Variability and Errors

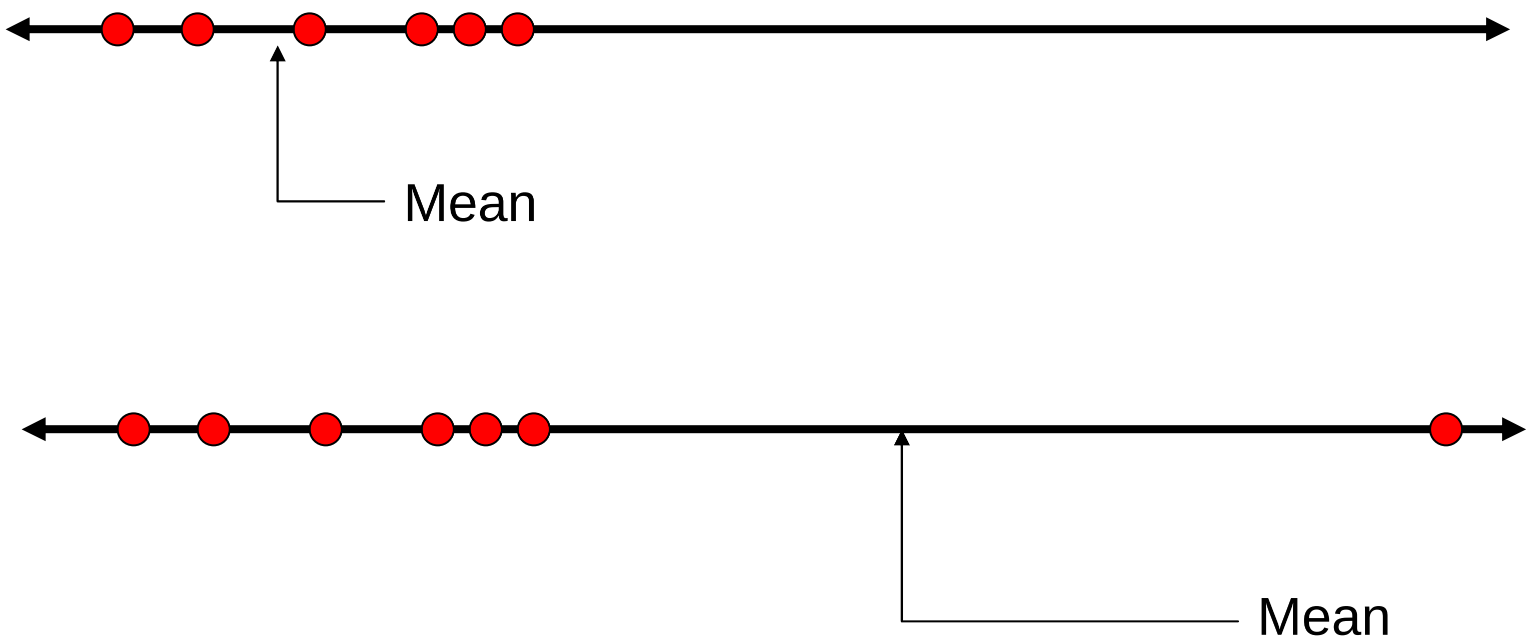
Per Hurtig

Today

- Variability
- Errors in measurements
  - Different types of errors
  - How to deal with them
- Comparing two sets of measurements

Variability

- A mean value is a single number that represents a possibly large pool of data
- Using only mean values are dangerous!



Quantifying variability

- Means hide information about variability
- How “spread out” are the values?
- How much spread relative to the mean?
- What is the shape of the distribution of values?

Quantifying variability

- There are several ways to illustrate the underlying distribution, to find how ”spread” out samples are.
  For example:
  - Histograms,
  - box plots,
  - sample variance,
  - standard deviation

