0529 Slides

MA 116

May 2025

Attendance

https://forms.gle/CfHqnqaRnusUzkpr8

Estimating a population mean

Idea is similar to estimating a population proportion.
To estimate a population mean, we may obtain a particular sample and calculate its sample mean $\overline{x}$. This value serves as a point estimator of population proportion.

Like before, we want to define an interval estimator of the form $\overline{x} \pm E$, where E is the margin of error.

Fix a level of confidence α. Our guess of definition of E to a level of confidence α would be

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where σ is the population standard deviation and n is the size of a particular sample.

Issue of not knowing σ

Like in the case of estimating p, we do not know σ. In the case of estimating p we solve this issue by claiming

$$\sigma_{\hat{p}} \cong \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Does this work for estimating $\overline{x}$?

If we tentatively define $E = z_{\alpha/2}\frac{s}{\sqrt{n}}$ where s is the standard deviation of a particular sample, we'll encounter a big issue that the true parameter μ does not lie in our interval $\overline{x} \pm E$, $E = z_{\alpha/2}\frac{s}{\sqrt{n}}$ at a rate that's acceptable. (Gosset, p.464.)

To solve this, let's introduce a new random variable t. We define

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}.$$

t is a new variable constructed from the variable $\overline{x}$ just like how we construct z from a change of variable.

We use this variable t to give a better definition of the margin of error E of μ.

Given a quantitative population, the equation

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

is defined with

- the symbol μ represents the population mean (true mean of the entire population)
- $\overline{x} - \mu$ measures how far your sample mean deviates from the assumed population mean
- sample size$=n$
- degree of freedom$=n-1$
- the symbol s represents

Properties of the variable t and its distribution

If the original population is normally distributed, the variable t follows the **Student's t-distribution** with $n-1$ degrees of freedom.

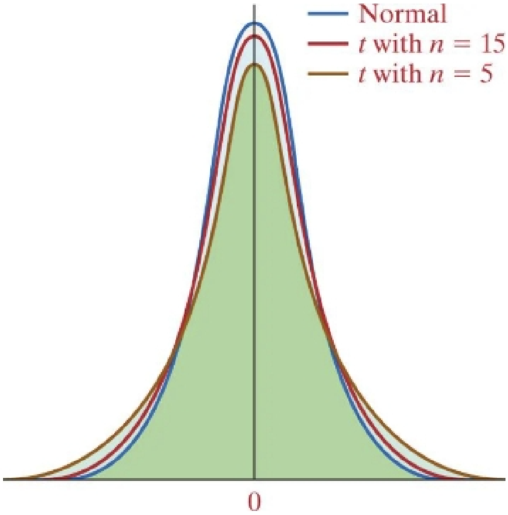
$$f(t) = \frac{\Gamma\left(\frac{df+1}{2}\right)}{\sqrt{df\pi}\,\Gamma\left(\frac{df}{2}\right)}\left(1+\frac{t^2}{df}\right)^{-\frac{df+1}{2}}$$

where,

- $\Gamma(.)$ is the gamma function
- df= Degrees of freedom

1. The Student's t-distribution is centered about 0 and **symmetric** about 0.
2. This is a good probability density function.
3. Approaching 0 at two ends.
4. Very similar to the standard normal curve! except at the two tails.

The Student's t-distribution degrees of freedom $n - 1$ influence tail heaviness, with smaller values yielding heavier tails.



Student's t-distribution

$$f(t) = \frac{\Gamma\left(\frac{df+1}{2}\right)}{\sqrt{df\pi}\,\Gamma\left(\frac{df}{2}\right)}\left(1+\frac{t^2}{df}\right)^{-\frac{df+1}{2}}$$

where,

- $\Gamma(.)$ is the gamma function
- df= Degrees of freedom

This model depends on t and df only. When $n > 30$ (i.e. $df \geq 30$), this model is approximately the standard normal distribution

$$\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-1}{2}z^2\right)$$

Be careful.

1. Student's t-distribution $\neq$ distribution of the variable t.
2. There are two cases one may say the distribution of the variable t can be approximated by the Student's t-distribution.
   1. The underlying population is approximately normally distributed.
   2. $n > 30$ so both the Student's t-distribution and the distribution of the variable t are approximately standard normal.

Student's t-distribution table

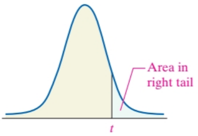


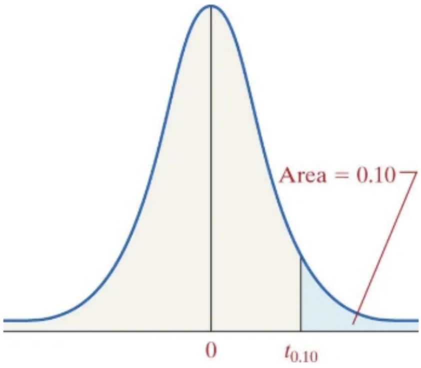
Table VII

***t*-Distribution**
Area in Right Tail

Degrees of Freedom	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.894	31.821	63.657	127.321	318.309	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850

Definition. (t_α.)

t_α represents the value on the horizontal axis of the Student's t-distribution, to the right of which the area under the Student's t-distribution curve is α.



Read the Student's t-distribution table

Figure 14

df	Area in Right Tail 0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.894	31.821	63.657	127.321	318.309	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.215	12.924
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015

The critical value of z with an area to the right of $0.10.$ is approximately 1.28, which is smaller than the critical value of t with 15 degrees of freedom. This is because the t-distribution has more spread than the z-distribution.

Note the table only contain t_α values for $0.005 \leq \alpha \leq 0.25$. In this section, we use α in the construction of level of confidence $(1-\alpha)100\%$, so we only use smaller α.

Definition. (Margin of error–estimating a population mean)

Sample size $=n$. Define $E = t_{\alpha/2}\dfrac{s}{\sqrt{n}}$ where $t_{\alpha/2}$ is with $n-1$ degrees of freedom.

Definition. (Confidence interval–estimating a population mean)

If we obtain a particular sample mean $\overline{x}$, sample standard deviation s, and sample size n. Pick a level of confidence $(1-\alpha)100\%$. We may then calculate $E = t_{\alpha/2}\dfrac{s}{\sqrt{n}}$. A confidence interval of confidence level $(1-\alpha)100\%$ of μ is $[\overline{x}-E, \overline{x}+E]$.

Example. Suppose the underlying population is normally distributed with unknown population mean μ. We want to estimate this μ. We obtained a sample of size 2 $\{1, 4\}$. Let's construct a confidence interval of confidence level 90%.