No.

Date

statistics:

A statistics is a function of the observable random variables in a sample and known constrains.

e.g. sample mean, sample median.

$\bar{X} = \bar{X}_s \qquad \bar{\sigma}_s = \frac{\sigma}{\sqrt{n}}$

Proof: Given $i = 1, 2, \cdots, n$: Given $X \in$ norm dist.

$\bar{X}_s = \frac{1}{N} \sum_{i=1}^{n} X_i$

$E(\bar{X}) = E\left(\frac{1}{N} \sum_{i=1}^{N} x_i\right) = \frac{1}{N} \sum\left(\sum_{i=1}^{N} x_i\right) = \frac{1}{N} \cdot N\mu = \mu.$

where μ is Expecation of X.

$\sigma_1^2 = \frac{1}{n^2} \sum_{i=1}^{n}\left[X_i - E(X_i)\right]^2 = \frac{1}{n^2}\, n \cdot \sigma^2 = \frac{\sigma^2}{n}.$

$\sigma = \frac{\sigma}{\sqrt{n}}$

Then $Z = \frac{\hat{X} - \bar{\mu}_s}{\sigma_s} = \frac{\hat{X} - \mu}{\sigma} \cdot \sqrt{n}$

$\chi^2 \rightarrow \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma}\right)^2$ where $Z_i = \frac{Y_i - \mu}{\sigma}$

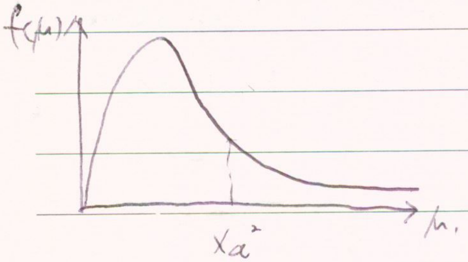
$Z_i^2 \in \chi^2$ distribution.

$P(\chi^2 > \chi_\alpha^2) = 1 - P(\chi^2 \leq \chi_\alpha^2) = \alpha.$

No.

Date RV: Random variable.

Typical χ^2 distribution: right-skewed.



find $P(Y<y_0) \to$ Pchisq (y_0, v)

find $x_\alpha^2 \to$ qchisq (P, v)

observe $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$

we have $\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{Y_i - \bar{Y}}{\sigma}\right)^2 = \sum_{i=1}^{n} z^2$.

So $\frac{(n-1)S^2}{\sigma^2}$ has a χ^2 distribution with $(n-1)$ df.

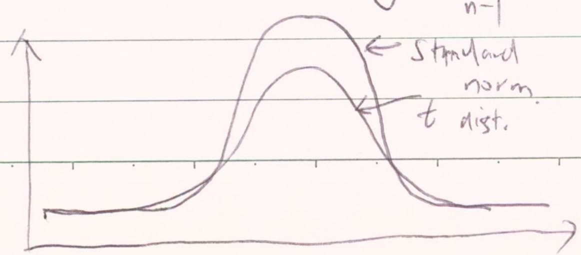
Where $\bar{Y}, s^2$ are independent RV.

let $Z \to$ std norm ~~the~~ random variable.

let $W \to \chi^2$ Random Variable.

$$T = \frac{Z}{\sqrt{\frac{W}{V}}} = \frac{\sqrt{n}\,\frac{(\bar{Y}-\mu)}{\sigma}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}} = \sqrt{n}\,\frac{(\bar{Y}-\mu)}{S}$$

$\boxed{\text{t distribution with df} \to (n-1)}$



No.

Date

$$\begin{cases} \text{Given } P(Y \le y_0) \xrightarrow{\text{Pth quantile}} qt(p, v) \\ \text{Given quantile} \xrightarrow{\text{Prob}} pt(q, v). \end{cases}$$

F-distribution : compare ~~Compute~~ var of two normal populations.

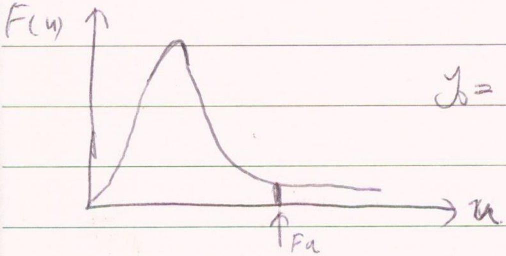
$$\frac{S_1^2}{\sigma_1^2} \Big/ \frac{S_2^2}{\sigma_2^2} \leftarrow \frac{W_1}{V_1} \Big/ \frac{W_2}{V_2} \rightarrow F\text{-}\cancel{\text{distribution}}\text{. distribution.}$$

with degree of freedom of $\boxed{V_1, V_2}$.

$$F = \frac{W_1/V_1}{W_2/V_2} \rightarrow \left(\frac{(n_1-1)S_1^2}{\sigma_1^2} \Big/ (n_1-1)\right) \Big/ \left(\frac{(n_2-1)S_2^2}{\sigma_2^2} \Big/ (n_2-1)\right)$$

$$\rightarrow \frac{S_1^2}{\sigma_1^2} \Big/ \frac{S_2^2}{\sigma_2^2}.$$

F distribution plot. — right-skew



y_0 = F score

$pf(y_0, V_1, V_2) \rightarrow P(Y \le y_0)$.

