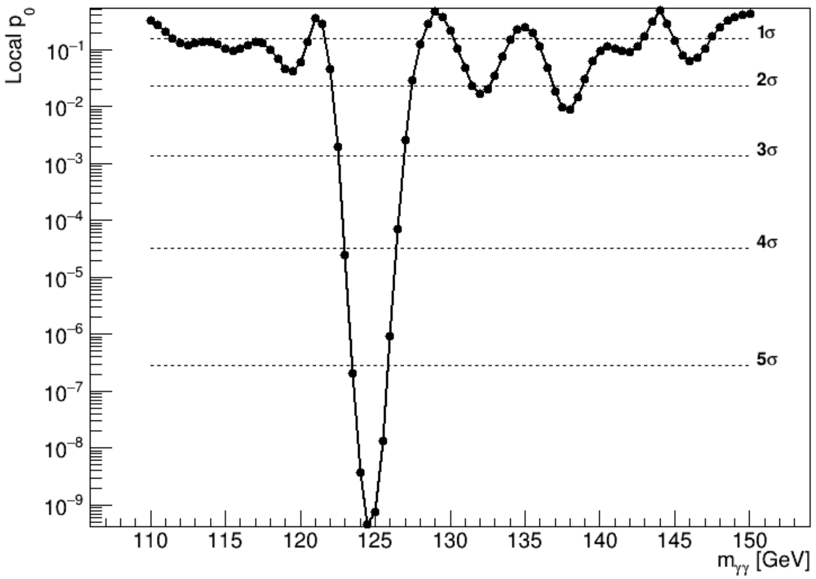
MIDDLE EAST TECHNICAL UNIVERSITY

Particle Data Analysis in High Energy Physics

Lecture 11

Statistics 2



Ahmet Bingül

METU, Physics

Apr 2026

Introduction

In this chapter, we will consider that we have normally distributed samples with sample size $n \geq 30$.

Content of the chapter is as follows:

Mathematics of Curve Fitting

Standard Normal Curve (Revised)

Confidence Interval Estimates

Hypothesis Tests

(Upper limit of neutrino mass and discovery of Higgs Boson)

Mathematics of
Curve Fitting

Least Square Fitting

(*This is alternative for Maximumum Likelihood Method. Both methods are implemented in ROOT*)

Let us suppose that measurements at n points, $\{x_1, x_2, \ldots, x_n\}$ are made of variable $\{y_1, y_2, \ldots, y_n\}$ with and error $\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$. It is desired to fit a function $f(x_i, \boldsymbol{\theta})$ to these data where $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_m\}$ are unknown parameters to be determined. Of course, $n > m$. The method of **least squares** states that the best values of θ_j are obtained by **minimizing** a chi-square function:

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{y_i - f(x_i, \boldsymbol{\theta})}{\sigma_i} \right)^2$$

To find the best values of θ_j one must solve

the system of *m*-equations:

$$\frac{\partial \chi^2}{d\theta_j} = 0 \quad \text{for } j = 1,2, \ldots, m$$

To estimate the errors on the parameters,

we form the covariance matrix, V_{ij}.

$$(V^{-1})_{ij} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial\theta_i \partial\theta_j}$$

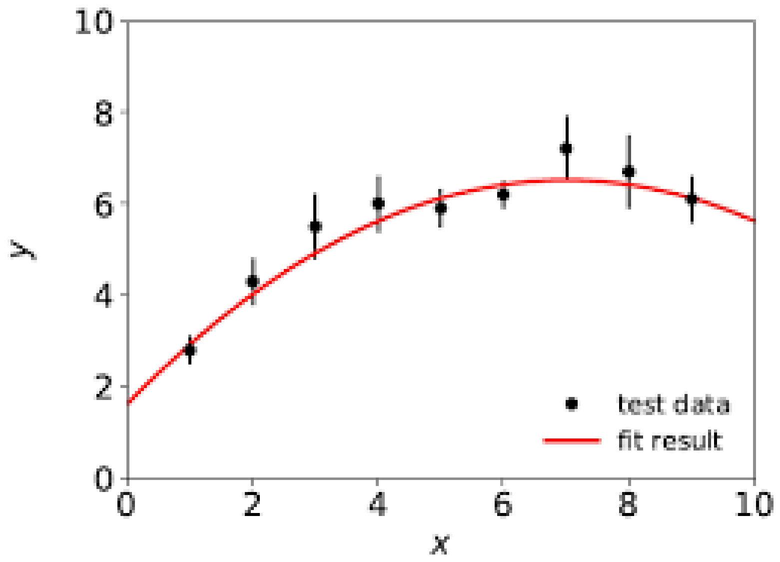
$$\underline{V} = \begin{pmatrix} \sigma_1^2 & \mathrm{cov}(1,2) & \mathrm{cov}(1,3) & \ldots \\ . & \sigma_2^2 & \mathrm{cov}(2,3) & \ldots \\ . & . & \sigma_3^2 & \ldots \\ . & . & . & \end{pmatrix}$$

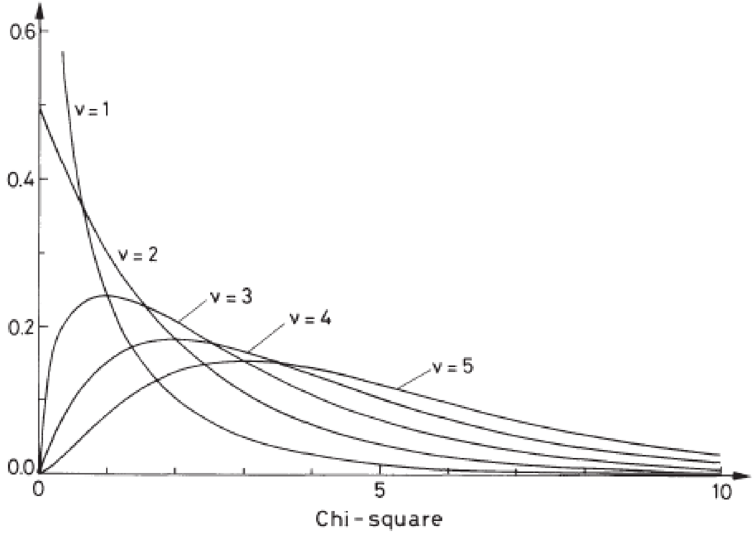
The diagonal elements are variances for θ_j

Quality of Fit

if the data correspond to the function and the deviations are Gaussian, χ^2 should be expected to follow a chi-square distribution, with mean value equal to the number of degrees of freedom (ndf), ν. Let $u = \chi^2$:

$$P(u)\,du = \frac{(u/2)^{(\nu/2)-1}\exp(-u/2)}{2\Gamma(\nu/2)}\,du$$





For a good fit we can expect χ^2 close to ν. Or

$$\frac{\chi^2_{\text{fit}}}{\nu} = \frac{\chi^2_{\text{fit}}}{n-m} \approx 1$$

A more rigorous test is to look at the probability of obtaining a χ^2 value greater than 0.05.

