Physics 307L

Spring 2021

Prof. Darcy Barron

Lecture 4: Error Analysis

Course webpage

Wiki is now ~ complete including remote lab procedures: https://ghz.unm.edu/juniorlab/

We will continue to use Schedule page to schedule experiments, post lecture info and assignments, and schedule talks.

https://ghz.unm.edu/juniorlab/index.php?title=Schedule_Spring_2021

Upcoming assignments

- Reminder: you will complete 6 experiments this semester, give 3 presentations, and write 3 full lab reports in the style of a scientific paper
- We will discuss scientific presentations and writing scientific papers next week (Feb 22)
- First student presentations will be during lecture on March 1 and March 8
    - 4 talk slots per class period, contact instructors with preferred date and which lab you will present
    - Ok to present on 'Lab 0' (Oscilloscope OR Chua Circuit)
- First written lab report is due Wed. March 24
    - Lab report cannot be on Lab 0
    - Will allow ONE experiment to overlap with presentation and written lab report (otherwise must pick one)

Lab notebooks

- Lab notebook for week 1 has been returned
- Lab notebook for week 2 will be graded and returned by Thursday
- Lab notebook for week 3 is due Wednesday at noon
- Lab notebook rubric was used to determine score, and comments are embedded in pdf.
- Lab notebooks for week 1 and week 2 only will be graded on a curve.
- Future lab notebooks will need to meet guidelines in rubric for full credit

Last lecture

- Review of basic error analysis
  - Lab manual from intro labs: https://ghz.unm.edu/education/juniorlab_pdfs/uncertainty.pdf
  - Chapters 1-3 of Taylor
- Intro to random vs systematic errors (Taylor Chapter 4)

Error analysis in labs

- The three labs you complete before spring break are exercises in error analysis and comparing data with a model
  - Balmer Series, Speed of Light, Poisson Statistics
  - Some relevant chapters in Taylor:
    - Chapter 6 – Rejection of Data
    - Chapter 7 – Weighted Averages
    - Chapter 8 – Least-Squares Fitting
    - Chapter 9 – Covariance and Correlation
    - Chapter 11 – The Poisson Distribution
- We can revisit some of these concepts in lecture after everyone has rotated through these labs

Today’s lecture

- Correlating data
- More on fitting data to a model
- Random errors and confidence interval
- Example: period of a pendulum

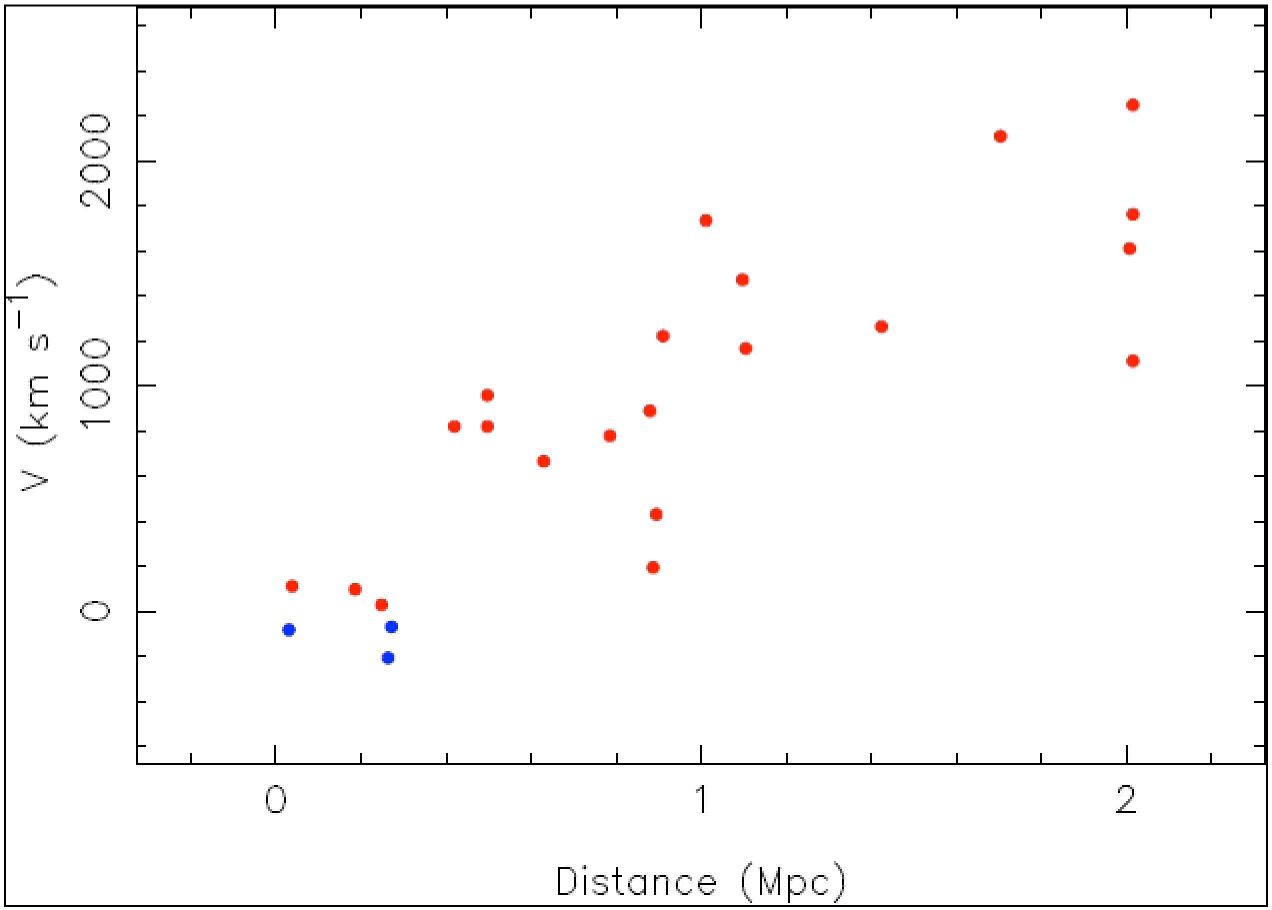
Correlating data

- When looking at new measurements, it is instinct to try to correlate it with other results
  - Checking if our measurements are reasonable
  - Checking if other results are reasonable
  - To test a hypothesis
  - Shot in the dark
- There are a few common traps to fall into when attempting to find correlations