Histograms

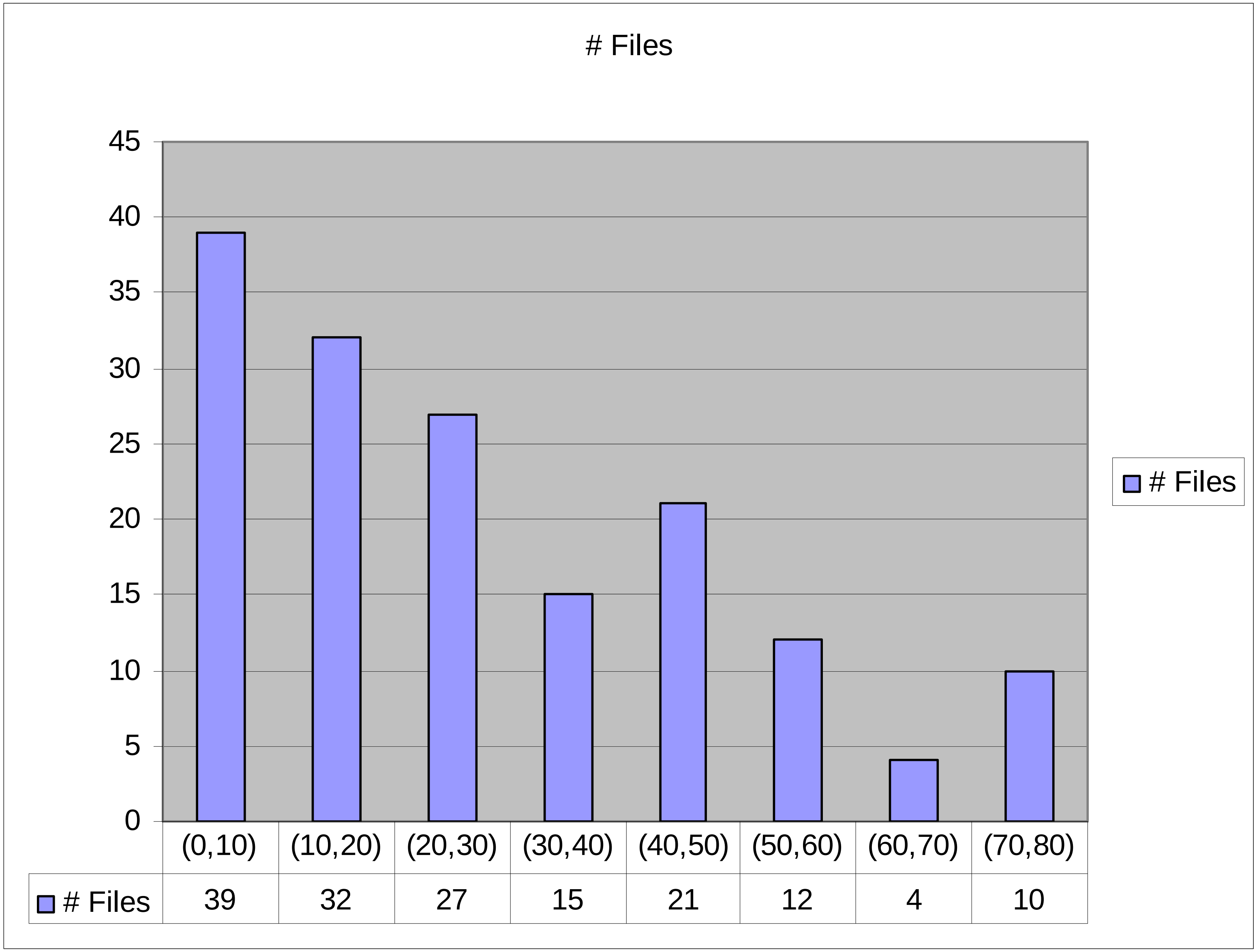
Let's say that we are interested in the size distribution of downloaded files from a webserver, over an hour (for example).

To construct a histogram that shows the filesize distribution you need to:

1. Record all downloaded filesizes
2. Make equally sized "buckets" of seen filesizes (at least 6-7 observations needed)
3. Put the number of observations seen in each bucket

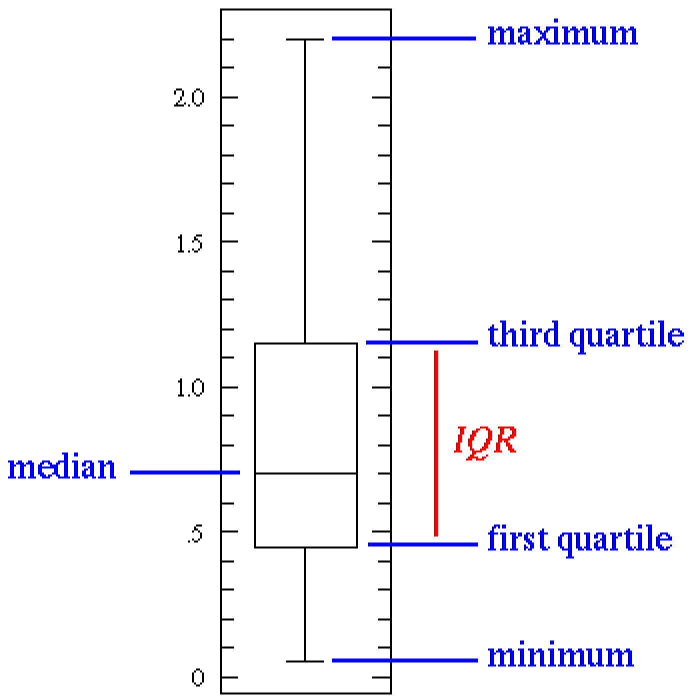
Filesize(KBytes)	# Files
$0 < x_i < 10$	39
$10 < x_i \leq 20$	32
$20 < x_i \leq 30$	27
$30 < x_i \leq 40$	15
$40 < x_i \leq 50$	21
$50 < x_i \leq 60$	12
$60 < x_i \leq 70$	4
$70 < x_i \leq 80$	10