So, if $P\left(\chi^2 > \chi^2_{\text{fit}}\right) > 0.05$ fit can be accepted.

In pyROOT: $p_{\text{value}} = P\left(\chi^2 > \chi^2_{\text{fit}}\right) \rightarrow$ `p_value = ROOT.Math.chisquared_cdf_c(chi2, ndf);`

Chatgpt (knows me and) says:

🔷 In particle physics (important for you)

We usually use:

✅ Goodness-of-fit

- Acceptable if:

$$0.05 \lesssim p \lesssim 0.95$$

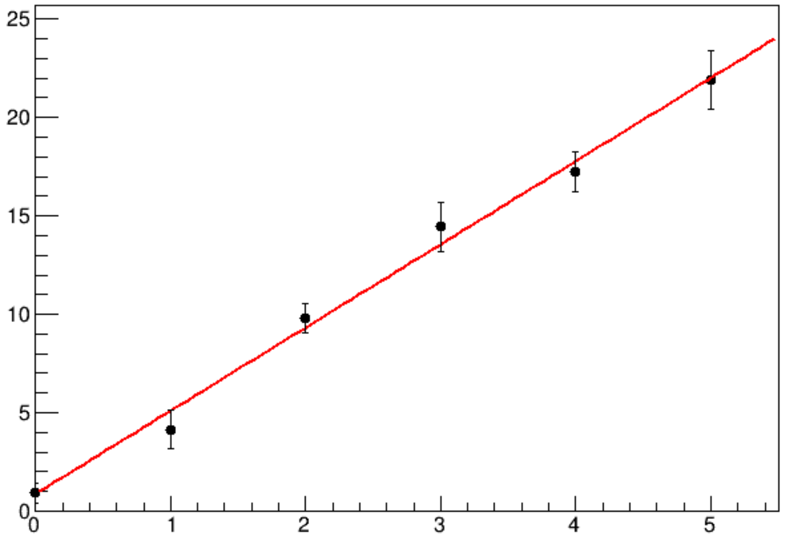
- Best region:

$$p \sim 0.1 - 0.9$$

Example

Find the best straight line ($y = ax + b$) through the following measured points:

x	0	1	2	3	4	5
y	0.92	4.15	9.78	14.46	17.26	21.90
σ	0.5	1.0	0.75	1.25	1.0	1.5



Solution

$a = 4.227$, $\sigma(a) = 0.044$

$b = 0.878$,, $\sigma(b) = 0.203$

$\mathrm{cov}(a, b) = -0.0629$

$\chi^2_{\mathrm{fit}} = 2.078$

$$\frac{\chi^2_{\mathrm{fit}}}{\nu} = \frac{2.078}{6-2} = 0.52$$

$p = P(\chi^2 > 2.078) = 0.721$ is well within **acceptable** limits, since $0.05 < p < 0.95$

```python
import ROOT
from array import array

x = array('d', [0,1,2,3,4,5])
y = array('d', [0.92,4.15,9.78,14.46,17.26,21.9])
s = array('d', [0.5,1.0,0.75,1.25,1.0,1.5])
ex= array('d', [0,0,0,0,0,0]) # no x errors

gr = ROOT.TGraphErrors(6, x, y, ex, s)
fit_fun = ROOT.TF1("fun","[0]*x+[1]",0,5)
fit_result = gr.Fit(fit_fun, "S")
gr.SetMarkerSize(1)
gr.SetMarkerStyle(20)
gr.SetMinimum(0)
gr.Draw("AP")

chi2    = fit_result.Chi2()
ndf     = fit_result.Ndf()
cov_mat = fit_result.GetCovarianceMatrix()
p_value = ROOT.Math.chisquared_cdf_c(chi2, ndf)

cov_mat.Print()
print("p_value  = ",p_value)
print("chi2/ndf = ",chi2/ndf)
input("Press enter to stop")
```

```
****************************************
Minimizer is Minuit2 / Migrad
Chi2                      =      2.07778
NDf                       =            4
Edm                       =   1.4774e-22
NCalls                    =           29
p0                        =      4.22694   +/-   0.209478
p1                        =     0.879203   +/-   0.450469
Info in <TCanvas::MakeDefCanvas>:  created default TCanvas

2x2 matrix is as follows

     |      0    |      1    |
-------------------------------
   0 |    0.04388    -0.06287
   1 |   -0.06287      0.2029

p_value  =  0.7214554414070304
chi2/ndf =  0.5194451992693816
```

Least Square vs Maximum Likelihood

We have seen:

$L(\theta) = \prod_{i=1}^{n} f(x_i, \boldsymbol{\theta})$, find best model parameters $(\hat{\boldsymbol{\theta}})$ by maximizing L or $L^* = \ln L \rightarrow$ solve $\frac{\partial(L^*)}{\partial\theta_j} = 0$

