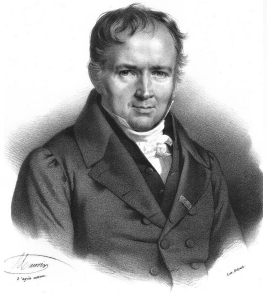
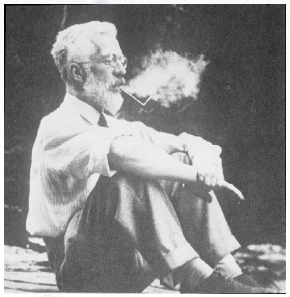
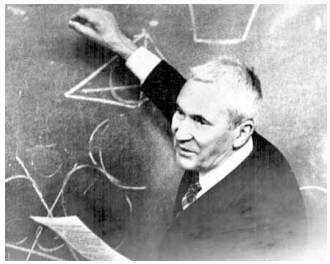
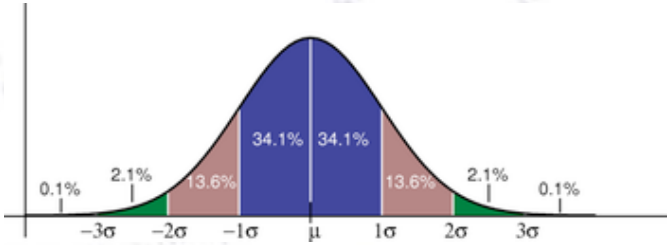
Applied Statistics

Why statistics?



Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense"

Why uncertainties?

Errors/Uncertainties

In physics there are various elements of uncertainty:

- Theory is not deterministic
  Examples: Quantum effects & chaos
- Random measurement errors
  Fluctuations are present even without quantum effects!
- Things we could know in principle but don't…
  e.g. from limitations in cost, time, etc.

We can quantify the uncertainty using **PROBABILITY**

Armed with the realisation of limitations, we can make better calculations/experiments and informed conclusions.

Example: Speed of Gravity

Imagine that you measured the speed of gravity, and got the following result:

$$v_{\text{gravity}} = 2.89 \times 10^8 \text{ m/s}$$

That would tell you...

Example: Speed of Gravity

Imagine that you measured the speed of gravity, and got the following result:

$$v_{\text{gravity}} = 2.89 \times 10^{8} \text{ m/s}$$

That would tell you...

Nothing!!!

Because you have no idea of the uncertainty.

Example: Speed of Gravity

Imagine that you measured the speed of gravity, and got the following result:

$$v_{\text{gravity}} = 2.89 \times 10^8 \text{ m/s}$$

Depending on the uncertainty, you might foresee three very different conclusions:

$$v_{\text{gravity}} = (2.89 \pm 9.21) \times 10^8 \text{ m/s}$$ **Could be anything,** even negative!

$$v_{\text{gravity}} = (2.89 \pm 0.09) \times 10^8 \text{ m/s}$$ **Consistent with c,** and not much else!

$$v_{\text{gravity}} = (2.89 \pm 0.01) \times 10^8 \text{ m/s}$$ **Inconsistent with c:** **New Discovery!!!**

(extreme) Conclusion:

Numbers without stated uncertainties are meaningless!

Why precision?

Newton's Law of Gravity

How well do we know Newton's Law of Gravity?

$$F = G\frac{mM}{r^2}$$

Newton's Law of Gravity

How well do we know Newton's Law of Gravity? Well, reasonably well, but...

Force central?

Valid for all masses?

$$F = G\frac{mM}{r^2}$$

Range of validity?

Square Law?

No other dependencies?

Newton's Law of Gravity

How well do we know Newton's Law of Gravity? Well, reasonably well, but...

Force central?

Seemingly...

Valid for all masses?

NO - not large ones!

Why is G so small?

$$F = G\frac{mM}{r^2}$$

Maybe not short ranges

Range of validity?

