Co-funded by the
Erasmus+ Programme
of the European Union

DeCAIR

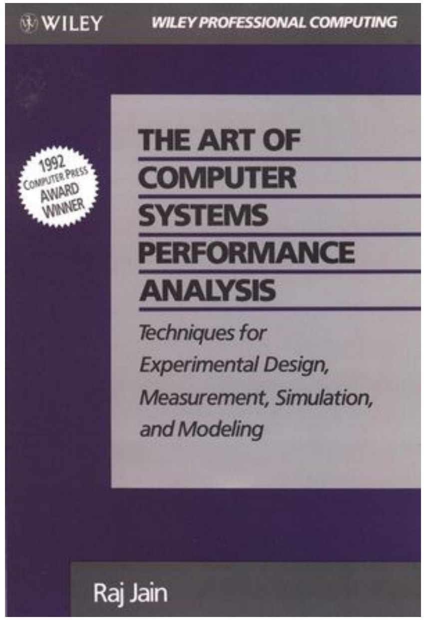
Comparing Systems Using Sample Data

Prof. Gheith Abandah

Developing Curricula for Artificial Intelligence and Robotics (DeCAIR)
618535-EPP-1-2020-1-JO-EPPKA2-CBHE-JP

References

- Raj Jain, **The Art of Computer Systems Performance Analysis**, Wiley, 1991.
  - Part I: An Overview of Performance Evaluation
  - Part II: Measurement Techniques and Tools
  - Part III: Probability Theory and Statistics
  - Part IV: Experimental Design and Analysis
  - Part V: Simulation



Outline

- Sample Versus Population
- Confidence Interval for the Mean
- Confidence Interval for Small Samples
- Approximate Visual Test
- Sample Size Selection

Sample Versus Population

- Generate several million random numbers (**population**) of unknown mean μ and standard deviation σ.
  - We use Greek letters for the population.
- Draw a sample of n observations to characterize the population.
- The **sample** has mean $\bar{x}$ and standard deviation s.
  - We use English letters for the sample.
- We use the **sample statistics** ($\bar{x}$ and s) **to estimate the population's characteristics** (μ and σ).
- The sample mean does not necessarily equal the population mean: $\bar{x} \neq \mu$.

Outline

- Sample Versus Population
- Confidence Interval for the Mean
- Confidence Interval for Small Samples
- Approximate Visual Test
- Sample Size Selection

