Sampling distribution

Population The group of individuals under study, i.e. collection of objects under study.

Sample : A finite subset of statistical individuals in a population.

Sampling The process of selection a sample is called sampling. The main objectives are

— to get maximum information about the population with the help of a sample with minimum input.

— to state the limits within which the parameters of the population are expected to lie with a specific degree of confidence.

Types of sampling

Purposive → selection with definite purpose
drawback → favouritism

Random → selection at random

Startified → Entire ~~homogeneous~~ heterogeneous population is divided into homogeneous group or 'strata' Selection will be made from each strata randomly.

<u>Parameters</u> Distribution of the variable in the population is known as parameter.
So, any statistical measure based on all units of population is called parameter.
e.g. μ, σ^2 etc.

<u>Statistic</u> Statistical measures computed from sample observations.
So any ~~value of~~ function $a(x_1, x_2 \cdots x_n)$ of sample values $x_1, x_2 \cdots x_n$ is called satistic.

<u>Sampling Distribution</u>

Probability distribution of statistic is sampling distribution.
e.g. $\bar{x}$ is distributed $N(\mu, \frac{\sigma^2}{n})$.

<u>Central Limit Theorem</u>

Let $X_1, X_2 \cdots X_n$ be a sequence iid RVs with mean μ and variance σ^2, i.e. $E(x_i) = \mu$, $var(x_i) = \sigma^2$
Then the random variable $S_n = X_1 + X_2 \cdots + X_n$ is asymtotically normal to $\mu = \sum_{i=1}^{n} \mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$.

i.e $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ is $N(0, 1)$ as $n \to \infty$.

- Generally $n \geqslant 30$, considered as large sample.
- Thus, according to CLT

  binomial $\longrightarrow$ normal as $n \longrightarrow \infty$

  poisson $\longrightarrow$ normal as $n \longrightarrow \infty$.

▨ Let $X_1, X_2 \cdots X_n$ be iid Poisson variates with parameter λ. Use CLT to estimate $P(120 \leq S_n \leq 160)$ where $S_n = X_1 + X_2 + \cdots + X_n$. Given $\lambda = 2$, $n = 75$

So, $E(S_n) = n\lambda = 150$ $\quad var(S_n) = n\lambda = 150$

$$\Rightarrow S_n \sim N(150, 150)$$

$$P(120 \leqslant S_n \leqslant 160) = P\left(\frac{120-150}{\sqrt{150}} \leqslant Z \leqslant \frac{160-150}{\sqrt{150}}\right)$$

$$= P(-2.45 \leqslant Z \leqslant .82)$$

$$= 1 - \phi(2.45) + \phi(.82) = .7868.$$

▨ Let a random sample of size 54 be taken from a discrete population with pmf $p(x) = \frac{1}{3}$, $x = 2, 4, 6$. Find the probability that the sample mean will lie between 4.1 and 4.4.

$\mu = \frac{1}{3}(2+4+6) = 4$ $\quad \Rightarrow \quad \dfrac{\bar{X} - 4}{\sqrt{8/3}/\sqrt{54}} \sim N(0,1)$

$\sigma^2 = 8/3$

$\hookrightarrow E(x^2) - \{E(x)\}^2$

$$P(4.1 \leqslant \bar{X} \leqslant 4.4) = P(.45 \leqslant Z \leqslant 1.8)$$

$$= .9641 - .6736$$

$$= .2905$$

$\Rightarrow$ 30% time sample mean will be within range.

Random sampling

Let $x_1, x_2 \cdots x_n$ be n independent and identically distributed random variables each having same propability distribution $f(x)$. Then we say $(x_1, x_2, \cdots x_n)$ is a random sample from populati

The joint distribution of $x_1, x_2 \cdots x_n$ is

$$f(x_1, x_2 \cdots x_n) = f(x_1)\, f(x_2) \cdots f(x_n)$$

Statistic

A function of random sample, i.e. $T = T(x_1, \cdots x_n)$

for example,

$$\bar{X} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) \longrightarrow \text{sample mean}$$

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \longrightarrow \text{sample variance}$$

$$\left.\begin{array}{l} X_{\left(\frac{n+1}{2}\right)} \text{ if } n \text{ is odd} \\[2ex] \dfrac{X_{\frac{n}{2}} + X_{\left(\frac{n}{2}+1\right)}}{2} \text{ if } n \text{ is even} \end{array}\right\} \longrightarrow \text{Sample median}$$

sample range: $X_{(n)} - X_{(1)}$.