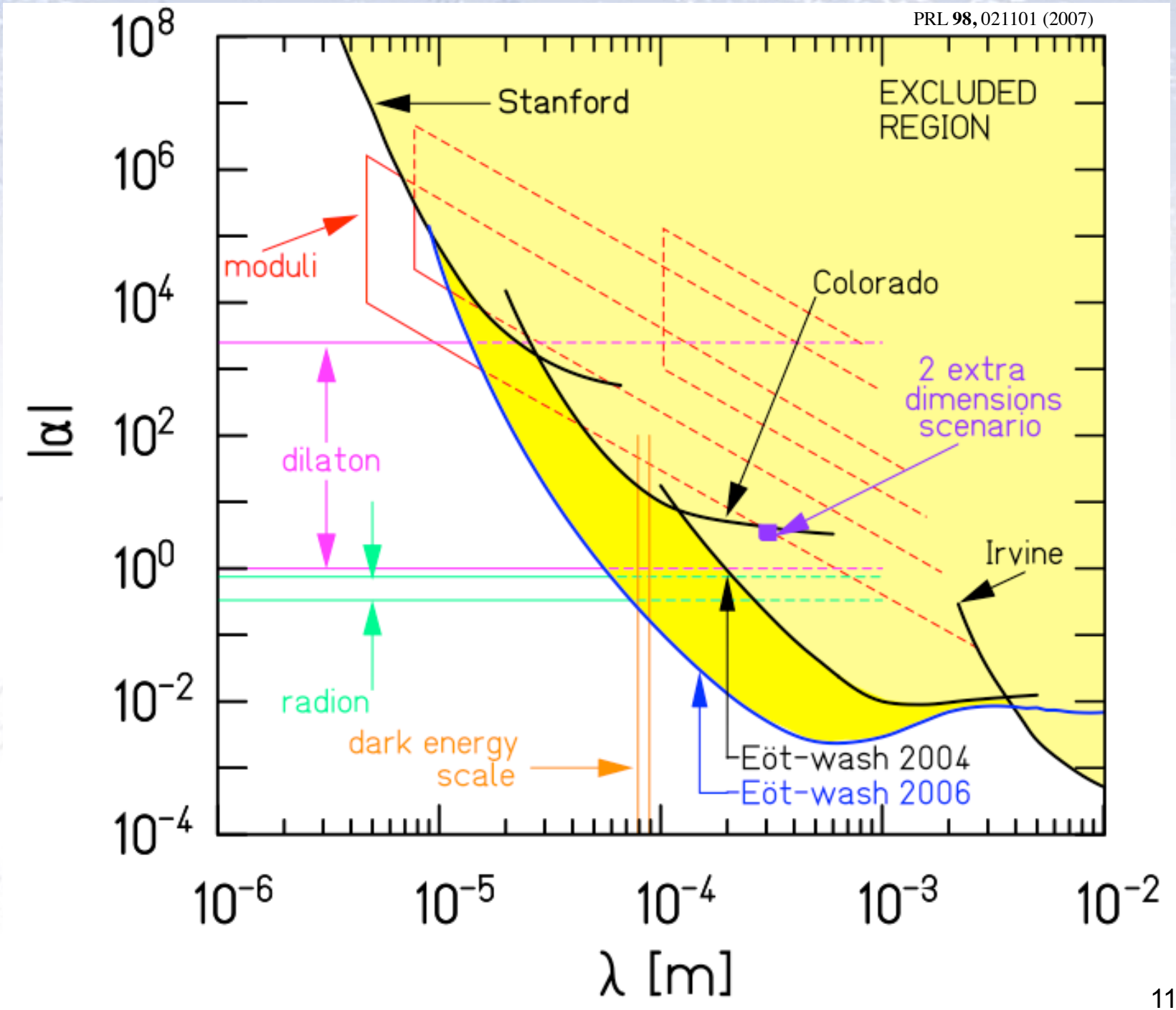
Being tested: Related to search for more dimensions

Square Law?

Yes, from generel relativity

No other dependencies?

Testing Gravity



Why statistics?!?

Why statistics in physics?

Experimental measurements are only **SAMPLES** of the reality, they can never represent the entire set of possibilities, so
→ they are affected by uncertainties
→ results can be expressed as probabilities

Theoretical calculations are mostly **APPROXIMATIONS** limited by finite resources to do the calculations or by imprecise input parameters, so
→ they are also affected by uncertainties
→ predictions can also be expressed in terms of probability

Statistics gives the understanding of uncertainty and probability in relating data and theory!!!

Why statistics in physics?

