COMP 190-088: Systems Performance Analysis

Analyzing Measured Data-II

Jasleen Kaur

Department of Computer Science
The University of North Carolina at Chapel Hill

Spring 2005

Analysis of Measured Data

- Summarizing the Average
- **Summarizing the Variability**
- Comparing Systems Using Sample Data
- Simple Linear Regression Models

Summarizing Variability

- Range:
  - Non-stable representative
    - Sensitive to outliers
  - If variable is bounded, gives best estimate of bounds
- Sample Variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
  - Representative if distribution is uni-modal and symmetric
  - Divided by (n-1)!
    - (n-1) degrees of freedom
  - Standard deviation:
    - Same units as mean
- Coefficient of Variation (COV): $= s/\bar{x}$
  - Unit-less manner of talking about the standard deviation

Summarizing Variability

- Percentiles:
  - 90-percentile:
    - Sort all values and report the 90% value
    - 90% of observations lie below this value
  - ☺:
    - Less prone to outliers than range
    - Good for non-symmetric distributions
  - Quartile: (25%, 50%, 75%)
    - 50-quartile = median
  - Inter-quartile range
    - Difference between 75% and 25% values
- Mean Absolute Deviation (MAD): $= \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$
  - Easy to compute with multiplication or square root

Analysis of Measured Data

- Summarizing the Average
- Summarizing the Variability
- **Comparing Systems Using Sample Data**
- Simple Linear Regression Models

Sample vs. Population

- Sample:
  - The measured data set, which is only a subset of all possible outcomes
  - eg, set of *n* execution times measured when program is run *n* times
- Population:
  - The complete set of all possible data elements
  - eg, set of all possible execution times that can occur if program is run infinite times
- Sample characteristics are only an estimation of population characteristics
  - Population characteristics are fixed
    - eg, population mean
  - Sample estimates are random variables
    - eg, sample means of two samples likely to be different

Confidence Interval for the Mean

- Finite number of finite-sized samples do not give perfect estimate
  - Can derive probabilistic bounds: Prob(c1 ≤ μ ≤ c2) = 1 – α
    - Confidence interval: (c1, c2)
    - Significance level: α
    - Confidence level: 100(1- α)
    - Confidence coefficient: (1- α)
- Determining confidence interval:
  - Take k different samples (with several observations each)
  - Find sample means of each
  - Sort in increasing order
  - Take the [1 + (α/2)(k-1)]th and [1 + (1-α/2)(k-1)]th elements

Need to gather <u>many</u> samples to get good accuracy!

Computing Single-sample Confidence Intervals

- Central limit theorem:
  - If {x1, …, xn} are independent observations of a sample, and come from the same population with mean μ and standard deviation σ,
    - Then sample mean for **<u>large</u>** samples is normally distributed

$$\bar{x} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

- Estimating a confidence interval of 100(1- α):
  - Use the (1- α/2) quantiles of a normal distribution as estimates
    - s: sample standard deviation

$$\left(\bar{x} - z_{1-\alpha/2}\, s/\sqrt{n},\ \bar{x} + z_{1-\alpha/2}\, s/\sqrt{n}\right)$$

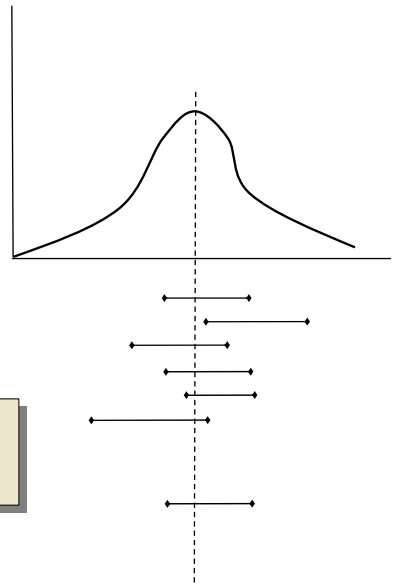
Example

- Sample mean = 3.90
- Standard deviation = 0.95
- n = 32
- 90% confidence interval for the mean:

$$= \bar{x} \mp z_{1-\alpha/2} \frac{s}{\sqrt{n}}$$

$$= 3.90 \mp (1.645)(0.95) / \sqrt{32}$$

$$= (3.62, 4.17)$$



Confidence intervals of 90 out of 100 samples would include the population mean

Standard Error

- Standard deviation of the sample mean: $= \frac{s}{\sqrt{n}}$
  - Different from the population standard deviation, σ
  - Decreases as the sample size (n) increases
- ☺:
  - Single sample can help estimate μ
- ☹:
  - Sample should be sufficiently large

What if only a small sample is available?

