Example: Let X denote the number of flaws in a 1 inch length of copper wire. The probability mass function of X is presented in the following table.

(1) Find the mean and variance of X.
(2) One hundred wires are sampled from this population. What is the distribution of the sample mean ($\bar{X}$) of the flaws per wire?
(3) What is the probability that the average number of flaws per wire in this sample is less than 0.5?

x	P(X=x)
0	0.48
1	0.39
2	0.12
3	0.01

Solution:

(1) The population mean and the population varance can be calculated as:

$$\mu = \sum_{i=1}^{4} x_i f(x_i) = 0.66;\ \sigma^2 = \sum_{i=1}^{4} (x_i - \mu)^2 f(x_i) = 0.5244$$

(2) Let $X_1, X_2, \ldots, X_{100}$ denote the number of flaws in the 100 wires sampled from the population. According to the central limit theorem, we have,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) = N(0.66, 0.005244)$$

(3) $P[\bar{X} < 0.5] = P\left[\frac{\bar{X}-0.66}{\sqrt{0.005244}} < \frac{0.5-0.66}{\sqrt{0.005244}}\right] = P[Z < -2.21] = 0.0136$

What is $P[X < 0.5]$? $P[X < 0.5] = P[X = 0] = 0.48$

Interval Estimation

Confidence interval: A $100(1-\alpha)\%$ confidence interval for a parameter θ is a random interval $[L_1, L_2]$ such that $P[L_1 \leq \theta \leq L_2] = 1-\alpha$, regardless of the value of θ. (typical value of $\alpha = 0.05, 0.01$).

Meaning of a confidence interval:

Example: $\theta = \mu, \alpha = 0.05, L_1 = 1.45, L_2 = 1.57$, the 95% confidence interval of the population mean μ is

$$P[1.45 \leq \mu \leq 1.57] = 0.95$$

Based on the given samples, we are 95% sure the true value of μ is within the interval $\mathbf{[1.45, 1.57]}$

Interval Estimation

Confidence interval of the mean: variance known

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution X with mean μ and variance σ^2. Find the 95% confidence interval on *mean* when the *variance* of the random variable is given.

From the central limit theorem, when the sample size n is large, we have,

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}), \text{ and } Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

From the z-table, we have:

$$P[-1.96 \le Z \le 1.96] = 0.95$$

For the 95% interval, we have,

$$P\left[-1.96 \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le 1.96\right] = 0.95 \Rightarrow P\left[-\frac{1.96\sigma}{\sqrt{n}} \le \bar{X} - \mu \le \frac{1.96\sigma}{\sqrt{n}}\right] = 0.95$$

$$\Rightarrow P\left[-\bar{X} - \frac{1.96\sigma}{\sqrt{n}} \le -\mu \le -\bar{X} + \frac{1.96\sigma}{\sqrt{n}}\right] = 0.95$$

$$\Rightarrow P\left[\bar{X} - \frac{1.96\sigma}{\sqrt{n}} \le \mu \le \bar{X} + \frac{1.96\sigma}{\sqrt{n}}\right] = 0.95$$

i.e., $L_1 = \bar{X} - \frac{1.96\sigma}{\sqrt{n}}$, $L_2 = \bar{X} + \frac{1.96\sigma}{\sqrt{n}}$, the 95% confidence interval of μ is $[\bar{X} - \frac{1.96\sigma}{\sqrt{n}}, \bar{X} + \frac{1.96\sigma}{\sqrt{n}}]$

Interval Estimation

Confidence interval of the mean: variance known

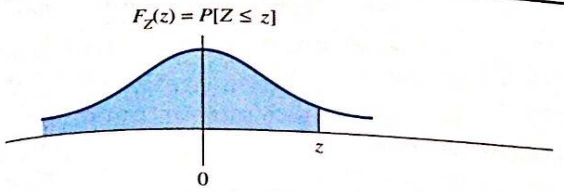
Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution X with mean μ and variance σ^2. The 95% confidence interval on *mean* when the *variance* of the random variable is given is:

$$[\bar{X} - \frac{1.96\sigma}{\sqrt{n}}, \bar{X} + \frac{1.96\sigma}{\sqrt{n}}]$$

The conditions to use this formula:

(1) The population variance is known
(2) When the population follows a normal distribution, the sample size does not need to be large
(3) Otherwise, the sample size n need to be large enough, i.e., $n > 30$

TABLE V
Cumulative distribution: Standard normal



z	$F_Z(z) = P[Z \le z]$ 0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1921	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Interval Estimation

Confidence interval of the mean: variance known

95% Confidence interval: $P\left[\bar{X} - \frac{1.96\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{1.96\sigma}{\sqrt{n}}\right] = 0.95$

Example: A new treatment is investigated to prolong the average survival time without affecting variability. Denote the survival time under the new treatment by X. We are assuming that X is normally distributed with $\sigma^2 = 9$ and μ unknown. When an experiment is conducted, the following result is observed: (months)

8.0, 13.6, 13.2, 13.6, 12.5, 14.2, 14.9, 14.5, 13.4, 8.6, 11.5, 16.0, 14.2, 19.0, 17.9, 17.0

Find the 95% confidence interval of μ.