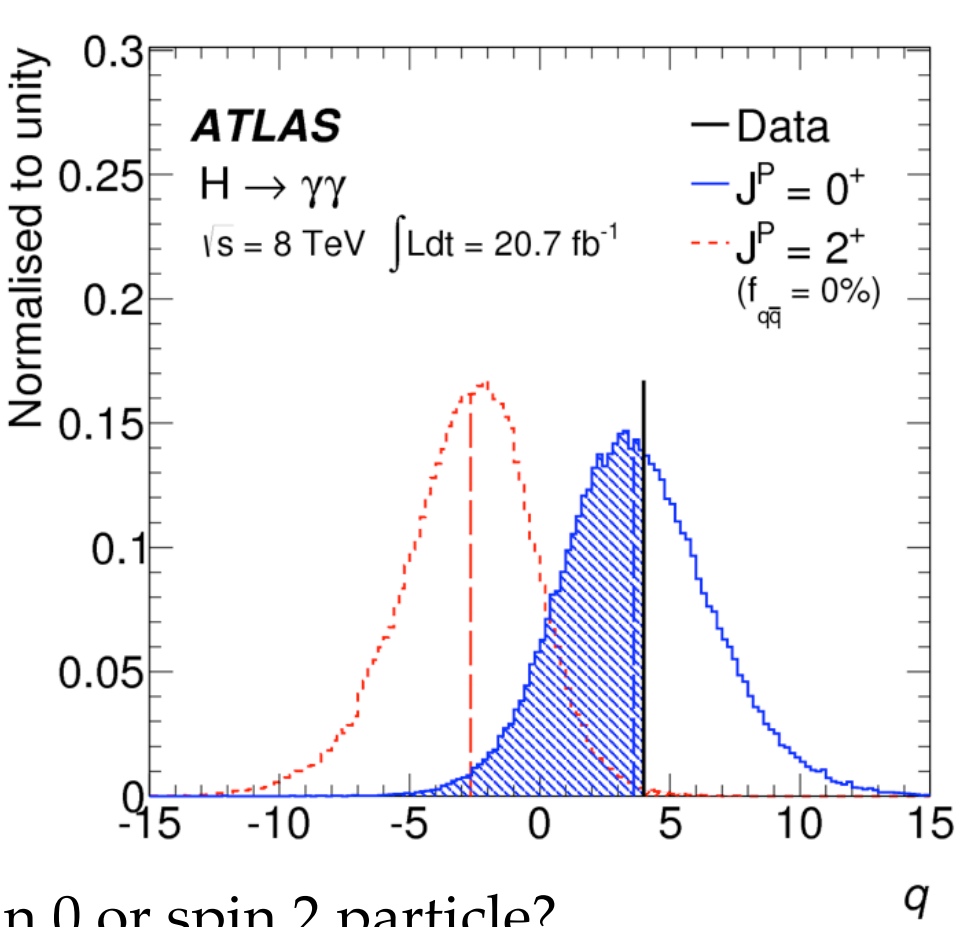
Statistics is about **hypothesis testing**, quantifying the answer to the question

"which theory matches the data best?"

Statistics is about collecting data and logically analysing it, not being fooled by coincidences and chance observations.

Statistics is about fitting trends in data, allowing for projections and predictions.

Statistics is about understanding data, and extracting the essential information from it in the most powerful way.



Is the Higgs a spin 0 or spin 2 particle?

Biases in statistics...

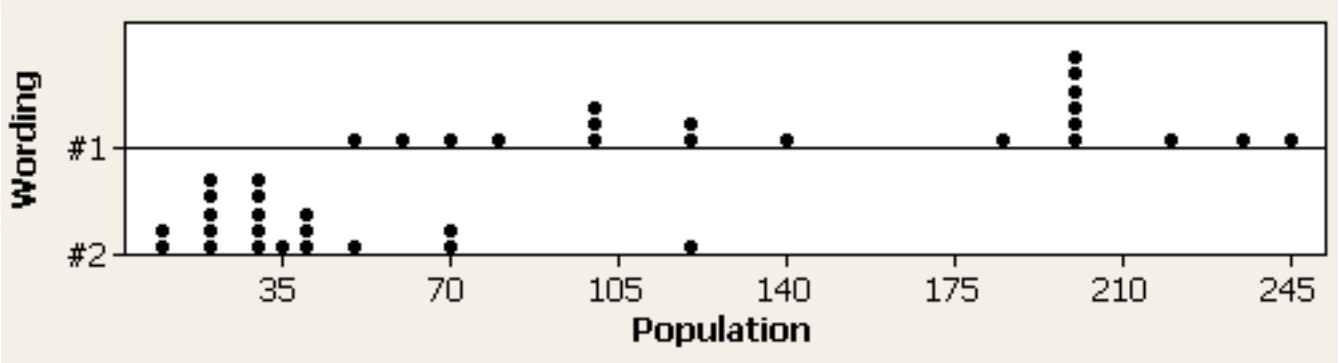
When ASKING people, one may introduce (deliberate?) biases:

- ***Wording 1*:** Pick a color: red or blue?
- ***Wording 2*:** Pick a color: blue or red?

Color Choice	Red	Blue
Wording 1	59 %	41 %
Wording 2	45 %	55 %

One may also bias answers by giving (ir-)relevant information:

- ***Wording 1*:** Knowing that the population of the U.S. is 270 million, what is the population of Canada?
- ***Wording 2*:** Knowing that the population of Australia is 15 million, what is the population of Canada?