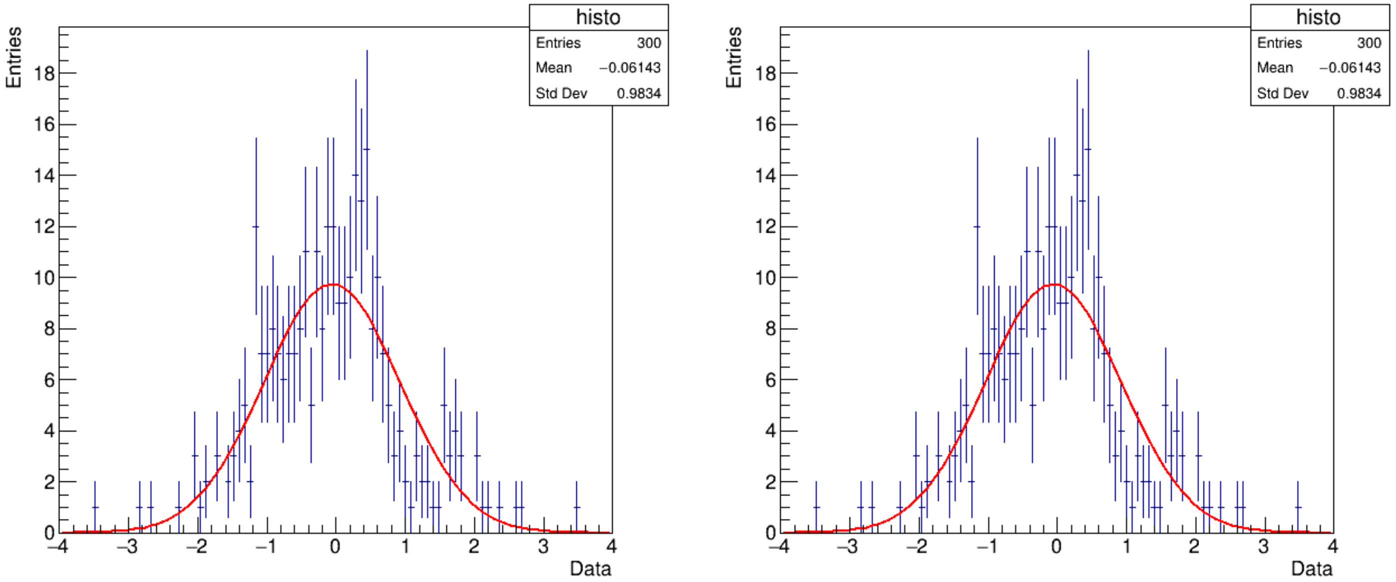
$\chi^2(\theta) = \sum_{i=1}^{n} \left(\frac{y_i - f(x_i,\boldsymbol{\theta})}{\sigma_i}\right)^2$ find best model parameters $(\hat{\boldsymbol{\theta}})$ by miminizing χ^2 $\rightarrow$ solve $\frac{\partial(\chi^2)}{\partial\theta_j} = 0$

If deviations are Gaussian, one can show that $\chi^2(\theta) = -2L^* + \text{constant}$

- ROOT uses the MINUIT package for the numerical optimization
- Default fit is chi2
- Maximum likelihood is better at low statistics, but may take longer

Example

```python
import ROOT

# booking
h = ROOT.TH1F("histo",";Data;Entries",100,-4,4)
h.FillRandom("gaus",300)

# Plotting
#ROOT.gStyle.SetOptFit(1111)
can = ROOT.TCanvas("c1","test",1000,600)
can.Divide(2,1)

can.cd(1)
h.Fit("gaus")      # chi2 fitting (default)
h.Draw("e")        # draw histogram with error bars

can.cd(2)
h.Fit("gaus","L") # likelihood fitting
h.Draw("e")        # draw histgoram with error bars

input('Press enter to quit')
```




```
************************************
Minimizer is Minuit2 / Migrad
Chi2                      =      54.0423
NDf                       =           55
Edm                       =  5.99245e-06
NCalls                    =           67
Constant                  =      9.32706   +/-   0.869228
Mean                      =    -0.197602   +/-   0.0624891
Sigma                     =     0.870928   +/-   0.0693346
************************************
Minimizer is Minuit2 / Migrad
MinFCN                    =      48.4609
Chi2                      =      96.9218
NDf                       =           97
Edm                       =  4.57551e-07
NCalls                    =           59
Constant                  =      9.73387   +/-   0.688961
Mean                      =   -0.0616516   +/-   0.0568191
Sigma                     =     0.983716   +/-   0.0403082
Press enter to quit
```

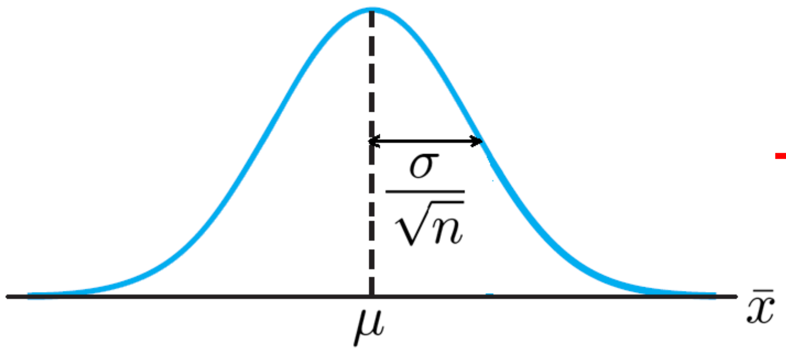
Fit Options

- " `W` " Set all weights to 1 for non empty bins; ignore error bars
- " `WW` " Set all weights to 1 including empty bins; ignore error bars
- " `I` " Use integral of function in bin instead of value at bin center
- " `L` " Use log likelihood method (default is chi-square method). To be used when the histogram represents counts
- " `WL` " Weighted log likelihood method. To be used when the histogram has been filled with weights different than 1.
- " `P` " Use Pearson chi-square method, using expected errors instead of the observed one given by `TH1::GetBinError` (default case). The expected error is instead estimated from the square-root of the bin function value.
- " `Q` " Quiet mode (minimum printing)
- " `V` " Verbose mode (default is between Q and V)
- " `S` " The result of the fit is returned in the `TFitResultPtr` .
- " `E` " Perform better errors estimation using the Minos technique
- " `M` " Improve fit results, by using the *IMPROVE* algorithm of TMinuit.
- " `R` " Use the range specified in the function range
- " `N` " Do not store the graphics function, do not draw
- " `0` " Do not plot the result of the fit. By default the fitted function is drawn unless the option " `N` " above is specified.
- " `+` " Add this new fitted function to the list of fitted functions (by default, the previous function is deleted and only the last one is kept)
- " `B` "Use this option when you want to fix one or more parameters and the fitting function is a predefined one, like `polN` , `expo` , `landau` , `gaus` . Note that in case of pre-defined functions some default initial values and limits are set.
- " `C` "In case of linear fitting, don't calculate the chisquare (saves time).
- " `F` "If fitting a linear function (e.g. `polN`), switch to use the default minimizer (e.g. `Minuit`). By default, `polN` functions are fitted by the linear fitter.

