GOSSET'S t DISTRIBUTION [DEVORE 9.2]

- **GOSSET'S t DISTRIBUTION (DEFINITION):**

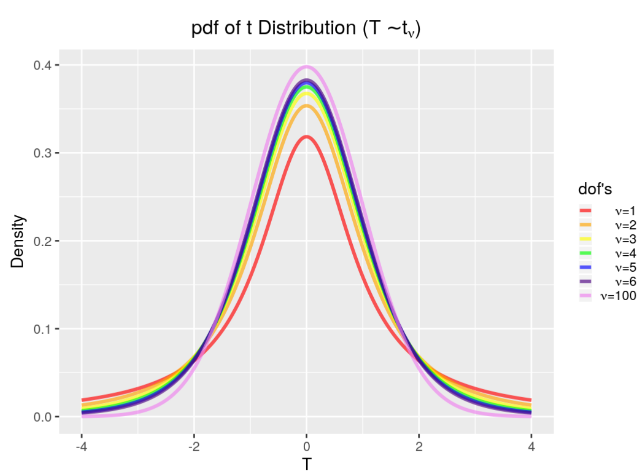
Notation	$T \sim t_\nu$
Parameters	$\nu \equiv \#$ Degrees of Freedom $(\nu = 1, 2, 3, \cdots)$
Support	$\text{Supp}(T) = (-\infty, \infty)$
pdf	$f_T(t;\nu) := \dfrac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\cdot\Gamma(\nu/2)} \cdot \dfrac{1}{[1+(t^2/\nu)]^{(\nu+1)/2}}$
cdf	$\Phi_t(t;\nu) = \dfrac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\cdot\Gamma(\nu/2)} \displaystyle\int_{-\infty}^{t} \dfrac{1}{[1+(\tau^2/\nu)]^{(\nu+1)/2}}\, d\tau$
Mean	$\mathbb{E}[T] = +\infty$, for $\nu = 1$ $\mathbb{E}[T] = 0$, for $\nu > 1$
Variance	$\mathbb{V}[T] = +\infty$, for $\nu = 1, 2$ $\mathbb{V}[T] = \nu/(\nu-2)$, for $\nu > 2$
Model(s)	(Used exclusively for Statistical Inference)

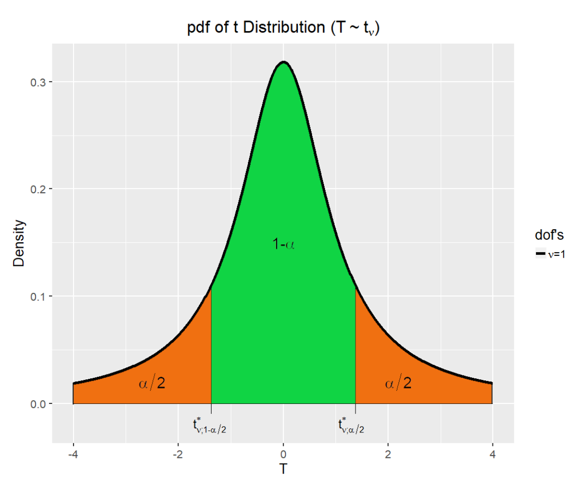
ν is the lowercase Greek letter "nu" $\quad$ τ is the lowercase Greek letter "tau"

- **GOSSET'S t DISTRIBUTION (PROPERTIES):**
  - The t_ν pdf curve is symmetric, bell-shaped and centered at zero.
  - The t_ν pdf curve is more spread out than the std normal pdf curve.
  - The spread of the t_ν pdf curve decreases as ν increases.
  - As $\nu \to \infty$, the t_ν pdf curves approaches the std normal pdf curve.
  - Let independent rv's $\begin{cases} Z \sim \text{StdNormal} \\ X \sim \chi^2_\nu \end{cases}$. Then $\dfrac{Z}{\sqrt{X/\nu}} \sim t_\nu$

- **t-CUTOFFS (AKA t CRITICAL VALUES):**
  - Value $t^*_{\nu;\alpha}$ is a **t-cutoff (AKA t critical value)** of t_ν distribution such that $\mathbb{P}(T > t^*_{\nu;\alpha}) = \alpha$
  - NOTE: Do not confuse t-cutoff $t^*_{\nu;\alpha}$ with t percentile $t_{\nu;\alpha}$: $\quad \mathbb{P}(T \le t_{\nu;\alpha}) = \alpha$
  - Lower-tail t-cutoffs can be found from upper-tail t-cutoffs: $\quad t^*_{\nu;1-\alpha} = -t^*_{\nu;\alpha}$ (this follows since t_v is symmetric)

- **GOSSET'S t DISTRIBUTION (EXAMPLE PDF PLOTS):**





INDEPENDENT t-TESTS & t-CI'S FOR $\mu_1 - \mu_2$ [DEVORE 9.2]

- **A STATISTIC RELATED TO THE t DISTRIBUTION:**

  Let $\mathbf{X} := (X_1, \ldots, X_{n_1})$ be a random sample from a Normal(μ_1, σ_1^2) population.

