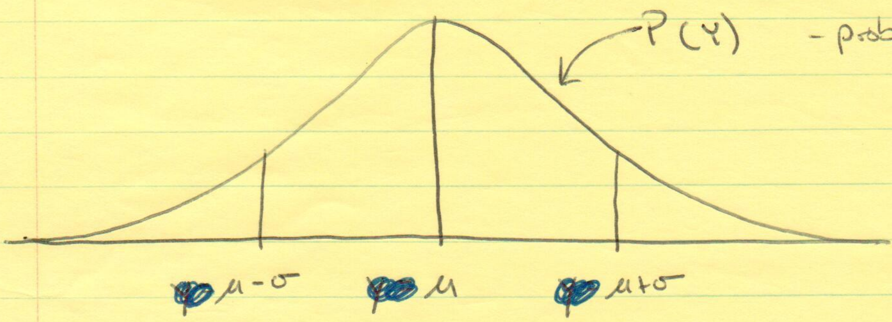
Confidence Intervals

Large sample statistics (# samples > 100)

When to use ~~[scribbled out]~~ Confidence Intervals

- don't have mathematical relationship between independent + dependent variables.
- cause + effect can only be determined from experimental data

Random variable Y has normal distribution
std deviation σ, mean μ



$$\int_{-\infty}^{\infty} P(y)\,dy = 1$$

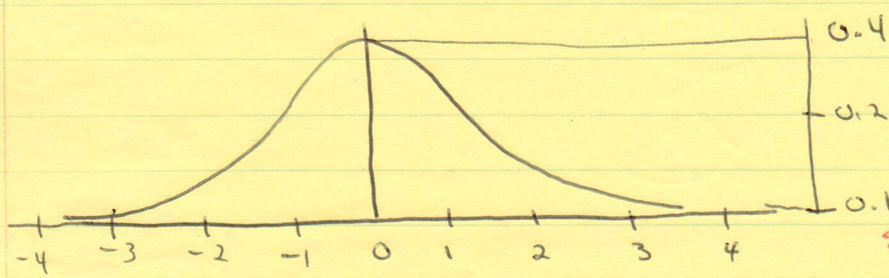
mean probability = 1

We can say:

$$P_r\{\mu-\sigma < y < \mu+\sigma\} = \int_{\mu-\sigma}^{\mu+\sigma} P(y)\,dy * 100 = 67\%$$

Probability y in above interval is 67%

Tables exist to give probability %, but are all based on area = 1



special case
Standard Normal Distribution
$\mu = 0$ $\sigma^2 = 1$

To normalize area + midpoint, define

$$z = \frac{y-\mu}{\sigma}$$

Main problem: must know true mean + std dev. to use.

Confidence intervals

Rarely have large sample size

Small sample statistics

estimate σ from finite data set

$$t = \frac{y - \mu}{s}$$

y = normally distributed random variable

μ = true mean of y

s = estimation of σ based on N data

↑ sample std. deviation

can't know μ either, so use $\bar{y}$ - sample mean

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad \bar{y} \to \mu \text{ as } N \to \infty$$

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2 \qquad s \to \sigma \text{ as } N \to \infty$$