Standard Normal Curve
(Revised)

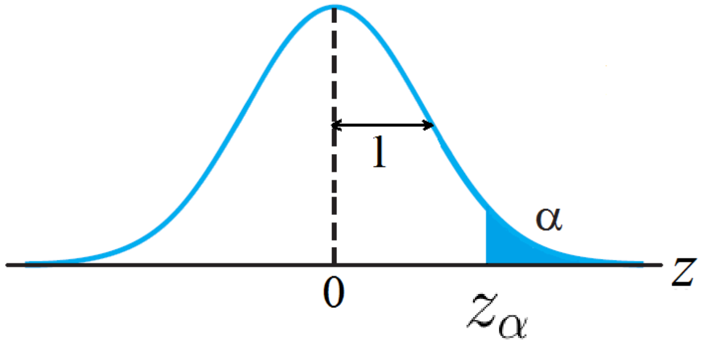
Standard Normal Function

From last week, we understand how observations of means are distributed:



The distribution of $\bar{x}$ values with large n.
(mean = μ, std.dev. = $\sigma/\sqrt{n}$)

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$



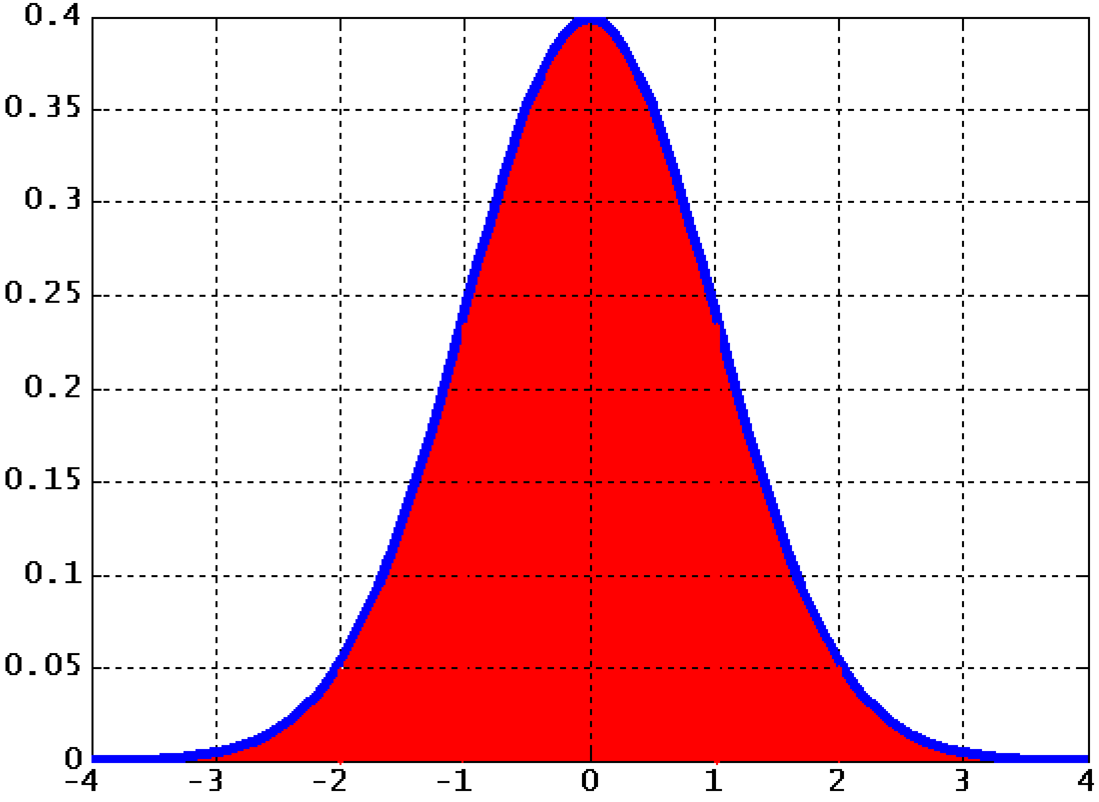
The z transformed distribution.
(mean = 0, std.dev. = 1)

$$\text{Shaded Area} = \alpha = P(z > z_\alpha) = \int_{z_\alpha}^{\infty} f(z)dz$$
$$= \text{ROOT.Math.normal\_cdf}(\infty) - \text{ROOT.Math.normal\_cdf}(z_\alpha)$$
$$= 1 - \text{ROOT.Math.normal\_cdf}(z_\alpha)$$

Total area under the standard normal curve is 1.

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$$\int_{-\infty}^{\infty} f(z)dz = 1$$



$$\int_a^b \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \Phi(b) - \Phi(a)$$

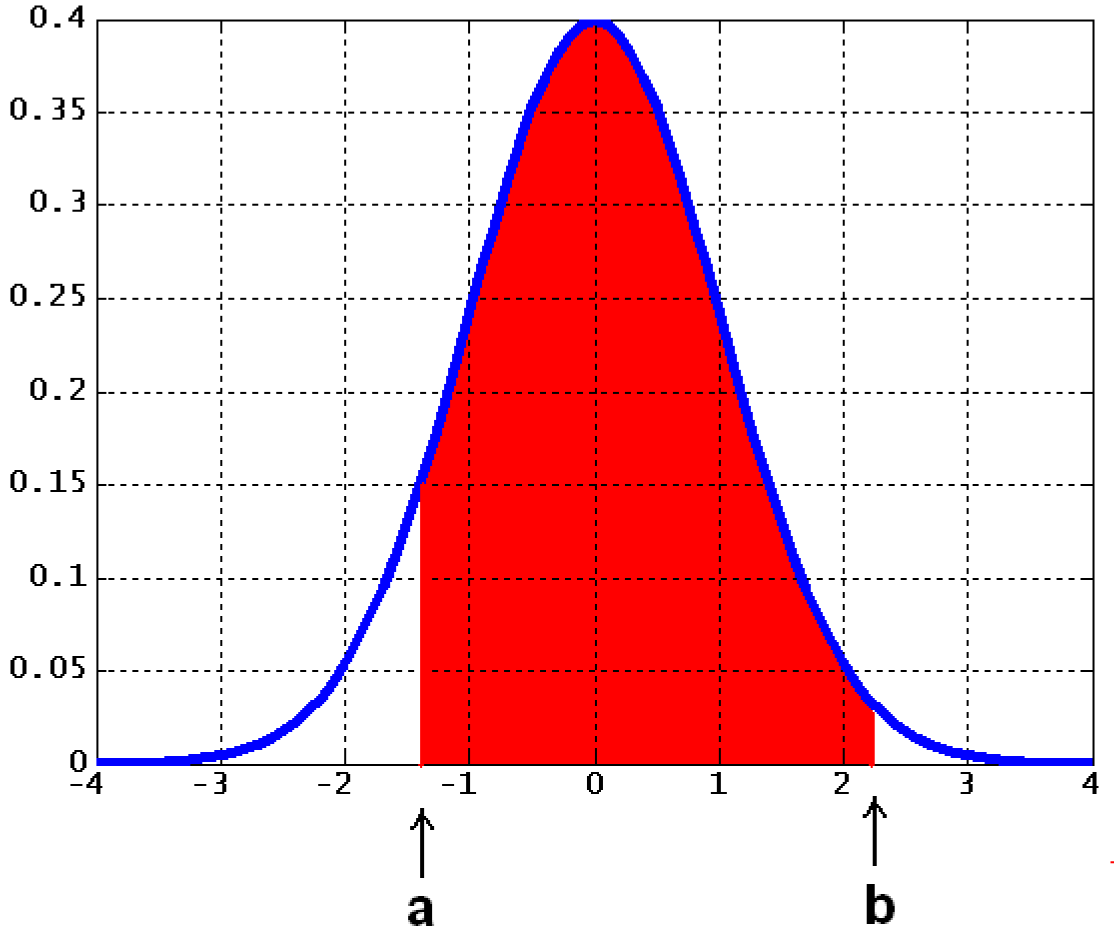
```
Area = ROOT.Math.normal_cdf(b)
     - ROOT.Math.normal_cdf(a)
```

