Statistics :

Sample vs population

(estimators) (parameters)

Sample variables might be biased estimators of the population parameters (the larger the sample, the smaller the bias)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \qquad \sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Covariances

$$s_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1} \qquad \sigma_{xy} = \frac{\sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)}{N}$$

Correlations

$$r = \frac{s_{xy}}{s_x s_y} \qquad \rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

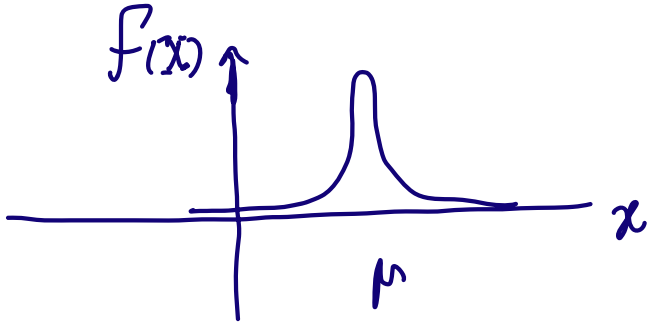
Central Limit Theorem (CLT)

Let's assume that x_{ij} represents the i-th independent variable in the j-th sample. Then, regardless of the underlying distribution of x_{ij}'s, we have:

x_{11}	x_{12}	...	x_{1m}
x_{21}	x_{22}		x_{2m}
x_{31}	x_{32}		x_{3m}
⋮	⋮		⋮
x_{n1}	x_{n2}	...	x_{nm}
$\bar{x}_1$	$\bar{x}_2$		$\bar{x}_m$

$\bar{X}$ (averages)

For large n: $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$



Confidence Intervals (CI) and the Margin of Error

Given a sample with size n:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \text{(Sample mean)}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \qquad \text{(Sample standard deviation)}$$

$$SE \equiv \frac{s}{\sqrt{n}} \qquad \text{(Standard error)}$$

If population variance (σ^2) is known or the sample size is large (eg. $n \gtrsim 40$)

$S \approx \sigma$ and $\bar{x} \approx \mu$ and we have

$$z \equiv \frac{X - \bar{x}}{\sigma/\sqrt{n}}$$

z-statistic $\sigma/\sqrt{n} \rightarrow SE$

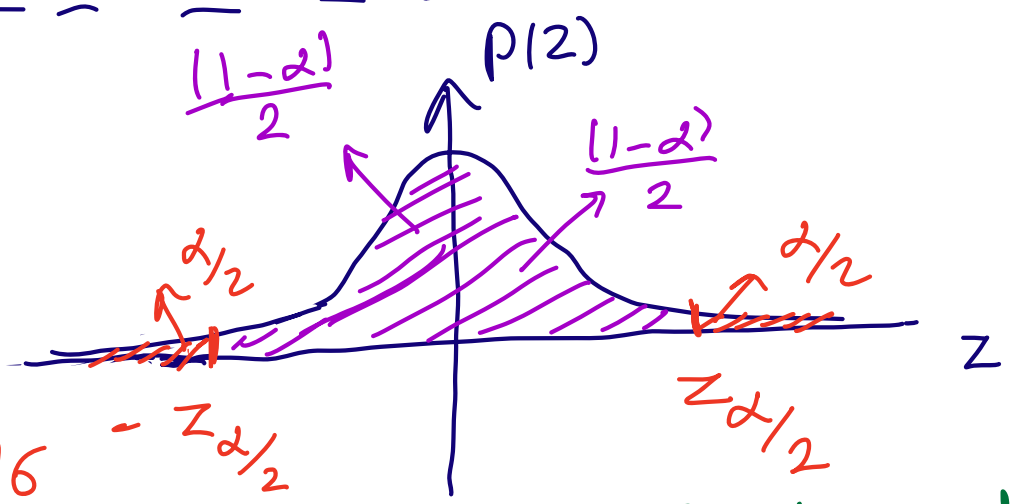
standard normal dist ($\bar{z} \simeq 0$, $\sigma_z \simeq 1$)

CI is defined as:

