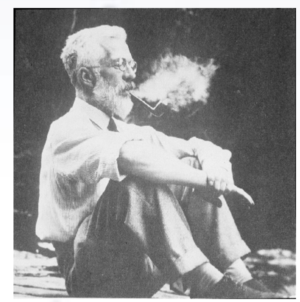
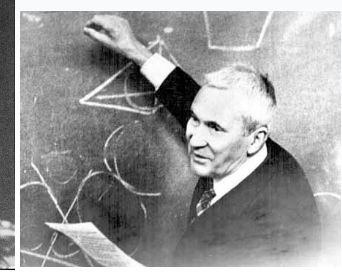
Applied Statistics

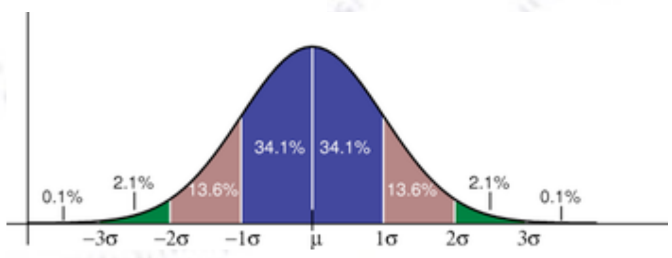
Expected abilities post-course





Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense"

Cognitive levels!

Take data using existing experiment

Obtain the ability to describe (logbook)

Understand an existing experiment

Be able to explain (report)

Recreate an experiment

Show ability to use (use description/logbook)

Test prediction using existing experiment

Check and evaluate (statistics and uncertainties)

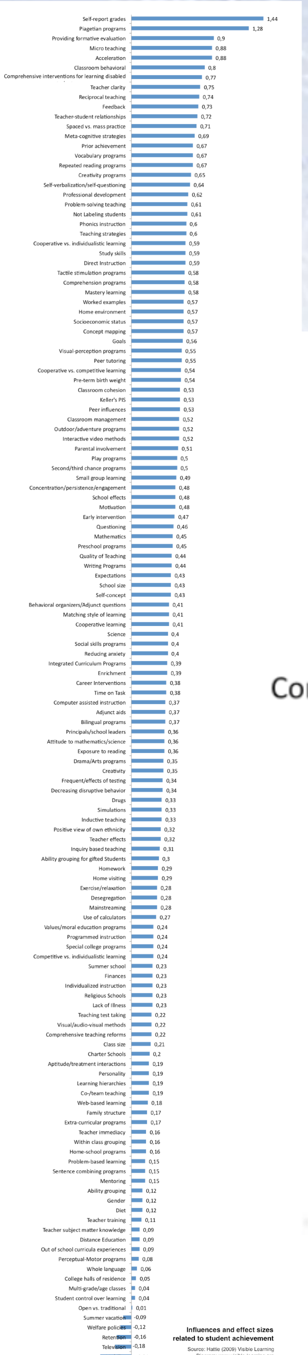
Devise and conduct experiment to test prediction

Make abstractions and combine (independency and simulation)