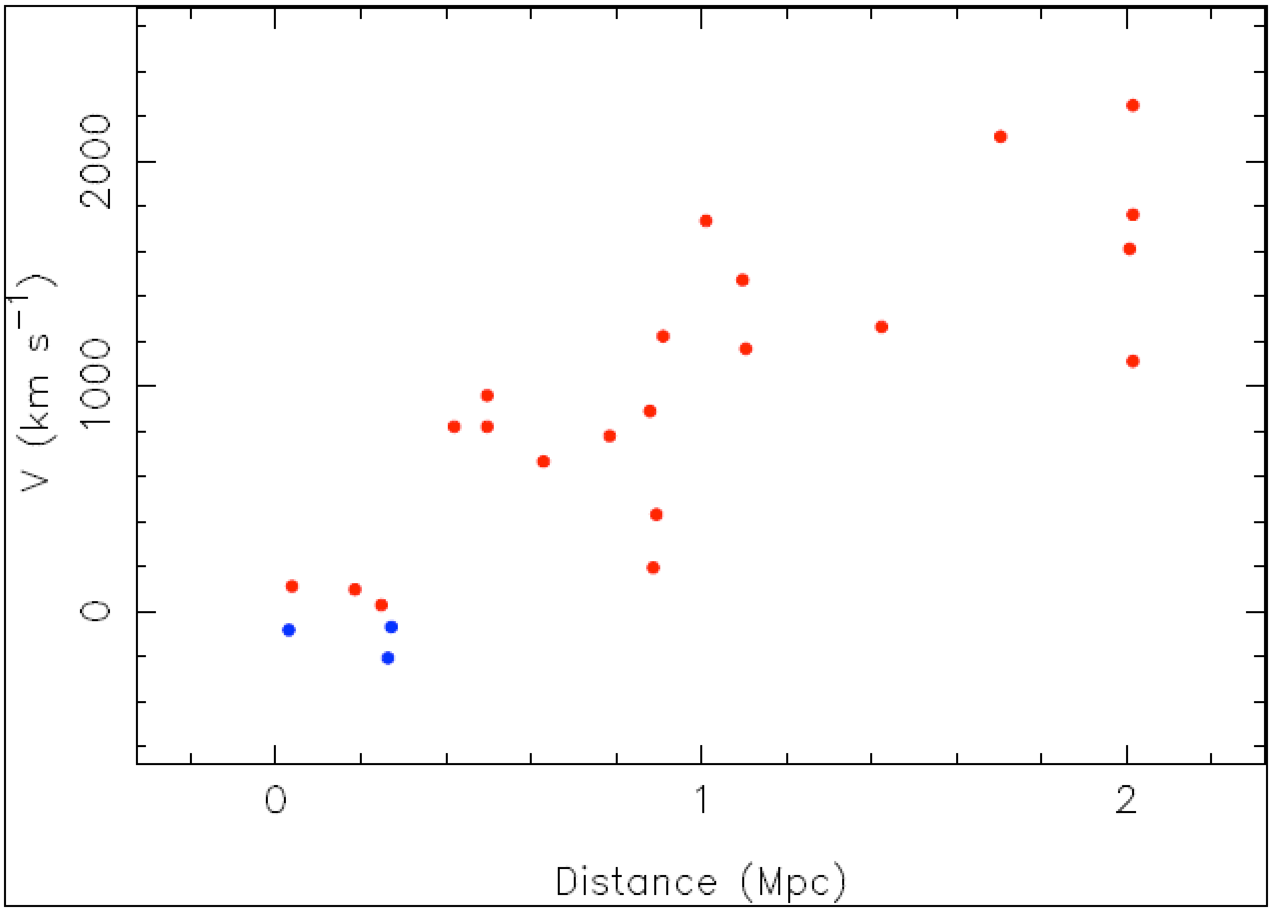
Fishing trips

- Correlation does not prove a causal connection!
  - Examples of correlations
    - Number of violent crimes in cities versus number of churches
    - The quality of student handwriting versus their height
    - Stock market prices and the sunspot cycle
    - Cigarette smoking vs lung cancer
    - Health vs alcohol intake
- Potential reasons
  - **Lurking third variables**
  - Similar time scales
  - Causal connection

Correlation?



Correlation?

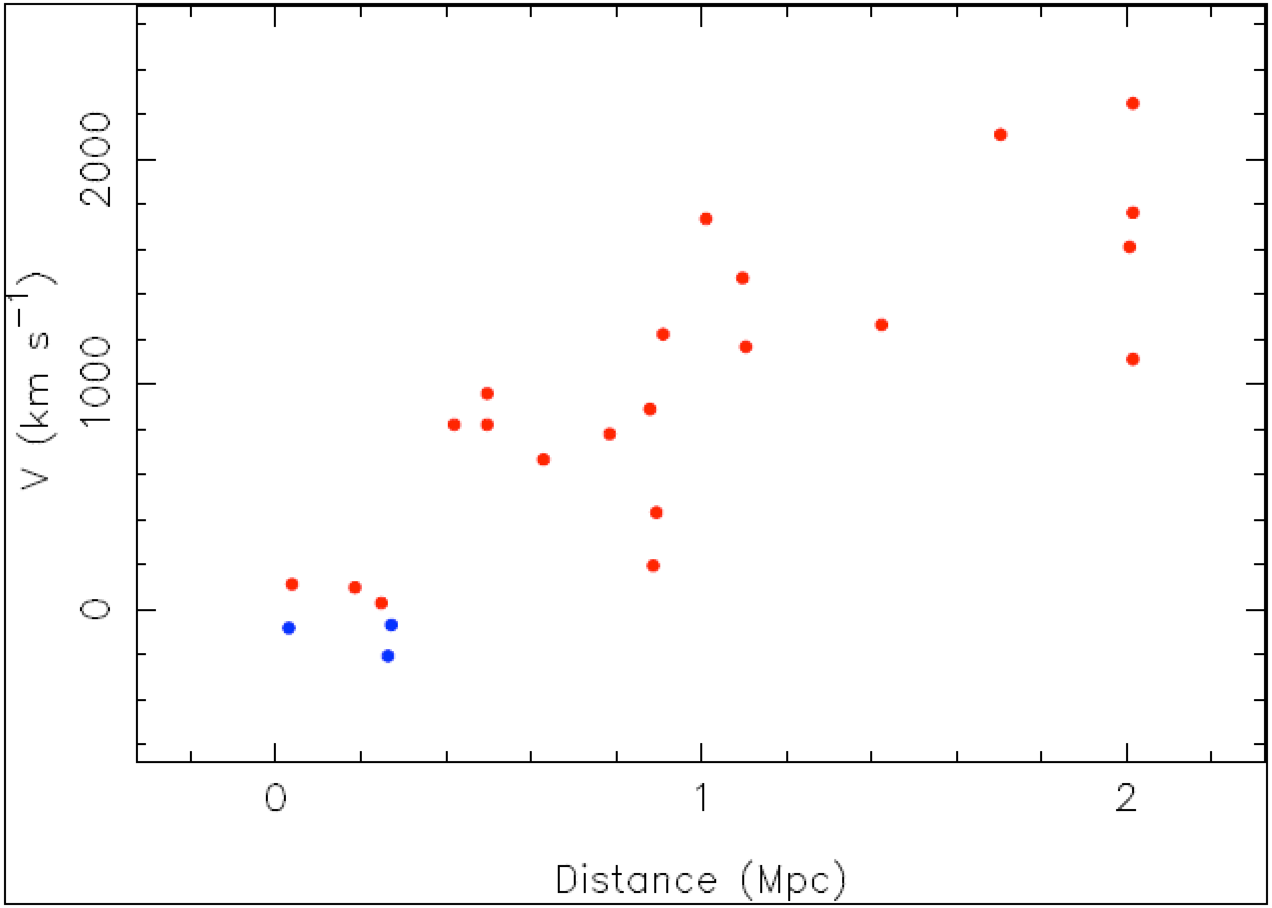


An early Hubble diagram, N=24 galaxies (1936)

