A neurologist is testing the effect of a drug on response time by injecting 100 rats with a unit dose of the drug, subjecting each to neurological stimulus, and recording its response time. The neurologist knows that the mean response time for rats not injected with the drug is 1.2 seconds. The mean of the 100 injected rats' response times is 1.05 seconds with a sample standard deviation of 0.5 seconds. Do you think that the drug has an effect on response time?

$n = 100$ $\mu = 1.2s$ $[S = 0.5s, \ \mu_{100} = 1.05s]$ sample standard deviation

reject 0.3%

① null hypothesis H_0 : Drug has no effect

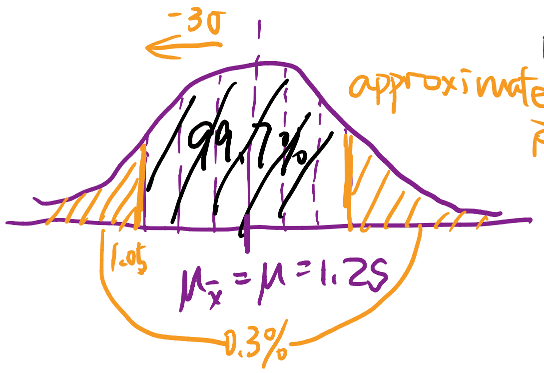
no effect

$\mu = 1.2s$ (even w/ drug)

② Alternative hypothesis H_1 : Drug has an effect

$\mu \neq 1.2s$ when the drug is given

Assume H_0 is true



$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{100}} \underset{\text{approximately}}{\approx} \frac{S}{\sqrt{100}} = \frac{0.5}{10} = 0.05$$

$$\hat{\sigma}_{\bar{x}} = 0.05$$

$$Z = \frac{1.2 - 1.05}{0.05} = \frac{0.15}{0.05} = 3$$

3 standard deviation away from the mean

Look up Z table $3 \rightarrow 99.7\%$

P-value : the probability of getting a result more extreme than this, given the null hypothesis called p-value

P-value = 0.003

If P-value < 0.005 ⇒ Reject the null hypothesis

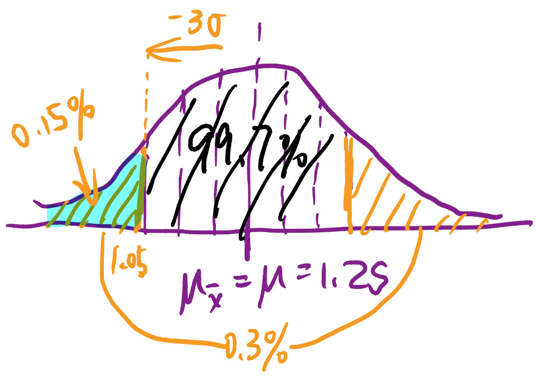
Two-tailed Test 双侧检验　　One-tailed Test 单侧检验

reject

H_0 : Drug has no effect $\mu_{drug} = 1.25$

one-tailed test

H_1 : Drug lowers response time $\mu_{drug} < 1.25$



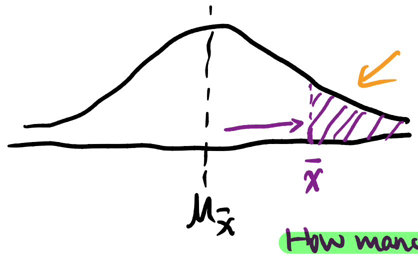
If H_0, P(result lower than 1.05s)

P-value = 0.0015

⇒ Reject H_0

z - statistic   t - statistic

① z - Statistic [n > 30]



How many standard deviation away from the actual mean

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \boxed{\frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}}$$

Unknown

≈ sample standard deviation

$$z \approx \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$$

(n > 30)