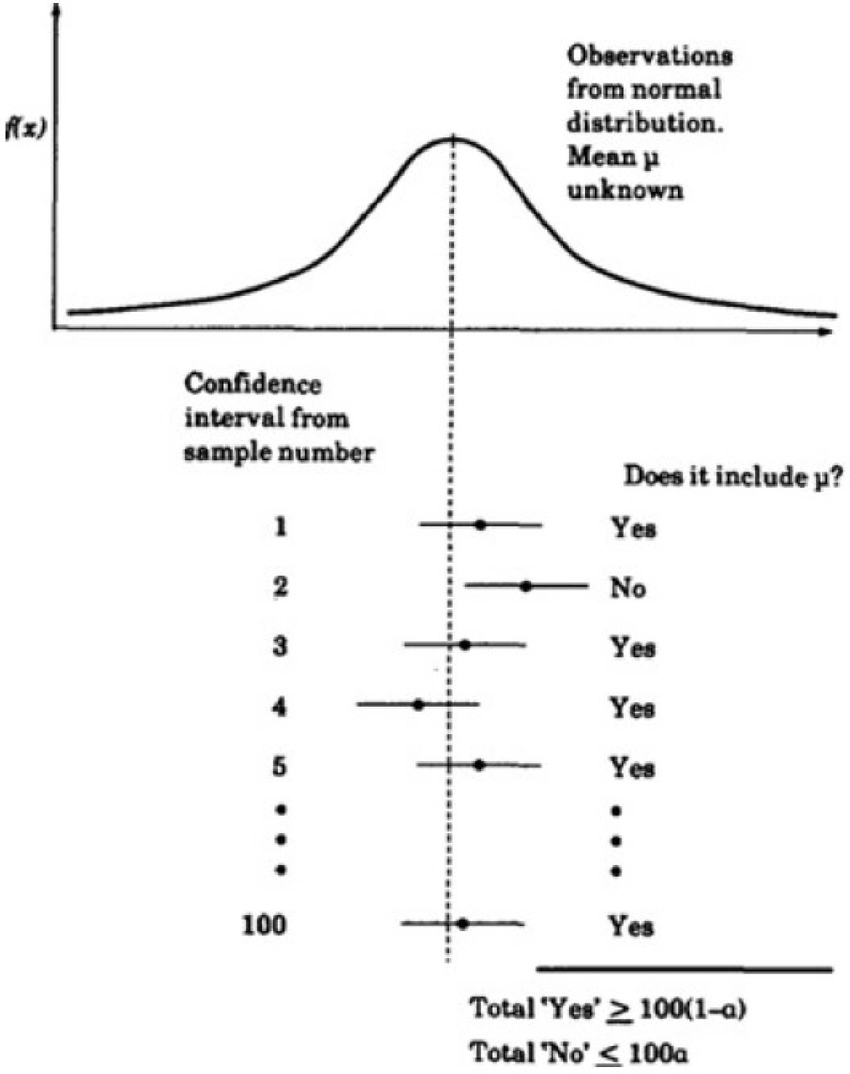
Confidence Interval for the Mean

- Drawing k samples implies getting k sample means
  - We don't get a single estimate of μ.
- We want to find the **confidence interval** (c_1, c_2) with $\text{Probability}\{c_1 \leq \mu \leq c_2\} = 1 - \alpha$

- **Confidence coefficient**: $1 - \alpha$, *e.g.*, 0.90.
- **Confidence level**: $100(1 - \alpha)$%, *e.g.*, 90%.
- **Significance level**: α, *e.g.*, 0.10.

Meaning of the Confidence Interval

- If we take 100 samples and construct confidence interval for each sample, **the interval will include the population mean μ** in $100(1-\alpha)$ cases.

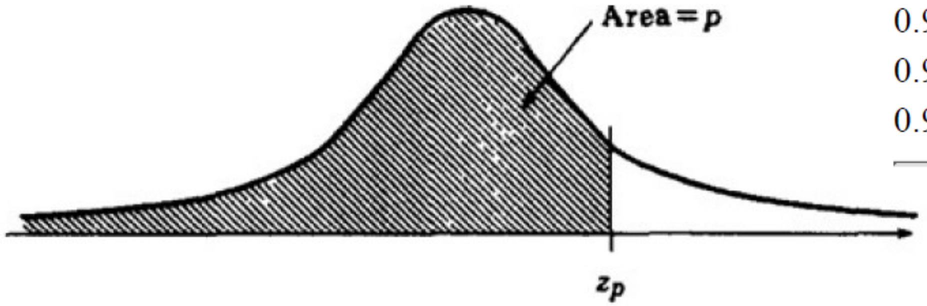


Determining the Confidence Interval

- **Central limit theorem**: The sample mean of independent observations has normal distribution: $\bar{x} \sim N(\mu, \sigma/\sqrt{n})$, where μ = population mean, σ = population standard deviation, and n = sample size.
- $\sigma/\sqrt{n}$ is called standard error.
- For $100(1-\alpha)\%$ **confidence level**, the **confidence interval** for μ is $\left(\bar{x} - z_{1-\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right), \bar{x} + z_{1-\frac{\alpha}{2}}\left(\frac{s}{\sqrt{n}}\right)\right)$
- where $z_{1-\frac{\alpha}{2}}$ = $(1-\frac{\alpha}{2})$-**quantile of the unit normal variate** $N(0,1)$

Table A.2

- Lists z_p for a given p.
- **Example**: For confidence interval at 90%, $\alpha = 0.10$, and $p = 1 - \frac{\alpha}{2} = 0.95$. The entry in the row labeled 0.95 and column labeled 0.000 gives $z_p = 1.645$.



p	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.90	1.282	1.287	1.293	1.299	1.305	1.311	1.317	1.323	1.329	1.335
0.91	1.341	1.347	1.353	1.359	1.366	1.372	1.379	1.385	1.392	1.398
0.92	1.405	1.412	1.419	1.426	1.433	1.440	1.447	1.454	1.461	1.468
0.93	1.476	1.483	1.491	1.499	1.506	1.514	1.522	1.530	1.538	1.546
0.94	1.555	1.563	1.572	1.580	1.589	1.598	1.607	1.616	1.626	1.635
0.95	1.645	1.655	1.665	1.675	1.685	1.695	1.706	1.717	1.728	1.739
0.96	1.751	1.762	1.774	1.787	1.799	1.812	1.825	1.838	1.852	1.866
0.97	1.881	1.896	1.911	1.927	1.943	1.960	1.977	1.995	2.014	2.034
0.98	2.054	2.075	2.097	2.120	2.144	2.170	2.197	2.226	2.257	2.290

p	0.0000	0.0001	0.0002	0.0003	0.0004	0.0005	0.0006	0.0007	0.0008	0.0009
0.990	2.326	2.330	2.334	2.338	2.342	2.346	2.349	2.353	2.357	2.362
0.991	2.366	2.370	2.374	2.378	2.382	2.387	2.391	2.395	2.400	2.404
0.992	2.409	2.414	2.418	2.423	2.428	2.432	2.437	2.442	2.447	2.452
0.993	2.457	2.462	2.468	2.473	2.478	2.484	2.489	2.495	2.501	2.506
0.994	2.512	2.518	2.524	2.530	2.536	2.543	2.549	2.556	2.562	2.569
0.995	2.576	2.583	2.590	2.597	2.605	2.612	2.620	2.628	2.636	2.644
0.996	2.652	2.661	2.669	2.678	2.687	2.697	2.706	2.716	2.727	2.737
0.997	2.748	2.759	2.770	2.782	2.794	2.807	2.820	2.834	2.848	2.863
0.998	2.878	2.894	2.911	2.929	2.948	2.968	2.989	3.011	3.036	3.062
0.999	3.090	3.121	3.156	3.195	3.239	3.291	3.353	3.432	3.540	3.719