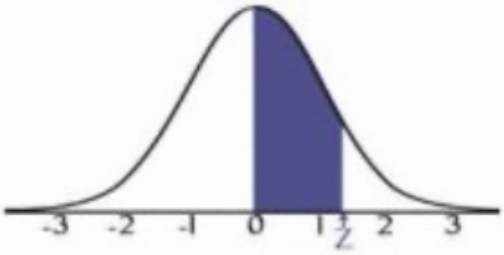
$$X \in \left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \;,\; \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

$1-\alpha$: confidence level (CL)

Ex:
95% CL
$\alpha = 0.05$
$z_{0.025} = 1.96$



(Sometimes called critical value)



STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for $z = 1.25$ the area under the curve between the mean (0) and z is 0.3944.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

Example: physic I Scores of 50 students (out of 100)

$\bar{x} = 81.56$

$\sigma \approx s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}\Big|_{n=50} = 7.68 \Rightarrow SE = \frac{\sigma}{\sqrt{n}} = \frac{7.68}{\sqrt{50}} = 1.086$

$$X \in \bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

1σ CI: $Z_{1\sigma} = 1 \Rightarrow$ $X \in [81.56 \pm 1.086]$
($\sim 68\%$ CL)

2σ CI: $Z_{2\sigma} = 2 \Rightarrow$ $X \in [81.56 \pm 2.172]$
($\sim 95\%$ CL)

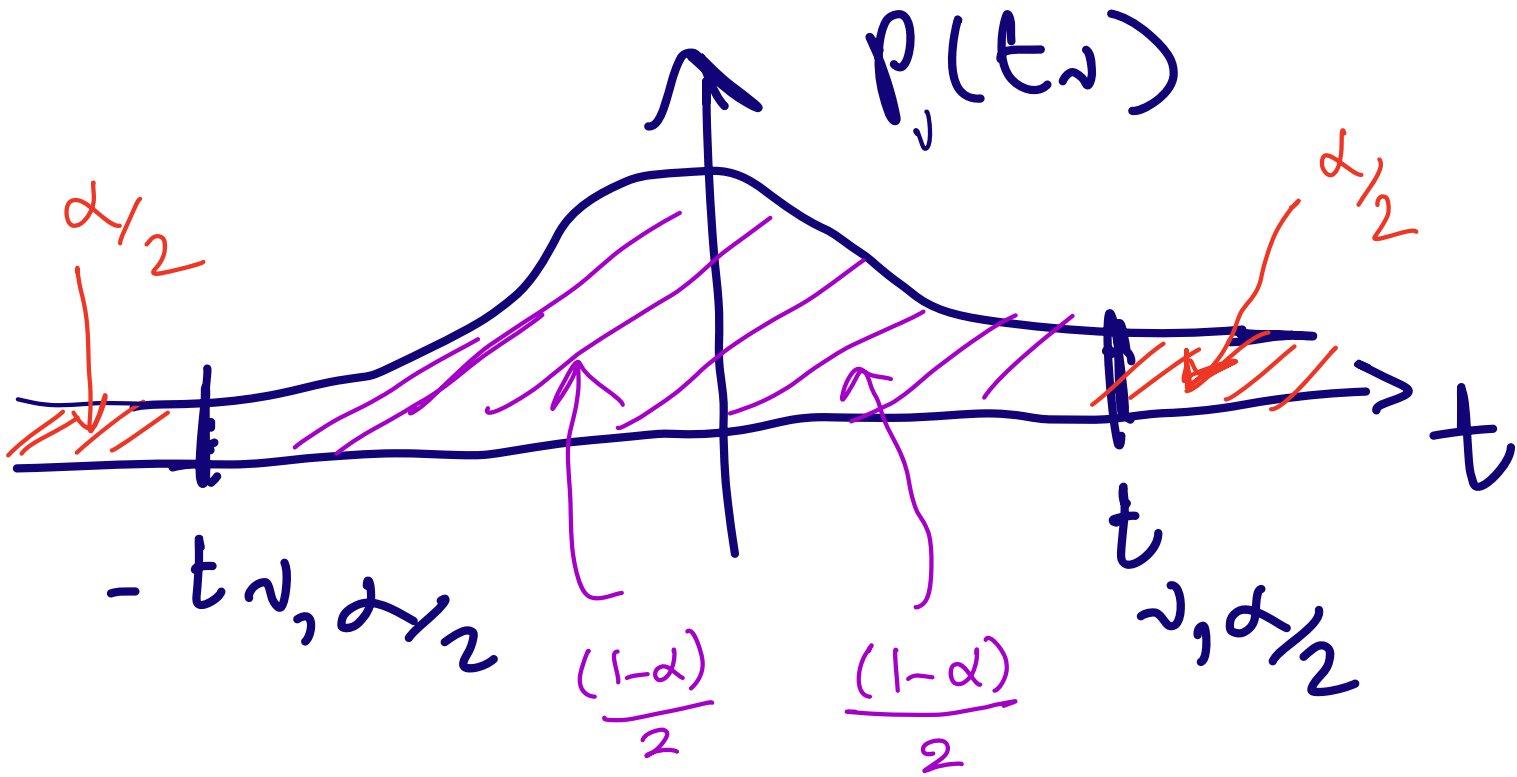
4σ CI $Z_{4\sigma} = 4 \Rightarrow$ $X \in [81.56 \pm 4.344]$
($\sim 100\%$ CL)

