Set 23: Student's T-Distribution

Stat 260: July 10, 2024

Recall: ***Central Limit Theorem (CLT):*** Let $X_1, X_2, \ldots, X_n$ be a random sample of size n, where n is large ($n \geq 30$) from **any** distribution, with mean μ and standard deviation σ. Then,

$$\overline{X} \sim \text{Normal}(\mu, \sigma/\sqrt{n}) \qquad \text{and} \qquad Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0, 1).$$

Recall: Confidences Intervals for μ

- **Case 1:** σ is known and population is near-normally distributed.

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

- **Case 2:** σ is unknown and $\underline{n \text{ is large } (n \geq 30)}$. ← CLT

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

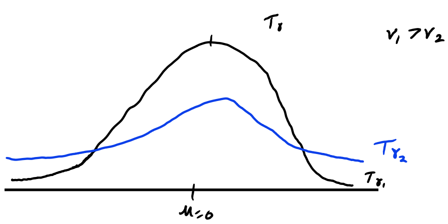
- ***What if σ is unknown and n is small ($n < 30$)?***

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a (near) normal distribution with mean μ and unknown variance. Then,

$T_\nu = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ follows a T-distribution with $\nu = n-1$ "degrees of freedom" (dof)

Properties of T Random Variables

- T random variables are continuous random variables.
- There are infinitely many T random variables.
  - ▷ Each identified by parameter $\nu > 0$, the ***degrees of freedom (dof)***.
  - ▷ T_ν denotes the T-distributed random variable with ν dof.
- The probability density function (pdf) for the T distribution is a symmetric bell curve, always centred at $\mu = 0$.



- The parameter ν dictates the shape of the T pdf:
  - ▷ As ν increases ⇒

    variance decreases (graph becomes taller in the middle)

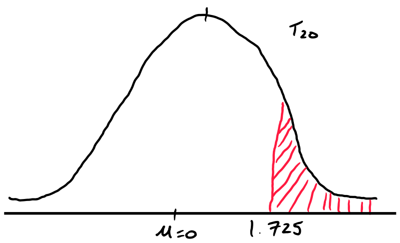
    more dof → more pinched (taller graph)

    as ν approaches ∞ : $T_\infty = Z$ (standard normal distribution)

$\boxed{\text{T-tables give values to the right}}$

Example 1: Determine the following:

(a) For $\nu = 20$, find $P(T_{20} \geq 1.725)$.



$P(T_{20} \geq 1.725) = 0.05$
from T-dist table

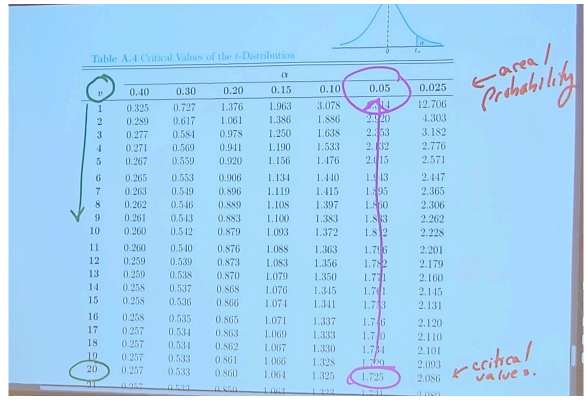
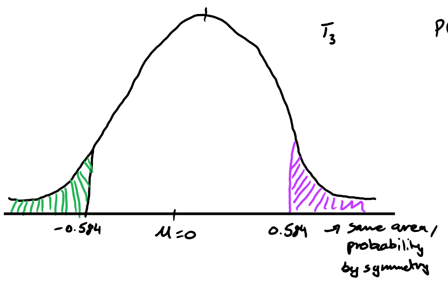


Table A.4 Critical Values of the t-Distribution

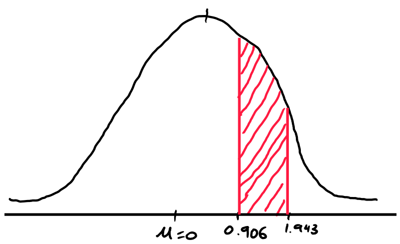
v	0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	[illegible]14	12.706
2	0.289	0.617	1.061	1.386	1.886	2.[illegible]20	4.303
3	0.277	0.584	0.978	1.250	1.638	2.[illegible]53	3.182
4	0.271	0.569	0.941	1.190	1.533	2.[illegible]32	2.776
5	0.267	0.559	0.920	1.156	1.476	2.[illegible]15	2.571
6	0.265	0.553	0.906	1.134	1.440	1.[illegible]43	2.447
7	0.263	0.549	0.896	1.119	1.415	1.[illegible]95	2.365
8	0.262	0.546	0.889	1.108	1.397	1.[illegible]60	2.306
9	0.261	0.543	0.883	1.100	1.383	1.8[illegible]3	2.262
10	0.260	0.542	0.879	1.093	1.372	1.8[illegible]2	2.228
11	0.260	0.540	0.876	1.088	1.363	1.79[illegible]6	2.201
12	0.259	0.539	0.873	1.083	1.356	1.7[illegible]2	2.179
13	0.259	0.538	0.870	1.079	1.350	1.7[illegible]1	2.160
14	0.258	0.537	0.868	1.076	1.345	1.7[illegible]1	2.145
15	0.258	0.536	0.866	1.074	1.341	1.7[illegible]3	2.131
16	0.258	0.535	0.865	1.071	1.337	1.7[illegible]6	2.120
17	0.257	0.534	0.863	1.069	1.333	1.7[illegible]0	2.110
18	0.257	0.534	0.862	1.067	1.330	1.7[illegible]4	2.101
19	0.257	0.533	0.861	1.066	1.328	[illegible]	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086