Example: *n* = 32

- For $\bar{x}$ = 3.90, s = 0.95, and n = 32.
- **A 90% confidence interval** for the mean = $3.90 \pm \frac{(1.645)(0.95)}{\sqrt{32}} = (3.62, 4.17)$
- We can state with 90% confidence that the population mean is between 3.62 and 4.17.
- **A 95% confidence interval** for the mean = $3.90 \pm \frac{(1.960)(0.95)}{\sqrt{32}} = (3.57, 4.23)$
- **A 99% confidence interval** for the mean = $3.90 \pm \frac{(2.576)(0.95)}{\sqrt{32}} = (3.46, 4.33)$

Outline

- Sample Versus Population
- Confidence Interval for the Mean
- Confidence Interval for Small Samples
- Approximate Visual Test
- Sample Size Selection

Confidence Interval for Small Samples

- $100(1-\alpha)\%$ confidence interval for $n < 30$ is:

$$(\bar{x} - t_{[1-\alpha/2;n-1]} s/\sqrt{n}, \bar{x} + t_{[1-\alpha/2;n-1]} s/\sqrt{n})$$

- where $t_{[1\text{-}\alpha/2;\, n\text{-}1]}$ = $(1-\alpha/2)$-**quantile of a** t**-variate with** $n-1$ **degrees of freedom**.

Table A.3

- Lists $t_{[p;\, n]}$.
- **Example**: the $t_{[0.95;\, 7]}$ for 90% confidence interval of the mean of a sample of 8 observation is 1.895.

	p							
n	0.6000	0.7000	0.8000	0.9000	0.9500	0.9750	0.9950	0.9995
1	0.325	0.727	1.377	3.078	6.314	12.706	63.657	636.619
2	0.289	0.617	1.061	1.886	2.920	4.303	9.925	31.599
3	0.277	0.584	0.978	1.638	2.353	3.182	5.841	12.924
4	0.271	0.569	0.941	1.533	2.132	2.776	4.604	8.610
5	0.267	0.559	0.920	1.476	2.015	2.571	4.032	6.869
6	0.265	0.553	0.906	1.440	1.943	2.447	3.707	5.959
7	0.263	0.549	0.896	1.415	1.895	2.365	3.499	5.408
8	0.262	0.546	0.889	1.397	1.860	2.306	3.355	5.041
9	0.261	0.543	0.883	1.383	1.833	2.262	3.250	4.781
10	0.260	0.542	0.879	1.372	1.812	2.228	3.169	4.587
11	0.260	0.540	0.876	1.363	1.796	2.201	3.106	4.437
12	0.259	0.539	0.873	1.356	1.782	2.179	3.055	4.318
13	0.259	0.538	0.870	1.350	1.771	2.160	3.012	4.221
14	0.258	0.537	0.868	1.345	1.761	2.145	2.977	4.140
15	0.258	0.536	0.866	1.341	1.753	2.131	2.947	4.073
16	0.258	0.535	0.865	1.337	1.746	2.120	2.921	4.015
17	0.257	0.534	0.863	1.333	1.740	2.110	2.898	3.965
18	0.257	0.534	0.862	1.330	1.734	2.101	2.878	3.922
19	0.257	0.533	0.861	1.328	1.729	2.093	2.861	3.883

Example: *n* = 8

- Sample: -0.04, -0.19, 0.14, -0.09, -0.14, 0.19, 0.04, and 0.09.
- Mean = 0, Sample standard deviation = 0.138.
- **For 90% interval**: $t_{[0.95;7]}$ = 1.895

- Confidence interval for the mean is

$$0 \pm 1.895 \times \frac{0.138}{\sqrt{8}} = 0 \pm 0.0925 = (-0.0925, 0.0925)$$

Example: Are Two Systems Different?

