

Data Analysis - 2020

Exercise sheet no 7:
Hypothesis testing

8. December 2020

Exercise 1: P-value examples (14 Points)

For the following situations decide whether the p-value should be one- or two-tailed and calculate it.

- (a) The standard particle physics theory predicts an electron-to-muon ratio R_K to be exactly 1.0. You measure $R_K = 0.83 \pm 0.06$. Calculate the p-value of your result with respect to the standard theory.
- (b) I measure earth's gravitational acceleration to be $g = 9.70 \pm 0.10 \text{ m/s}^2$ and you measure it to be $g = 9.90 \pm 0.09 \text{ m/s}^2$ km. Calculate the compatibility p-value of the two measurements.
- (c) You run an experiment designed to look for a new physics signal. An average of 1.5 background events is predicted during the experiment run period. You observe 6 events, four times the background prediction. Calculate the p-value of the background-only hypothesis.
- (d) *Gun crime has gone up by 20% in the UK this year!* claims a tabloid newspaper. These figures are based on 50 gun incidents in 2019 and 60 gun incidents in 2020. Calculate the p-value that the gun crime level has actually stayed constant.
- (e) A COVID-19 vaccine trial is being performed. During the trial period, without a vaccine, the average infection rate is 3000 people per 1 million citizens. A group of 8924 people are given the vaccine and 3 are infected with COVID-19 during the trial period. Calculate the p-value with respect the hypothesis that the vaccine is not effective.
- (f) The height of a hockey team and a soccer team are compared in the table below. If the standard deviation (σ) is unknown, what is the p-value that these samples are compatible with each other? What is the p-value if the standard deviation of both parent height distributions are known to be $\sigma = 5 \text{ cm}$?

Volleyball players	187 cm	185 cm	183 cm	176 cm	190 cm						
Football players	170 cm	174 cm	186 cm	178 cm	185 cm	176 cm	182 cm	184 cm	179 cm	189 cm	177 cm

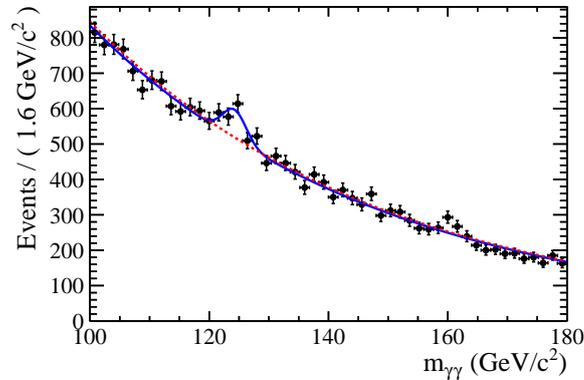
Exercise 2: Higgs discovery (6 Points)

A maximum likelihood fit is performed to measurements of the di-photon mass between 110 and 180 GeV/c^2 , which is representative of what was used to discover the Higgs Boson in 2012. The fit is a background only fit, using an exponential PDF. The negative-log-likelihood of the background fit is $\text{NLL} = -98137.25$ and the χ^2 value of a histogram of 50 bins shown in the plot below with the red-dashed PDF is $\chi^2 = 68.1$.

Secondly, a Gaussian PDF is added to the background PDF to make a new background+signal PDF. The Gaussian PDF has a fixed mean at $\mu = 125 \text{ GeV}/c^2$ and width at $\sigma = 5 \text{ GeV}/c^2$. The size of the Gaussian relative to the exponential background is allowed to vary in the fit.

Data Analysis (PHY231) HS20 - Exercise sheet no 7: Hypothesis testing

The negative-log-likelihood of this signal and background fit is $\text{NLL} = -98150.44$ and the χ^2 value of the histogram on the plot below with 50 bins with the blue curve is $\chi^2 = 49.5$.



- Use the relevant χ^2 value to calculate a p-value that the background-only (null) hypothesis is correct.
- Use Wilk's theorem to calculate the p-value of the null hypothesis with respect to the signal+background hypothesis.
- In reality, the Higgs boson mass of $125 \text{ GeV}/c^2$ was not known when it was discovered. Describe how would this change the p-value of the signal+background hypothesis? Explain your reasoning.

Deadline for submission: Friday, 18. December 2020 14:00

Form: Please submit your solutions to da@physik.uzh.ch. The solutions should be submitted as a single python script with answers to specific questions in the comments.