"OK", if n > 30

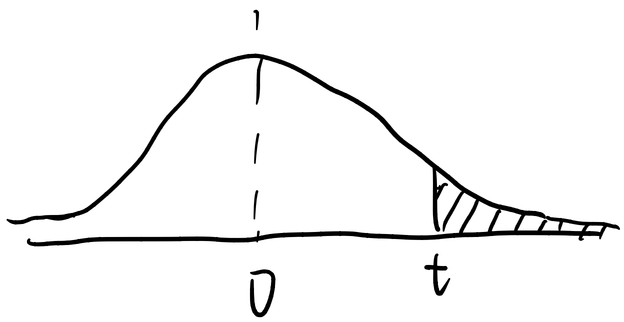
∵ normally distributed

∴ Z-statistic

∵ central limit theorem ⇒ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Then look up in a Z-table or in a normal distribution table

② t-statistic [n < 30]



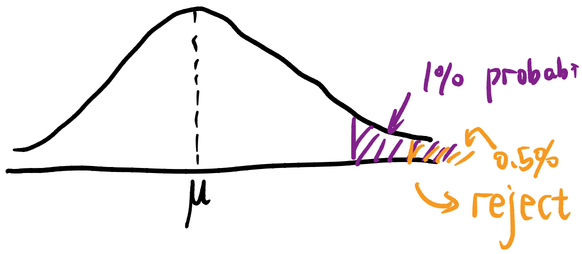
$$t \approx \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$$

(n < 30)

Look up into a t-table

Type I Error : rejecting H_0 even though it's true

Assume H_0 true



means there are 0.5% chance to make a mistake of Type I Error

The mean emission of all engines of a new design needs to be below 20 ppm if the design is to meet new emission requirements. Ten engines are manufactured for testing purposes, and the emission level of each is determined. The emission data is:

15.6 16.2 22.5 20.5 16.4 19.4 16.6 17.9 12.7 13.9 $\bar{X} = 17.17$ $s = 2.98$

$n = 10$ ∴ use t-statistic

Does the data supply sufficient evidence to conclude that this type of engine meets the new standard? Assume we are willing to risk a Type I error with probability = 0.01

$H_0: \mu = 20\text{ppm}$ $H_1: \mu < 20\text{ ppm}$

Assume H_0 true, If $P(\bar{x} = 17.17) < 0.01 \Rightarrow$ reject H_0

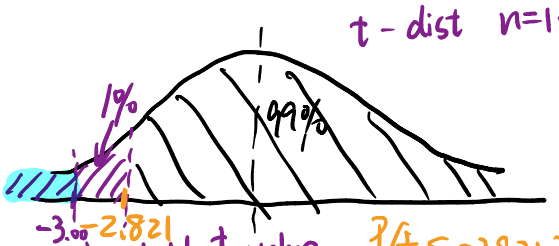
$$t = \frac{17.17 - 20}{\frac{2.98}{\sqrt{10}}} = -3.00$$

∵ $t < -2.821$

∴ reject H_0

t-dist $n=10$ or $10-1=9$ d.f

degrees of freedom



threshold t-value

界限t值

$P(t < -2.821) = 1\%$

***t* Table**

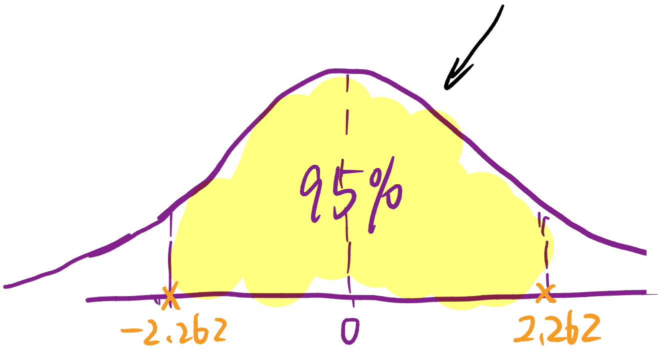
cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587

The mean emission of all engines of a new design needs to be below 20 ppm if the design is to meet new emission requirements. Ten engines are manufactured for testing purposes, and the emission level of each is determined. The emission data is:

15.6 16.2 22.5 20.5 16.4 19.4 16.6 17.9 12.7 13.9 $\bar{x} = 17.17$ $s = 2.98$

Q: 95% confidence interval for the actual mean emission for this new engine design

A: $n = 10 < 30 \Rightarrow t\text{-dist}$ 9 degrees of freedom



t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	$\boxed{0.05}$	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	$\boxed{2.262}$	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587

95% chance

$$-2.262 < t < 2.262$$

$$-2.262 < t = \frac{17.17 - \mu}{\frac{2.98}{\sqrt{10}}} < 2.262$$

过程略

$$19.3 > \mu > 15.04 \rightarrow 95\% \text{ chance}$$

We want to test the hypothesis that more than 30% of U.S. households have Internet access (with a significance level of $\boxed{5\%}$). We collect a sample of 150 households and find that 57 have access.