C++

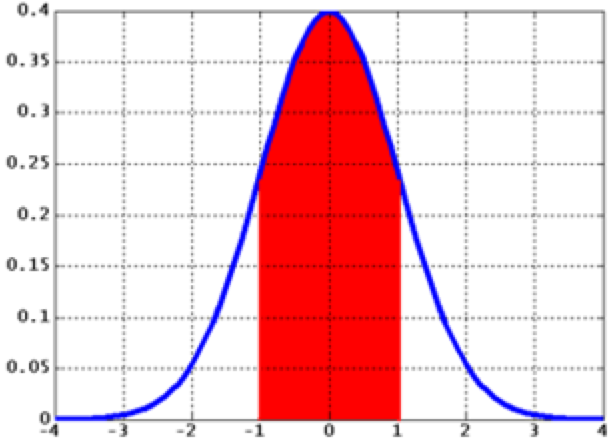
$\Phi(x) =$ **`ROOT::Math::normal_cdf(x)`**

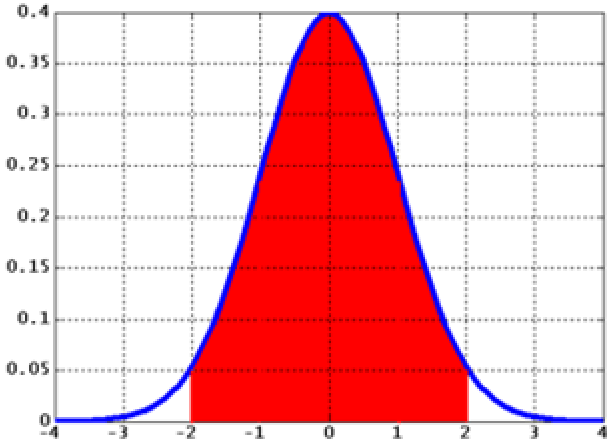
Python

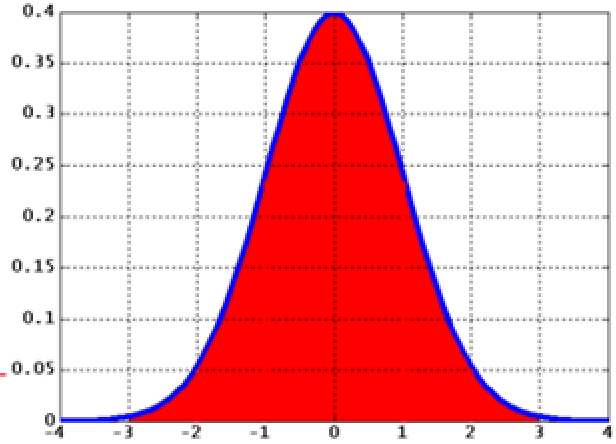
$\Phi(x) =$ **`ROOT.Math.normal_cdf(x)`**



CL	Shaded Area
68.27%	$\int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0.6827$
95.45%	$\int_{-2}^{2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0.9545$
99.73%	$\int_{-3}^{3} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0.9973$







Confidence Interval Estimates of Population Parameters

Confidence Intervals

In particle physics, we examine only a small part of a large population, which is called a **sample**. Any quantity obtained from a sample for the purpose of estimating a population parameter is called a **statistic**.

Let μ_S and σ_S be the mean and standard error of the sampling distribution of statistic S.
Then, if the sampling distribution of S is approximately normal (this is true for many statistics for sample size $n \geq 30$), we can expect to find S lying in the intervals:

- $\mu_S - \sigma_S$ to $\mu_S + \sigma_S$ about 68.27% of time
- $\mu_S - 2\sigma_S$ to $\mu_S + 2\sigma_S$ about 95.45% of time
- $\mu_S - 3\sigma_S$ to $\mu_S + 3\sigma_S$ about 99.73% of time

Confidence Interval Estimates of Population Parameters

- Numbers (1, 2, 3) are called 68.27%, 95.45%, and 99.73% *confidence limits.* *(The percentage confidence is often called the confidence level)*
- Similarly, $\mu_S \pm 1.96\sigma_S$ and $\mu_S \pm 2.58\sigma_S$ are 95% and 99% *confidence levels*.
- The numbers 1, 1.96, 2, 2.58, 3 etc., in the confidence limits are *critical values* denoted by z_c.

Confidence levels used in practice:

Confidence Level	99.73%	99%	98%	96%	95.45%	95%	90%	80%	68.27%	50%
z_c	3.00	2.58	2.33	2.05	2.00	1.96	1.645	1.28	1.00	0.6745

Example

$$\frac{1}{\sqrt{2\pi}}\int_{-z_c}^{+z_c} e^{-z^2/2}dz = \Phi(z_c) - \Phi(-z_c)$$

```python
# CL vs critical values
import ROOT

z_c = 0.5
while z_c < 3.5:
  conf_level = ROOT.Math.normal_cdf(+z_c) \
             - ROOT.Math.normal_cdf(-z_c)
  print ("CL = %.4f  zc = %.1f" %(conf_level, z_c))
  z_c += 0.1
```