Histograms



Files
45
40
35
30
25
20
15
10
5
0
Files
(0,10) (10,20) (20,30) (30,40) (40,50) (50,60) (60,70) (70,80)
Files 39 32 27 15 21 12 4 10

Box Plots


maximum
2.0
1.5
third quartile
1.0
IQR
median
.5
first quartile
minimum
0

Histograms & Box plots

- However, histograms and box plots are only visual aids, and can’t be used in calculations
- Can we represent the variability mathematically as well?
- Sure, ...

Variance

One way to represent how the samples relate to the mean is to calculate the sample variance:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

By calculating the squared difference between all samples and the mean we can get the ”average” variance (or spread) of our samples.

The reason why we don’t divide by *n*, but instead *n-1*, is that only *n-1* of the differences are independent. The last could be calculated using the others. We therefore say that the *degrees of freedom* in this equation is *n-1*

(If we know the mean and the first *n-1* samples, the n_{th} sample can be deduced)

Variance: Example

Let's say that we are investigating how long it takes to download a certain web page from a specific site.

We have downloaded the page 8 times and the time it took to download it was: 4, 5, 1, 3, 7, 9, 9, 15, (sec)

To calculate the variance, we first need the mean:

$$\bar{x}_A = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}\left(4+5+1+3+7+9+9+15\right) = 6.625$$

Then we simply insert the mean into the "variance formula":

$$s_A{}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{(4-6.625)^2 + (5-6.625)^2 + \ldots + (9-6.625)^2 + (15-6.625)^2}{8-1} = 19.410...$$

Standard Deviation

Instead of using variance, which doesn't have the same dimension as the mean value, we can use the standard deviation:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

If we apply this formula on our example, we find that the standard deviation when downloading the webpage is:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{19.410...} = 4.405...$$

Half-way Summary

- Mean values
  - A single value that tries to capture the average ”behavior” of a set of samples
  - Dangerous, as it hides the variability
- Variability
  - The underlying sample distribution
  - Quantify variability
    - Histograms, Box plots, Variance, Standard Deviation

Errors in Experimental Measurements

- In real life, nothing is perfect!
- So measurements on something ”real” are, by definition, NOT perfect!
  - Why do we have variability?
- What to do?
  - Avoid errors that can be avoided, and
  - quantify the other errors mathematically!

Experimental errors

- Errors → *noise* in measured values
- *Systematic* errors
  - Result of an experimental "mistake"
  - Typically produce constant or slowly varying bias
- Controlled through skill of experimenter
- Examples
  - Temperature change causes clock drift
  - Forget to clear cache before timing run

Experimental errors

- *Random* errors
  - Unpredictable, non-deterministic
  - Unbiased → equal probability of increasing or decreasing measured value
- Result of
  - Limitations of measuring tool
  - Observer reading output of tool
  - Random processes within system
- Typically cannot be controlled
  - Use statistical tools to characterize and quantify