$n = 150$ $\qquad \bar{p} = 57/150 = 0.38$

↳ Bionomial dist, Z-statistic $\quad np > 5,\ n(1-p) > 5$

approximate it a normal dist

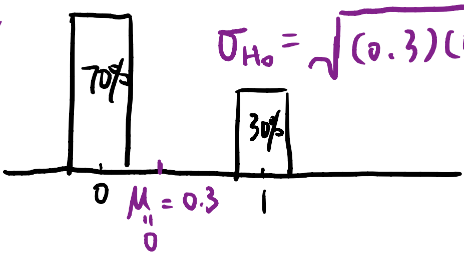
$H_0: P \leq 30\%$ $\qquad H_1: P > 30\%$

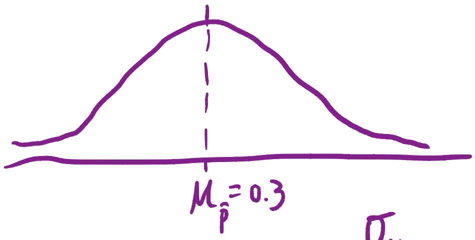
Assume H_0 true

$P_{H_0} = 0.3$

Bernoulli dist

$\sigma_{H_0} = \sqrt{(0.3)(0.7)} = \sqrt{0.21}$





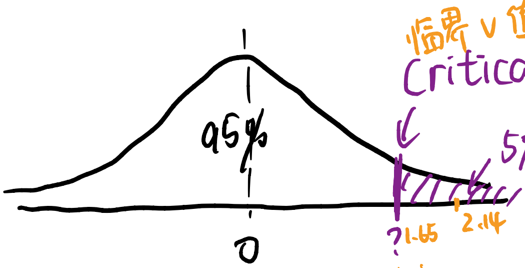
$$\sigma_{\bar{p}} = \frac{\sigma_{H_0}}{\sqrt{n}} = \frac{\sqrt{0.21}}{\sqrt{150}} = 0.037$$

$$Z = \frac{\bar{p} - \mu_{\bar{p}}}{\sigma_{\bar{p}}} = \frac{0.38 - 0.3}{0.037} = \boxed{2.14}$$

2.14 standard deviations above the mean

临界值

Critical z value = 1.65



the next digit

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817

∴ In any normal distribution, the propability of being less than 1.65 standard deviations away from the mean is going to be 95%

∵ Actual Z-value = 2.14 > 1.65

∴ Reject H_0

Deal with sums and differences of random variables

X Y independent random variables

$E(X) = \mu_X$ $E(Y) = \mu_Y$

方差 variance 方差 variance

$Var(X) = E((X-\mu_X)^2) = \sigma_X^2$ $Var(Y) = E((Y-\mu_Y)^2) = \sigma_Y^2$

$Z = X + Y$ $E(Z) = E(X+Y) = E(X) + E(Y)$ $\mu_Z = \mu_X + \mu_Y$

$A = X - Y$ $E(A) = E(X-Y) = E(X) - E(Y)$ $\mu_A = \mu_X - \mu_Y$

$= \mu_{X-Y}$

$Var(Z) = Var(X) + Var(Y)$

$\sigma_Z^2 = \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$

$\sigma_A^2 = \sigma_{X-Y}^2 = \sigma_{X+(-Y)}^2 = \sigma_X^2 + \sigma_{-Y}^2$

$\sigma_{-Y}^2 = Var(-Y) = E((-Y-E(-Y))^2$

$-Y-E(-Y) = (-1)^2(Y+E(-Y))$

$\because E(-Y) = -E(Y)$

$\therefore \sigma_{-Y}^2 = E((Y-E(Y))^2) = \sigma_Y^2$

$\sigma_A^2 = \sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$