If σ is unknown or Sample Size is Small, one can use t-statistic:

$$t_\nu = \frac{\bar{X} - \bar{x}}{s/\sqrt{n}} \qquad \nu = \text{dof} = n-1$$

(n: sample size)

$$CI: \bar{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$$



Student's T distribution

Example: physic I scores of 10 students (out of 100)

$\bar{x} = 82$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = 9.6 \Rightarrow SE = \frac{s}{\sqrt{10}} \simeq 3.04$$

$n = 10$

$\nu = n - 1 = 9$

95% CI ?

From t-tables, the corresponding critical value, $t_{9, 0.025}$, is 2.262. Therefore.

$$X \in [82 \pm 6.88] \quad 95\% \text{ CI}$$

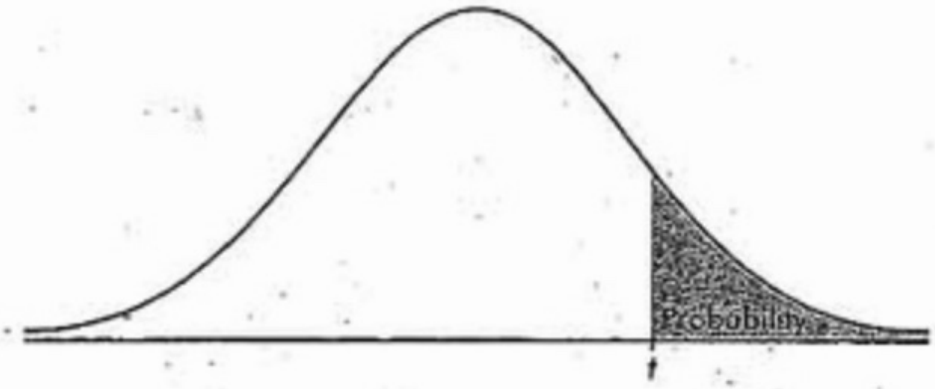


TABLE B: *t*-DISTRIBUTION CRITICAL VALUES

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725

CI for two samples

Dependent Samples:

Example: Before & after of a drug given to people.

Before	after	Difference
x_1	y_1	$x_1 - y_1$
x_2	y_2	$x_2 - y_2$
$\vdots$	$\vdots$	$\vdots$
x_N	y_N	$x_N - y_N$

average difference: $\bar{d}$

Sample std of differences: s_d

(should be calculated directly)

$$CI: \bar{d} \pm t_{n-1, \alpha/2} \frac{s_d}{\sqrt{n}}$$

Independent samples (σ's known or Samples are large)

Test Scores (Economics Class)

	Engineering x	Management y
Size	100	70
Sample mean	58	65
Sample std	10	5

z-statistic: $\bar{d} \pm z_{\alpha/2} SE_d$

$\bar{d} = \bar{x} - \bar{y}$, $SE_d = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$

CI: $\bar{d} \pm z_{\alpha/2} SE_d$: $(-9.27, -4.73)$

$\bar{d}$: -7 ; $z_{\alpha/2}$: 1.96 ; SE_d: 1.16

95% CI

Independent Samples (unknown population variances assumed equal)