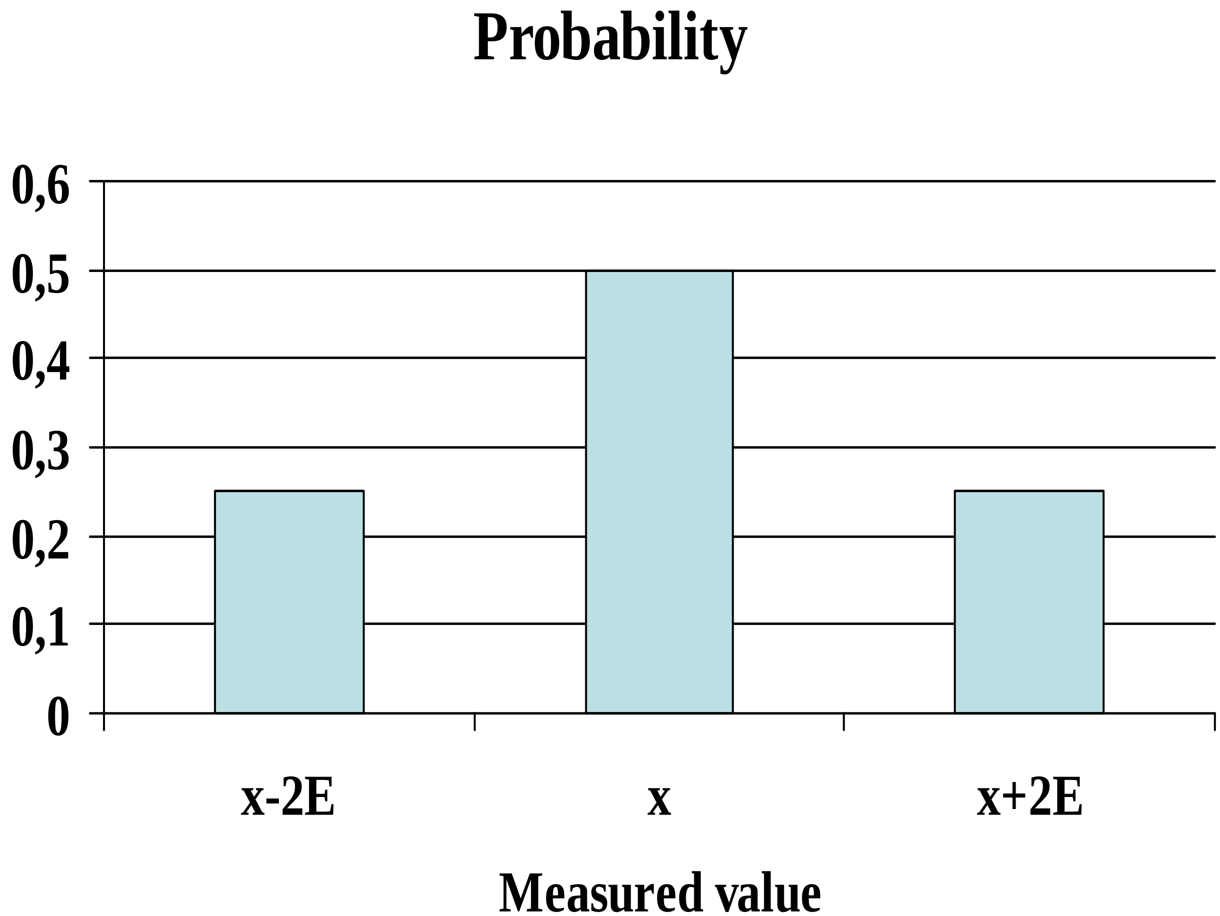
A Model of Errors

Error	*Measured value*	*Probability*
-E	x – E	½
+E	x + E	½

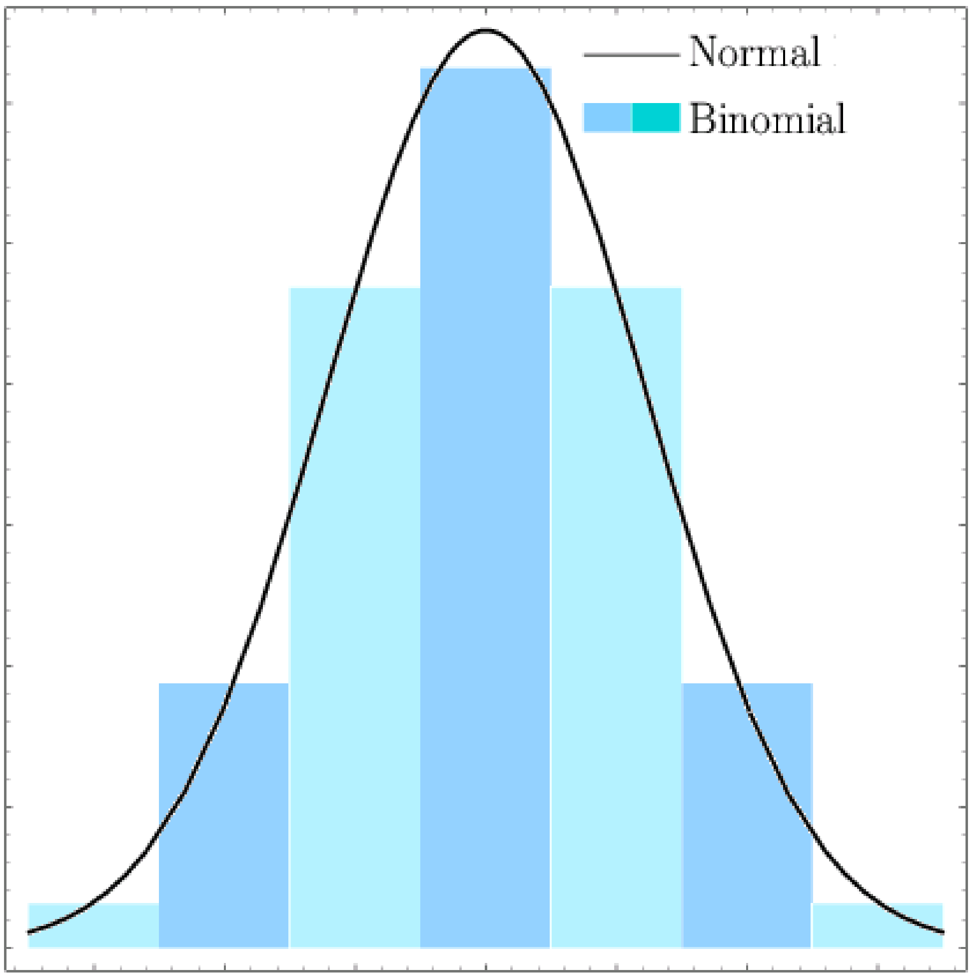
A Model of Errors

Error 1	*Error 2*	*Measured value*	*Probability*
-E	-E	x – 2E	¼
-E	+E	x	¼
+E	-E	x	¼
+E	+E	x + 2E	¼