Measure something new

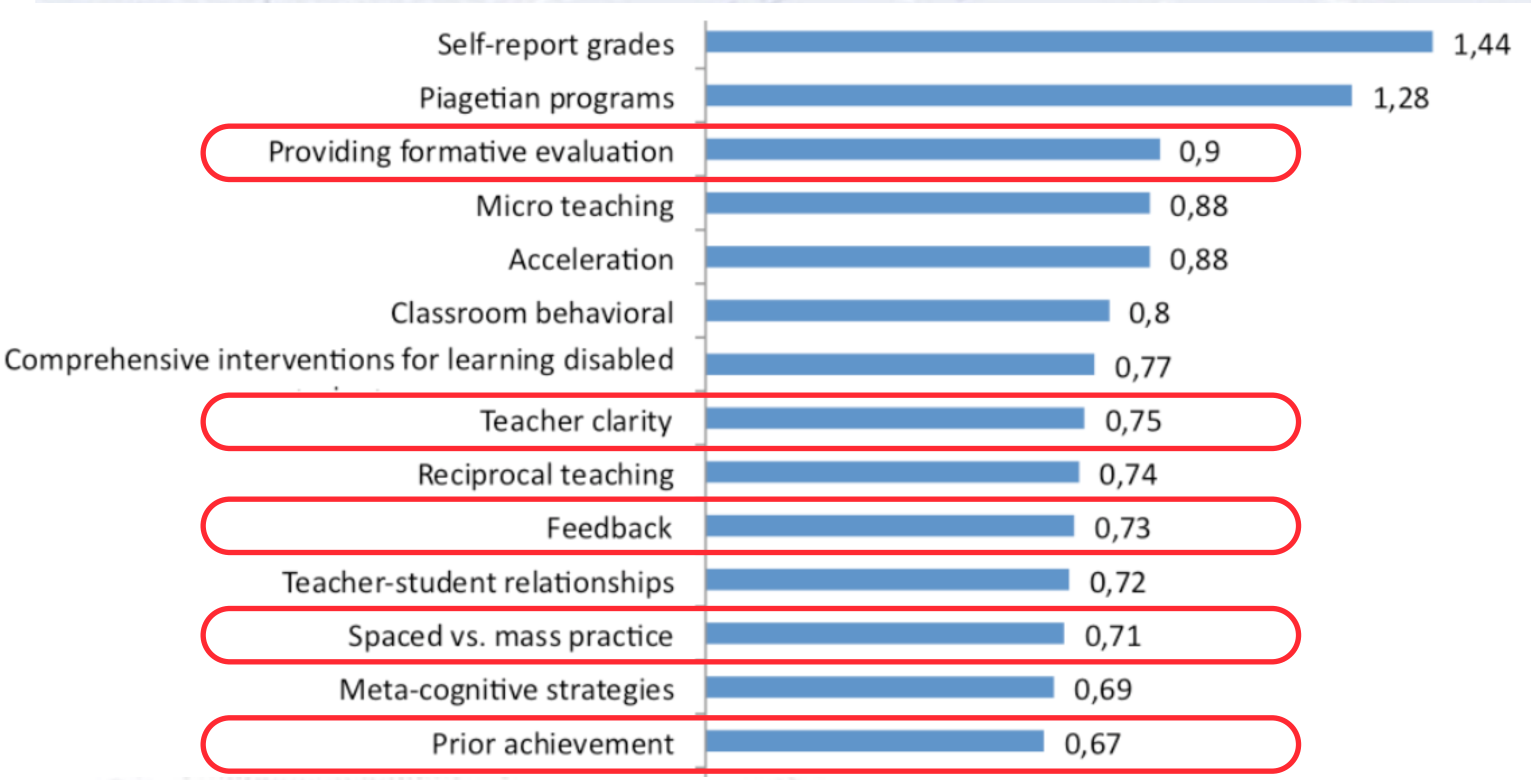
Nobel prize (…)

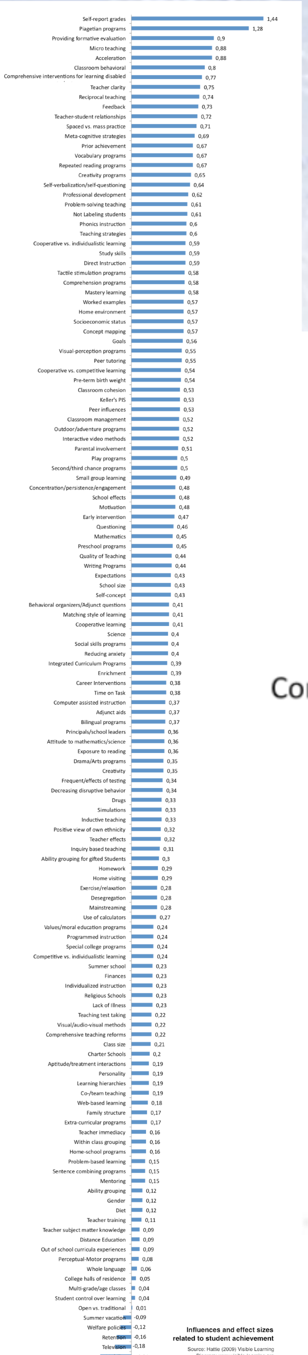
What influences learning?



What influences learning?