It is easy to calculate that $\bar{X}$=13.88. We also know: $n = 16, \sigma = 3$. Then,

$$L_1 = \bar{X} - \frac{1.96\sigma}{\sqrt{n}} = 13.88 - 1.96 \times \frac{3}{4} = 12.41; L_2 = \bar{X} + \frac{1.96\sigma}{\sqrt{n}} = 15.35$$

i.e., $P[12.41 \leq \mu \leq 15.35] = 95\%$

This means, there is 95% chance that true μ lies within the interval $[12.41, 15.35]$

Q: How does the confidence interval change when the sample size n changes?

Q: If α value of 0.01 is used, based on the same samples, how will the interval change? Will it get wider or narrower?

Interval Estimation

Confidence interval of the mean: variance known

Practice example: Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution X with mean μ and variance σ^2. Find the 99% confidence interval on *mean* when the *variance* of the random variable is given.

From the central limit theorem, when the sample size is large, we have, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$, and $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

From the z-table, we have:

$$P[-2.58 \leq Z \leq 2.58] = 0.99$$

For the 99% interval, we have,

$$P\left[-2.58 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq 2.58\right] = 0.99 \Rightarrow P\left[-\frac{2.58\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq \frac{2.58\sigma}{\sqrt{n}}\right] = 0.99$$

$$\Rightarrow P\left[-\bar{X}-\frac{2.58\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X}+\frac{2.58\sigma}{\sqrt{n}}\right] = 0.99$$

$$\Rightarrow P\left[\bar{X}-\frac{2.58\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}+\frac{2.58\sigma}{\sqrt{n}}\right] = 0.99$$

i.e., $L_1 = \bar{X} - \frac{2.58\sigma}{\sqrt{n}}$, $L_2 = \bar{X} + \frac{2.58\sigma}{\sqrt{n}}$, the 99% confidence interval of μ is $[\bar{X} - \frac{2.58\sigma}{\sqrt{n}}, \bar{X} + \frac{2.58\sigma}{\sqrt{n}}]$

Interval Estimation of Variance

Definition of sample variance:
Let $X_1, X_2, \dots, X_n$ be a random sample of size n from the distribution of X with mean μ and variance σ^2. Then the statistic:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Is called the *sample variance*. Where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the *sample mean*.
s^2 is an unbiased estimator of the population variance σ^2.

Theorem 8.1.1: Distribution of $(n-1)\frac{s^2}{\sigma^2}$
Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. The random variable

$$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \mathrm{X}^2{}_{n-1}$$

i.e., it has a chi-squared distribution with $n-1$ degree of freedom.

Theorem 8.1.1: Distribution of $(n-1)\frac{s^2}{\sigma^2}$

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. The random variable

$$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \mathrm{X}^2{}_{n-1}$$

Proof: Let $W = \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2}$. We already proved that $W \sim \mathrm{X}_n^2$

$$W = \sum_{i=1}^{n}\frac{(X_i-\mu)^2}{\sigma^2} = \sum_{i=1}^{n}\frac{((X_i-\bar{X})+(\bar{X}-\mu))^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{(X_i-\bar{X})+(\bar{X}-\mu)}{\sigma}\right)^2$$

$$= \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 + \sum_{i=1}^{n}\left(\frac{\bar{X}-\mu}{\sigma}\right)^2 + 2\left(\frac{\bar{X}-\mu}{\sigma}\right)\sum_{i=1}^{n}\frac{X_i-\bar{X}}{\sigma}$$

$$= \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 + \sum_{i=1}^{n}\left(\frac{\bar{X}-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 + \frac{n(\bar{X}-\mu)^2}{\sigma^2}$$

$$= \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 + \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 = Y + Y_1$$

Interval Estimation of Variance

Proof: (continued)

$$W = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2 = Y + Y_1$$

From the central limit theorem, we have, $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$. Then, $Y_1 = \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \mathrm{X}_1^2$

The moment generating function of W is $m_W(t) = (1 - 2t)^{-n/2}$

The moment generating function of Y_1 is $m_{Y_1}(t) = (1 - 2t)^{-1/2}$

Then, under the condition that S^2 *and* $\bar{X}$ are independent (this is true for normal distribution), we have,

$$m_W(t) = m_Y(t) m_{Y_1}(t) \Rightarrow m_Y(t) = (1 - 2t)^{-(n-1)/2}$$

Hence,

$$Y = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \sim \mathrm{X}_{n-1}^2$$