- **Paired performance**: {(5.4, 19.1), (16.6, 3.5), (0.6, 3.4), (1.4, 2.5), (0.6, 3.6), (7.3, 1.7)}
- **Differences**: {-13.7, 13.1, -2.8, -1.1, -3.0, 5.6}
- Sample mean = -0.32
- Sample standard deviation = 9.03
- The 0.95-quantile of a t-variate with five degrees of freedom is 2.015.
- **90% confidence interval** for the mean is

$$-0.32 \pm 2.015 \times \frac{9.03}{\sqrt{6}} = -0.32 \pm 7.43 = (-7.75, 7.11)$$

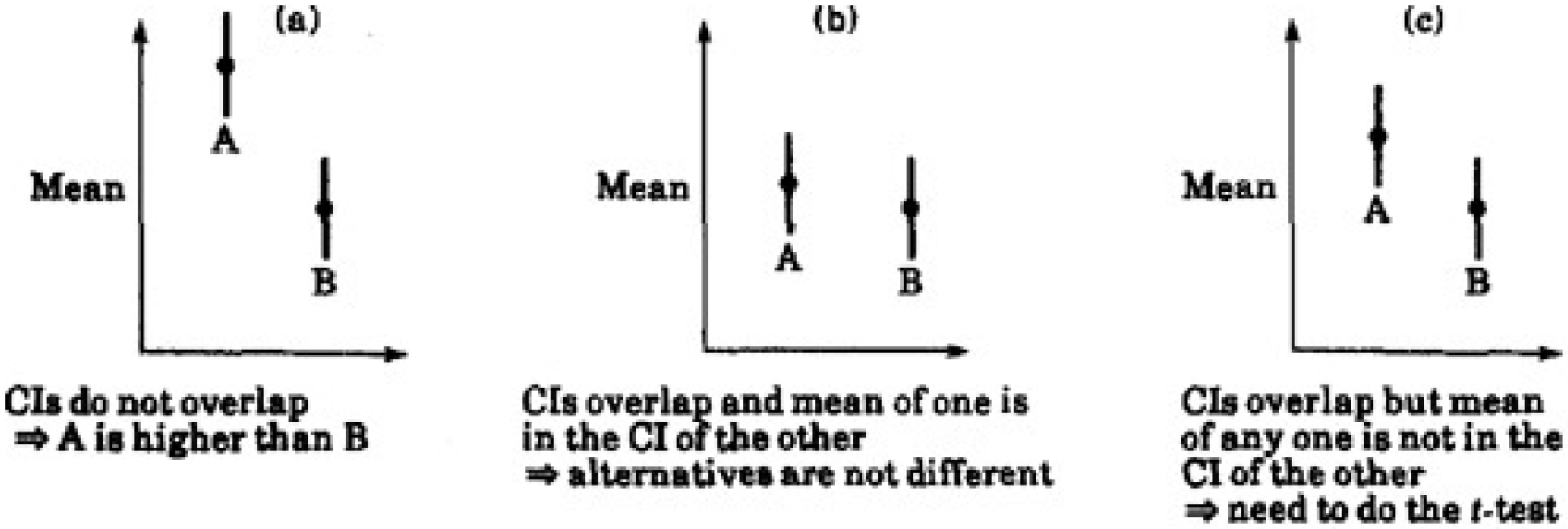
- The confidence interval **includes zero**. Therefore, **the two systems are not different**.

Outline

- Sample Versus Population
- Confidence Interval for the Mean
- Confidence Interval for Small Samples
- Approximate Visual Test
- Sample Size Selection

Approximate Visual Test

- **To compare two unpaired samples** by simply computing the confidence interval for each alternative separately.



Outline

- Sample Versus Population
- Confidence Interval for the Mean
- Confidence Interval for Small Samples
- Approximate Visual Test
- Sample Size Selection

Sample Size Selection

- Larger sample $\longrightarrow$ Narrower confidence interval $\longrightarrow$ Higher confidence
- **How many observations** n are needed to get an **accuracy** of $\pm r\%$ and a **confidence level** of $100(1-\alpha)\%$?

$$\mathrm{CI} = (\bar{x}(1 - r/100), \bar{x}(1 + r/100))$$

$$\bar{x} \mp z\frac{s}{\sqrt{n}} = \bar{x}\left(1 \mp \frac{r}{100}\right)$$

$$z\frac{s}{\sqrt{n}} = \bar{x}\frac{r}{100}$$

$$n = \left(\frac{100zs}{r\bar{x}}\right)^2$$

Example: Sample size

- Based on a preliminary test, the sample mean of the response time is 20 seconds, and the sample standard deviation is 5. How many repetitions are needed to get the response time accurate within 1 second at 95% confidence?
- Required accuracy of 1 in 20 seconds → 5%
- Here, $\bar{x} = 20$, $s = 5$, $z = 1.960$, and $r = 5$.
- $n = \left(\frac{100zs}{r\bar{x}}\right)^2 = \left(\frac{100\times1.960\times5}{5\times20}\right)^2 = 96.04 \rightarrow 97$

Summary

- Sample Versus Population
- Confidence Interval for the Mean
- Confidence Interval for Small Samples
- Approximate Visual Test
- Sample Size Selection