  Let $\mathbf{Y} := (Y_1, \ldots, Y_{n_2})$ be a random sample from a Normal(μ_2, σ_2^2) population.

  Moreover, suppose random samples $\mathbf{X}$ & $\mathbf{Y}$ are independent of each other.

  Then the following statistic has a t distribution: $$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{approx}{\sim} t_{\nu^*}, \text{ where } \nu^* = \left\lfloor \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \right\rfloor$$

- **INDEPENDENT t-TESTS FOR $\mu_1 - \mu_2$:**

Population:	Two Normal Populations with unknown σ_1, σ_2
Realized Samples:	$\mathbf{x} := (x_1, x_2, \cdots, x_{n_1})$ with mean $\overline{x}$, std dev s_1 $\mathbf{y} := (y_1, y_2, \cdots, y_{n_2})$ with mean $\overline{y}$, std dev s_2 Samples $\mathbf{x}$ & $\mathbf{y}$ are independent of each other
Test Statistic Value: $W(\mathbf{x}, \mathbf{y}; \delta_0)$	$t = \frac{(\overline{x} - \overline{y}) - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$, $\nu^* = \left\lfloor \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \right\rfloor$

HYPOTHESIS TEST:	REJECTION REGION AT LVL α:	P-VALUE DETERMINATION:
$H_0 : \mu_1 - \mu_2 = \delta_0$ vs. $H_A : \mu_1 - \mu_2 > \delta_0$	$t \geq t^*_{\nu^*;\alpha}$	P-value $= 1 - \Phi_t(t; \nu^*)$
$H_0 : \mu_1 - \mu_2 = \delta_0$ vs. $H_A : \mu_1 - \mu_2 < \delta_0$	$t \leq t^*_{\nu^*;1-\alpha}$	P-value $= \Phi_t(t; \nu^*)$
$H_0 : \mu_1 - \mu_2 = \delta_0$ vs. $H_A : \mu_1 - \mu_2 \neq \delta_0$	$t \leq t^*_{\nu^*;1-\alpha/2}$ or $t \geq t^*_{\nu^*;\alpha/2}$	P-value $= 2 \cdot [1 - \Phi_t(\lvert t \rvert; \nu^*)]$

DECISION RULE:
If P-value $\leq \alpha$ then reject H_0 in favor of H_A
If P-value $> \alpha$ then accept H_0 (i.e. fail to reject H_0)

- **INDEPENDENT t-CI's FOR $\mu_1 - \mu_2$:**

  Given two normal populations with means μ_1 and μ_2 and unknown σ_1, σ_2.

  Let $x_1, x_2, \cdots, x_{n_1}$ be a sample taken from the 1^{st} population with mean $\overline{x}$ and variance s_1^2.

  Let $y_1, y_2, \cdots, y_{n_2}$ be a sample taken from the 2^{nd} population with mean $\overline{y}$ and variance s_2^2.

  Moreover, suppose samples $\mathbf{x}$ & $\mathbf{y}$ are independent of each other.

  Then the $100(1-\alpha)\%$ **independent t-CI for** $\mu_1 - \mu_2$ is $$\left((\overline{x} - \overline{y}) - t^*_{\nu^*;\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \ \ (\overline{x} - \overline{y}) + t^*_{\nu^*;\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

  OR WRITTEN MORE COMPACTLY: $$(\overline{x} - \overline{y}) \pm t^*_{\nu^*;\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{where} \quad \nu^* = \left\lfloor \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \right\rfloor$$

POOLED t-TESTS & t-CI'S FOR $\mu_1 - \mu_2$ [DEVORE 9.2]

- **A POOLED STATISTIC RELATED TO THE t DISTRIBUTION:**

Let $\mathbf{X} := (X_1, \ldots, X_{n_1})$ be a random sample from a Normal(μ_1, σ^2) population.

Let $\mathbf{Y} := (Y_1, \ldots, Y_{n_2})$ be a random sample from a Normal(μ_2, σ^2) population.

Moreover, suppose random samples $\mathbf{X}$ & $\mathbf{Y}$ are independent of each other.

Then:

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{S^2_{pool} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2} \quad \text{where} \quad S^2_{pool} := \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

S^2_{pool} is the weighted average of the two sample variances. This means the sample with more data provides more information about the population variance σ^2 and, hence, its sample variance has more weight in the average.