Sampling distribution of mean

We assume that the concerned population has a mean μ and variance σ^2.

If $(x_1, x_2 \cdots x_n)$ is a random sample, then

$$\text{Sample mean} \longrightarrow \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$E(\bar{x}) = E\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) = \frac{1}{n}\sum_i E(x_i) = \mu.$$

variance of sample mean is

$$\begin{aligned}
var(\bar{X}) &= E(\bar{X} - \mu)^2 \\
&= E\left[\frac{x_1 + x_2 + \cdots + x_n}{n} - \mu\right]^2 \\
&= \frac{1}{n^2}\left\{E\left((x_1 + x_2 + \cdots + x_n) - n\mu\right)^2\right. \\
&= \frac{1}{n^2} E\left((x_1 - \mu) + (x_2 - \mu) + \cdots + (x_n - \mu)\right)^2 \\
&= \frac{1}{n^2}\left\{E(x_1 - \mu)^2 + E(x_2 - \mu)^2 + \cdots + E(x_n - \mu)^2\right\}
\end{aligned}$$

since $E(x_i - \mu)(x_j - \mu) = 0$, $i \neq j$ because sampling is done with replacement.

$$= \frac{1}{n^2}\left\{\sigma^2 + \sigma^2 + \cdots + \sigma^2\right\} = \frac{\sigma^2}{n}.$$

$Var(\bar{x})$ is called as standard error of $\bar{x}$.

The distribution of a population RV X is given by $P(X=0) = P(X=1) = \frac{1}{2}$. A random sample (x_1, x_2, x_3, x_4) of size 4 is taken from the population. Show that the sampling distribution of statistic $t = x_1 + x_2 + x_3 + x_4$ is a binomial $(4, \frac{1}{2})$ distribution.

RV T corresponding to statistic t is

$$T = X_1 + X_2 + X_3 + X_4$$

X_i has same distribution as X & X_is are ind.

$$P(T=0) = \prod_{i=1}^{4} \{ P(X_i = 0) \} = \left(\tfrac{1}{2}\right)^4 = {}^4C_0 \left(\tfrac{1}{2}\right)^4$$

$$P(T=1) = P(X_1=0, X_2=0, X_3=0, X_4=1) + \cdots$$
$$+ P(X_1=1, X_2=0, X_3=0, X_4=0)$$
$$= \left(\tfrac{1}{2}\right)^4 + \left(\tfrac{1}{2}\right)^4 + \left(\tfrac{1}{2}\right)^4 + \left(\tfrac{1}{2}\right)^4$$
$$= 4\left(\tfrac{1}{2}\right)^4 = {}^4C_1 \left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right)^3$$

$$P(T=2) = 6 \cdot \left(\tfrac{1}{2}\right)^4 = {}^4C_2 \left(\tfrac{1}{2}\right)^2 \left(\tfrac{1}{2}\right)^2$$

$$P(T=3) = {}^4C_3 \left(\tfrac{1}{2}\right)^4 \qquad P(T=4) = {}^4C_4 \left(\tfrac{1}{2}\right)^4$$

$$\Rightarrow T \sim \text{bin}\left(4, \tfrac{1}{2}\right).$$

CLT (Cont.)

Let $X_{11}, X_{12} \cdots X_{1n_1}$ be iid RVs with mean μ_1 and var σ_1^2

& $X_{21}, X_{22} \cdots X_{2n_2}$ be iid RVs with mean μ_2 and var σ_2^2

If $\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1i}$, $\overline{X}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} x_{2i}$ Then

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \longrightarrow N(0,1) \text{ as } \begin{array}{l} n_1 \to \infty \\ n_2 \to \infty . \end{array}$$

The TV picture tubes of manufacturer A have a mean life time of 6.5 years and s.d. .9 years. Those from manufacturer B have a mean life 6 yrs and s.d. .8 yrs. What is the probability that a random sample of 36 tubes from A will have a mean life that is atleast 1 year more than the mean life of a sample 49 tubes from B?