OUTPUT

```
CL = 0.3829   zc = 0.5
CL = 0.4515   zc = 0.6
CL = 0.5161   zc = 0.7
CL = 0.5763   zc = 0.8
CL = 0.6319   zc = 0.9
CL = 0.6827   zc = 1.0
CL = 0.7287   zc = 1.1
CL = 0.7699   zc = 1.2
CL = 0.8064   zc = 1.3
CL = 0.8385   zc = 1.4
CL = 0.8664   zc = 1.5
CL = 0.8904   zc = 1.6
CL = 0.9109   zc = 1.7
CL = 0.9281   zc = 1.8
CL = 0.9426   zc = 1.9
CL = 0.9545   zc = 2.0
CL = 0.9643   zc = 2.1
CL = 0.9722   zc = 2.2
CL = 0.9786   zc = 2.3
CL = 0.9836   zc = 2.4
CL = 0.9876   zc = 2.5
CL = 0.9907   zc = 2.6
CL = 0.9931   zc = 2.7
CL = 0.9949   zc = 2.8
CL = 0.9963   zc = 2.9
CL = 0.9973   zc = 3.0
CL = 0.9981   zc = 3.1
CL = 0.9986   zc = 3.2
CL = 0.9990   zc = 3.3
CL = 0.9993   zc = 3.4
```

Confidence Intervals for Means

- Standard error of means: $e = s/\sqrt{n}$ where s is the sample standard deviation.
- Remember for large values of n, we have $s = \sigma$. We'll use σ.
- The confidence limits for the population mean are given by:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}}$$

sample mean

critical value

number of samples

sample standard deviation

Therefore, we can be confident of finding the population mean μ in the intervals:

$$\bar{x} - z_c \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_c \frac{\sigma}{\sqrt{n}}$$

Confidence Intervals for Sum & Diffrences

For large n values, the condidence limits of sum and differences of two samples are given by:

$$\bar{x}_1 + \bar{x}_2 \pm z_c \sigma_{\bar{x}_1 + \bar{x}_2} = \bar{x}_1 + \bar{x}_2 \pm z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\bar{x}_1 - \bar{x}_2 \pm z_c \sigma_{\bar{x}_1 - \bar{x}_2} = \bar{x}_1 - \bar{x}_2 \pm z_c \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Example

In an experiment, a detector measures the energy of a particle produced in high-energy collisions. A total of 64 independent events are recorded. The measurements yield:

Sample mean: $\bar{x} = 91.25\ \text{GeV}$ and Sample std: $\sigma = 2.40\ \text{GeV}$

Construct a 90%, 95% 99% confidence intervals for the true mean energy.

Solution

Standard error: $e = \frac{\sigma}{\sqrt{n}} = \frac{2.4}{\sqrt{64}} = 0.3\ \text{GeV}$ and $\bar{x} - z_c \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_c \frac{\sigma}{\sqrt{n}}$

$z_c = 1.645$ for 90% confidence interval:

$91.25 - 1.654(0.3) \leq \mu \leq 91.25 + 1.645(0.3)$

$90.76 \leq \mu \leq 91.74$

$\mu = 91.25 \pm 0.49\ \text{GeV}$

z_c	CL	Interval	Alternative notation
1.645	90%	$90.76 \leq \mu \leq 91.74$	$91.25 \pm 0.49\ \text{GeV}$
1.96	95%	$90.66 \leq \mu \leq 91.84$	$91.25 \pm 0.59\ \text{GeV}$
2.58	99%	$90.48 \leq \mu \leq 92.02$	$91.25 \pm 0.77\ \text{GeV}$

Example

In a test beam study at the CERN, a detector is used to measure the time-of-flight of a fast particle traveling through a known distance. Previous calibration studies indicate that the standard deviation of the time measurements $\sigma = 50\,\mathrm{ps}$. A physicist wants to estimate the mean time-of-flight with a maximum statistical error of $10\,\mathrm{ps}$. How large a sample size n is required in order to be:

(a) 95% confident (b) 99% confident.

Solution

(a) The error of the estimate being $e = 1.96\frac{\sigma}{\sqrt{n}}$ or $\sqrt{n} = \frac{1.96\sigma}{e} = \frac{1.96(50)}{(10)} \rightarrow n = 97$

we can be 95% confident that the error in the estimate will be less than 10 ps if n is 97 or larger.

(b) Similarly, we can be 99% confident that the error will be less than 10 ps if $n \geq 167$.

Example

In an experiment, two different detector systems (Detector A and Detector B) are used to measure the energy of the same type of particle produced in collisions.

The results are:

Detector A

Sample size $n_1 = 180$

Mean energy: $\overline{x_1} = 52.4 \text{ GeV}$

Standard deviation: $\sigma_1 = 6.5 \text{ GeV}$

Detector B

Sample size $n_2 = 220$

Mean energy: $\overline{x_2} = 49.8 \text{ GeV}$

Standard deviation: $\sigma_2 = 5.2 \text{ GeV}$

Find the confidence intervals for the difference of the population means at 95% and 99% CI.

Ans:

$\text{CI}(95\%) = 2.60 \pm 1.17 \text{ GeV}$

$\text{CI}(99\%) = 2.60 \pm 1.54 \text{ GeV}$

Hypothesis Tests

Statistical Hypotheses

Very often in practice we make decisions about populations on the basis of sample information.
To reach a decision, it is useful to make assumptions (**hypotheses**) about the population involved.

A **null hypothesis** (H_0) is a fundamental statistical statement assuming there is no effect.

Any hypothesis that differs from a given null hypothesis is called an **alternative hypothesis** (H_1).

- H_0 : probability of getting head is $p = 0.5$
- H_1 : $p = 0.7$ or $p \neq 0.5$, or $p > 0.5$

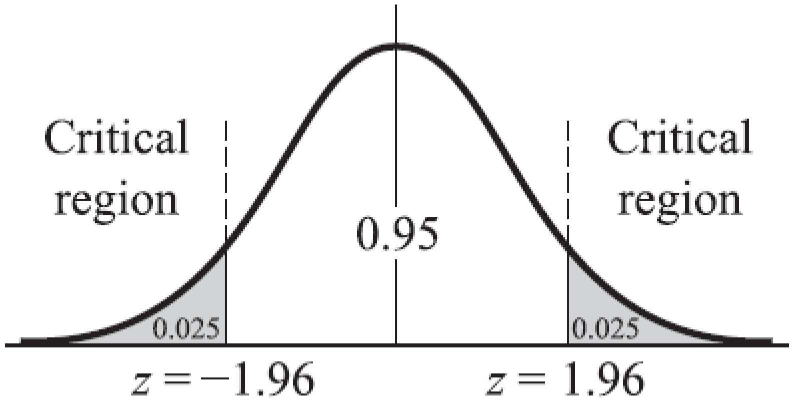
- Procedures that enable us to decide whether to accept or reject hypotheses are called **tests of hypotheses**.
- If we reject a hypothesis when it happens to be true, we say that a **Type I error** has been made. If we accept a hypothesis when it should be rejected, we say that a **Type II error** has been made. In either case a wrong decision or error in judgment has occurred.