- **POOLED t-TESTS FOR $\mu_1 - \mu_2$:**

Population:	Two Normal Populations with unknown $\sigma_1 = \sigma_2$
Realized Samples:	$\mathbf{x} := (x_1, x_2, \cdots, x_{n_1})$ with mean $\overline{x}$, std dev s_1 $\mathbf{y} := (y_1, y_2, \cdots, y_{n_2})$ with mean $\overline{y}$, std dev s_2 Samples $\mathbf{x}$ & $\mathbf{y}$ are independent of each other
Pooled Sample Variance	$s^2_{pool} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$
Test Statistic Value: $W(\mathbf{x}, \mathbf{y}; \delta_0)$	$t_{pool} = \frac{(\overline{x} - \overline{y}) - \delta_0}{\sqrt{s^2_{pool} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \quad \nu_{pool} = n_1 + n_2 - 2$

HYPOTHESIS TEST:	REJECTION REGION AT LVL α :	P-VALUE DETERMINATION:
$H_0 : \mu_1 - \mu_2 = \delta_0$ vs. $H_A : \mu_1 - \mu_2 > \delta_0$	$t_{pool} \geq t^*_{\nu_{pool};\alpha}$	P-value $= 1 - \Phi_t(t_{pool}; \nu_{pool})$
$H_0 : \mu_1 - \mu_2 = \delta_0$ vs. $H_A : \mu_1 - \mu_2 < \delta_0$	$t_{pool} \leq t^*_{\nu_{pool};1-\alpha}$	P-value $= \Phi_t(t_{pool}; \nu_{pool})$
$H_0 : \mu_1 - \mu_2 = \delta_0$ vs. $H_A : \mu_1 - \mu_2 \neq \delta_0$	$t_{pool} \leq t^*_{\nu_{pool};1-\alpha/2}$ or $t_{pool} \geq t^*_{\nu_{pool};\alpha/2}$	P-value $= 2 \cdot [1 - \Phi_t(\lvert t_{pool} \rvert; \nu_{pool})]$

DECISION RULE: If P-value $\leq \alpha$ then reject H_0 in favor of H_A
If P-value $> \alpha$ then accept H_0 (i.e. fail to reject H_0)

- **POOLED t-CI's FOR $\mu_1 - \mu_2$:**

Given two normal populations with means μ_1 and μ_2 and unknown σ_1, σ_2.

Let $x_1, x_2, \cdots, x_{n_1}$ be a sample taken from the 1^{st} population with mean $\overline{x}$ and variance s_1^2.

Let $y_1, y_2, \cdots, y_{n_2}$ be a sample taken from the 2^{nd} population with mean $\overline{y}$ and variance s_2^2.

Moreover, suppose samples $\mathbf{x}$ & $\mathbf{y}$ are independent of each other.

Then the $100(1-\alpha)\%$ **pooled t-CI for** $\mu_1 - \mu_2$ is $\quad (\overline{x} - \overline{y}) \pm t^*_{\nu_{pool};\alpha/2} \cdot \sqrt{s^2_{pool} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

where $\quad s^2_{pool} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \quad$ and $\quad \nu_{pool} = n_1 + n_2 - 2$

GOSSET'S t-CUTOFFS, $t^*_{\nu;\alpha}$ $\quad \mathbb{P}(T > t^*_{\nu;\alpha}) = \alpha, \quad t^*_{\nu;1-\alpha} = -t^*_{\nu;\alpha}$

