\right)$$

To get a confidence interval with 95% confidence, we go to the official tables, page 40, noting that the table is organised differently from the Normal distribution. It gives the percentage points. The t-distribution is symmetric, so we can use the 2.5% percentage point. There are 9 degrees of freedom: the point is 2.262 (compare with 1.96!)

$$\sqrt{10}\left(\frac{1.62-\mu}{0.92}\right) \in [-2.262, 2.262]$$

with 95% confidence.

$$\mu \in 1.62 \mp \frac{.92}{\sqrt{10}} \times 2.262$$
$$1.62 \mp 0.658$$
$$\mu \in [0.962, 2.278]$$