Two-Tailed Test

In testing a given hypothesis, the maximum probability with which we would be willing to risk a Type I error is called the **level of significance** of the test denoted by α.

In practice, people use $\alpha = 0.10$, $\alpha = 0.05$ or $\alpha = 0.01$.

Suppose that under a given hypothesis the sampling distribution of a statistic S is a normal distribution with mean $\bar{x}$ and standard deviation s.
Also, suppose we decide to reject the hypothesis if S is either too small or too large.

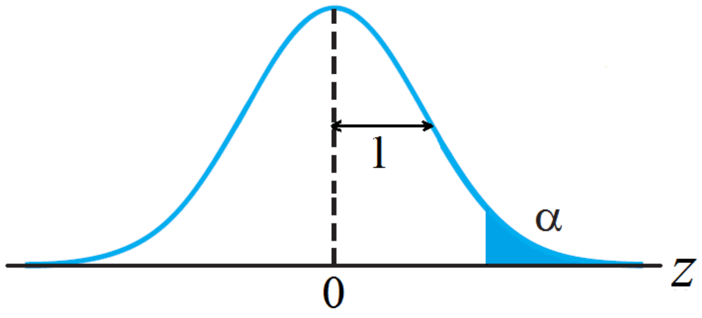


As indicated in the figure, we can be 95% confident that, if the hypothesis is true, the z score of an actual sample statistic S will lie between −1.96 and 1.96.

- Reject the hypothesis at a 0.05 level of significance if the z score of the statistic S lies <u>outside</u> the range −1.96 to 1.96
- Accept the hypothesis otherwise.

One-Tailed Test

We may be interested only in extreme values to one side of the mean. In such cases the critical region is a region to one side of the distribution, with area equal to the level of significance, α.



Critical values of z for both one-tailed and two-tailed tests at various levels of significance,

Level of Significance α	0.10	0.05	0.01	0.005	0.002
Critical Values of z for One-Tailed Tests	-1.28 *or* 1.28	-1.645 *or* 1.645	-2.33 *or* 2.33	-2.58 *or* 2.58	-2.88 *or* 2.88
Critical Values of z for Two-Tailed Tests	-1.645 *and* 1.645	-1.96 *and* 1.96	-2.58 *and* 2.58	-2.81 *and* 2.81	-3.08 *and* 3.08

Example

```python
# Critical values of z for One-Tailed Tests
import ROOT

# practical level of significance
los = [0.10, 0.05, 0.01, 0.005, 0.002, 0.001]

print ("alpha   1-alpha   critical value")
for alpha in los:
  zc = ROOT.Math.normal_quantile(1-alpha,1.0)
  print ("%.3f   %.3f     %.3f" % (alpha, 1-alpha,zc))
```

OUTPUT

```
alpha    1-alpha    critical value
0.100    0.900      1.282
0.050    0.950      1.645
0.010    0.990      2.326
0.005    0.995      2.576
0.002    0.998      2.878
0.001    0.999      3.090
```

```python
# normal_quantile() is the inverse function of normal_cdf()
# The following call returns 1.645 which is critical value for
# standard normal curve corresponding to 0.95 level of significance
sigma = 1.0
x = ROOT.Math.normal_quantile(0.95, sigma)
```

p Value Approach

Consider a null hypothesis for the mean of a population;

$$H_0 : \mu = \mu_0$$

We can then form an alternative hypothesis that disagrees with the null hypothesis;

$$H_1 : \mu < \mu_0 \quad \text{or} \quad H_1 : \mu > \mu_0$$

We then determine if our observation (sample mean) $\bar{x}$ is statistically consistent with H_0 or not.
We can ask:

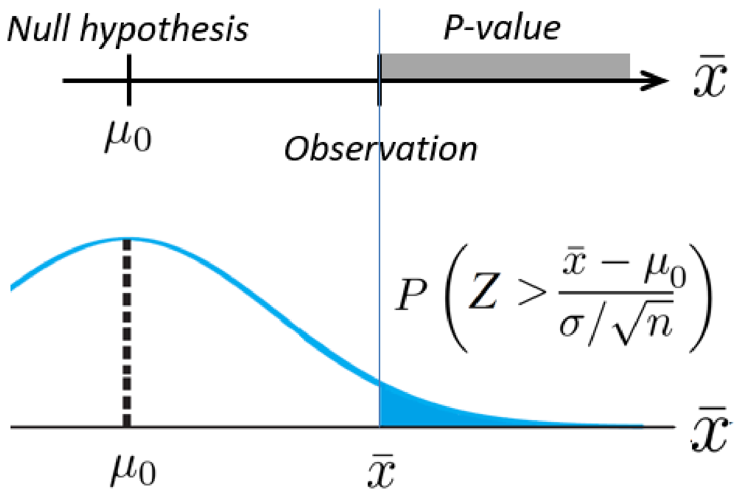
How likely is this observation is more extreme, in the direction away from the null hypothesis?
This is called the *p*-value.

p Value

The p-value is the observed statistical significance of an experiment.

An observation that is statistically "far" from the null hypothesis will give a small p-value.

- If the p-value is small, then H_0 is rejected in favor of the H_1
- If the p-value is large, then we do not reject H_0.
  (this does not mean that H_0 is true)



Procedure for Hypothesis Testing

1. Form the null hypothesis H_0 , and the alternative hypothesis H_1.
The choices of the null and alternative hypotheses will be made by scientist.

Example $H_0\colon \mu = 8\ \mathrm{mm}$ $H_1\colon \mu > 8\ \mathrm{mm}$

2. Make relevant observations and form the *p*-value:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

p-value = $P\left(Z > \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)$

3. Based on the *p*-value, determine **how strongly** the alternative hypothesis is suggested; and reject, or not, the null hypothesis.

P-value ≤ 5% “strongly suggests”;
P-value ≤ 1% “very strongly suggests”.

Example

The mass of an unstable particle is normally distributed with a mean of 5.0 GeV and a standard deviation of 0.12 GeV due to detector effects. A student selects a sample containing 100 particles from a large set and finds the mean of the sample distribution to be 5.027 GeV. Is the assumption that the mass of the particle is 5 GeV correct? Use one–tailed test at a 95% confidence level.