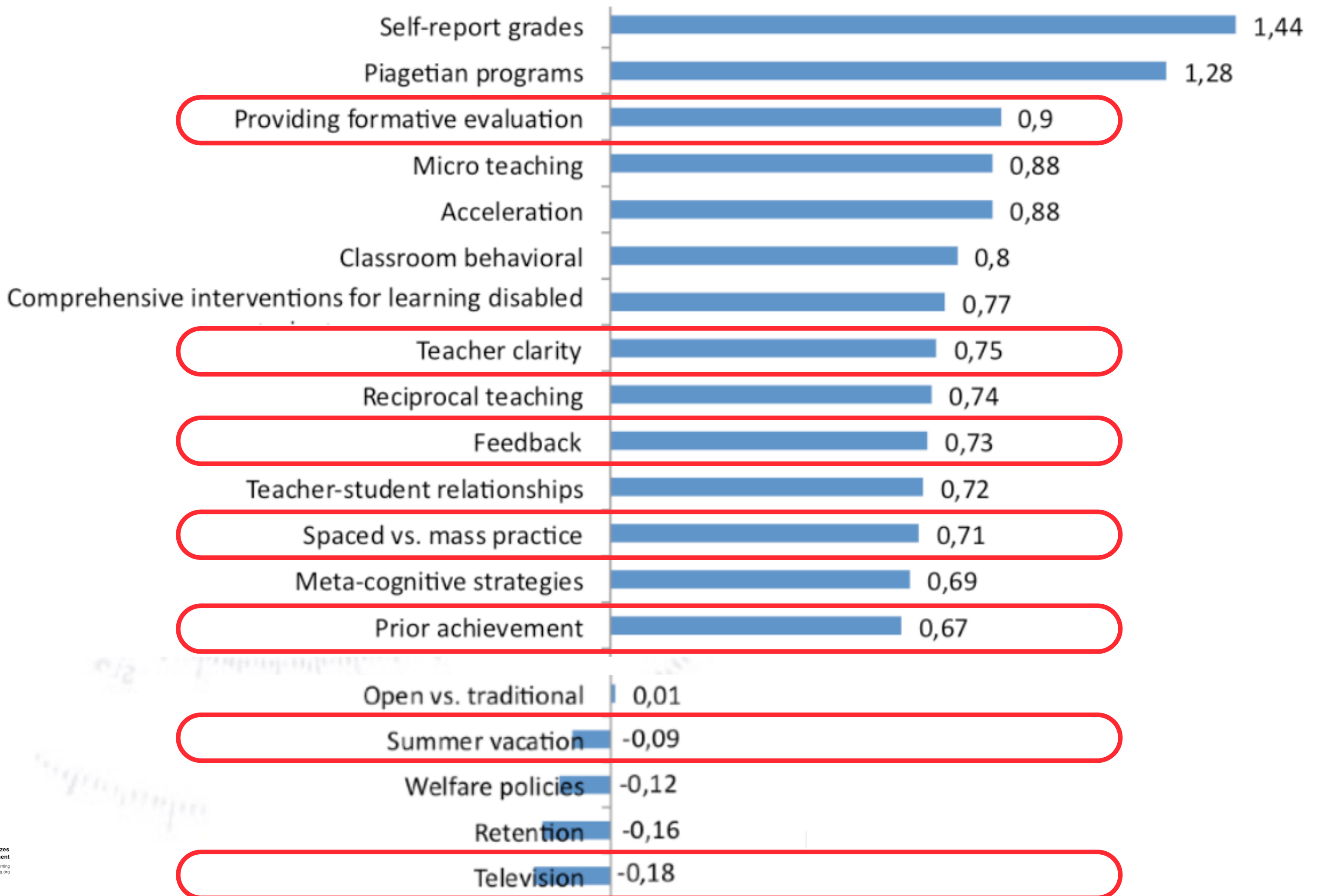
There are studies of this, one result shown below:





What influences learning?

There are studies of this, one result shown below:



Expected learning outcome

Course description in faculty webpages (http://kurser.ku.dk/course/nfyk13011u/2018-2019) gives the following description of the learning outcome:

Skills
The student should in the course obtain the following skills:
Determining mean, width, uncertainty on mean and correlations.
Understanding how to use probability distribution functions.
Be able to calculate, propagate and interpret uncertainties.
Be capable of fitting data sets and obtain parameter values.
Know the use of simulation in planing experiments and data analysis.

Knowledge
The student will obtain knowledge about statistical concepts and procedures, more specifically:
Binomial, Poisson and Gaussian distributions and origins.
Error propagation formula and how to apply it.
ChiSquare as a measure of Goodness-of-fit.
Calculation and interpretation of ChiSquare probability.

Competences
This course will provide the students with an understanding of statistical methods and knowledge of data analysis, which enables them to analyse data in ALL fields of science. The students should be capable of handling uncertainties, fitting data, applying hypothesis tests and extracting conclusions from data, and thus produce statistically sound scientific work.

It would be painful to go through it all in details, but for reference, I've listed everything I expect you to learn (and know for the exam!) in the next pages.

Learning Objectives

The learning objectives of this course can be summarised very simply!
After having followed this course, you should be able to…

"Take raw data and through statistical analysis provide quantified conclusions."

