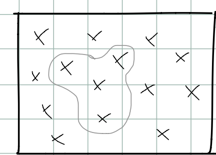
# Statistical Inference

## Def
Statistics is the science of collecting, summarizing and analyzing experimental data to provide the basis for inference of decisions concerning the true nature of the population at study.

## Ex
Height of a Chalmers female student.
Pick 10 students at random, measure their heights



Results: 182, 171, 177, 174, 186, 187, 193, 172, 180, 181

## Def
A random sample of size $n$ from the distribution of $X$ is a collection of $n$ independent r.v's, each with the same distribution as $X$.

## Notes
$\mu$ = the "true" average heights?
$\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 179.9$

## Def
A parameter of probability distribution (of population) is a number associated with the distribution (and in some way descriptive of that distribution).

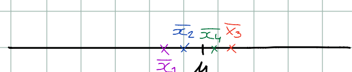
## Ex
$\mu_X$ - expectation and $\sigma_X^2$ - variance

## Def
A statistic is some specified numerical function of observed sample values $x_1, \dots, x_n$.

## Def
A statistic used to approximate/estimate a distribution parameter $\theta$ is called point estimator for $\theta$ and is denoted $\hat{\theta}$. A numerical value obtained on a given data is called estimate

## Def
An estimator $\hat{\theta}$ is an unbiased estimator of parameter $\theta$ iff $E[\hat{\theta}] = \theta$

## Picture




## Def
Let $X_1 \dots X_n$ be a random sample of size $n$ from the distribution of $X$. Then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is called sample mean.

## Proposition
$\bar{X}$ is an unbiased estimator of $\mu_X$ with $\sigma_{\bar{X}}^2 = \frac{\sigma^2_X}{n}$, $n$ = number of observations.

## Proof

$$E[\overline{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E[X_i]}_{=\mu_x} = \frac{1}{n}\cdot n \cdot \mu_x = \mu_x.$$

$$Var[\overline{X}] = Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} Var\left[\sum_{i=1}^{n} X_i\right] = \left\{\begin{matrix} X_1 \dots X_n \text{ are} \\ \text{independent} \end{matrix}\right\} = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{Var[X_i]}_{Var[X] = \sigma_x^2} = \frac{1}{n^2}\cdot n \cdot \sigma_x^2 = \frac{\sigma_x^2}{n}$$

## Ex

$\overline{X}' = X_1$

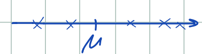
$E[\overline{X}'] = E[X_1] = \mu_x$

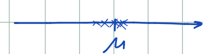
$Var[\overline{X}'] = Var[X_1] = \sigma_x^2$

$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$E[\overline{X}] = \mu_x$

$Var[\overline{X}] = \frac{\sigma_x^2}{n}$

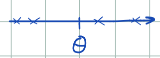


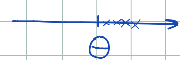




$E[\hat{\theta}_1] = \theta$

$E[\hat{\theta}_2] - \theta = b > 0$

$Var[\hat{\theta}_1] \gg Var[\hat{\theta}_2]$




## Intuition

We need to estimate $E[X-\mu_x]^2 = \sigma_x^2$

$E[X-\overline{X}]^2$

$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ - estimater fer $\sigma_x^2$

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}\left((X_i-\mu_x)-(\overline{X}-\mu_x)\right)^2\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}\left((X_i-\mu_x)^2 - 2(X_i-\mu_x)(\overline{X}-\mu_x) + (\overline{X}-\mu_x)^2\right)\right] =$$

$$\frac{1}{n}E\left[\sum_{i=1}^{n}(X_i-\mu_x)^2 - 2(\overline{X}-\mu_x)\sum_{i=1}^{n}(X_i-\mu_x) + n(\overline{X}-\mu_x)^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i-\mu_x)^2 - 2n(\overline{X}-\mu_x)^2 + n(\overline{X}-\mu_x)^2\right] =$$

$$\frac{1}{n}E\left[\sum_{i=1}^{n}(X_i-\mu_x)^2 - n(\overline{X}-\mu_x)^2\right] = \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i-\mu_x)^2 - n E[(\overline{X}-\mu_x)^2]\right] = \frac{1}{n}(n\sigma_x^2 - \sigma_x^2) = \frac{n-1}{n}\cdot\sigma_x^2$$

## Def

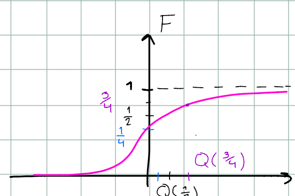
Let $X_1 \dots X_n$ be a random sample of size $n$ from the distribution $X$. The statistic $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is called a sample variance

For computing: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - n(\overline{X})^2$

## Def

Given a CDF $F$ we can define the quantile function $Q(p) = \inf\{x : F(x) \geq p\}$ for $p \in (0,1)$

## Pic




$P(\frac{1}{2})$ $Q(\frac{1}{2})$ = median

$P(\frac{3}{4})$ $Q(\frac{3}{4})$ = upper quartile

$P(\frac{1}{4})$ $Q(\frac{1}{4})$ = lower quartile