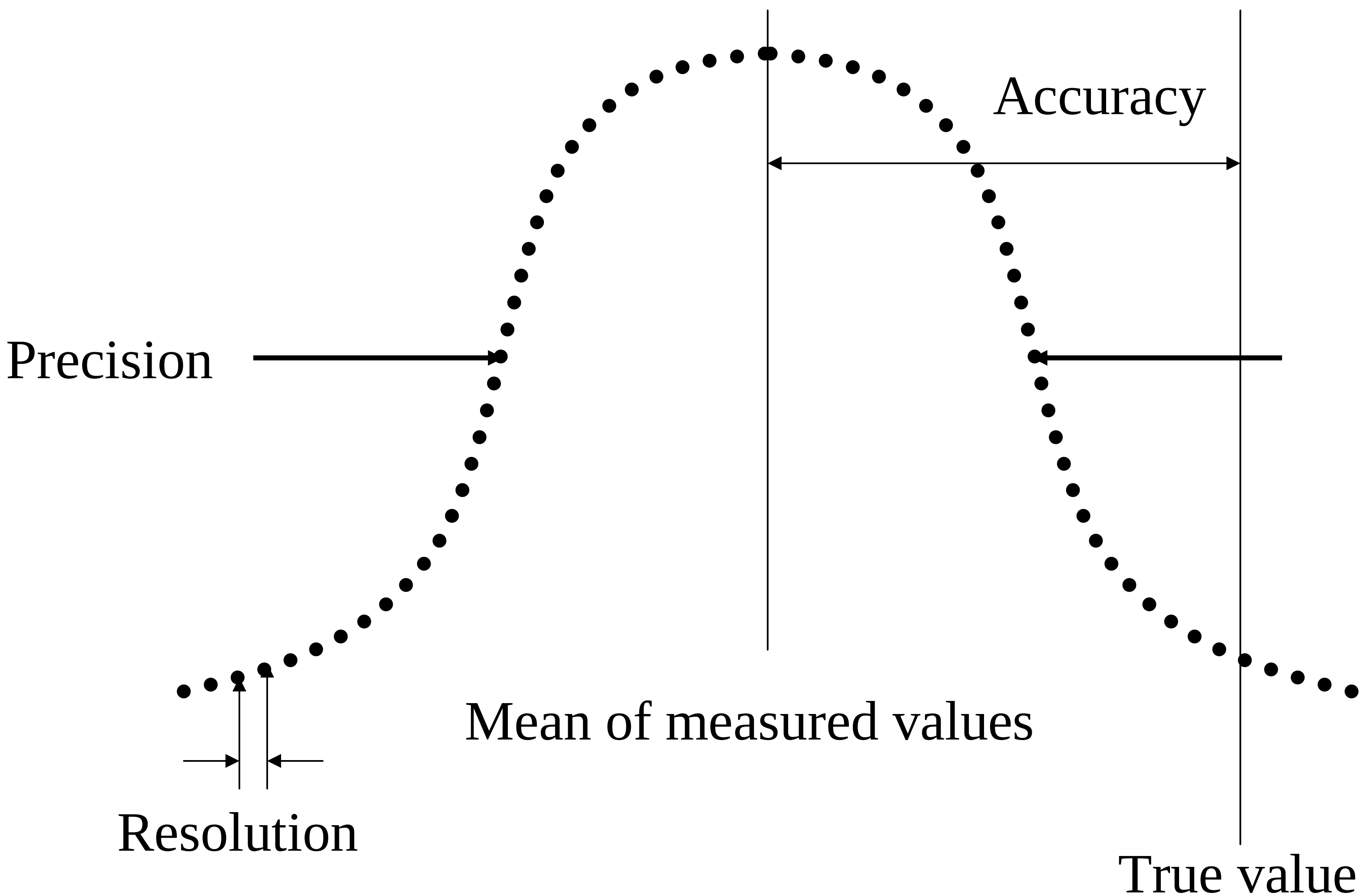
A Model of Errors



A Model of Errors



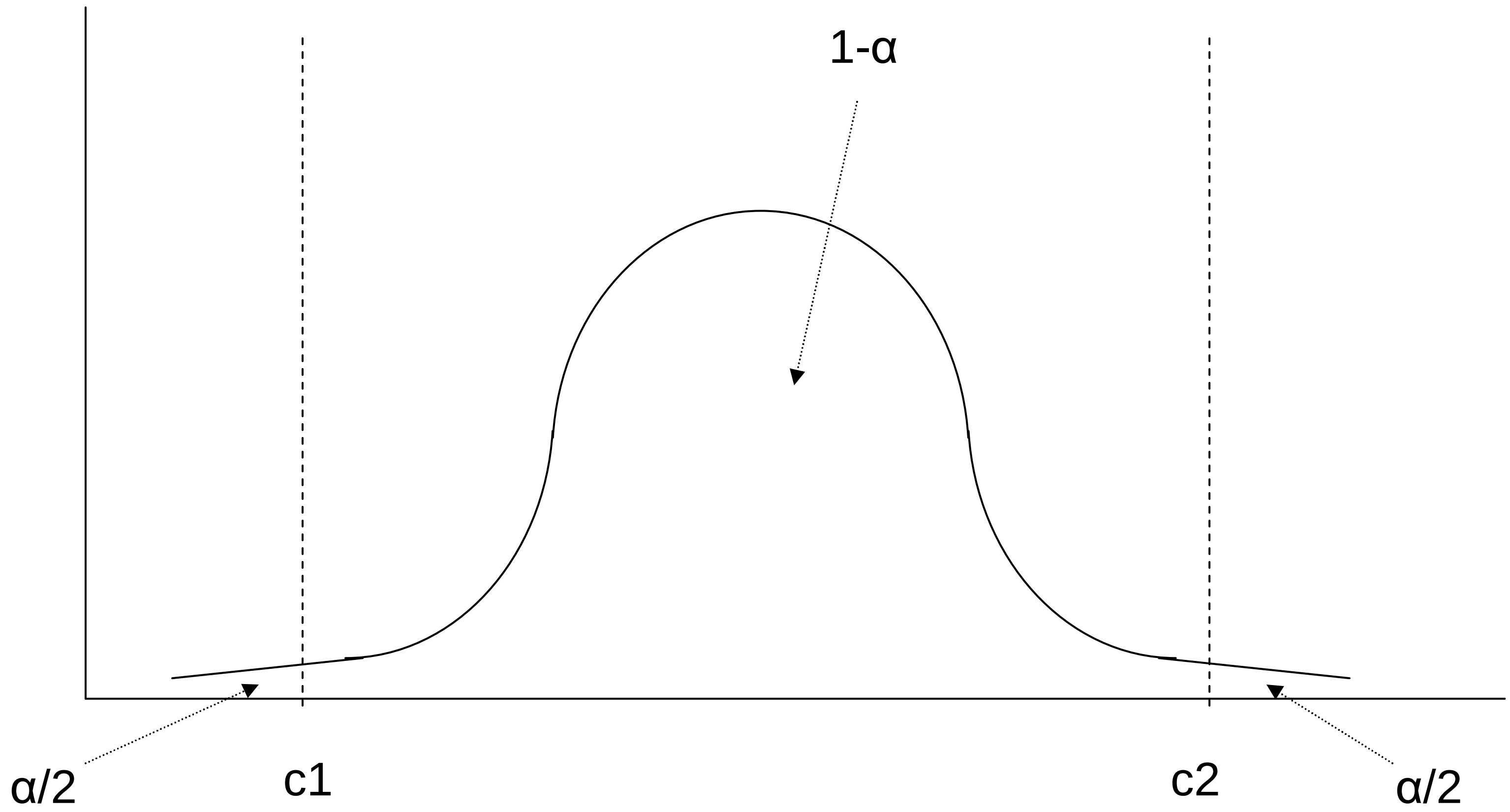
A Model of Errors



Errors

- Systematic Errors
  - For example, Accuracy
  - Should be removed, controlled, or at least have an understandable bias on the results
- Random Errors
  - For example, precision.
  - We can quantify these through statistical methods (e.g. confidence intervals for mean values)