We need to find $P\{(\overline{x}_1 - \overline{x}_2) > 1\}$

Here $\mu_1 - \mu_2 = .5$ $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = .189$

$$\Rightarrow P(\overline{x}_1 - \overline{x}_2 > 1)$$

$$= P\left(\frac{\overline{x}_1 - \overline{x}_2 - .5}{.189} > \frac{1 - .5}{.189}\right)$$

$$= P(Z > 2.65) = .004.$$

Distributions arising from Normal dist^n:

If $X \sim N(\mu, \sigma^2)$, $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

$$Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2 \text{ is a chi-square with 1 d.f.}$$

If X_i $(i = 1, 2 \ldots n)$ are n independent normal variates

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \text{ is a chi-square variate with } n\text{d.}$$

M.G.F. of $\chi^2_{(1)}$

$$E\left(e^{tX}\right) = E\left(e^{tZ^2}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx^2} e^{-\frac{x^2}{2}} dx.$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{1-2t}{2}\right)x^2} dx$$

$$= (1-2t)^{-\frac{1}{2}} \cdot \frac{1}{\sqrt{2\pi}\,(1-2t)^{-\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}(1-2t)} dx$$

$$= (1-2t)^{-\frac{1}{2}}.$$

M.G.F of $\chi^2_{(n)}$

$$E\left(e^{tX}\right) = E\left(e^{t(Z_1^2 + Z_2^2 + \cdots + Z_n^2)}\right)$$

$$= (1-2t)^{-\frac{n}{2}}$$

$$M_{\chi^2}(t) = (1-2t)^{-\frac{n}{2}}.$$

Again this is M.G.F. of $\gamma\left(\frac{n}{2}, \frac{1}{2}\right)$

$$\Rightarrow \chi^2 \text{ is identical to } \gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

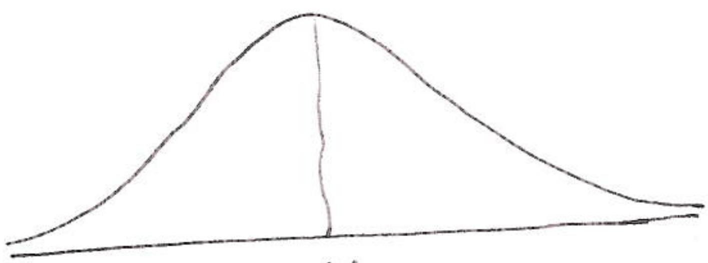
So, density of χ^2 is

$$f(x) = \frac{\frac{1}{2} e^{-\frac{x}{2}} \left(\frac{x}{2}\right)^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)}$$

The t-distribution

$$t_{(n)} = \frac{Z}{\sqrt{\frac{\chi^2_{(n)}}{n}}}$$ is t-distributed with n d.f.

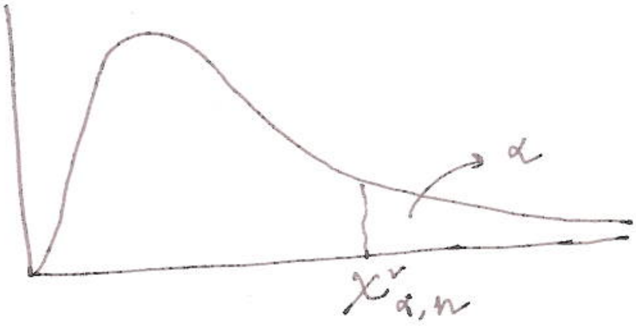
where Z and $\chi^2_{(n)}$ are independent.



Mean $= 0$

Variance $= \frac{n}{n-2}$, $n > 2$.

χ^2-distribution



$$\text{Prob}\left(x \geq \chi^2_{\alpha,n}\right) = \alpha.$$

□ To prove t is a sampling distribution.

Let $X_1, X_2 \cdots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1).$$

$$\Rightarrow \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\Rightarrow \frac{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}{\sqrt{\frac{(n-1)s^2}{\sigma^2(n-1)}}} \sim t_{n-1}$$

$$\Rightarrow \frac{\sqrt{n}(\bar{X}-\mu)}{s} \sim t_{n-1}$$

Stirling's Approximation

$$\Gamma(p+1) \cong \sqrt{2\pi}\, e^{-p} \cdot p^{p+\frac{1}{2}}$$

for large p.

