הרצאה 2

מטרה: נאמדת הנשוי של נאמדים

מדגם: יקיר של נמדדים "אקראי"

$$x = \{x_1, \dots, x_m\}$$

דוגמה:
* אמת DC רועש
$n=5$
$n=500$
מה יהיה קרוב ל-μ?
$\bar{x}_5$ לא $\bar{x}_{500}$

יהיה נמוכים = יהיה מדויק

1. **נאמדת**

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

נמדדים — מספר

יש תוחלת "אמיתית" לפי יודע

μ

כמה נמדדים

הכללה: $\bar{x} \sim \frac{\sigma}{\sqrt{n}}$ — σ מה μ

שגיאה כי $\bar{x}$ קטנים

לפי: $\sqrt{n}/1$ ($n^{-1/2}$)

תזכורת תמ' סדר: נתון, תמ' אמת

דוגמה: משכורת תמ'ות — סדרים — ממוצעים

$$x_{(1)} \le \cdots \le x_{(n)}$$

sample median

$$\text{median}(x) = \begin{cases} x_{(\frac{n+1}{2})}, & n \text{ odd}, \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & n \text{ even} \end{cases}$$

Mode סדרה הכי נפוצה

שונות Sample variance ← שמסמן σ^2 יש תוחלת לפי יודע

unbiased
$$s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

biased
$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

S מסיר תקין

נתוח ע"י נימוקים חוצים

תהה: שנה אופן סדרים מלים

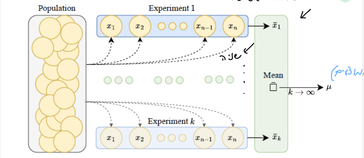
פה מדם לוקחים n סדרים שבחרו באקן אקראי.

n סדרים מדם ראשונה

דוגמה: נמדדת אמת DC

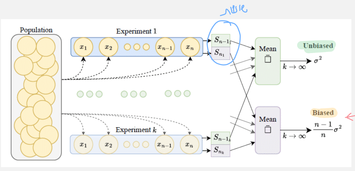
פה מדם 5 מדידות ← $\bar{x}$

ממוצע של



k סדרים (ממוצעים) — התפלגות — ממוצעים ← סדר k אמת

סדרים k מביאים חישוב ממוצע של k סדרים



Python, בדיר הנחה סטטיסטיקה

בדיר מהיר

סטה שלמה מוכנה כפל סדר n לעומת

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

Matlab, הנחה לאות הבנה

$$\text{MSE}_{n-1} = \frac{1}{k}\sum_{j=1}^{k}\left(\underbrace{s_{n-1_j}}_{\text{סדר שלם}} - \sigma^2\right)^2 \to \frac{2}{n-1}\sigma^4$$

$$\text{MSE}_n = \frac{1}{k}\sum_{j=1}^{k}\left(s_{n_j} - \sigma^2\right)^2 \to \frac{2n-1}{n^2}\sigma^4$$

$$\text{MSE}_{n-1} \geq \text{MSE}_n$$

mean-squared error (MSE)

דוגמה לסבירות:

- True distribution: $X \sim \mathcal{N}(0,1)$, so $\mu = 0$, $\sigma^2 = 1$.
- Sample sizes: $n \in \{5, 20, 100\}$.
- Number of replicates: $k = 5000$.

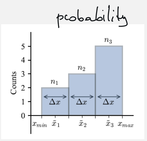
נמדד תמיד לפי תשוב של $\bar{X}$, S^2 ←

n	$\hat{\mu}_n$	Empirical $\text{Var}[\bar{x}]$	Theoretical $\text{Var}[\bar{x}] = 1/n$	$\hat{\sigma}^2_{\text{biased}}$	$\hat{\sigma}^2_{\text{unbiased}}$	$\widehat{\text{MSE}}_{\text{biased}}$	$\widehat{\text{MSE}}_{\text{unbiased}}$
5	0.001	0.198	0.200	0.79	0.99	0.36	0.50
20	−0.000	0.049	0.050	0.95	1.00	0.10	0.11
100	0.000	0.010	0.010	0.99	1.01	0.02	0.02

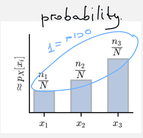
מתמלאים אוכלוסיה תשובה

גרפים

histogram

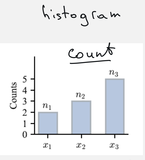


תוצאה רגיל



1 = סכום

0 פחות N תוצאה



פרק

תוצאה בדידה

* מחלקים לתחומים ברוחב אחיד
* סופרים כמה נתונים

Boxplot

מטרה: השוואה של כמה מדידות

הקדמה מבוסס Quartile

- Q_1 (first quartile, 25th percentile) — the median of the lower half of the data. ← 25% מתחת, 75% מעל
- Q_2 (second quartile, 50th percentile) — the *median* of the entire data set. ← חצי/חצי
- Q_3 (third quartile, 75th percentile) — the median of the upper half of the data. ← 75% מתחת, 25% מעל

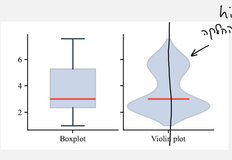
$$\text{IQR} = Q_3 - Q_1$$

interquartile range (IQR)

הגדרה: $[Q_1 - 1.5\,\text{IQR},\ Q_3 + 1.5\,\text{IQR}]$ כל מחוץ לתחום הוא "חריג" Outlier

Violin Plot

מטרה: שילוב boxplot ו-histogram



histogram עם הנחה