So you think your data is correlated

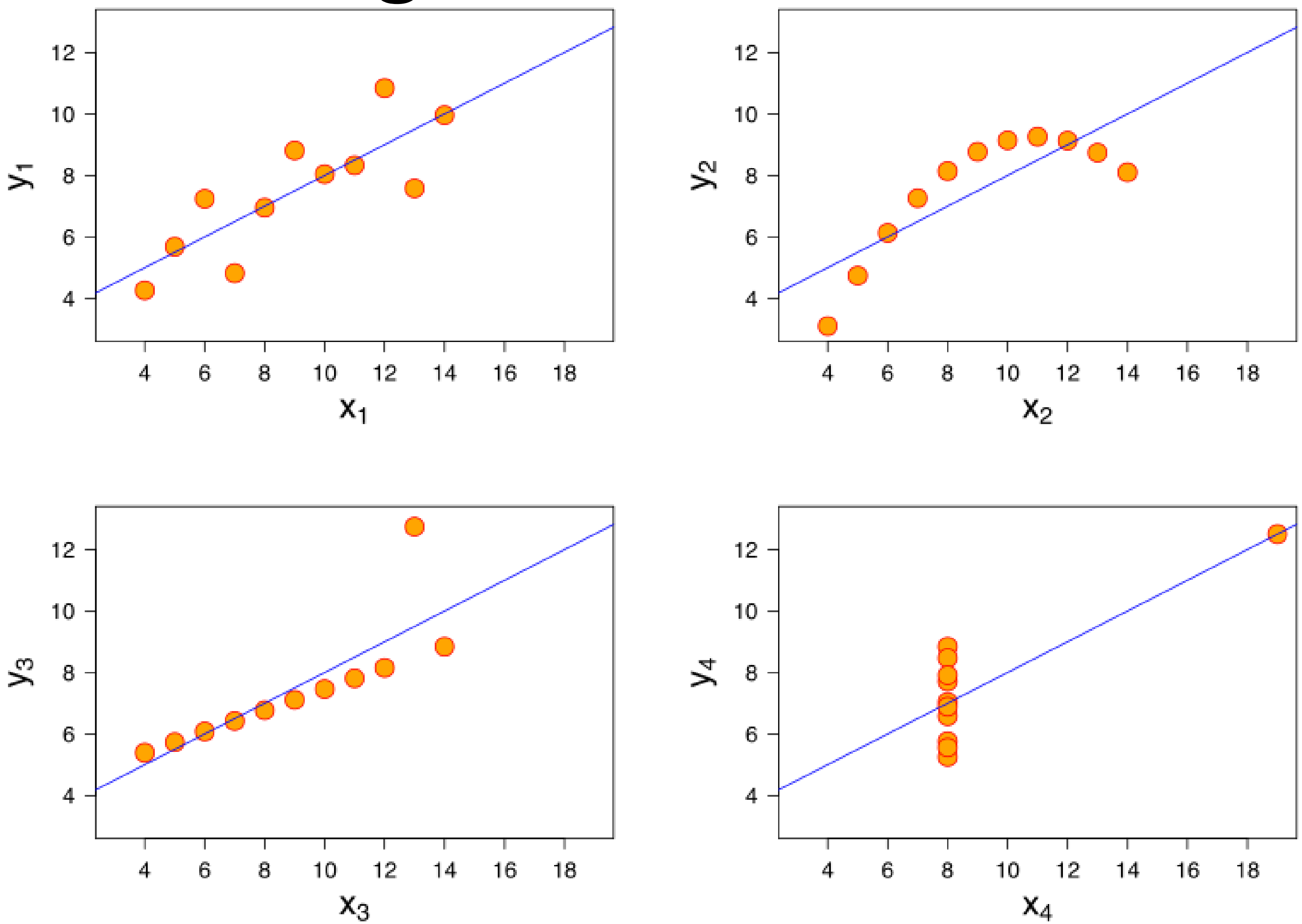
- Time to fit the data to a model!
- “All models are wrong, but some models are useful.”
    - George E. P. Box

Fitting data to a model



An early Hubble diagram, N=24 galaxies (1936)

Fitting data to a model



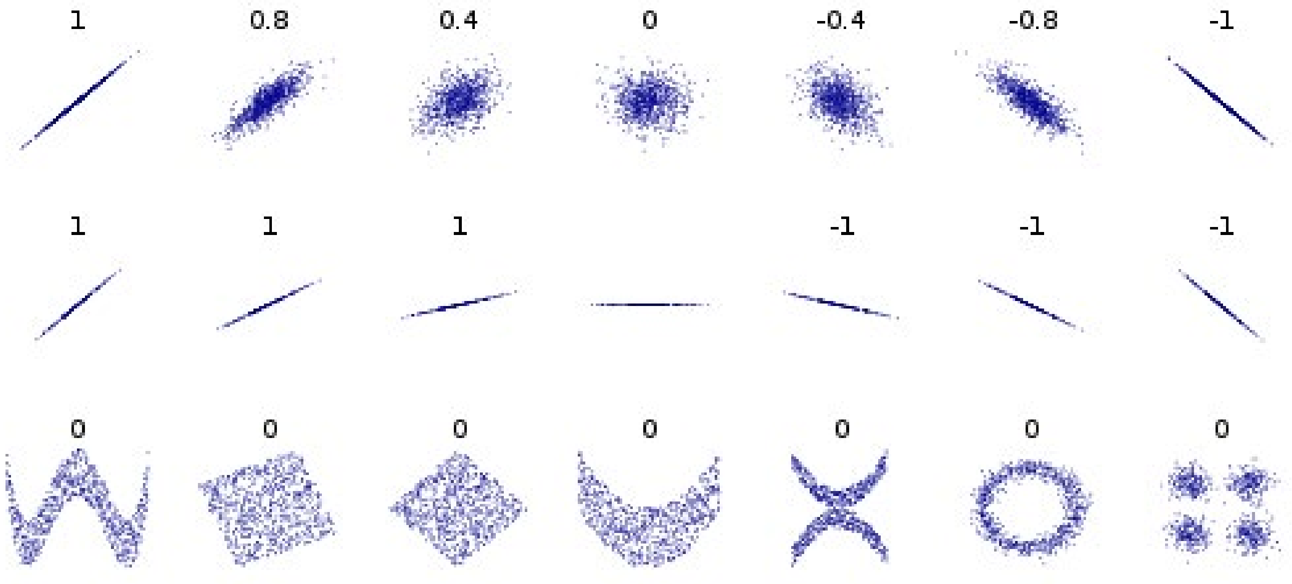
Error bars?