Confidence Intervals for Small Samples

- Can be constructed **only** if sample observations come from a normally-distributed population
  - Sample mean follows a "t-variate with (n-1) degrees of freedom"

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

Comparing Two Alternatives

- To test whether the population mean is (not) A
  - See if A lies in the confidence interval for the mean
  - Sample mean may not be equal to A!
- To compare two systems
  - Test to see if difference in population means is zero or not
- If samples are "paired"
  - eg, if the i-th test of each system corresponds to running the same benchmark i on the system
  - Generate a sample of performance difference
    - Test to see if a zero mean is likely

Comparing Systems With Paired Samples

Workload	System A	System B	Performance Difference
1	5.4	19.1	-13.7
2	16.6	3.5	13.1
3	0.6	3.4	-2.8
4	1.4	2.5	-1.1
5	0.6	3.6	-3.0
6	7.3	1.7	5.6

- Sample mean = -0.32
- Sample variance = 81.62
- 90% confidence interval = -0.32 +/- (2.015)*(81.62/6)$^{1/2}$
  - = (-7.75, 7.11)
  - Includes zero
  - The two systems are not significantly different

Comparing Systems With Unpaired Samples

- t-test
  - Compute sample standard deviations, s_a and s_b
  - Compute the mean difference: $\bar{x}_a - \bar{x}_b$
  - Compute standard deviation of the mean difference: $s = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$
  - Compute the effective number of degrees of freedom

$$\nu = \frac{\left(s_a^2/n_a + s_b^2/n_b\right)^2}{\frac{1}{n_a+1}\left(\frac{s_a^2}{n_a}\right)^2 + \frac{1}{n_b+1}\left(\frac{s_b^2}{n_b}\right)^2} - 2$$

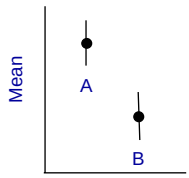
  - Compute confidence interval for the mean difference

$$(\bar{x}_a - \bar{x}_b) \mp t_{[1-\alpha/2;\nu]} s$$

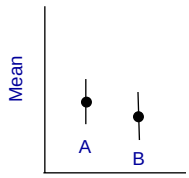
If confidence interval includes 0, systems are not significantly different

Visual Test

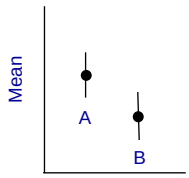
- Simpler test to compare two unpaired samples



A is higher than B



Alternatives are not different



Need to do the t-test

Do not perform complex analysis if simpler tests work for your data

What Confidence Level to Use?

- Depends on loss incurred if parameters is outside the range, and gain achieved if it is inside the range
  - Need not always be 90% or 95%
  - eg, how many lottery tickets to buy?
    - Ticket cost is $1, prize is $5M, and 10M tickets are sold.
    - Probability of winning on 1 ticket = 10^{-7}.
    - To win with 90% confidence, need to buy 9M tickets!
    - For most people, 0.01% confidence would be fine.

Confidence Intervals for Proportions

- Proportions useful for summarizing categorical variables
  - eg, what fraction of the population is of type A?
    - Sample proportion = $p = n_A/n$
    - Variance for the sample proportion = $p(1-p)/n$
    - Confidence interval for the proportion = $p \mp z_{[1-\alpha/2]}\sqrt{p(1-p)/n}$
  - Valid only if: $np \geq 10$
- Special case: comparing two systems with "paired" observations
  - Extend test for zero means to test for proportions
  - Eg, in 26 out of 40 repetitions of same experiment, A was found better than B
    - Can we say with 99% confidence that A is better?
    - p = 26/40 = 0.65
    - 99% confidence interval = (0.46, 0.84)
      - Includes point of equality, 0.5
    - 90% confidence interval = (0.53, 0.77)
      - We can say with 90% confidence that A is better

Determining Sample Size

- The confidence level drawn from a sample depends on its size
  - Larger the sample, higher is the confidence
  - Collecting larger samples requires more effort and resources
    - Find the smallest sample size that will produce the desired confidence
- What is the sample size required to achieve a given level of accuracy and confidence?
  - First conduct a small set of preliminary measurements to estimate the variance
  - Then use this estimate to determine the required sample size

Sample Size for Determining Mean

- How big a sample do we need to estimate the mean system performance with:
  - An accuracy of: ± r%
  - A confidence level of: 100(1-a)
- For sample size n, confidence interval is = $\bar{x} \mp z_{[1-\alpha/2]} s/\sqrt{n}$
- For r% accuracy, confidence interval should be = $\bar{x}\left(1 \mp r/100\right)$
- Equating and solving, we get:

$$n = \left(100\, z_{[1-\alpha/2]} s / r\bar{x}\right)^2$$

Sample Size for Determining Proportions

- How big a sample do we need to estimate a proportion with:
  - An accuracy of: ± r
  - A confidence level of: 100(1-a)
- For sample size n, confidence interval is = $p \mp z_{[1-\alpha/2]}\sqrt{p(1-p)/n}$
- For r accuracy, confidence interval should be = $p \mp r$
- Equating and solving, we get:

$$n = z^2_{[1-\alpha/2]}\, p(1-p)/r^2$$

Sample Size for Comparing Two Alternatives

- How big a sample do we need to state with 100(1- a) confidence that system A is better?
  - Use the requirement of non-overlapping intervals
- Example: loss rates of router queue management algorithms
  - Preliminary analysis:
    - Algorithm A loses 0.5% of packets
    - Algorithm B loses 0.6% of packets
  - How many packets do we need to observe to state with 95% confidence that A is better?
  - Mean loss rate with A, p_A, should be less than mean loss rate with B, p_B
    - Confidence intervals for p_A should be less than that for p_B

$$p_A \mp z_{[1-\alpha/2]}\sqrt{p_A(1-p_A)/n} \le p_B \mp z_{[1-\alpha/2]}\sqrt{p_B(1-p_B)/n}$$

$$\boxed{n \ge 84340}$$