# Terascale Statistics School 2023

# Report of Contributions

https://indico.desy.de/e/38671

Contribution ID: **0** Type: **not specified**

# Welcome

*Monday 3 July 2023 10:30 (15 minutes)*

**Presenters:** Dr MELZER-PELLMANN, Isabell (CMS (CMS-Experiment)); BEHNKE, Olaf (CMS (CMS Fachgruppe TOP))

Contribution ID: **1** Type: **not specified**

# Basics Part I

*Monday 3 July 2023 10:45 (45 minutes)*

Probability, Frequentist and Bayesian. Confidence. Bayes theorem. Priors and posteriors

## Summary

**Presenter:** BARLOW, Roger (Huddersfield)

Contribution ID: **2** Type: **not specified**

# Basics Part II

*Monday 3 July 2023 13:30 (45 minutes)*

Probability distributions (Binomial and Poisson) and Probability distribution functions (Gaussian). Expectation values. Hypothesis testing

## Summary

**Presenter:** BARLOW, Roger (Huddersfield)

Contribution ID: **3** Type: **not specified**

# Higgs Analysis Walk Through Tutorial Part 1

*Monday 3 July 2023 15:00 (3 hours)*

DESY Terascale Statistics School - July 2023
What: description analysis walkthrough session
Who: Oliver Rieger, Zef Wolffs, Ivo van Vulpen

Course description

The analysis walkthrough is a hands-on session in which we will address, in a 'real life'example, several of the statistics topics that were covered in the lectures. More concretely: we will study the four-muon invariant mass distribution in the Higgs boson decay to four charged muons to explore three different topics:
1. Significance (Poisson distribution, p-values, optimizations)
2. Likelihood fits (parameter estimation, side band fits, background uncertainty)
3. Hypothesis testing (test statistic, toy data-sets, limits)

Goal of the exercises is to guide participants through the various steps without using the standard toolkits, i.e. we'll focus on the concepts and program as much as we can ourselves.

Set-up and computing requirements:
After an introduction lecture all exercises and related background information can be accessed through a dedicated website.
As participant you can use the Root set-up on your laptop (Option 1 below),
or the DESY computing cluster (Option 2 below) for which guest accounts will be provided by the workshop organisers.
Through a GitLab repository all participants have access to all the material: exercises, data-sets, scripts, background information and, not unimportant, the answers to the exercises. Details on the GitLab repository and the website will be provided later.

---

**Option 1: Run Root locally on the laptop**

We will run the software locally on the laptop. This requires that you have
Root installed. If that is not the case you can follow the instructions
here: https://root.cern/install/

---

**Option 2: Run on DESY cluster from laptop**

We have rented guest accounts on the DESY Cluster, where
tutorials can be run. The only requirement for participants is a
laptop that allows an SSH connection.
- For Linux and MacOS this can be done directly from the terminal
- For Windows machines you need an SSH client (Putty) or chrome extension
secure SSH or setup via VS code and SSH keys & config.

ssh -Y schoolxx@naf-cms.desy.de with xx in range [00,79]
pwd: p9FqFt7f

## Summary

**Presenter:** VAN VULPEN, Ivo (NIKHEF)

Contribution ID: **4** Type: **not specified**

# Basics Part III

*Tuesday 4 July 2023 09:30 (45 minutes)*

Basics Estimation. Maximum likelihood. Least squares. Fitting histograms. Chi squared and goodness of fit. p-value

## Summary

**Presenter:** BARLOW, Roger (Huddersfield)

Contribution ID: **5** Type: **not specified**

# Confidence Interval Estimation Part I

*Tuesday 4 July 2023 11:15 (45 minutes)*

Intricately linked with the estimation of parameters is the question on how to obtain meaningful uncertainties on the obtained parameters. In these two lectures, we will look at this problem in detail and discuss confidence intervals on extracted parameters. We will start with a simple case of a counting experiment following Poissonian statistics. This will lead us to the coverage of uncertainties, which we will use to study different estimators for the statistical uncertainty. We will continue with the more general case of how to obtain confidence intervals for a true parameter, given a measurement of a quantity related to this parameter. The lecture will cover confidence intervals close to physical boundaries and limit setting with the CLs method.

**Presenter:** KOGLER, Roman (DESY FH, CMS)

Contribution ID: **6** Type: **not specified**

# Confidence Interval Estimation Part II

*Tuesday 4 July 2023 14:00 (45 minutes)*

**Presenter:** KOGLER, Roman (DESY FH, CMS)

Contribution ID: **7** Type: **not specified**

# Higgs Analysis Walk Through Tutorial Part 2

*Tuesday 4 July 2023 15:30 (3 hours)*

**Presenter:** VAN VULPEN, Ivo (NIKHEF)

Contribution ID: **8** Type: **not specified**

# Machine Learning Part 2

*Wednesday 5 July 2023 15:30 (2h 30m)*

## Summary

**Presenter:** PROSPER, Harrison (FSU)

Contribution ID: **9** Type: **not specified**

# Systematics Part 1

*Wednesday 5 July 2023 09:30 (45 minutes)*

Estimation and checks

## Summary

**Presenter:** BARLOW, Roger (Huddersfield)

Contribution ID: **10** Type: **not specified**

# Machine Learning Part 1

*Wednesday 5 July 2023 11:15 (45 minutes)*

Title: An Introduction to Machine Learning

Abstract:

The three lecture/tutorials cover the basic principles of machine learning, which is the underlying technology of artificial intelligence. Lecture 1 covers the mathematical foundations with a focus on supervised learning. Lectures 2 and 3 cover a few of the widely used machine learning models, including boosted decision trees, deep feed forward neural networks, convolutional neural networks, auto-encoders and if time permits, transformers (the model underlying ChatGPT). The models will be illustrated with simples example from particle physics, astronomy, and calculus.