Simplifying correlations

- Linear correlation
  - $y=mx+b$
  - Multidimensional: $z = mx + ny + b$
- **Linear correlations are easy to plot and examine**
- Can linearize your data to make it a linear correlation
  - Example: Surface brightness of a disk
    - $I(r) = I_0 e^{-r/h}$
    - Linearized form: $\ln(I) = \ln(I_0) - r/h$
  - Also straightforward to do for power laws

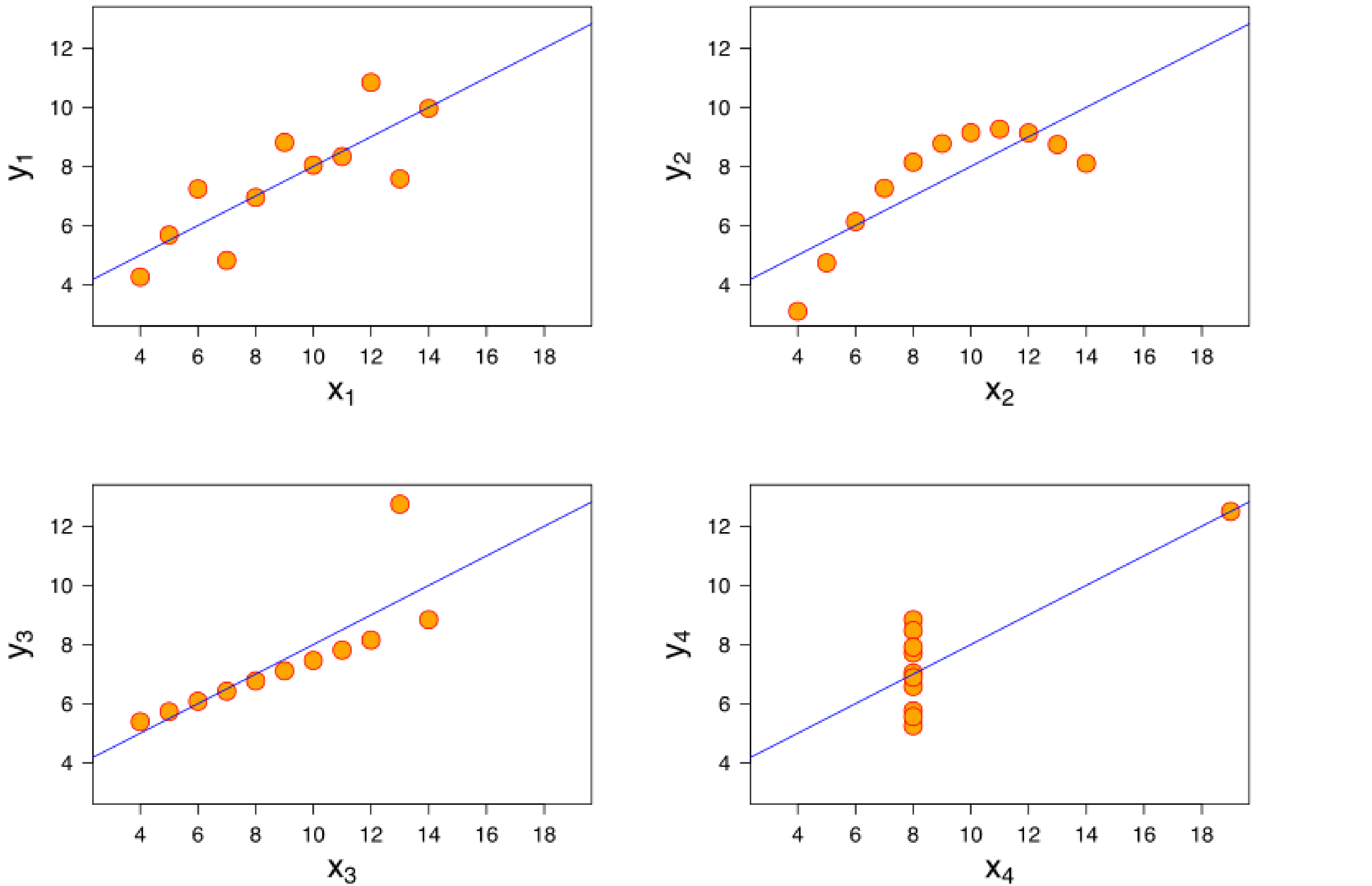
Pearson's correlation coefficient

https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

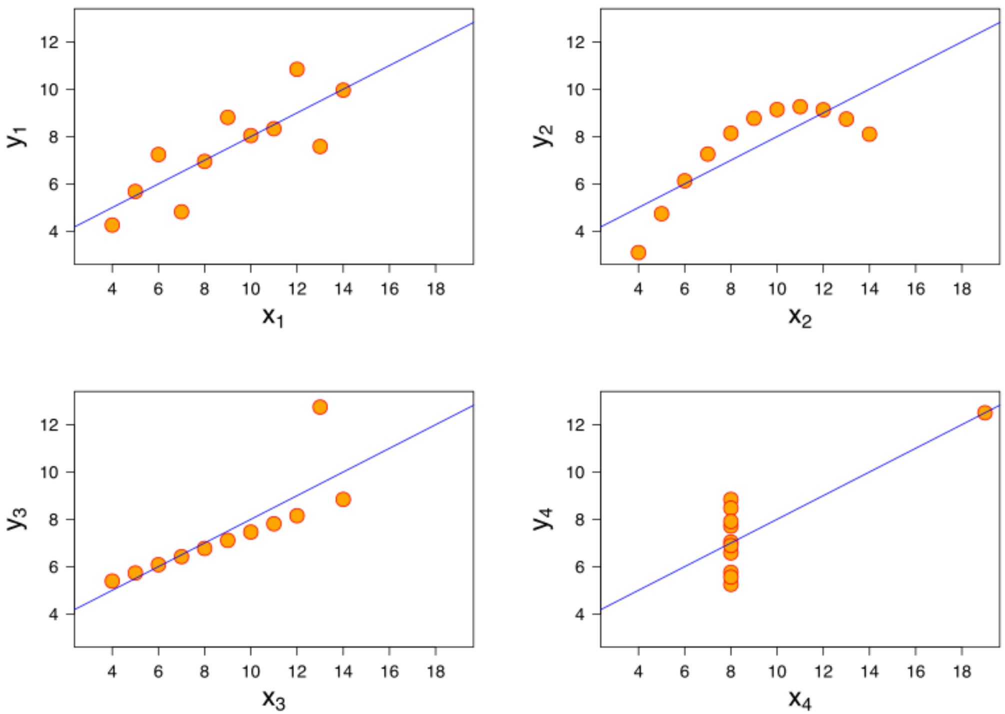


- Pearson's correlation coefficient, r, measures the linear correlation between two variables