Confidence Intervals for the Mean



Normalize *x*

$$z = \frac{\bar{x} - x}{s / \sqrt{n}}$$

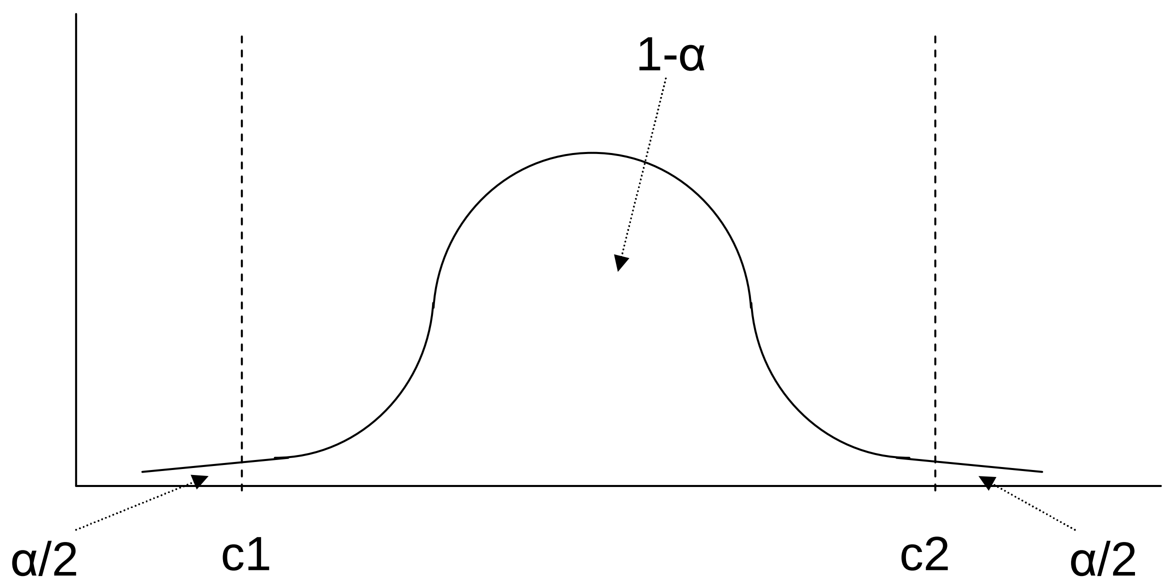
$$n = \text{number of measurements}$$

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

$$s = \text{standard deviation} = \frac{\sum_{\text{i}=1}^{\text{n}} (x_i - \bar{x})^2}{n-1}$$

Confidence Interval for the Mean

- Normalized *z* follows a Student's *t* distribution
  - (n-1) degrees of freedom, if $n \geq 30$ we approximate with an infinite degree of freedom (the normal distribution)
  - Tabulated values for *t* (check handout)



Confidence Interval for the Mean

$$c_1 = \bar{x} - t_{1-\alpha/2;n-1} \frac{s}{\sqrt{n}}$$

$$c_2 = \bar{x} + t_{1-\alpha/2;n-1} \frac{s}{\sqrt{n}}$$

Then,

$$\Pr(c_1 \le x \le c_2) = 1 - \alpha$$

Example

- Handouts of two examples has been given in class
- They can also be found at the course website

How many measurements?

- We want to control the width of the interval!
  - By specifying how large the error (*e*) is allowed to be
- Thus, the interval endpoints (c_1, c_2), should only allow an error of max ± *e*

$$(c_1, c_2) = \left[(1-e)\bar{x}, (1+e)\bar{x}\right]$$

How many measurements?

$$c_2 = \begin{cases} \bar{x} + z_{1-\alpha/2} \dfrac{s}{\sqrt{n}} \\ (1+e)\bar{x} \end{cases}$$

$$\Rightarrow$$

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

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

$$n = \left(\frac{z_{1-\alpha/2} s}{\bar{x} e} \right)^2$$

How many measurements?

- But n depends on knowing mean and standard deviation!
- Estimate s with small number of measurements first
- Use this s to find n needed for desired interval width

How many measurements?

- We first make some measurements
- Mean = 7.94 s
- Standard deviation = 2.14 s
- Want 90% confidence mean is within 7% of actual mean.

How many measurements?

- We first make some measurements
- Mean = 7.94 s
- Standard deviation = 2.14 s
- Want 90% confidence mean is within 7% of actual mean.
- *$1-\alpha = 0.90$*
- *$(1-\alpha/2) = 0.95$*
- *Error = ± 3.5%*
- *$e = 0.035$*

How many measurements?