$qf(p, V_1, V_2) \rightarrow$ quantile

Central limit theorem: Let $Y_1, Y_2, Y_3, \dots, Y_n$ i.i.d.

$U_n = \dfrac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}}$ where $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ (样本分布).

$$\lim_{n \to \infty} P(U_n \le u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \quad (n > 30).$$

No.

Date

Proof: $U_n = \dfrac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n} \cdot \sum_{i=1}^{n} Z_i$

$m_{\sum Z_i}(t) = m_{Z_1}(t) \cdot m_{Z_2}(t) \cdots m_{Z_n}(t) = [m_{Z_1}(t)]^n$

(if $A \cap B = \phi$ $P(AB) = P(A) \cdot P(B)$).

$m_{U_n}(t) = \sqrt{n}\ m_{Z_1}(t) = [m_{Z_1}(\frac{t}{\sqrt{n}})]^n$.

By Taylor's theorem:

$$\begin{cases} m_{Z_1}(t) = m_{Z_1}(0) + m'_{Z_1}(0) + m''_{Z_1}(\xi)\frac{t^2}{2} \quad (0<\xi<t) \\ m_{Z_1}(0) = E(e^{0 \cdot Z_1}) = E(1) = 1. \\ m'_{Z_1}(0) = (1)' = 0 \end{cases}$$

$\longrightarrow m_{Z_1}(t) = 1 + \dfrac{m''_{Z_1}(\xi)}{2} t^2$.

$m_{U_n}(t) = \left[1 + \dfrac{m''_{Z_1}(\xi_n)}{2} \cdot (\frac{t}{\sqrt{n}})^2\right]^n$

$m_{U_n}(t) = \left[1 + \dfrac{m''_{Z_1}(\xi_n) t^2/2}{n}\right]^n$.

Recall that: $\lim_{n\to\infty} b_n = b$ then $\lim_{n\to\infty} (1+\frac{b_n}{n})^n = e^b$

then $\lim_{n\to\infty} m_{U_n}(t) = \lim_{n\to\infty} \left[1 + \dfrac{m''_{Z_1}(\xi_n) t^2/2}{n}\right]$

$= e^{\frac{t^2}{2}} \longrightarrow Z$ dist

~~Estima~~ Estimator:

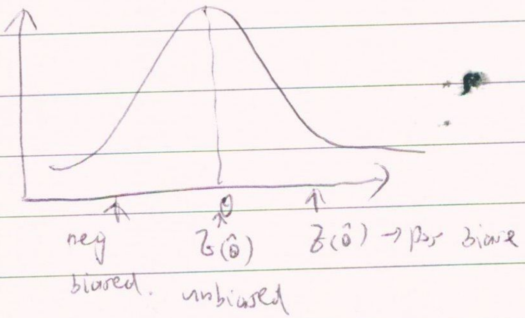
An estimator is a rule (formular) that tells how to calculate the value of an estimate based on the measurements contained in a sample.

θ : A population parameter

$\hat{\theta}$: point estimate of θ.

$E(\hat{\theta}) = \theta \rightarrow$ point estimator ~~a~~ unbiased.

$E(\hat{\theta}) > \theta \rightarrow$ positively biased.



Bias of $\hat{\theta} = B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

$$MSE = E\left[(\hat{\theta}-\theta)^2\right] = V(\hat{\theta}) + B^2(\hat{\theta})$$

standard error : $\sigma_{\hat{\theta}} \Rightarrow$ the std of the sampling distribution of the estimator $\hat{\theta}$.

Some of the example of Expected values / std errs. (given sample size n)

No.

Date

target Para	Point estimator	$E(\hat{\theta})$	std err.
Gaussian → μ	$\bar{Y}$	μ	$\frac{\sigma}{\sqrt{n}}$
binomial → p	$\hat{p} = \frac{Y}{n}$	$np/n = p$	$\frac{\sqrt{p(1-p)}}{\sqrt{n}}$
$\mu_1 - \mu_2$	$\bar{Y}_1 - \bar{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$p_1 - p_2$	$\frac{Y_1}{n_1} - \frac{Y_2}{n_2}$	$p_1 - p_2$	$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

(p_1, p_2 assumed to be independent).

Prove S'^2 is biased but S^2 not, where

$$\begin{cases} S'^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \\ S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \end{cases}$$

It can be shown that : $\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2$

$$E\left[\sum_{i=1}^{n} (Y_i - \bar{Y})^2\right] = E\left(\sum_{i=1}^{n} Y_i^2\right) - nE(\bar{Y}^2) \quad = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2$$

$$= \sum_{i=1}^{n} E(Y_i^2) - nE(\bar{Y}^2)$$

(note : Y_i is const)

since $E(Y^2) = V(Y) + [E(Y)]^2$

$$E(Y_i^2) = V(Y_i) + [E(Y_i)]^2 = \sigma^2 + \mu^2$$

$$E(\bar{Y}^2) = V(\bar{Y}) + [E(\bar{Y})]^2 = \frac{\sigma^2}{n} + \mu^2$$

$$E\left[\sum_{i=1}^{n} (Y_i - \bar{Y})\right]^2 = n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)$$

$$= (n-1)\sigma^2$$