Anscombe's quartet



Anscombe’s quartet



Property	Value	Accuracy
Mean of *x*	9	exact
Sample variance of *x*	11	exact
Mean of *y*	7.50	to 2 decimal places
Sample variance of *y*	4.125	±0.003
Correlation between *x* and *y*	0.816	to 3 decimal places
Linear regression line	$y = 3.00 + 0.500x$	to 2 and 3 decimal places, respectively
Coefficient of determination of the linear regression	0.67	to 2 decimal places

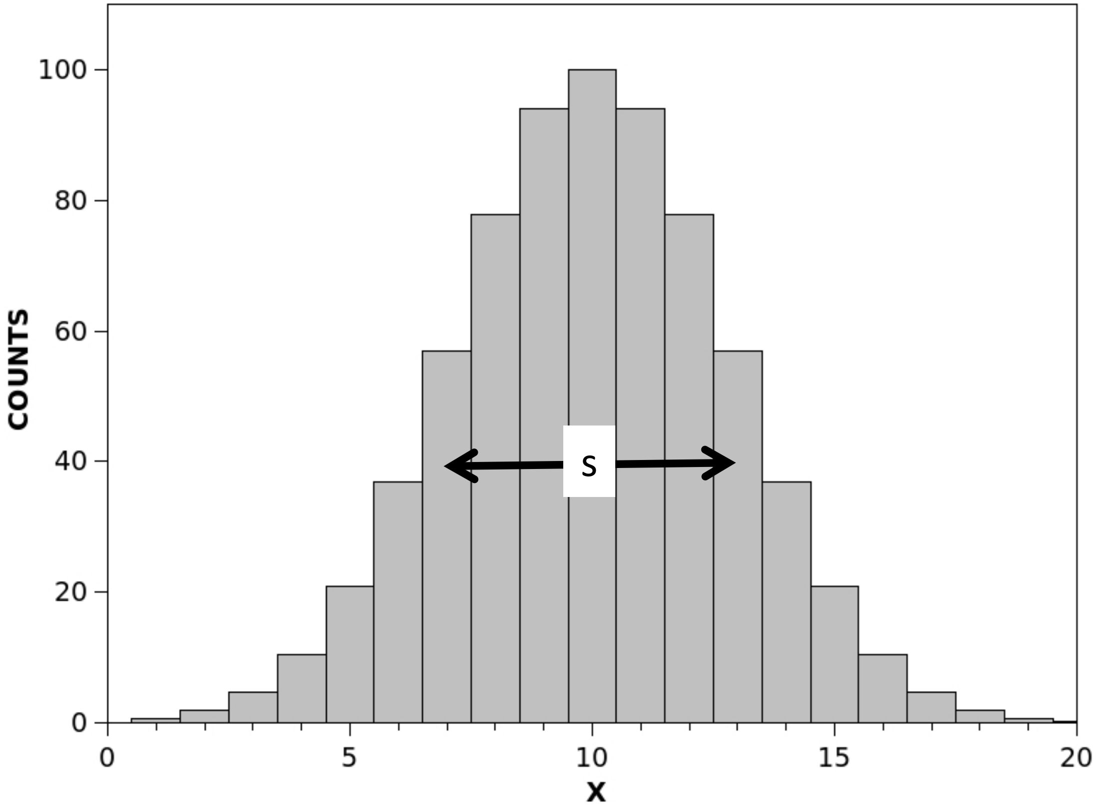
https://en.wikipedia.org/wiki/Anscombe%27s_quartet

Modeling uncertainty

- Least squares fitting
  - Chapter 8 of Taylor
  - A straightforward output of Python/Matlab/Excel/etc
  - **Assumes uncorrelated Gaussian statistics**
  - Can get different results depending on the exact algorithm, especially for data with a small number of samples, or data with outliers
- Weighted least squares fitting – uses uncertanties
- Other ways to check uncertainty
  - Jack-knife
    - Go through data and toss out data points, and recalculate
    - Common to split data in half (e.g. first-half vs second-half), or split by (supposedly identical) data runs
  - Bootstrap
    - Go through dataset picking N points at random, recalculate and look at variation

PROBABILITY DISTRIBUTION

If the value of x is random: **GAUSSIAN** distribution



EXAMPLE

Most probable value: x = 10 (Mean)