City 1 apples	City 2 apples
3.8	3.02
3.76	3.22
3.87	3.24
3.99	3.02
4.02	3.06
4.25	3.15
4.13	3.81
3.98	3.44
3.99	
3.62	
Size: 10	8
Sample mean: 3.94	3.25
Sample std: 0.18	0.27

please check calculations!

pooled variance formula:

$$s_p^2 = \frac{(n_x - 1)\, s_x^2 + (n_y - 1)\, s_y^2}{n_x + n_y - 2}$$

t-statistic:

$$CI: (\bar{x} - \bar{y}) \pm t_{n_x + n_y - 2,\, \alpha/2} \sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}$$

$\Rightarrow$ CI: 95% (0.47, 0.92)

Independent Samples (unknown population variances assumed different)

$$CI: \quad \bar{x} - \bar{y} \pm t_{\nu, \alpha/2} \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$$

$$\nu \approx \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\left(\frac{s_x^2}{n_x}\right)^2 / (n_x - 1) + \left(\frac{s_y^2}{n_y}\right)^2 / (n_y - 1)}$$

Hypothesis Testing

Null hypothesis (H_0): usually refers to status quo
(to be tested)

Alternative hypothesis (H_1): Usually refers to change/innovation

α: Significance level: The (maximum acceptable) probability of rejecting the null hypothesis while it is true (the probability of making this error)

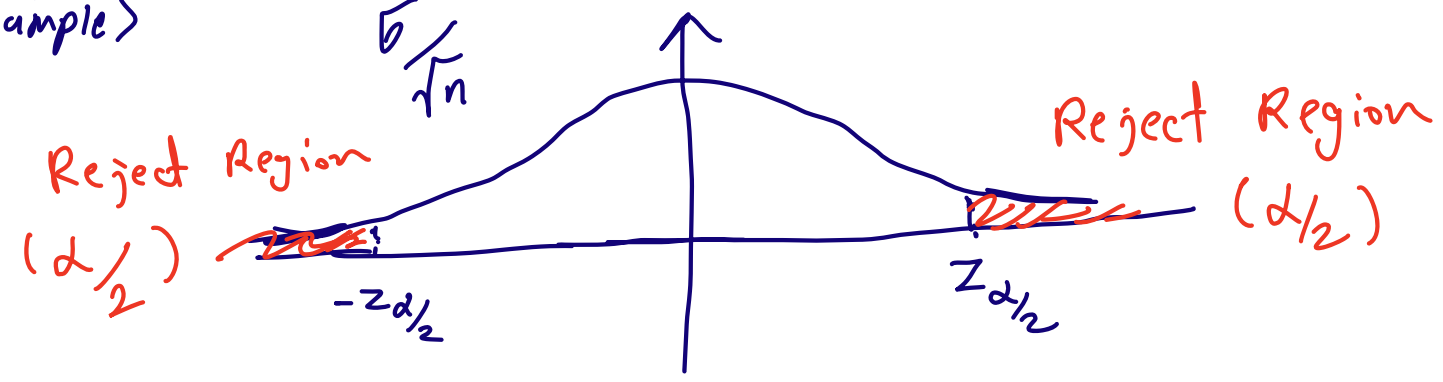
max acceptable

Common choices [$\alpha = 0.01, 0.05, 0.1$]

For example: two-sided test

(large sample) $Z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

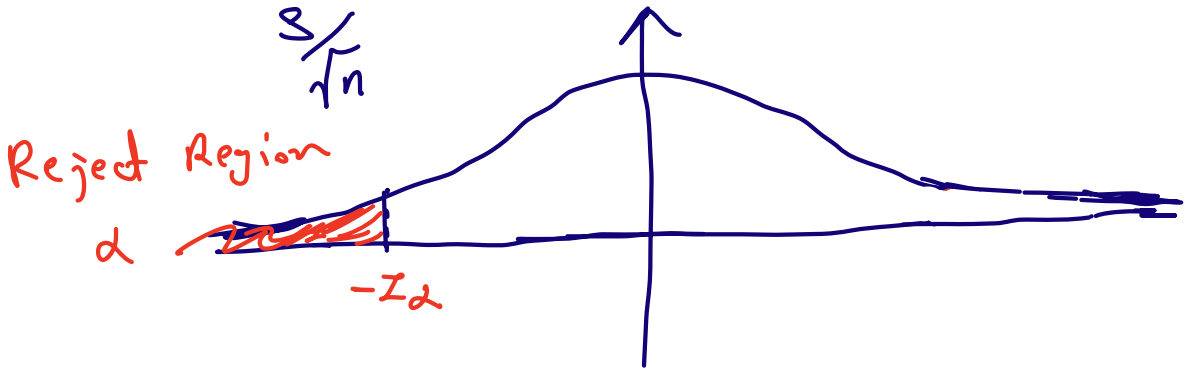
$\mu_0 \rightarrow$ Hypothesized mean (null)