χ^2 is sampling distribution

$$\underbrace{\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}_{W} = \underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2}_{W_1} + \underbrace{\frac{n}{\sigma^2}(\bar{x}-\mu)^2}_{W_2}$$

Here, $\frac{x_i-\mu}{\sigma} \sim N(0,1) \Rightarrow \boxed{W \sim \chi^2_n}$

Again, $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$\Rightarrow \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \Rightarrow \boxed{W_2 \sim \chi^2_1}$$

$$M_W(t) = M_{W_1}(t)\, M_{W_2}(t) \Rightarrow M_{W_1}(t) = \frac{M_W(t)}{M_{W_2}(t)}$$

$$\Rightarrow M_{W_1}(t) = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}}$$

$$= (1-2t)^{-\frac{n-1}{2}} \quad , t < \frac{1}{2}$$

$$\Rightarrow W_1 = \frac{\Sigma(x_i-\bar{x})^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

$\Rightarrow$ Sample variance is distributed with χ^2_{n-1}

Student's t-distribution

Let X and Y be two independent RVs, $X \sim N(0,1)$, $Y \sim \chi^2_n$. Then $T = \dfrac{X}{\sqrt{Y/n}}$ is said to be student's t distⁿ with n degree of free.

Joint pdf. of X and Y is

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{2^{n/2}\Gamma\frac{n}{2}} e^{-\frac{y}{2}} y^{\frac{n}{2}-1}$$

$$-\infty < x < \infty, \ y > 0$$

Consider the transformation

$$T = \frac{\sqrt{n}\,X}{\sqrt{Y}}, \quad U = Y \Rightarrow x = \sqrt{\frac{u}{n}}\,T, \ y = u$$

$$\text{Jacobian } J = \begin{vmatrix} \sqrt{\frac{u}{n}} & \frac{t}{2\sqrt{nu}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{u}{n}}.$$

$$f_{T,U}(t,u) = \frac{1}{2^{\frac{n+1}{2}}\sqrt{\pi n}\,\Gamma\frac{n}{2}} e^{-\frac{u}{2}\left(1+\frac{t^2}{n}\right)} \cdot u^{\frac{n+1}{2}-1}$$

$$-\infty < t < \infty$$
$$u > 0$$

Marginal pdf of T is

$$f_T(t) = \int_0^\infty f_{T,U}(t,u)\,du$$

$$= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\frac{n}{2}\sqrt{\pi n}} \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

$-\infty < t < \infty$

$$= \boxed{\frac{1}{\sqrt{n}\, B\left(\frac{n}{2},\frac{1}{2}\right)} \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}}.$$

$\Rightarrow$ • T is symmetric at $t=0$.

• Odd order moments vanish

• Even order moments exist $< n$.

• $E(T) = 0$, $E(T^2) = \frac{n}{n-2}$.

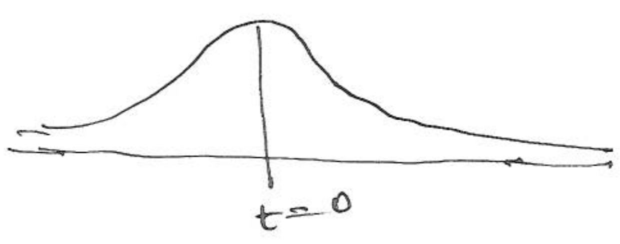
$$\boxed{E(T^k) = \frac{n^{k/2}\,\Gamma\frac{k+1}{2}\,\Gamma\frac{n-k}{2}}{\Gamma\frac{1}{2}\,\Gamma\frac{n}{2}}}$$

• $\mu_4 = E(T^4) = \frac{3n^2}{(n-2)(n-4)}$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{6}{n-4} > 0.$$

As $n \to \infty$ $\beta_2 \to 0 \Rightarrow$ approches to normal.

in general, it is leptokurtic.