(b) $P(T_3 \leq -0.584)$.



$P(T \leq -0.581) = P(T_3 \geq 0.584)$
$= 0.3$ → from table

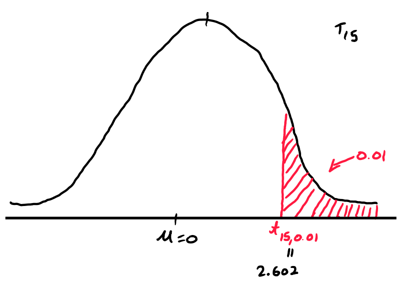
(c) $P(0.906 \leq T_6 \leq 1.943)$ $= P(T_6 \geq 0.906) - P(T_6 \geq 1.943)$



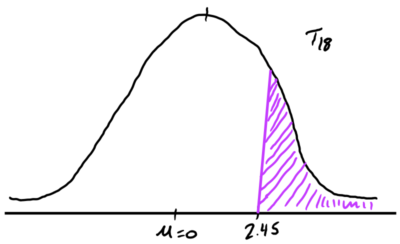
$= 0.2 - 0.05$

$= 0.15$ ← can never be negative so you know answer is wrong if its negative

(d) For $\nu = 15$, determine the critical value: $t_{\underset{\nu}{15},\underset{\alpha}{0.01}}$



(e) Determine $P(T_{18} > 2.45)$



$P(T_{18} > 2.45) \Rightarrow$ not in table
So you take two closest values from table and give it as a bound ↓

$0.01 < P(T_{18} > 2.45) < 0.015$

(f) Determine the value c such that $P(-c < T_{10} < c) = 0.90$.

post

Confidence Intervals for the mean μ

- **Case 3:** σ is unknown and n is small $(n < 30)$.

  When the population is near-normal, with unknown standard deviation and a small sample size, the $(100 - \alpha)$ CI is given by:

$$\underbrace{\bar{X}}_{\text{point estimator}} \pm \underbrace{t_{r,\frac{\alpha}{2}}}_{\text{critical values}} \cdot \underbrace{\frac{s}{\sqrt{n}}}_{\text{estimated standard error}}$$

sample mean

Example 2: The temperatures from 8 weather stations from around Victoria are measured, and a sample temperature of 14.2°C and sample standard deviation of 2.1°C are found. Find the 99% CI for the true mean temperature μ.

(reasonable to assume temp readings from same region are normally distributed)

$\bar{x} = 14.2°c$ $s = 2.1°c$ $n = 8$

Since n is small ($n < 30$), σ is unknown, and population is near-normal use a $t_{\nu, \alpha/2}$ critical value.

$99\%\ CI = 1 - \alpha = 0.99 \rightarrow \alpha = 0.01 \rightarrow \alpha/2 = 0.005$ $(\nu = n-1 = 8-1 = 7)$

$t_{7, 0.005} = 3.499$

$14.2 \pm (3.499)\frac{2.1}{\sqrt{8}} = 14.2 \pm 2.60°c$ or $[L_1, L_2]$ form

$\rightarrow [11.6°c, 16.8°c]$

Extra Example 1: Seven tomato plants are treated regularly with a standard commercial 5-10-10 fertilizer (5% Nitrogen, 10% Phosphorus, and 10% Potassium). At the end of the growing season, their total fruit yields are recorded below (in kg). Determine the 90% for the true mean fruit yield of the fertilized plants.

5.5 4.0 4.1 4.7 5.2 5.8 4.3

Ans: $[4.28, 5.32]$

Readings: Swartz 6.1.1 [EPS bottom half of p. 204 – end of 5.5]
Practice problems: EPS 5.9, 5.11
Devore 7ed: Read §7.2 to pg 265 [we use n ≥ 30 as large] and §7.3. Practice Problems §7.2: 13, and §7.3: 29, 33, 35a, 37a, 39b, 41.