2. Confidence interval on μ, σ^2 *Unknown*

If $\bar{x}$ and s are the mean and standard deviation of a random sample of size n from a normal population with unknown variance $\boldsymbol{\sigma^2}$, a $100(1-\alpha)\%$ confidence interval for $\boldsymbol{\mu}$ is given by

$$\bar{X} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

Where $t_{\alpha/2,n-1}$ is the t-value with n-1 (Degree of Freedom), leaving an area of $\alpha/2$ to the right and equal to $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$

Example: the contents of 7 similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2 and 9.6 Lıters.

- Find a 95% confidence interval for the mean contents of all such Containers, assuming an approximately normal distribution.
- Find a 95% Lower and Upper bound for the mean contents of all such Containers.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{9.8 + 10.2 + 10.4 + 9.8 + 10 + 10.2 + 9.6}{7} = \frac{70}{7} = 10$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$
$$= \sqrt{\frac{(9.8-10)^2 + (10.2-10)^2 + (10.4-10)^2 + (9.8-10)^2 + (10-10)^2 + (10.2-10)^2 + (9.6-10)^2}{7-1}}$$
$$= 0.283$$

$$Two\ sided: p\left(\bar{X} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

$$p\left(10 - 2.447\frac{0.283}{\sqrt{7}} < \mu < 10 + 2.447\frac{0.283}{\sqrt{7}}\right) = 0.95$$

$$p(9.74 < \mu < 10.26) = 0.95$$

$$Upper: p\left(\mu < \bar{X} + t_{\alpha,n-1}\frac{s}{\sqrt{n}}\right) = 1-\alpha \rightarrow p\left(\mu < 10 + 1.943\frac{0.283}{\sqrt{7}}\right) = 0.95 \rightarrow p(\mu < 10.2078) = 0.95$$

$$Lower: p\left(\mu > \bar{X} - t_{\alpha,n-1}\frac{s}{\sqrt{n}}\right) = 1-\alpha \rightarrow p\left(\mu > 10 - 1.943\frac{0.283}{\sqrt{7}}\right) = 0.95 \rightarrow p(\mu > 9.7922) = 0.95$$

3. Confidence interval the difference between mean of two populations

3-1 confidence interval for $\mu_1 - \mu_2$, σ_1^2 and σ_2^2 known

If $\bar{x}_1$ and $\bar{x}_2$ are means of independent random samples of size n_1 and n_2 from populatıons with known variances σ_1^2 and σ_2^2, respectively, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Where $z_{\alpha/2}$ is the z value leaving an area of $\alpha/2$ to the right .

Example: A study was conducted in which two types of engines, A and B, were compared. Gas mileage, in miles per gallon, was measured. 50 experiments were conducted using engine type A and 75 experiments were done with engine type B. The gasoline used and other conditions were hold constant. The average gas mile age was 36 miles per gallon for engine A and 42 miles per gallon for engine B.

- Find a 96% confidence interval on $\mu_B - \mu_A$. where μ_A and μ_B are population mean gas mileages for engines A and B, respectively. Assume that the population standard deviations are 6 and 8 for engines A and B, respectively.
- Find a 96% Upper and Lower bounds on $\mu_B - \mu_A$

$$n_A = 50, n_b = 75, \bar{x}_A = 36, \bar{x}_B = 42, \sigma_A = 6, \sigma_B = 8, 1 - \alpha = 0.96$$

$$p\left((\bar{x}_B - \bar{x}_A) - z_{\alpha/2}\sqrt{\frac{\sigma_B^2}{n_B} + \frac{\sigma_A^2}{n_A}} < \mu_B - \mu_A < (\bar{x}_B - \bar{x}_A) + z_{\alpha/2}\sqrt{\frac{\sigma_B^2}{n_B} + \frac{\sigma_A^2}{n_A}} \right) = 1 - \alpha$$

$$p\left((42 - 36) - 2.05\sqrt{\frac{64}{75} + \frac{36}{50}} < \mu_B - \mu_A < (42 - 36) + 2.05\sqrt{\frac{64}{75} + \frac{36}{50}} \right) = 0.96$$

$$p(3.43 < \mu_B - \mu_A < 8.57) = 0.96$$

$$upper:\ p\left(\mu_B - \mu_A < (\bar{x}_B - \bar{x}_A) + z_\alpha\sqrt{\frac{\sigma_B^2}{n_B} + \frac{\sigma_A^2}{n_A}} \right) = 1 - \alpha$$

$$p\left(\mu_B - \mu_A < (42 - 36) + 1.75\sqrt{\frac{64}{75} + \frac{36}{50}} \right) = 0.96 \ \rightarrow p(\mu_B - \mu_A < 8.88) = 0.96$$

$$lower:\ p\left(\mu_B - \mu_A > (\bar{x}_B - \bar{x}_A) - z_\alpha\sqrt{\frac{\sigma_B^2}{n_B} + \frac{\sigma_A^2}{n_A}} \right) = 1 - \alpha$$

$$p\left(\mu_B - \mu_A > (42 - 36) - 1.75\sqrt{\frac{64}{75} + \frac{36}{50}} \right) = 0.96 \ \rightarrow p(\mu_B - \mu_A > 3.12) = 0.96$$

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
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

3-2 Confidence interval for $\mu_1 - \mu_2$, $\sigma_1^2 \neq \sigma_2^2$ but unknown

If $\bar{x}_1$ and $\bar{x}_2$ are means of independent random samples of size n_1 and n_2 , respectively, from aproximately normal population with known and unequal variances, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