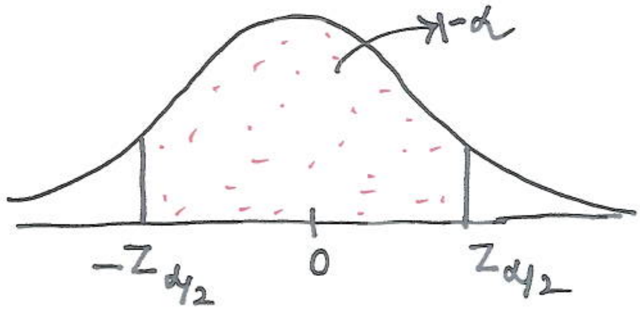


Given the sample $(2.3, -0.2, -0.4, -0.9)$, if is drawn from a normal population with $\sigma = 3$. Estimate the paramater μ (population mean)

We know,

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Here, $\bar{x} = 0.2$



$$\boxed{P\left(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}\right) = 1-\alpha.}$$

If $\alpha = \cdot 05$, Then $Z_{\alpha/2} = 1.96$.

$$P\left(-1.96 \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq 1.96\right) = \cdot 95$$

$$\text{or, } P\left(\cdot 2 - \frac{3}{\sqrt{4}} \times 1.96 \leq \mu \leq \cdot 2 + \frac{3}{\sqrt{4}} \times 1.96\right) = \cdot 95$$

$$\text{or, } P\left(-2.74 \leq \mu \leq 3.14\right) = \cdot 95.$$

95% confidence interval for μ is $[-2.74, 3.14]$ where σ is known

The population of math scores in a test is known to have s.d. 5·2. If a random sample of size 20 shows a mean 16·9 find 95% confidence limits for the mean score of the population assuming that the population is normal.

Here, suitable statistic is $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$.

$$P\left(-Z_{\alpha/2} \le Z \le Z_{\alpha/2}\right) = 1 - \alpha$$

$$\Rightarrow P\left(-1\cdot96 \le \frac{16\cdot9 - \mu}{5\cdot2/(20)^{1/2}} \le 1\cdot96\right) = \cdot95$$

$\Rightarrow$ 95% confidence int for μ is $[14\cdot62, 19\cdot18]$

Confidence interval for parameters of normal dist^n

Confidence interval for μ

σ is known

$\left(\bar{x} - \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}, \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{\alpha/2}\right)$ is a $100(1-\alpha)\%$ confidence interval of μ.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

where $s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$

Then

$$\frac{(\bar{X}-\mu)/\sigma/\sqrt{n}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}\cdot\frac{1}{(n-1)}}} \sim t_{n-1}$$

provided $\bar{X}$ and s^2 are independent

$$\Rightarrow \boxed{\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}}$$

$\left(\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha/2, n-1}, \ \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha/2, n-1}\right)$ is $100(1-\alpha)\%$ confidence interval of μ when σ is unknown.

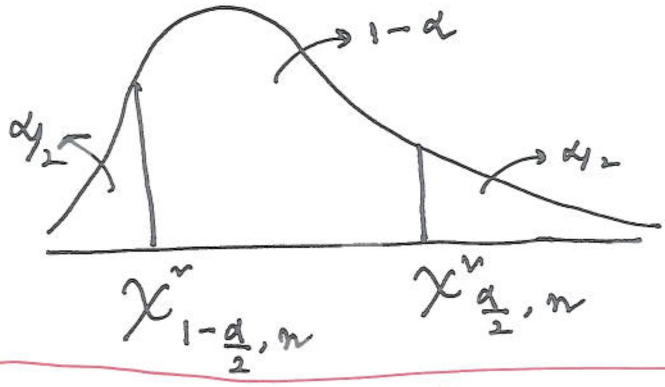
Confidence interval for σ^2

(μ is known)

$$Z_i = \frac{x_i - \mu}{\sigma} \sim N(0,1)$$

$$W = \sum_i Z_i^2 = \frac{\sum_i (x_i - \mu)^2}{\sigma^2} \sim \chi^2_{(n)}$$

$$P\left(\chi^2_{1-\frac{\alpha}{2}, n} \leq W \leq \chi^2_{\frac{\alpha}{2}, n}\right) = 1 - \alpha$$



$$\left(\frac{\sum_i (x_i - \mu)^2}{\chi^2_{\frac{\alpha}{2}, n}}, \frac{\sum_i (x_i - \mu)^2}{\chi^2_{1-\frac{\alpha}{2}, n}} \right) \text{ is } 100(1-\alpha)\%$$

confidence interval for σ^2.