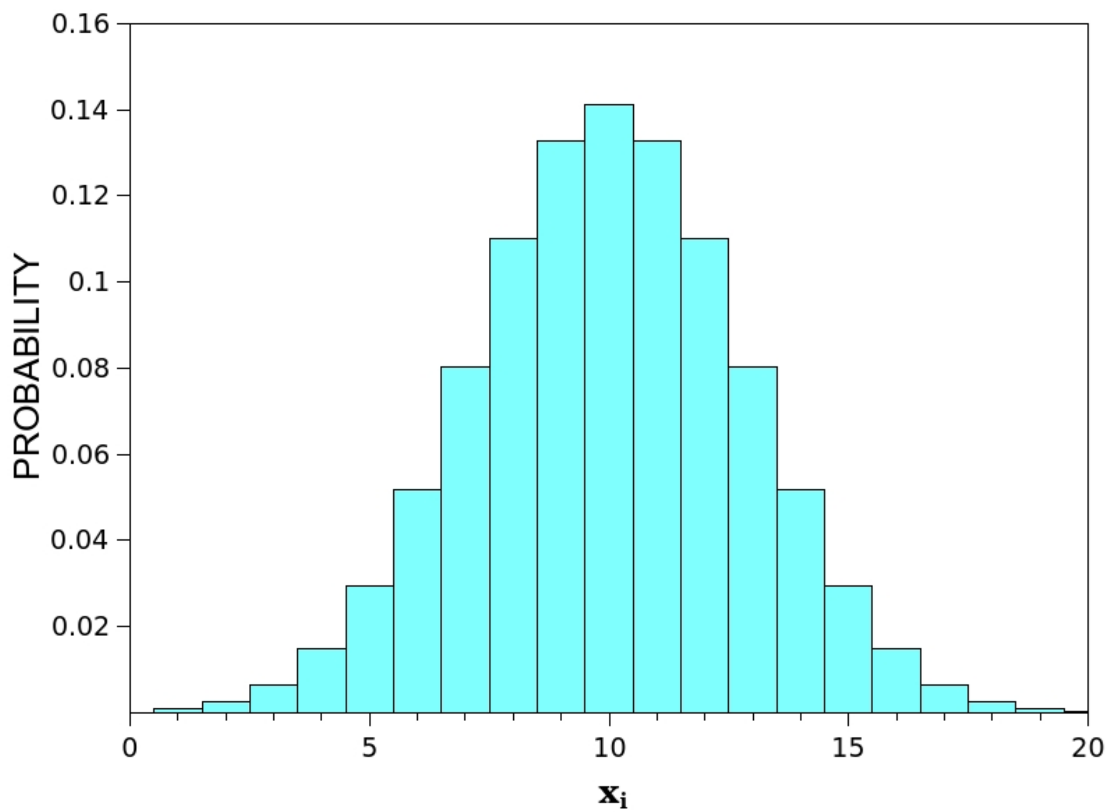
Variance: $s^2 = 8$

Standard deviation: $s = 2.82$

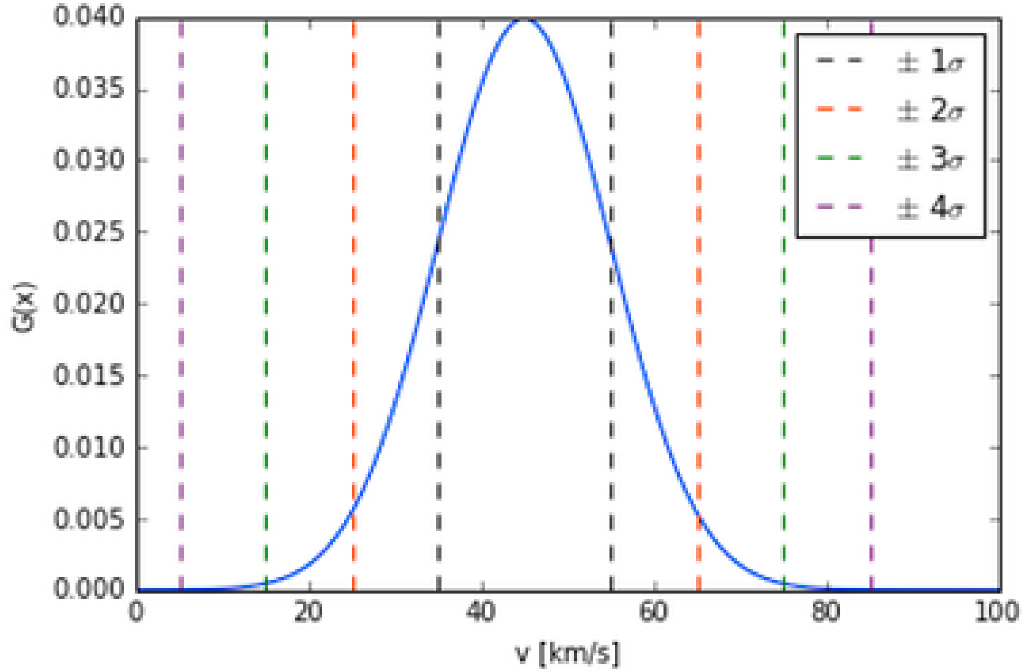
Probability p_i that x will have a specific value x_i

Probabilities must sum to 1: $\sum_i^N p_i = 1$

Expectation value: $\langle x \rangle = \sum_i^N x_i p_i = \overline{x}$



Uncertainties and confidence intervals

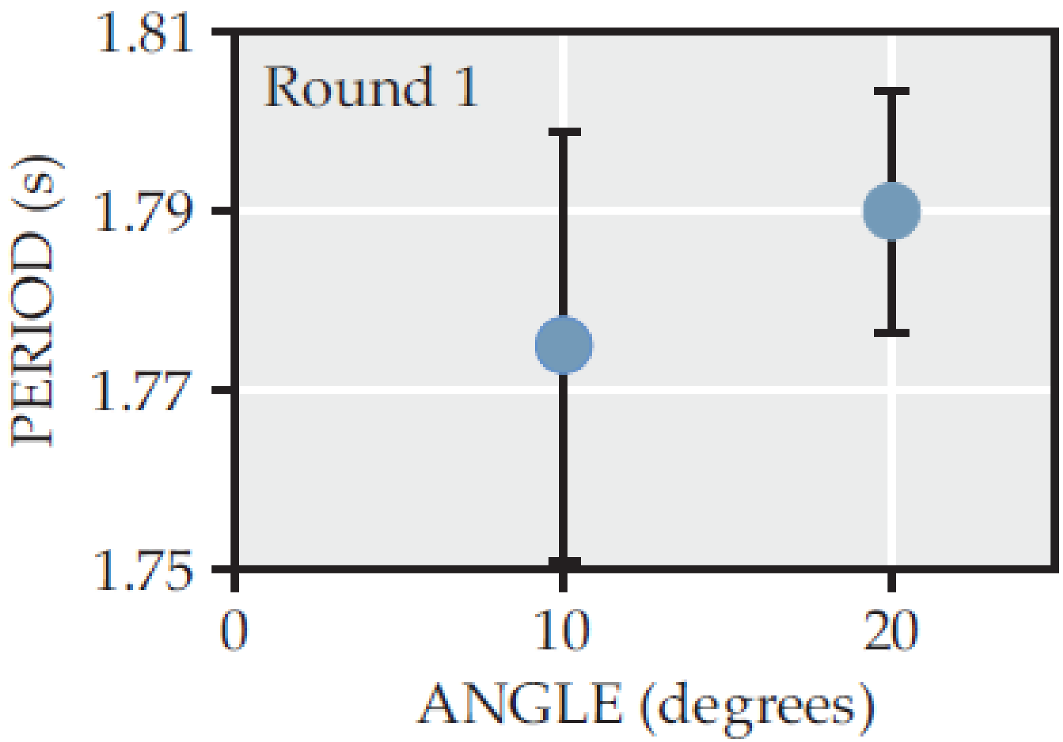


- Chapter 5 of Taylor – The Normal Distribution
- http://astro.physics.uiowa.edu/ITU/courses/general-astronomy/error-analysis/reporting-uncertainties.html