Link to the software for the Computer Tutorials:
https://github.com/hbprosper/Terascale

## Summary

**Presenter:** PROSPER, Harrison (FSU)

Contribution ID: **11** Type: **not specified**

# Systematics Part II

*Wednesday 5 July 2023 14:00 (45 minutes)*

Correlation, building and handling the matrices

## Summary

**Presenter:** BARLOW, Roger (Huddersfield)

Contribution ID: **12** Type: **not specified**

# Machine Learning Part 3

*Thursday 6 July 2023 09:30 (45 minutes)*

**Presenter:** PROSPER, Harrison (FSU)

Contribution ID: **13** Type: **not specified**

# Special Q&A session

*Thursday 6 July 2023 14:00 (45 minutes)*

We will discuss questions collected at:
https://docs.google.com/document/d/1id8UczLtlOWlhI7LMXK8cU37AE_wAbIQzyo1nodgxWI/edit

## Summary

Contribution ID: **14** Type: **not specified**

# Introduction to R

*Thursday 6 July 2023 11:15 (45 minutes)*

An introduction to the R language. The aim of this is to explain the minimum about R that everybody needs to know, and hopefully encourage those that would benefit from learning and using the language to do so.

For Lecture 6, it will be helpful (though not essential) to download R beforehand from https://cran.r-project.org/ [cran.r-project.org] or, if you prefer working in an IDE, R-studio from https://posit.co/download/rstudio-desktop/ [posit.co]

## Summary

**Presenter:** BARLOW, Roger (Huddersfield)

Contribution ID: **15** Type: **not specified**

# Artificial Intelligence Today and Tomorrow

*Thursday 6 July 2023 15:00 (45 minutes)*

Recent advances in the field of artificial intelligence (AI) have given us a glimpse of a potentially thrilling, even civilization-changing, future. From that optimistic viewpoint, after a brief historical introduction, I survey some of the recent advances in AI. Then, I speculate about what the impact of these advances might be on the nature of research in particle physics over the next few decades. I end by acknowledging the fact that every technology, however benign it may at first appear, can be used for good or for ill. It is, therefore, appropriate to sound a cautionary note, not so much to be a doomsayer, which I'm not, but rather to remind you that the future remains yours to make.

## Summary

**Presenter:** PROSPER, Harrison

Contribution ID: **16** Type: **not specified**

# Bayesian inference

**Presenter:** Prof. KRÖNINGER, Kevin (TU Dortmund)

Contribution ID: **17** Type: **not specified**

# GOF- and Likelihood-Ratio tests for improved data modelling

**Presenter:** Dr BEHNKE, Olaf (CMS (CMS Fachgruppe TOP))

Contribution ID: **18** Type: **not specified**

# Goodbye

Contribution ID: **19** Type: **not specified**

# Tutorial

Contribution ID: **20** Type: **not specified**

# Discussion Time

*Monday 3 July 2023 11:30 (30 minutes)*

Contribution ID: **21** Type: **not specified**

# Discussion time

*Monday 3 July 2023 14:15 (15 minutes)*

Contribution ID: **22** Type: **not specified**

# Discussion time

*Tuesday 4 July 2023 10:15 (30 minutes)*

Contribution ID: **23** Type: **not specified**

# Discussion time

*Tuesday 4 July 2023 12:00 (30 minutes)*

Contribution ID: **24** Type: **not specified**

# Discussion time

*Tuesday 4 July 2023 14:45 (15 minutes)*

Contribution ID: **25** Type: **not specified**

# Discussion time

*Wednesday 5 July 2023 10:15 (30 minutes)*

Contribution ID: **26** Type: **not specified**

# Discussion time

*Wednesday 5 July 2023 12:00 (30 minutes)*

Contribution ID: **27** Type: **not specified**

# Discussion time

*Wednesday 5 July 2023 14:45 (15 minutes)*

Contribution ID: **28** Type: **not specified**

# Discussion time

Contribution ID: **29** Type: **not specified**

# Discussion time

*Thursday 6 July 2023 10:15 (30 minutes)*

Contribution ID: **30** Type: **not specified**

# Discussion time

*Thursday 6 July 2023 12:00 (30 minutes)*

Contribution ID: **31** Type: **not specified**

# Discussion time

*Thursday 6 July 2023 14:45 (15 minutes)*

Contribution ID: **32** Type: **not specified**

# Discussion time

*Thursday 6 July 2023 15:45 (15 minutes)*

Contribution ID: **33** Type: **not specified**

# Closing/Good bye

*Thursday 6 July 2023 16:00 (10 minutes)*

**Presenters:** MELZER-PELLMANN, Isabell (CMS (CMS-Experiment)); BEHNKE, Olaf (CMS (CMS Fachgruppe TOP))

Contribution ID: **34** Type: **not specified**

# Registration

*Monday 3 July 2023 10:00 (30 minutes)*