In other words get from data to results, as indicated by the title of the course!

In order to be able to do this, you need to know about the subjects listed on the next page.

Learning Objectives

In order to be able to fulfil the ultimate learning goal, the learning objectives required are listed below. Thus you should know about:

1. Central Limit Theorem, Means, RMS, covariance, and correlation.
2. Probability Density Functions, Binomial, Poisson, and Gaussian relation.
3. Uncertainties and their propagation both analytical and with simulation.
4. Systematic errors and their interpretation. How to obtain uncertainties.
5. Estimators, biases, residuals, and pull plots.
6. Least squares and Chi-Square fits, Ndof, Chi2 probability and interpretation.
7. Principle of Maximum Likelihood, relation with Chi-Square, and use.
8. Bayes' Theorem and usage, confidence intervals, and confidence limits.
9. Hypothesis testing, Errors of type I and II, ROC curves and likelihood ratio.
10. One/Two sample, Chi2, Kolmogorov, runs, and Fisher's exact test.
11. Producing random numbers according to any distribution.
12. Evaluate experimental design prior to execution and results.
13. Knowledge of MultiVariate Analysis and its applications.

While this list is nearly complete, there are many details to it, which I've tried to list in the next page.

Learning Objectives

1. **Central Limit Theorem, Means, RMS, covariance, and correlation.**
   - Central Limit Theorem, its application and implications.
   - Significant digits in results and their uncertainty
   - Arithmetic, Geometric, Harmonic, and Truncated means, mode and median.
   - Variance and Root-Mean-Square (RMS), and alternative names.
   - Weighted mean and the lack of RMS for weighted data.
   - Uncertainty on mean, skewness and kurtosis.
   - Correlation, covariance, covariance matrix, linear correlation coefficient.
2. **Probability Density Functions, Binomial, Poisson, and Gaussian relation.**
   - Probability Density Function (PDF) as concept, their use and interpretation.
   - Binomial (and Multinomial) distribution, its formula and when it is applicable.
   - Poisson distribution, its formula, and when it is applicable.
   - Gaussian distribution, its formula, and when it is applicable.
   - Uniform and exponential distributions, their formula, and applicability.
   - Student's t,Cauchy, and Chi-Square distribution, and when they are applicable.
   - PDFs in several dimensions and how to handle this conceptually.
3. **Uncertainties and their propagation both analytical and with simulation.**
   - The use and interpretation of uncertainties.
   - How to propagate uncertainties analytically.
   - How to propagate uncertainties using simulation.
   - How to propagate uncertainties in the presence in (linear) correlation.
4. **Systematic errors and their interpretation. How to obtain uncertainties.**
   - What systematic errors are conceptually and how they stand out.
   - Knowledge of methods to detector systematic errors.
   - Knowledge of calibration and use of control channels for investigating and minimising systematic errors.
   - How to obtain statistical uncertainties from repeated measurements.
   - How to estimate statistical uncertainties from range with known PDF.
5. **Estimators, biases, residuals, and pull plot.**
   - The concept of estimators in statistics.
   - How to estimate means, widths, correlations, etc.
   - Knowledge of the meaning of bias and efficiency of estimators.
   - How to calculate residuals and their use.
   - How to produce a pull plot and its use.
6. **Least squares and Chi-Square fits, Degrees of Freedom, Chi2 probability and interpretation.**
   - Principle of least squares.
   - Chi-Square, its calculation, use, limitation, interpretation, pitfalls, and impact.
   - Number of degrees of Freedom and how to obtain this for any problem.
   - How to calculate and interpret Chi-Square probability.
   - Fit arbitrary, many parameter, possibly discontinuous 1D functions.
7. **Principle of Maximum Likelihood, relation with Chi-Square, and use.**
   - Principle of Maximum Likelihood and that most things in statistics can be derived from this principle.
   - The relation between likelihood and Chi-Square when fitting.
   - How to interpret difference in Chi-Square between two models.
   - The importance of initial values (i.e. the "art" of fitting).
   - The principle behind obtaining a Goodness-of-Fit value for likelihood fits.
   - How uncertainties are obtained from (-2ln) likelihood fit.
8. **Bayes' Theorem and usage, confidence intervals, and confidence limits.**
   - Bayes' Theorem and (some of) its interpretations.
   - How to use Bayes' Theorem, and how it changes probabilities.
   - The use and impact of prior probabilities.
   - Classification of
   - The concept of confidence intervals and how to calculate these.
   - The concept of confidence limits and how to calculate these (for simple cases).
9. **Hypothesis testing, Errors of type I and II, ROC curves and likelihood ratio.**
   - Hypothesis testing as a concept and the principle behind this.
   - Errors of type I and type II and their difference.
   - The definition, calculation, and interpretation of ROC-curves.
   - The Neyman-Pearson test and likelihood ratio.
   - The concept of Null and Alternative hypotheses.
10. **One/Two sample, Chi2, Kolmogorov, runs, and Fisher's exact test.**
    - One and two sample tests along with paired test.
    - Chi-Square test.
    - Kolmogorov-Smirnov test.
    - Wald-Wolfowitz runs test.
    - Fisher's exact test for contingency tables.
11. **Producing random numbers according to any distribution.**
    - Random numbers and the applications of simulation.
    - Transformation method and requirements.
    - Accept-Reject method, possible combination and requirements.
    - The use of random numbers for complex error propagation and integration.
12. **Evaluate experimental design prior to execution and results.**
    - Evaluate experiments possibly through simulation.
    - Chauvenet's Criterion and how to evaluate/reject single measurements.
    - Experience with real data and the challenges that this poses.
13. **Knowledge of MultiVariate Analysis and its applications.**
    - Understanding of the concept of MultiVariate Analysis.
    - Fisher discriminant, how it is calculated and used.
    - Conceptual understanding of Neural Networks and Boosted Decision Trees.
    - Understand overtraining and the use of training and testing samples.