One-sided test

$$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$\mu_0 \rightarrow$ Hypothesized mean (null)



Example:

Single population, known variance or large sample (Z-test)

Salary for a specific job

Sample mean: 100200

$\sigma \approx s = 15000$

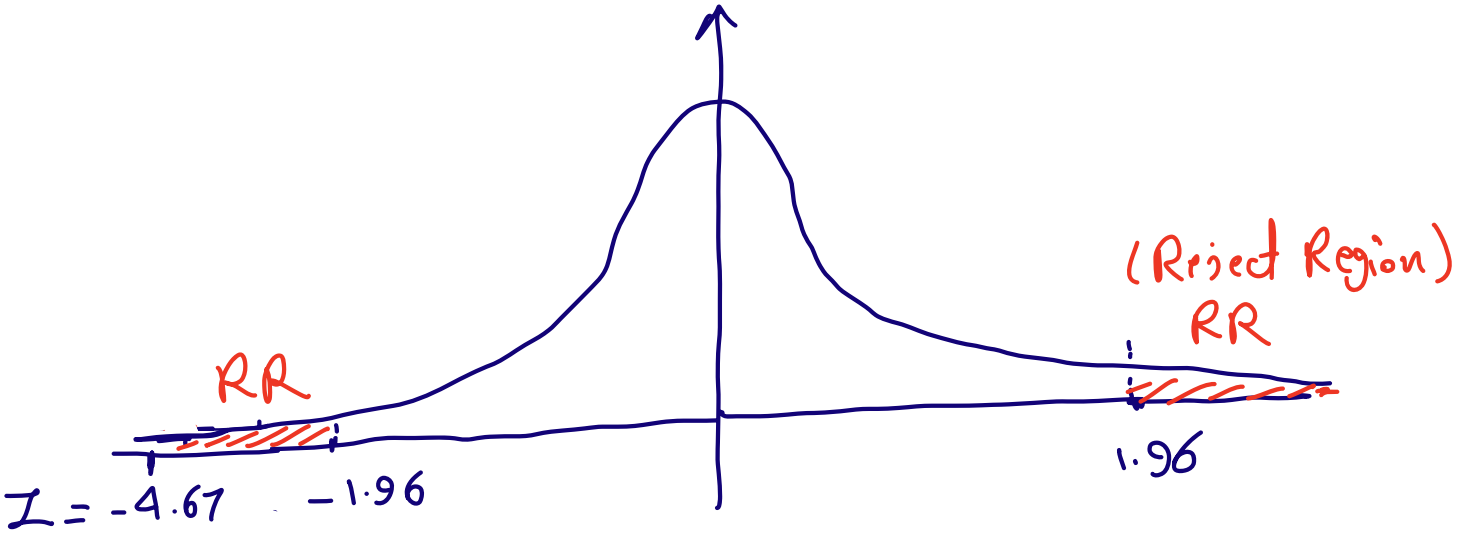
Standard error 2739 ($n=30$)

Glassdoor Report: 113000

$H_0 = \mu_0 = 113000$ (null hypothesis)

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{100200 - 113000}{2739} = -4.67$$