10 bearings made from a certain process have a mean diameter 0·0506 cm and a s.d. ·004 cm. Assuming that the data may be looked upon as a random sample from a normal population, construct a 95% confidence interval for the actual average diameter of bearings made by this process.

We know. $t_{.025,9} = 2.262$

$$\bar{x} \pm \frac{s}{\sqrt{n}} t_{0.025,9} = \cdot 0506 \pm \frac{\cdot 004}{\sqrt{10}} \times 2.262$$

$$= 0.0506 \pm 0.0028612$$

$\Rightarrow (0.0477, \cdot 0535) \rightarrow 95\%$ conf. int. of μ.

The calories of 10 samples of ghee of one ounce were recorded from a large collection as 105, 107, 112, 110, 115, 100, 108, 98, 106, 113. Assuming that the calories of ghee one ounce having normal population, find 98% confidence interval for mean calory and its standard deviation.

$\bar{x} = 107.4$, $n = 10$, $s = 5.46097$,

$$t_{\alpha/2, 9} = t_{.01,9} = 2.821$$

$\Rightarrow (102.53, 112.27)$

(When μ is unknown)

We know,

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$\boxed{\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}, n-1}} , \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \right) \text{ is } 100(1-\alpha)\% \text{ confidence interval of } \sigma^2}$$

Previous problem ...

$$\chi^2_{.99, 9} = 2.088$$

$$\chi^2_{.01, 9} = 21.666$$

so, conf. int for σ^2 is

$$(3.52 , 11.34).$$

If 31 measurements of boiling point of sulphur have a s.d. ·83°c. construct a 98% conf. Int. for The true s.d. of such measurements.

$$\chi^2_{.01,30} = 50.89$$

$$\chi^2_{0.99,30} = 14.95$$

$$\left(\sqrt{\frac{30 \cdot s^2}{50.89}}, \sqrt{\frac{30 \cdot s^2}{14.95}}\right) = (\cdot 6373, 1\cdot 1756)$$

is 98% conf Int of σ.

Brief

CLT

$X_1, X_2 \cdots X_n \xrightarrow{iid}$ mean μ, variance σ^2

$$S_n = X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2) \quad \text{as } n \to \infty$$

$$\bar{X} = \frac{S_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{as } n \to \infty$$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{as } n \to \infty$$

Ext^n:

$X_{11}, X_{12} \cdots X_{1n_1} \overset{iid}{\sim}$ mean μ_1, variance σ_1^2

$X_{21}, X_{22} \cdots X_{2n_2} \overset{iid}{\sim}$ mean μ_2, variance σ_2^2

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,1) \quad \text{as } \begin{array}{l} n_1 \to \infty \\ n_2 \to \infty \end{array}$$

Sample mean $= \frac{1}{n} \sum_i x_i = \bar{x}$

Sample variance $= \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 = s^2$

Is $\bar{x}$ true representative of population mean μ?

A statistic is unbiased estimator of population parameter if

$$E(\hat{\theta}) = \theta$$

$\hat{\theta}$: Statistic, θ: parameter

A statistic is consistent estimator if

$$\hat{\theta} \longrightarrow \theta \quad \text{as} \quad n \longrightarrow \infty$$

or, for every $\epsilon > 0$

$$\lim_{n \to \infty} \text{Prob}(|\hat{\theta} - \theta| < \epsilon) = 1$$

- Sample mean is unbiased estimate of the population mean, provided population mean exists.
- Sample mean is consistent estimate of population mean, provided population s.d. exists.
- Again $s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$ is unbiased estimate of population variance σ^2. As

$$E(s^2) = \frac{1}{n-1} \left\{ \sum_i E(x_i^2) - n \{E(\bar{x}^2)\} \right\}$$

$$= \frac{1}{n-1} (n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2) = \sigma^2$$

where, $E(x_i^2) = \mu^2 + \sigma^2$ & $E(\bar{x}^2) = \mu^2 + \frac{\sigma^2}{n}$.

