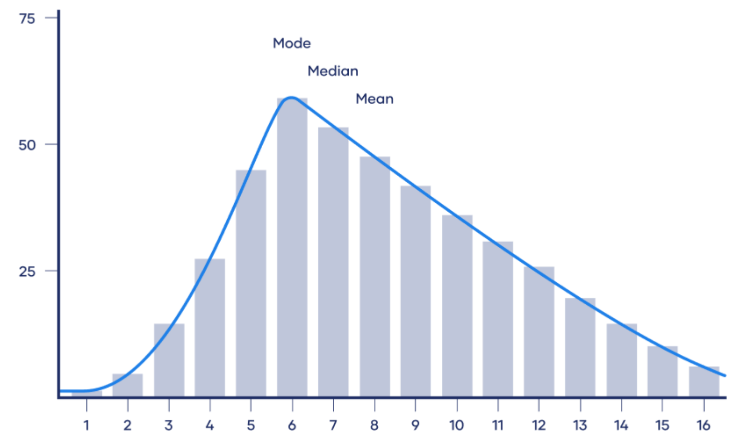
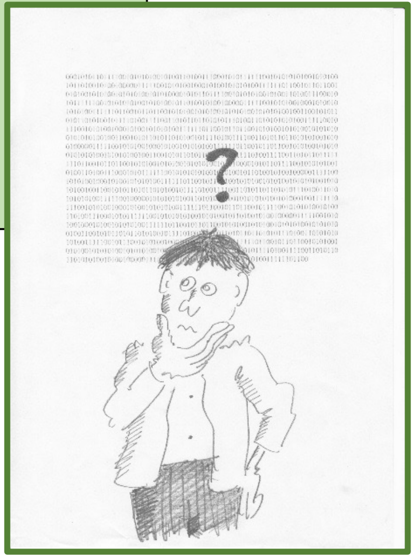
Confidence Intervals for Medians



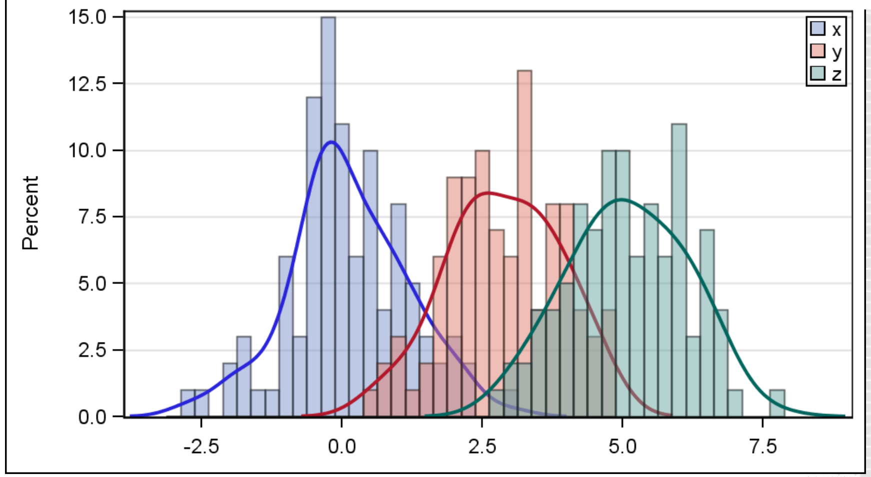
Statistics and data analysis

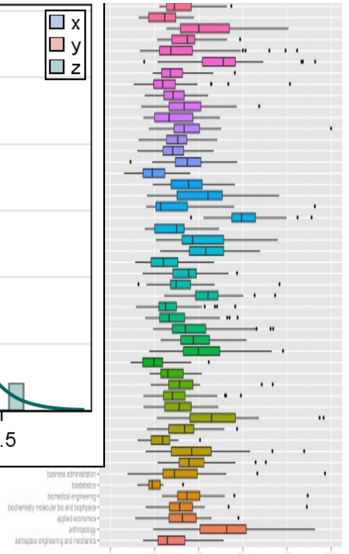
Zohar Yakhini

IDC, Herzeliya



0010011101010100101010100100100010
1010100010101111101011010011001001
1110101010011010010110010110110010





These slides follow DeGroot

Setting up

- Let X be a random variable with $\mathrm{median}(X) \; = \; m$.
- Take n samples from X: $x_1, \dots, x_n$.

  Order them and denote: $a_1 \le \cdots \le a_n$
- Let $c = \frac{n+1}{2}$. Our estimate of the median will be $\widehat{m} = a_c$
- Let $k \geq 1$. We want to estimate

$$P(m \in [a_{c-k}, a_{c+k}]) = P\big((a_1, \dots, a_{c-k} \le m) \wedge (a_{c+k}, \dots, a_n \ge m)\big)$$

Confidence interval

- Denote $A_k = (a_1, \ldots, a_{c-k} \leq m) \wedge (a_{c+k}, \ldots, a_n \geq m)$
- Let J_i be a RV such that $\begin{cases} 1, & x_i \leq m \\ 0, & else \end{cases}$. Note: $J_i \sim Bernoulli(0.5)$.
- $a_1, \ldots, a_{c-k} \leq m \iff \sum_{i=1}^{n+1} J_i \geq c - k$
- $P(A_k) = P\left(c - k \leq \sum_{i=1}^{n+1} J_i \leq c + k\right)$
- Note that $\sum_{i=1}^{n+1} J_i \sim Binom(n + 1, 0.5)$
- Given $\alpha > 0$, find k that is the smallest to satisfy:
  $1 - \alpha \leq P(c - k \leq B \leq c + k)$,
  where $\mathrm{B} \sim Binom(n + 1, 0.5)$
- Then, $P(m \in [a_{c-k}, a_{c+k}]) \geq 1 - \alpha$, and k is minimal w this property

Confidence interval

We just computed an interval of values: $[a_{c-k}, a_{c+k}]$,

which is the

<u>$1-\alpha$ confidence interval</u>

for the mean of the sampled distribution.