Biases in statistics...

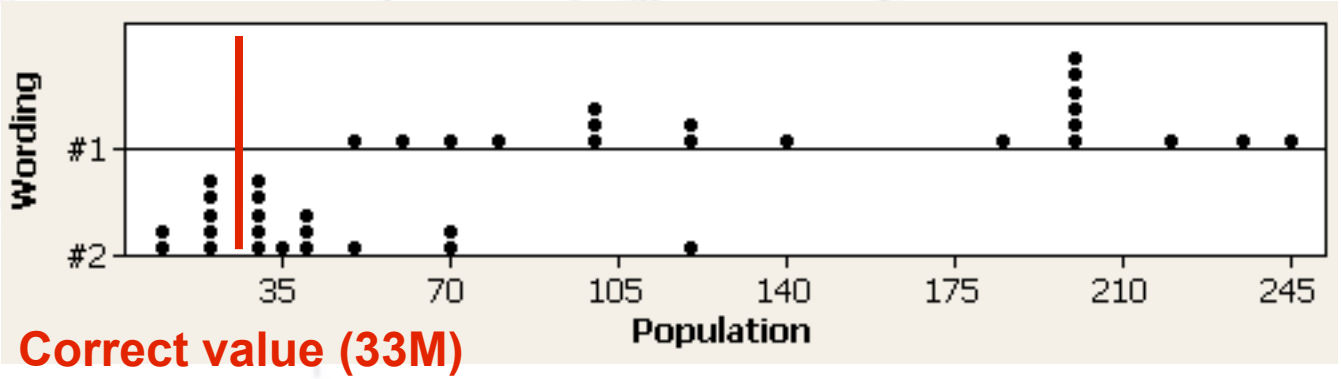
When ASKING people, one may introduce (deliberate?) biases:

- ***Wording 1*:** Pick a color: red or blue?
- ***Wording 2*:** Pick a color: blue or red?

Color Choice	Red	Blue
Wording 1	59 %	41 %
Wording 2	45 %	55 %

One may also bias answers by giving (ir-)relevant information:

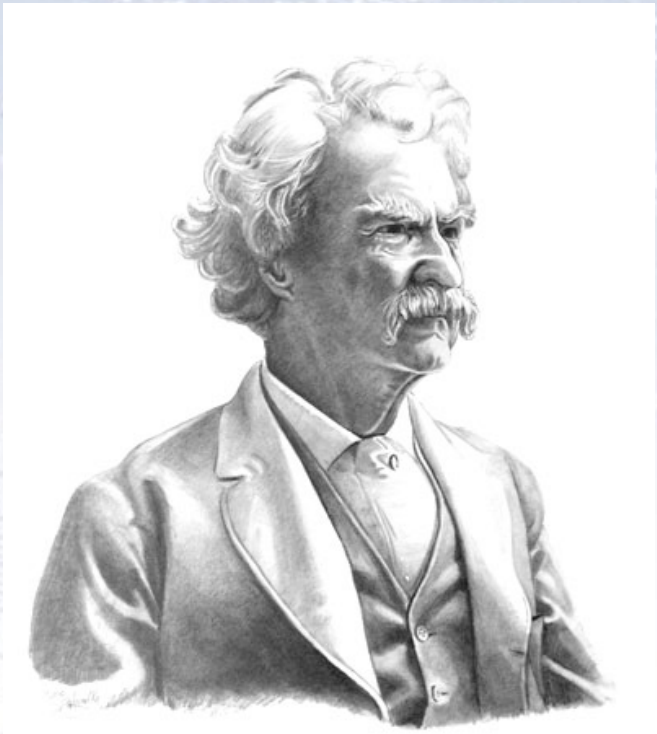
- ***Wording 1*:** Knowing that the population of the U.S. is 270 million, what is the population of Canada?
- ***Wording 2*:** Knowing that the population of Australia is 15 million, what is the population of Canada?



Mark Twain:

"There are three kinds of lies:

lies, damned lies, and statistics."



My opinion:

"The only way to convey accurate

information is by statistics."

THE WALL STREET JOURNAL.

Big Data's High-Priests of Algorithms

'Data Scientists' Meld Statistics and Software for Find Lucrative High-Tech Jobs

8th of August 2014

While a six-figure starting salary might be common for someone coming straight out of a doctoral program, data scientists with just two years' experience can earn between $200,000 and $300,000 a year, according to recruiters.

“Fysik er et brødløst studium”.

[Studenterhåndbogen AU, 1954]

“I keep saying the sexy job in the next ten years will be statisticians.”

[Hal Varian in 2009, Chief economist of Google, Berkeley professor]