ν \ α	**0.2**	**0.1**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.001**	**0.0005**
1	1.376	3.078	6.314	12.706	15.895	31.821	63.657	318.309	636.619
2	1.061	1.886	2.920	4.303	4.849	6.965	9.925	22.327	31.599
3	0.978	1.638	2.353	3.182	3.482	4.541	5.841	10.215	12.924
4	0.941	1.533	2.132	2.776	2.999	3.747	4.604	7.173	8.610
5	0.920	1.476	2.015	2.571	2.757	3.365	4.032	5.893	6.869
6	0.906	1.440	1.943	2.447	2.612	3.143	3.707	5.208	5.959
7	0.896	1.415	1.895	2.365	2.517	2.998	3.499	4.785	5.408
8	0.889	1.397	1.860	2.306	2.449	2.896	3.355	4.501	5.041
9	0.883	1.383	1.833	2.262	2.398	2.821	3.250	4.297	4.781
10	0.879	1.372	1.812	2.228	2.359	2.764	3.169	4.144	4.587
11	0.876	1.363	1.796	2.201	2.328	2.718	3.106	4.025	4.437
12	0.873	1.356	1.782	2.179	2.303	2.681	3.055	3.930	4.318
13	0.870	1.350	1.771	2.160	2.282	2.650	3.012	3.852	4.221
14	0.868	1.345	1.761	2.145	2.264	2.624	2.977	3.787	4.140
15	0.866	1.341	1.753	2.131	2.249	2.602	2.947	3.733	4.073
16	0.865	1.337	1.746	2.120	2.235	2.583	2.921	3.686	4.015
17	0.863	1.333	1.740	2.110	2.224	2.567	2.898	3.646	3.965
18	0.862	1.330	1.734	2.101	2.214	2.552	2.878	3.610	3.922
19	0.861	1.328	1.729	2.093	2.205	2.539	2.861	3.579	3.883
20	0.860	1.325	1.725	2.086	2.197	2.528	2.845	3.552	3.850
21	0.859	1.323	1.721	2.080	2.189	2.518	2.831	3.527	3.819
22	0.858	1.321	1.717	2.074	2.183	2.508	2.819	3.505	3.792
23	0.858	1.319	1.714	2.069	2.177	2.500	2.807	3.485	3.768
24	0.857	1.318	1.711	2.064	2.172	2.492	2.797	3.467	3.745
25	0.856	1.316	1.708	2.060	2.167	2.485	2.787	3.450	3.725
26	0.856	1.315	1.706	2.056	2.162	2.479	2.779	3.435	3.707
27	0.855	1.314	1.703	2.052	2.158	2.473	2.771	3.421	3.690
28	0.855	1.313	1.701	2.048	2.154	2.467	2.763	3.408	3.674
29	0.854	1.311	1.699	2.045	2.150	2.462	2.756	3.396	3.659
30	0.854	1.310	1.697	2.042	2.147	2.457	2.750	3.385	3.646
31	0.853	1.309	1.696	2.040	2.144	2.453	2.744	3.375	3.633
32	0.853	1.309	1.694	2.037	2.141	2.449	2.738	3.365	3.622
33	0.853	1.308	1.692	2.035	2.138	2.445	2.733	3.356	3.611
34	0.852	1.307	1.691	2.032	2.136	2.441	2.728	3.348	3.601
35	0.852	1.306	1.690	2.030	2.133	2.438	2.724	3.340	3.591
36	0.852	1.306	1.688	2.028	2.131	2.434	2.719	3.333	3.582
37	0.851	1.305	1.687	2.026	2.129	2.431	2.715	3.326	3.574
38	0.851	1.304	1.686	2.024	2.127	2.429	2.712	3.319	3.566
39	0.851	1.304	1.685	2.023	2.125	2.426	2.708	3.313	3.558
40	0.851	1.303	1.684	2.021	2.123	2.423	2.704	3.307	3.551
60	0.848	1.296	1.671	2.000	2.099	2.390	2.660	3.232	3.460
120	0.845	1.289	1.658	1.980	2.076	2.358	2.617	3.160	3.373
500	0.842	1.283	1.648	1.965	2.059	2.334	2.586	3.107	3.310

EX 9.2.1: A manufacturing plant produces a certain type of brake pad used in big rig trucks.

A sample of 10 day-shift workers showed that the mean # of brake pads produced was 52 with a std dev of 4.4.

A sample of 13 night-shift workers showed that the mean # of brake pads produced was 55 with a std dev of 4.6.

Assume that the two samples are independent and the two corresponding populations are normally distributed.

Do these samples suggest that the average # brake pads produced by the two shifts differ?

(Use significance level $\alpha = 0.01$)

(a) State the appropriate null hypothesis H_0 & alternative hypothesis H_A.

(b) Compute the independent t-test statistic value for this hypothesis test.

(c) Identify the two appropriate independent t-cutoffs and lookup/compute their values.

(d) Compute the appropriate P-value using software.

(e) Using the computed independent t-cutoffs, render the appropriate decision.

(f) Using the computed P-value, render the appropriate decision.

(g) Construct the approximate 99% independent t-CI for $\mu_1 - \mu_2$.

EX 9.2.2: A manufacturing plant produces a certain type of brake pad used in big rig trucks.

A sample of 10 day-shift workers showed that the mean # of brake pads produced was 52 with a std dev of 4.4.

A sample of 13 night-shift workers showed that the mean # of brake pads produced was 55 with a std dev of 4.6.

Assume that the two samples are independent and the two corresponding populations are normally distributed.

Do these samples suggest that the average # brake pads produced by the two shifts differ?

(Use significance level $\alpha = 0.01$)

(a) State the appropriate null hypothesis H_0 & alternative hypothesis H_A.

(b) Compute the pooled t-test statistic value for this hypothesis test.

(c) Identify the two appropriate pooled t-cutoffs and lookup/compute their values.

(d) Compute the appropriate P-value using software.

(e) Using the computed pooled t-cutoffs, render the appropriate decision.

(f) Using the computed P-value, render the appropriate decision.

(g) Construct the 99% pooled t-CI for $\mu_1 - \mu_2$.