Learning Objectives

1. **Central Limit Theorem, Means, RMS, covariance, and correlation.**
   - Central Limit Theorem, its application and implications.
   - Significant digits in results and their uncertainty
   - Arithmetic, Geometric, Harmonic, and Truncated means, mode and median.
   - Variance and Root-Mean-Square (RMS), and alternative names.
   - Weighted mean an
   - Uncertainty on me
   - Correlation, covari
2. **Probability Dens**
   - Probability Density
   - Binomial (and Mul
   - Poisson distributio
   - Gaussian distributi
   - Uniform and expor
   - Student's t,Cauchy
   - PDFs in several dir
3. **Uncertainties and**
   - The use and interp
   - How to propagate
   - How to propagate
   - How to propagate
4. **Systematic errors**
   - What systematic er
   - Knowledge of metl
   - Knowledge of calib
     minimising system
   - How to obtain stat
   - How to estimate st
5. **Estimators, biase**
   - The concept of esti
   - How to estimate m
   - Knowledge of the
   - How to calculate residuals and their use.
   - How to produce a pull plot and its use.
6. **Least squares and Chi-Square fits, Degrees of Freedom, Chi2 probability and interpretation.**
   - Principle of least squares.
   - Chi-Square, its calculation, use, limitation, interpretation, pitfalls, and impact.
   - Number of degrees of Freedom and how to obtain this for any problem.
   - How to calculate and interpret Chi-Square probability.
   - Fit arbitrary, many parameter, possibly discontinuous 1D functions.
7. **Principle of Maximum Likelihood, relation with Chi-Square, and use.**
   - Principle of Maximum Likelihood and that most things in statistics can be derived from this principle.
   - The relation between likelihood and Chi-Square when fitting.
   - How to interpret difference in Chi-Square between two models.

r likelihood fits.

nfidence limits.

ities.

hese.

e (for simple cases).

and likelihood ratio.

his.

ves.

exact test.

on.

nents.

n and integration.

12. **Evaluate experimental design prior to execution and results.**
    - Evaluate experiments possibly through simulation.
    - Chauvenet's Criterion and how to evaluate/reject single measurements.
    - Experience with real data and the challenges that this poses.
13. **Knowledge of MultiVariate Analysis and its applications.**
    - Understanding of the concept of MultiVariate Analysis.
    - Fisher discriminant, how it is calculated and used.
    - Conceptual understanding of Neural Networks and Boosted Decision Trees.
    - Understand overtraining and the use of training and testing samples.

However…

Remember that it is not only knowing each single thing on this list, but rather their combination, and the understanding of how to do this, which enables you to analyse data.

The power of their combination is greater, than the sum of their parts.

Study-Life-Balance

A 7.5 ECTS course should involve 187.5-225 hours of work, i.e. **23-28 hours/week**. Classes covers **11 hours/week** (lectures / exercises), and so you should aim to put at least another **12 hours/week!** (Well, at least 8 to be realistic).

I know that this course is hard work (it is also for me!), so let us try to put in our best efforts, while still retaining a healthy work-study-life-balance.

Bonus slides

What influences learning?