- $\chi^2_{(n)} = \sum_i Z_i^2$ where $Z_i \sim N(0,1)$

- $t_{(n)} = \dfrac{Z}{\sqrt{\dfrac{\chi^2_{(n)}}{n}}}$ where $Z \sim N(0,1)$

- $\dfrac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$ or t_n

- $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$ or χ^2_{n-1}

$100(1-\alpha)\%$ confidence interval

For population mean μ

- σ known $\left[\bar{x} - Z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}} \, , \, \bar{x} + Z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}\right]$

- σ unknown $\left[\bar{x} - t_{\frac{\alpha}{2}, n-1} \dfrac{s}{\sqrt{n}} \, , \, \bar{x} + t_{\frac{\alpha}{2}, n-1} \dfrac{s}{\sqrt{n}}\right]$

For population variance σ^2

- $\left[\dfrac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \, , \, \dfrac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\right]$

F distribution

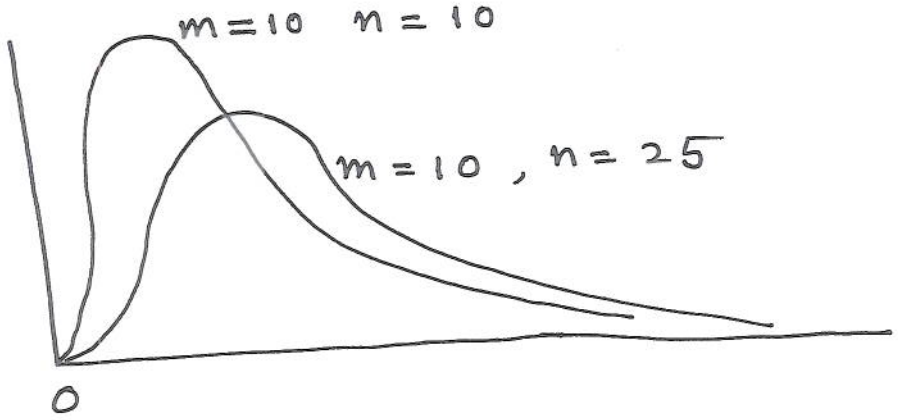
Let Y_1 and Y_2 are independently distributed

$$Y_1 \sim \chi^2_m \quad \& \quad Y_2 \sim \chi^2_n$$

Then $U = \dfrac{Y_1/m}{Y_2/n} = \dfrac{\chi^2_m/m}{\chi^2_n/n}$ have F-distribution with (m, n) degree of freedom.

Important property

$$F_{m,n,\alpha} = \frac{1}{F_{n,m,1-\alpha}}$$



$$\text{Prob}\left[F_{1-\alpha/2,\, m,\, n} \leq F \leq F_{\alpha/2,\, m,\, n}\right] = 1-\alpha$$

Let $X_1 \sim N(\mu_1, \sigma_1)$ & $X_2 \sim N(\mu_2, \sigma_2)$

Two random samples of size n_1 & n_2 be taken, s_1^2 and s_2^2 are sample variances

Then $\underset{(n_2-1,\ n_1-1)\ d.f.}{F} = \dfrac{\dfrac{(n_2-1)s_2^2}{\sigma_2^2} \Big/ (n_2-1)}{\dfrac{(n_1-1)s_1^2}{\sigma_1^2} \Big/ (n_1-1)} = \dfrac{s_2^2/\sigma_2^2}{s_1^2/\sigma_1^2}$

$$P\left[F_{1-\alpha/2,\, n_2-1,\, n_1-1} \leq F \leq F_{\alpha/2,\, n_2-1,\, n_1-1}\right] = 1-\alpha$$

$$\text{or, } P\left[\frac{s_1^2}{s_2^2} F_{1-\alpha/2,\, n_2-1,\, n_1-1} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} F_{\alpha/2,\, n_2-1,\, n_1-1}\right] = 1-\alpha$$

$100(1-\alpha)\%$ confidence interval for $\dfrac{\sigma_1^2}{\sigma_2^2}$ is

$$\left[\frac{s_1^2}{s_2^2} F_{1-\alpha/2,\, n_2-1,\, n_1-1}\ ,\ \frac{s_1^2}{s_2^2} \cdot F_{\alpha/2,\, n_2-1,\, n_1-1}\right]$$

Inference for Two population.

$x_1, x_2 \cdots x_{n_1} \longrightarrow N(\mu_1, \sigma_1^2)$

$y_1, y_2 \cdots y_{n_2} \longrightarrow N(\mu_2, \sigma_2^2)$

σ_1 & σ_2 are known

$$\bar{X} - \bar{Y} \sim N(\mu, \sigma)$$

where $\mu = \mu_1 - \mu_2$, $\sigma^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$.