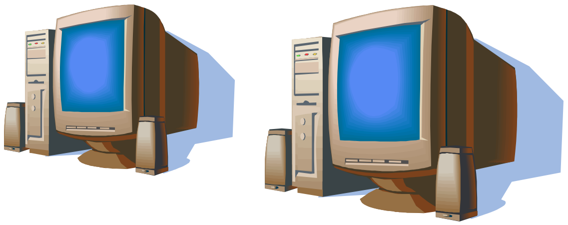
$$n = \left(\frac{z_{1-\alpha/2} s}{\bar{x} e} \right)^2 = \left(\frac{1.895(2.14)}{(7.94)0.035} \right)^2 = 212.9$$

- 213 measurements

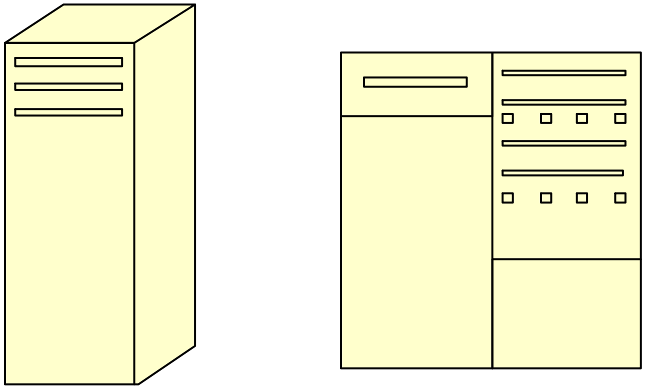
→ 90% chance true mean is within ± 3.5% interval

Comparing Two Alternatives

1. *Before-and-after*
   Did a change to the system have a statistically significant impact on performance?



2. *Non-corresponding measurements*
   Is there a statistically significant difference between two different systems?



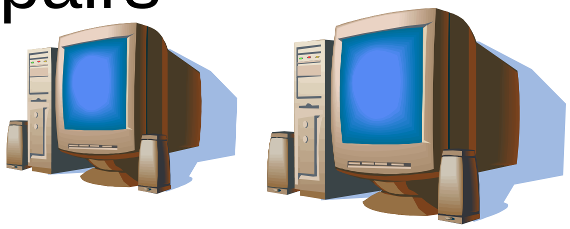
Before-and-After Comparison

- Assumptions
  - Before-and-after measurements are not independent
  - Variances in two sets of measurements may not be equal

→ Measurements are related

  - Form obvious corresponding pairs

- Find *mean of differences*



Before-and-After Comparison

$$b_i = \text{before measurement}$$

$$a_i = \text{after measurement}$$

$$d_i = a_i - b_i$$

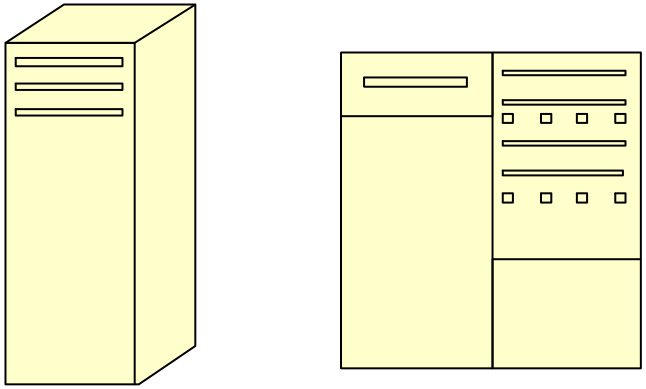
$$\bar{d} = \text{mean value of } d_i$$

$$s_d = \text{standard deviation of } d_i$$

$$(c_1, c_2) = \bar{d} \mp t_{1-\alpha/2;n-1} \frac{s_d}{\sqrt{n}}$$

Noncorresponding Measurements

- No direct correspondence between pairs of measurements
- *Unpaired* observations
- n_1 measurements of system 1
- n_2 measurements of system 2



Confidence Interval for Difference of Means

1. Compute means
2. Compute difference of means
3. Compute standard deviation of difference of means
4. Find confidence interval for this difference
5. No statistically significant difference between systems if interval includes 0

Confidence Interval for Difference of Means

$$\text{Difference of means:}$$

$$\bar{x} = \bar{x}_1 - \bar{x}_2$$

$$\text{Combined standard deviation:}$$

$$s_x = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Number of Degrees of Freedom

$$\text{Not simply } n_{df} = n_1 + n_2 - 2$$

$$n_{df} = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(s_1^2 / n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2 / n_2\right)^2}{n_2 - 1}}$$

Summary

- Use confidence intervals *to quantify precision*
- Confidence intervals for
  - Mean of *n* samples
- Confidence level
  - Pr(actual mean within computed interval)
- Compute number of measurements needed for desired interval width
- It's possible to use confidence intervals for comparison between measurements
- This lecture has partially covered Chapters 4-5.1.3