↗ Sample variance

$\nu = N - 1$ Degrees of freedom

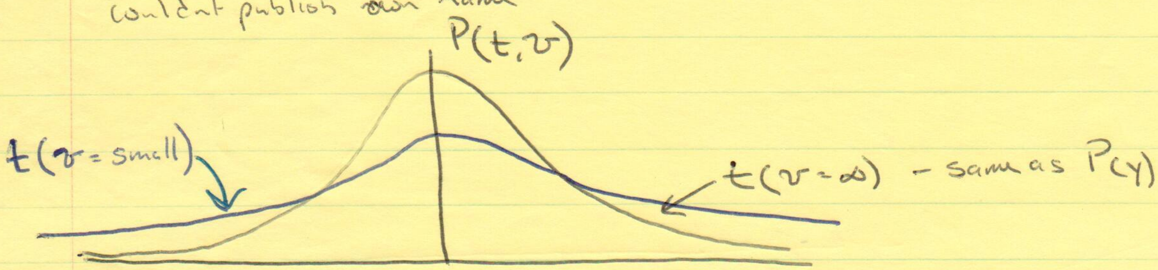
* Break for Activity *

- activity review on "key" form

Introduce need for probability statements dependent on sample size

Students T-statistic - ~~[illegible]~~ Guinness employee ~ 1908

couldn't publish own name

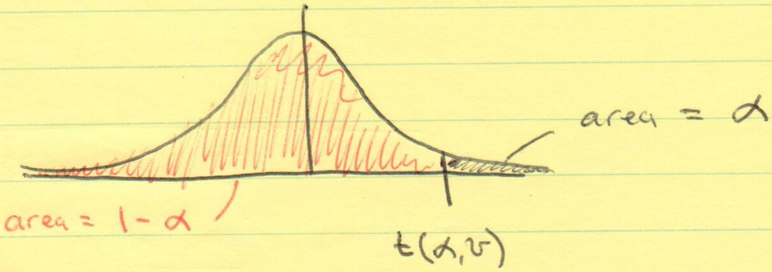


$$\int_{-\infty}^{\infty} P(t,\nu)\,dt = 1 \qquad \int_{-\infty}^{t(\alpha,\nu)} P(t,\nu)\,dt = 1 - \alpha$$

$t = f(\alpha, \nu)$ defined on pg 640 Appendix II

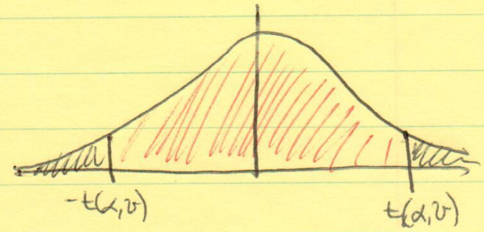
↗ confidence ↑ degrees of freedom

Confidence Interval



this says:

$$Pr\{t < t(\alpha,\nu)\} = (1-\alpha)*100\%$$



$$Pr\{-t(\alpha,\nu) < t < t(\alpha,\nu)\} = (1-2\alpha)*100\%$$

but $t = \frac{y-\mu}{s}$

$$Pr\{-t(\alpha,\nu) < \frac{y-\mu}{s} < t(\alpha,\nu)\} = (1-2\alpha)*100\%$$

algebra

$$Pr\{\mu - t(\alpha,\nu)\cdot s < y < \mu + t(\alpha,\nu)\cdot s\} = (1-2\alpha)\cdot 100\%$$

$\alpha = 0.025 \rightarrow$ 95% confidence

says for given confidence y is between mean − (modifier)(sample st. dev)
and mean + (modifier)(sample st. dev)

look @ a few points on table + effect on equation

$\alpha = 0.025$ – 95% confident

$t \approx 2 \rightarrow$ like 2σ look @ $\nu = \infty$

look @ ν = low #'s

next time, will look @ performance data closer

Sr. des. students stay for survey → respond by email
by Friday afternoon.

II. Percentage Points of the t Distribution[a]

ν \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.727	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.019	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.261	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.260	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	.260	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.259	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.259	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.258	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.258	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.258	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.257	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.257	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.257	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.257	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.257	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	.256	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.256	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.256	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.256	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.256	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.256	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.256	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.256	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.256	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.255	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.254	.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	.254	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.253	.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

ν = degrees of freedom.

[a] Adapted with permission from *Biometrika Tables for Statisticians*, Vol. 1, 3rd edition, by E. S. Pearson and H. O. Hartley, Cambridge University Press, Cambridge, 1966.

ν	.995	.990
1	0.00 +	0.00 +
2	0.01	0.02
3	0.07	0.11
4	0.21	0.30
5	0.41	0.55
6	0.68	0.87
7	0.99	1.24
8	1.34	1.65
9	1.73	2.09
10	2.16	2.56
11	2.60	3.05
12	3.07	3.57
13	3.57	4.11
14	4.07	4.66
15	4.60	5.23
16	5.14	5.81
17	5.70	6.41
18	6.26	7.01
19	6.84	7.63
20	7.43	8.26
25	10.52	11.52
30	13.79	14.95
40	20.71	22.16
50	27.99	29.71
60	35.53	37.48
70	43.28	45.44
80	51.17	53.54
90	59.20	61.75
100	67.33	70.06

ν = degrees of freedom

[a] Adapted with permission fro
H. O. Hartley, Cambridge Uni