$100(1-\alpha)\%$ confidence interval for $(\mu_1 - \mu_2)$

$$\left[(\bar{X}-\bar{Y}) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \ , (\bar{X}-\bar{Y}) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

σ_1 & σ_2 are unknown

But S_1^2 and S_2^2 are known

obtain combined or pooled variance

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1+n_2-2)}$$

Then $\dfrac{(\bar{X}-\bar{Y}) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

$$\left[(\bar{X}-\bar{Y}) - t_{\alpha/2, n_1+n_2-2}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \ , (\bar{X}-\bar{Y}) + t_{\alpha/2, n_1+n_2-2}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

Tensile strength tests were performed on two different grades of aluminium spars used in manufacturing wing of a commercial transport aircraft. From past experience with spar manufacturing process and the testing procedure, the s.d.s of tensile strengths are assumed to be known. The data are shown below. Find 90% confidence interval on the difference in mean tensile strength.

Spar Grade	Sample size	Sample Mean TS (kg/mm²)	s.d. (kg/mm²)
1	$n_1 = 10$	$\bar{x}_1 = 87.6$	$\sigma_1 = 1.0$
2	$n_2 = 12$	$\bar{x}_2 = 74.5$	$\sigma_2 = 1.5$

$$\left[(\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \; , \; (\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

$$\equiv \left[13.1 - 1.645\sqrt{\frac{1}{10} + \frac{1.5^2}{12}} \; , \; 13.1 + 1.645\sqrt{\frac{1}{10} + \frac{1.5^2}{12}}\right]$$

$$\equiv [12.22 \; , \; 13.98]$$

$$\bar{X} = 29.38 \quad , \quad \bar{Y} = 28.6$$

$$Z_{.05} = 1.65$$

$$\left[(\bar{X} - \bar{Y}) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \;,\; (\bar{X} - \bar{Y}) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

$$= \left[(29.38 - 28.6) - 2.04 \;,\; (29.38 - 28.6) + 2.04\right]$$

Use previous data (given that s.d.s are <u>unknown</u> but equal). Find 90% Conf. int for $(\mu_1 - \mu_2)$

$$s_1^2 = \frac{1}{7}\sum (x_i - \bar{x})^2 = 9.125$$

$$s_2^2 = \frac{1}{9}\sum (Y_i - \bar{Y})^2 = 4.267$$

$$S_p = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}} = 2.53$$

$$t_{.05, 16} = -1.75 \qquad t_{.95, 16} = 1.75$$

conf. int $\rightarrow$ $[-1.32, \; 2.88]$.

The weights of two types of mice in a psychology research lab are normally distributed. Type 1 mice have mean weight 28 gms and s.d. 3 gm. Type 2 mice have mean weight 28 gms and s.d. 2 gms.

A new diet is designed to increase the average weight of each type. The gm weights of 8 type 1 mice under new diet are

29, 28, 30, 31, 26, 32, 25, 34.

Type 2 mice under new diet are

27, 31, 30, 28, 29, 25, 31, 30, 29, 26

Find a 90% conf. int for $\mu_1 - \mu_2$.

In a batch chemical process used for etching printed circuit boards, two different catalysts are being compared to determine whether they require different emersion times for removal of identical quantities of photoresist material. Twelve batches were run with catalyst 1, resulting in a sample mean emersion time of $\bar{x}_1 = 24.6$ mins and s.d. $s_1 = .85$ min. Fifteen batches were run with catalyst 2, resulting in a mean emersion time of $\bar{x}_2 = 22.1$ mins and a s.d. $s_2 = .98$ min. Find 95% conf. int. on difference in means.
Find 90% conf. int. on the ratio of variances σ_1^2/σ_2^2.

conf. int. of $(\mu_1 - \mu_2) \rightarrow [1.76, 3.24]$

$$\frac{s_1^2}{s_2^2} F_{.95,14,11} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} F_{.05,14,11}$$

$$\frac{1}{F_{.05,11,14}}$$

$$\frac{(0.85)^2}{(.98)^2} \times 0.39 \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{(.85)^2}{(.98)^2} \times 2.74$$

$\rightarrow$ conf int of $\frac{\sigma_1^2}{\sigma_2^2}$ is $[.29, 2.06]$