$Z = -4.67$

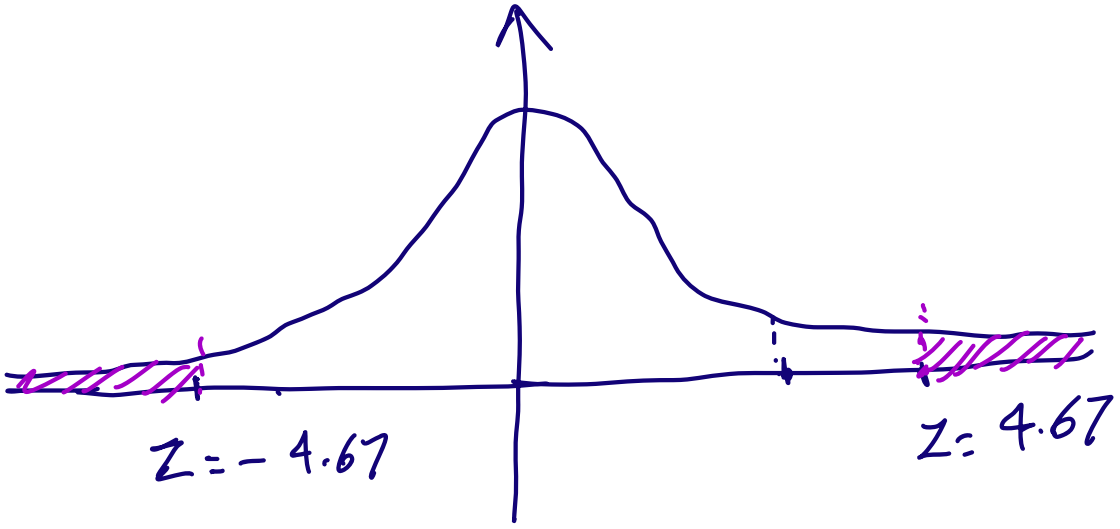


Assuming $\alpha = 0.05 \rightarrow Z_{0.025} = 1.96$

$|Z| > 1.96 \Rightarrow$ falls into the reject region

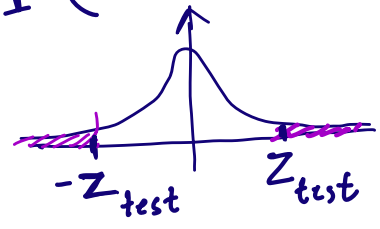
$\Rightarrow$ At 5% Significance level there is no statistical evidence that the mean Salary is 113000.

Equivalently, we can use p-value:



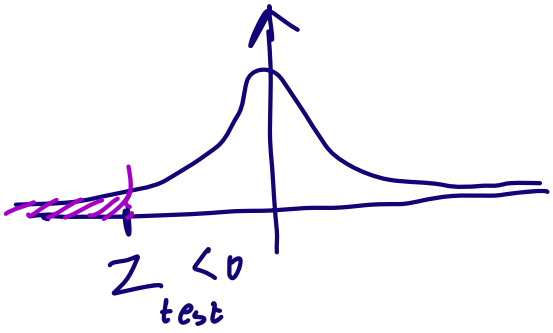
p-value: The probability of obtaing the observed results or more extreme results by mistake if the null hypothesis is true.

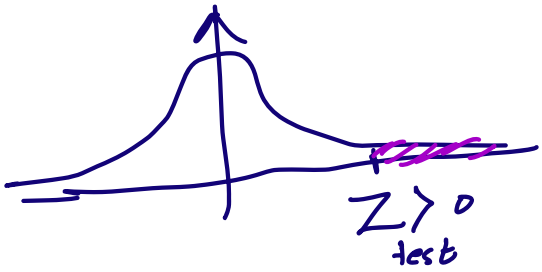
two-sided tests: $p = 2P(z > |Z_{test}|)$



one-sided tests:

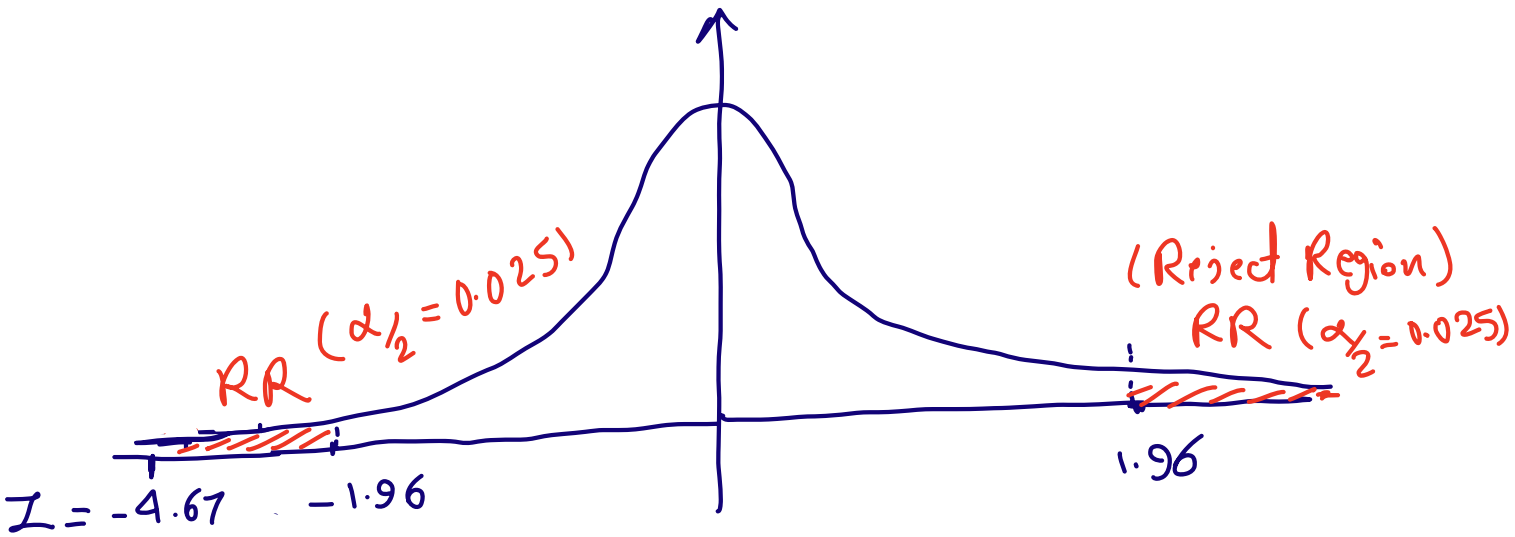
$p = P(z < Z_{test})$ or $p = P(z > Z_{test})$





We reject the null hypothesis if

$p < \alpha$ (most often $\alpha = 0.05$)



p-value: $2\,P(z > |-4.67|) < 0.05$

(Reject)

Single sample, unknow σ: T-test

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \rightarrow$$

$p = 2P_\nu(t > |T|)$ (two-sided)

$p = P_\nu(t > T)$ or (one-sided)

$p = P_\nu(t < T)$

P_ν: Student's T distribution with $\nu = n-1$ degrees of freedom.

Reject if $p < \alpha$

Two samples, dependent:

Example: Mg concentration in blood before (x) and after (y) a drug (or conductivity of a metal before and after a process)

Before	After
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_n	y_n
averages: $\bar{x}$	$\bar{y}$

$H_0: \bar{x} \geq \bar{y}$ (null)

$H_1: \bar{x} < \bar{y}$ (Alternative)

$\bar{D} = \bar{y} - \bar{x}$

$S_{\bar{D}}$: direct calculation from differences

$\bar{D}_0: \bar{y} - \bar{x} \leq 0$

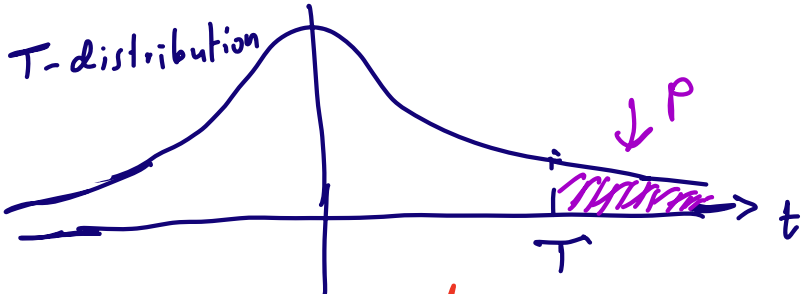
$\bar{D}_1: \bar{y} - \bar{x} > 0$

$$T = \frac{\bar{D} - \mu_0}{S_{\bar{D}}/\sqrt{n}} \rightarrow p = P_\nu(t > T)$$