Where $t_{\alpha/2}$ is the t value with

Fisher approximation of $v = \dfrac{(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2})^2}{\frac{1}{n_1-1}(\frac{s_1^2}{n_1})^2+\frac{1}{n_2-1}(\frac{s_2^2}{n_2})^2}$

degrees of freedom. leaving an area of $\alpha/2$ to the right .

example: A study was conducted by the department of zoology at the Virginia Tech to estimate the difference in the amounts of the chemical orthophosphorous measured at two different stations on the James River. Orthophosphorous was measured in milligrams per liter. One sample with 15 observations were collected from station 1 and another sample with 12 observations were collected from station 2. First sample had an average orthophospgorous content of 3.84 milligrams per liter and a standard deviation of 3.07 milligrams per liter, while second sample had an average orthophospgorous content of 1.49 milligrams per liter and a standard deviation of 0.80 milligrams per liter.

- Find a 95% confidence interval for the difference in the true average orthophosphorous contents at these two stations, assuming that the observations came from normal populations with different variances.
- Find a 90% uper and lower bounds on the difference between the population means for the two samples

$$v = \frac{(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2})^2}{\frac{1}{n_1-1}(\frac{s_1^2}{n_1})^2+\frac{1}{n_2-1}(\frac{s_2^2}{n_2})^2} = \frac{(\frac{3.07^2}{15}+\frac{0.8^2}{12})^2}{\frac{1}{15-1}(\frac{3.07^2}{15})^2+\frac{1}{12-1}(\frac{0.8^2}{12})^2} = 16.3 \cong 16$$

$$p\left((\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,v}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right) = 1-\alpha$$

$$p\left((3.84-1.49) - 2.12\sqrt{\frac{3.07^2}{15}+\frac{0.8^2}{12}} < \mu_1 - \mu_2 < (3.84-1.49) - 2.12\sqrt{\frac{3.07^2}{15}+\frac{0.8^2}{12}}\right) = 0.95$$

$$p(0.6 < \mu_1 - \mu_2 < 4.1) = 0.95$$

$$upper: p\left(\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+t_{\alpha,v}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right)=1-\alpha$$

$$p\left(\mu_1-\mu_2<(3.84-1.49)+1.337\sqrt{\frac{3.07^2}{15}+\frac{0.8^2}{12}}\right)=0.9\rightarrow p(\mu_1-\mu_2<3.45)=0.9$$

$$lower: p\left(\mu_1-\mu_2>(\bar{x}_1-\bar{x}_2)-t_{\alpha,v}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right)=1-\alpha$$

$$p\left(\mu_1-\mu_2>(3.84-1.49)-1.337\sqrt{\frac{3.07^2}{15}+\frac{0.8^2}{12}}\right)=0.9\rightarrow p(\mu_1-\mu_2>1.25)=0.9$$

3-3 Confidence interval for $\mu_1-\mu_2$, $\sigma_1^2=\sigma_2^2$ but unknown

If $\bar{x}_1$ and $\bar{x}_2$ are means of independent random samples of size n_1 and n_2 , respectively, from aproximately normal population with unknown but equal variances, a $100(1-\alpha)\%$ confidence interval for $\mu_1-\mu_2$ is given by

$$(\bar{x}_1-\bar{x}_2)-t_{\alpha/2,v}\ s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+t_{\alpha/2,v}\ s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

Wheres_p is the pooled estimate of the population standard deviation and $t_{\alpha/2}$ is the t value with $v=n_1+n_2-2$ degrees of freedom. leaving an area of $\alpha/2$ to the right .

Example: one article published in an journal, to determine the relationship between different factors. Two independent sampling stations were chosen to this study, one located downstream from the acid mine discharge point and the other located upstream. For 12 monthly onservations collected at the downstream station, the species diversity index had a mean value $\bar{x}_1=3.11$ and a standard deviation $s_1=0.771$, while 10 monthly observations collected at the upstream station had a mean index value $\bar{x}_2=2.04$ and a standard deviation $s_2=0.448$.

- Find a 90% confidence interval for the difference between the population means for the two locations, assuming that the populations are approximately normally distributed with equal variances.
- Find a 90% uper and lower bounds on the difference between the population means for the two locations
-

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(12-1)0.771^2 + (10-1)0.448^2}{12 + 10 - 2} = 0.417$$

$$p\left((\bar{x}_1 - \bar{x}_2) - t_{\alpha/2,v}\ s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,v}\ s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) = 1 - \alpha$$

$$p\left((3.11 - 2.04) - 1.725 \times \sqrt{0.417} \sqrt{\frac{1}{12} + \frac{1}{10}} < \mu_1 - \mu_2 < (3.11 - 2.04) + 1.725 \times \sqrt{0.417} \sqrt{\frac{1}{12} + \frac{1}{10}} \right) = 0.90$$

$$p(0.593 < \mu_1 - \mu_2 < 1.547) = 0.90$$

$$upper{:}\, p\left(\mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2,v}\ s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) = 1 - \alpha$$

$$upper{:}\, p\left(\mu_1 - \mu_2 < (3.11 - 2.04) + 1.325 \times \sqrt{0.417} \sqrt{\frac{1}{12} + \frac{1}{10}} \right) = 0.90 \rightarrow p(\mu_1 - \mu_2 < 1.426) = 0.90$$

$$lower{:}\, p\left(\mu_1 - \mu_2 > (3.11 - 2.04) - 1.325 \times \sqrt{0.417} \sqrt{\frac{1}{12} + \frac{1}{10}} \right) = 0.90 \rightarrow p(\mu_1 - \mu_2 > 0.714) = 0.90$$