Example: period of a pendulum

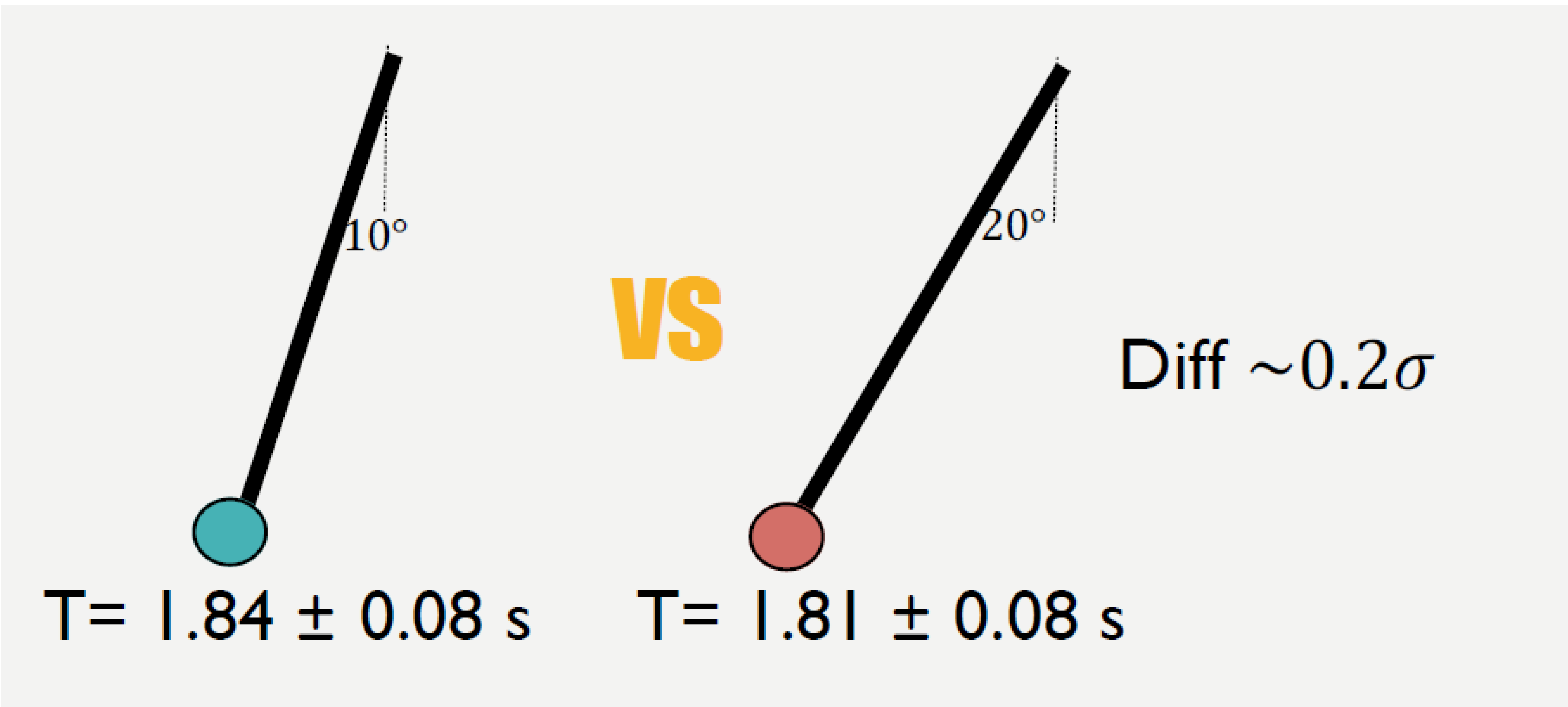
- Does the period of a pendulum differ when released from different amplitudes (10 degrees vs 20 degrees)?

Example: period of a pendulum



What are some sources of random error?
What are some sources of systematic error?

Example: period of a pendulum

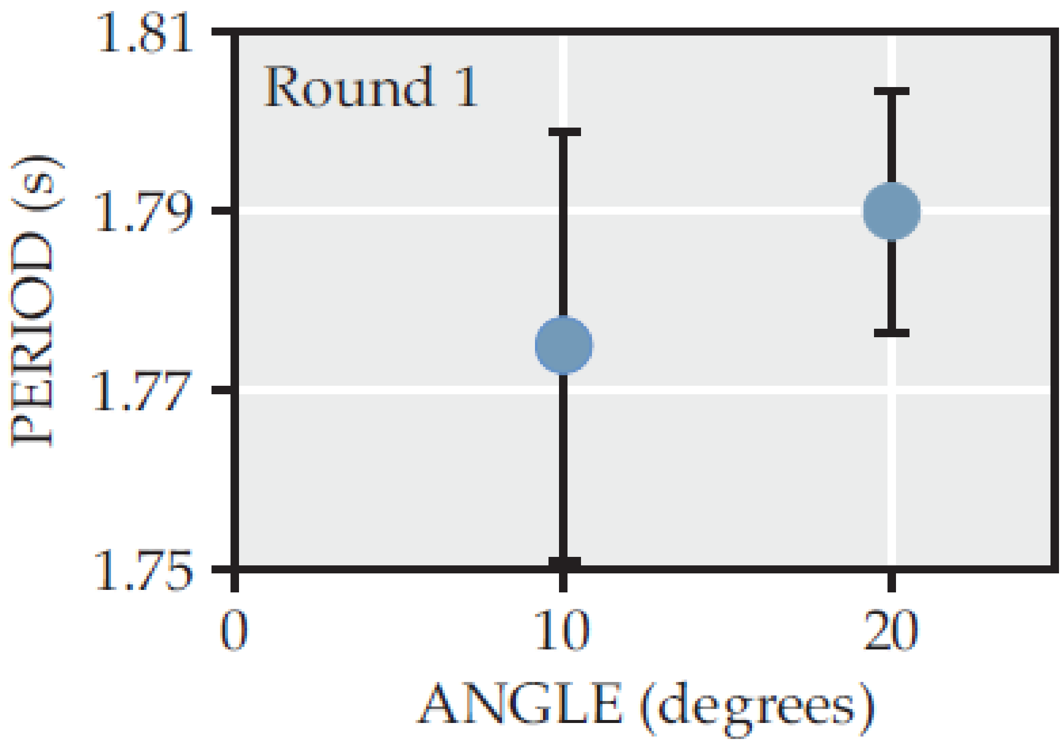


Example: period of a pendulum

$$t' = \frac{T_{10^\circ} - T_{20^\circ}}{Uncertainty}$$

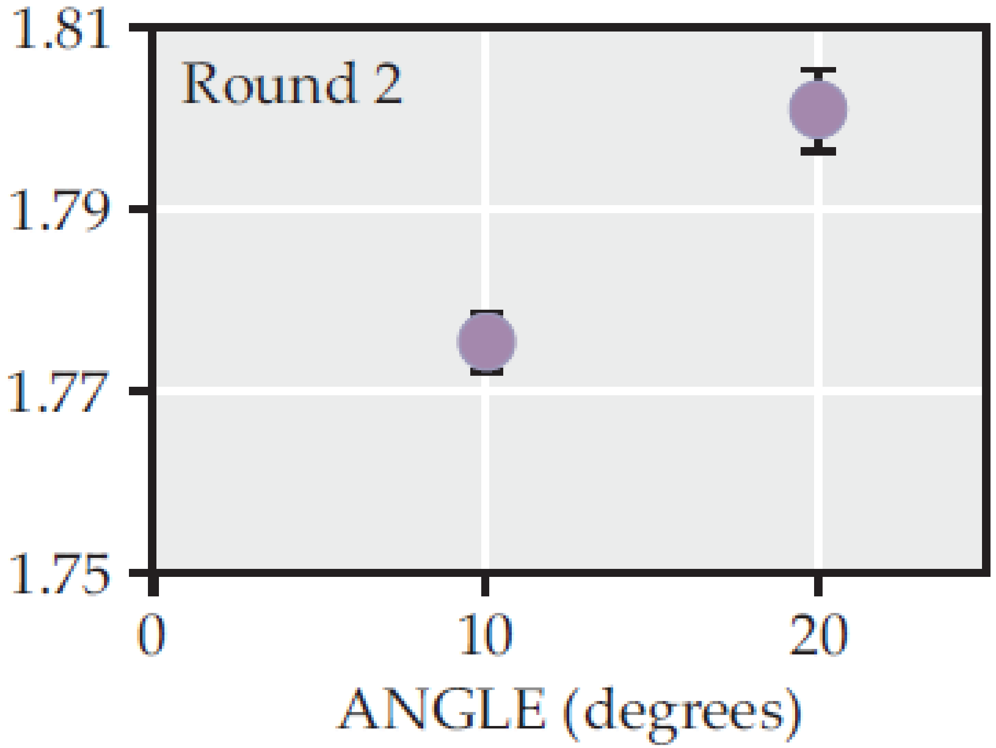
- T-score: a simple way to quantify agreement between measurements
- What does a low t-score mean?
- What does a high t-score mean?