($\bar{D}_0$: $\mu_0 = 0$)

(P_ν: one-sided)

Student's T distribution with $\nu = n-1$ degrees of freedom



one-sided test

If $p < \alpha \rightarrow$ Reject

(α is usually 0.05)

Test Scores (Economics Class)

	Engineering x	Management y
Size	100	70
Sample mean	58	65
Sample std	10	5

$H_0: \bar{x} - \bar{y} = -4$ $\quad (\alpha = 0.05)$

$H_1: \bar{x} - \bar{y} \neq -4$

$$Z = \frac{\bar{d} - \mu_0}{SE_{\bar{d}}} = \frac{-7+4}{1.16} = -2.59$$

$$SE_{\bar{d}} = \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \quad \Rightarrow \quad p = 2P(z > |Z|)$$

$$= 2P(z > 2.59) < 0.05$$

Reject

Independent Samples, unknow population Variances (assumed equal)

	City 1 apples (x)	City 2 apples (y)
Size :	10	8
Sample mean:	3.94	3.25
Sample std:	0.18	0.27

pooled variance formula: $s_p^2 = \frac{(n_x-1)s_x^2 + (n_y-1)s_y^2}{n_x+n_y-2}$

$H_0 = \bar{x} - \bar{y} = 0$, $H_1 = \bar{x} - \bar{y} \neq 0$

$(\bar{d} = \bar{x} - \bar{y})$

$T = \frac{\bar{d} - M_0}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} = 6.53$ → T-distribution

$\Rightarrow p = 2 P_\nu (t > 6.53)$

$\nu = n_x + n_y - 2 = 16$

$p \ll 0.05 \rightarrow$ Reject

chi-square (χ^2) test

A statistical test for comparing observed counts with expected counts (to check if there is a meaningful difference).

The relevant statistic is defined as

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i}.$$

[If $X = \sum_{i=1}^{k} Z_i^2$ where Z_i are independent, standard normal variables, then X obeys a chi-square distribution with k degrees of freedom, $X \sim \chi^2_k$]

One of its main uses is goodness-of-fit test

How consistent is the observed distribution with a claimed distribution?

Goodness of fit example:

Let's assume that we roll a die 60 times. the results are the following:

observations:

Face	Count
1	5
2	8
3	12
4	15
5	10
6	10

Is the die fair?

Null Hypothesis: It is fair $\Rightarrow$ we expect that each face has equal counts $\Rightarrow$ $E_1 = E_2 = \cdots = E_6 = 60/6 = 10$

$$\Rightarrow \chi^2 = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} = \frac{25}{10} + \frac{4}{10} + \frac{4}{10} + \frac{25}{10}$$

$$+ 0 + 0 = 5.8$$

We assume that χ^2 obeys a chi-square distribution with $k = N - 1 = 6 - 1 = 5$ degrees of freedom $\Rightarrow$ we can calculate the p-value corresponding to $\chi^2 = 5.8$

$$p = \underset{k=5}{P}(\chi^2 > 5.8)$$

$p = 0.33 \gg \alpha = 0.05 \Rightarrow$ Do not reject

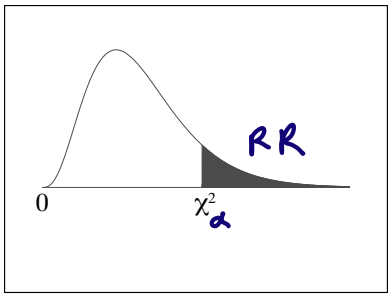
Equivalently:

we can find $\chi^2_{0.05}$ $\left(\underset{k=5}{P}(\chi^2 > \chi^2_{0.05}) = 0.05 \right)$ from tables and then compare our calculated χ^2 with $\chi^2_{0.05}$:

If $\chi^2 > \chi^2_{0.05}$ Reject

If $\chi^2 < \chi^2_{0.05}$ Do not reject

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

If $\chi^2 > \chi^2_\alpha$ Reject

($\alpha = 0.05$ in many cases)

In our case: $\chi^2 < \chi^2_\alpha$

$(5.8 < 11.07)$

$\Rightarrow$ We do not Reject H_0

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

https://math.arizona.edu/~jwatkins/chi-square-table.pdf