Example: period of a pendulum



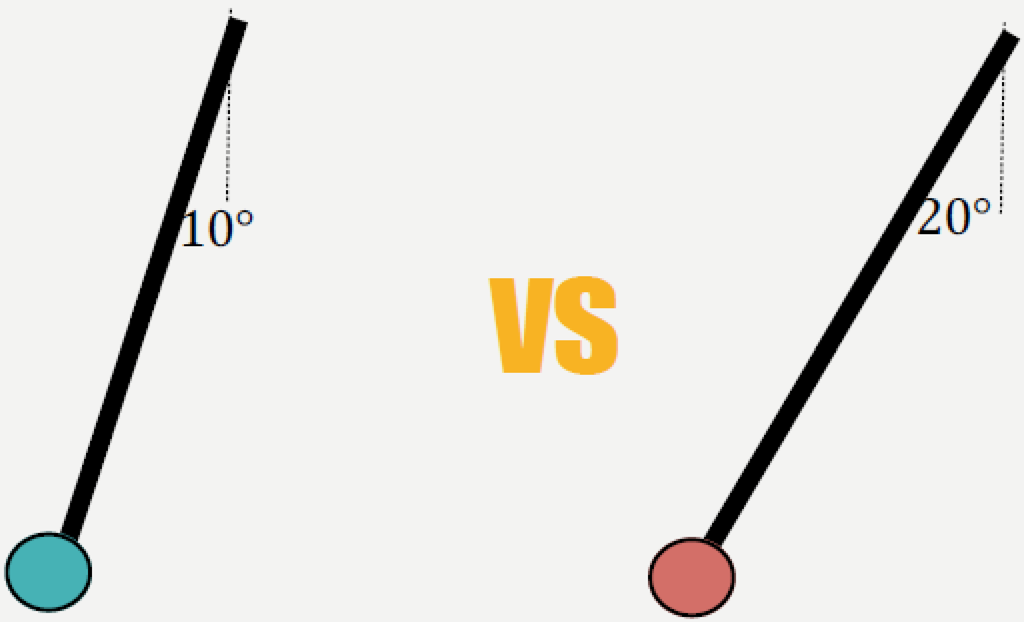
What are some sources of random error?
What are some sources of systematic error?
How can we improve uncertainties?

Example: period of a pendulum



Reduce statistical uncertainty by measuring more periods of the pendulum (more trials)

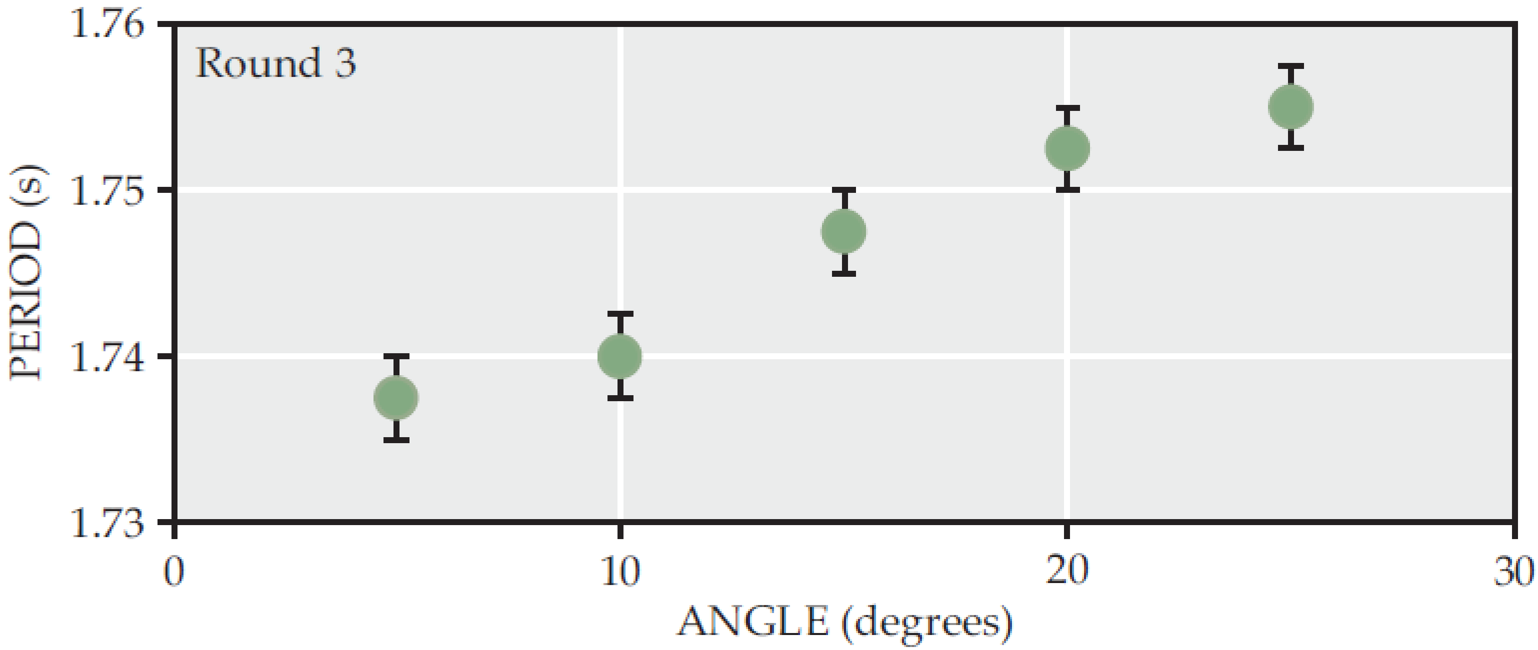
Example: period of a pendulum



$t' \sim 3.7\sigma$

T= 1.830 ± 0.004 s T= 1.851 ± 0.004 s

Example: period of a pendulum



References

- Physics Today **71**, 1, 38 (2018)
  - https://doi.org/10.1063/PT.3.3816
- http://astro.physics.uiowa.edu/ITU/courses/general-astronomy/error-analysis/reporting-uncertainties.html