Solution

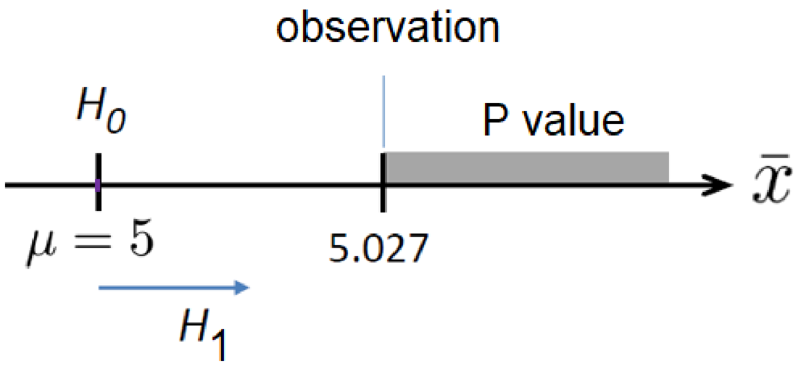
Null Hypothesis $H_0 : \mu = 5 \text{ GeV}$

Alternative Hypothesis $H_1 : \mu > 5 \text{ GeV}$

Population std.dev. $\sigma = 0.12 \text{ GeV}$

Sample mean $\bar{x} = 5.027 \text{ GeV}$

Sample size $n = 100$



Standart error $e = \frac{\sigma}{\sqrt{n}} = \frac{0.12}{\sqrt{100}} = 0.012 \text{ GeV}$

Z score $z = \frac{\bar{x} - \mu}{e} = \frac{5.027 - 5.0}{0.012} = 2.25$

p value (one-tailed) $p = P(z > 2.25) = 1 - \Phi(2.25) = 0.0122$

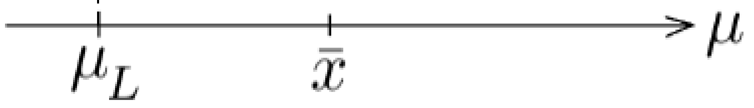
Result:
Since p = 0.0122 < 0.05, we reject H_0.
Discussion:
Given a true mean value of 5 GeV, it is unlikely (but possible) to obtain 5.027 or greater. So, we reject H_0 in favour of H_1.

***What if** you could use two-tailed test ($\mu \neq 5$)?*
P = 0.0244 < 0.05, hence we reject H_0.

Determination of Upper Limits

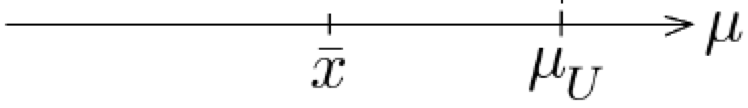
One can determine upper bound of a certain parameter as follows:

$$Area\left(\mu > \bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

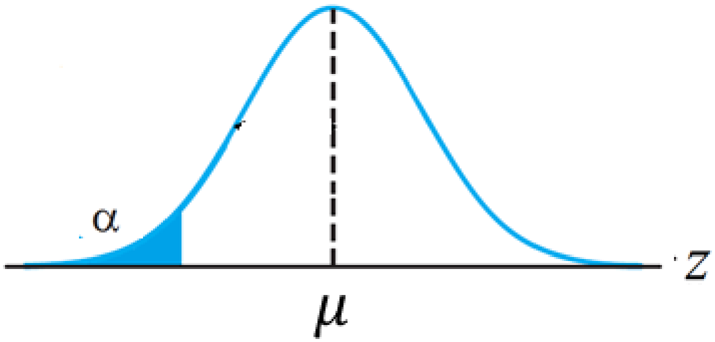


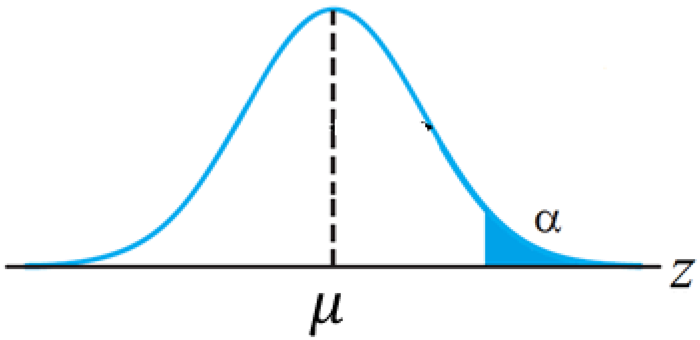
Lower one-sided bound on μ.

$$Area\left(\mu < \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$



Upper one-sided bound on μ.





Summary

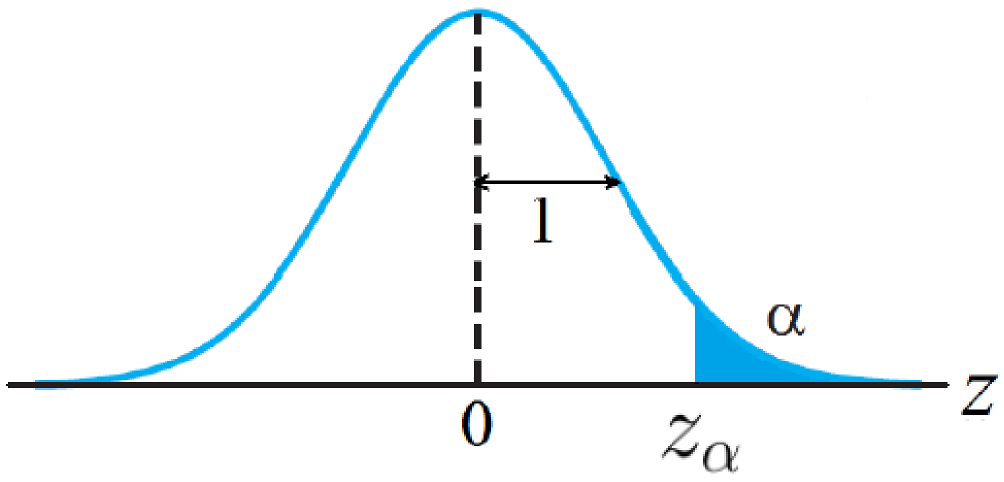
P-value: $$p = P(z > z_\alpha) = \int_{z_\alpha}^{\infty} f(z)dz = \Phi(\infty) - \Phi(z_\alpha) = 1.0 - \Phi(z_\alpha)$$

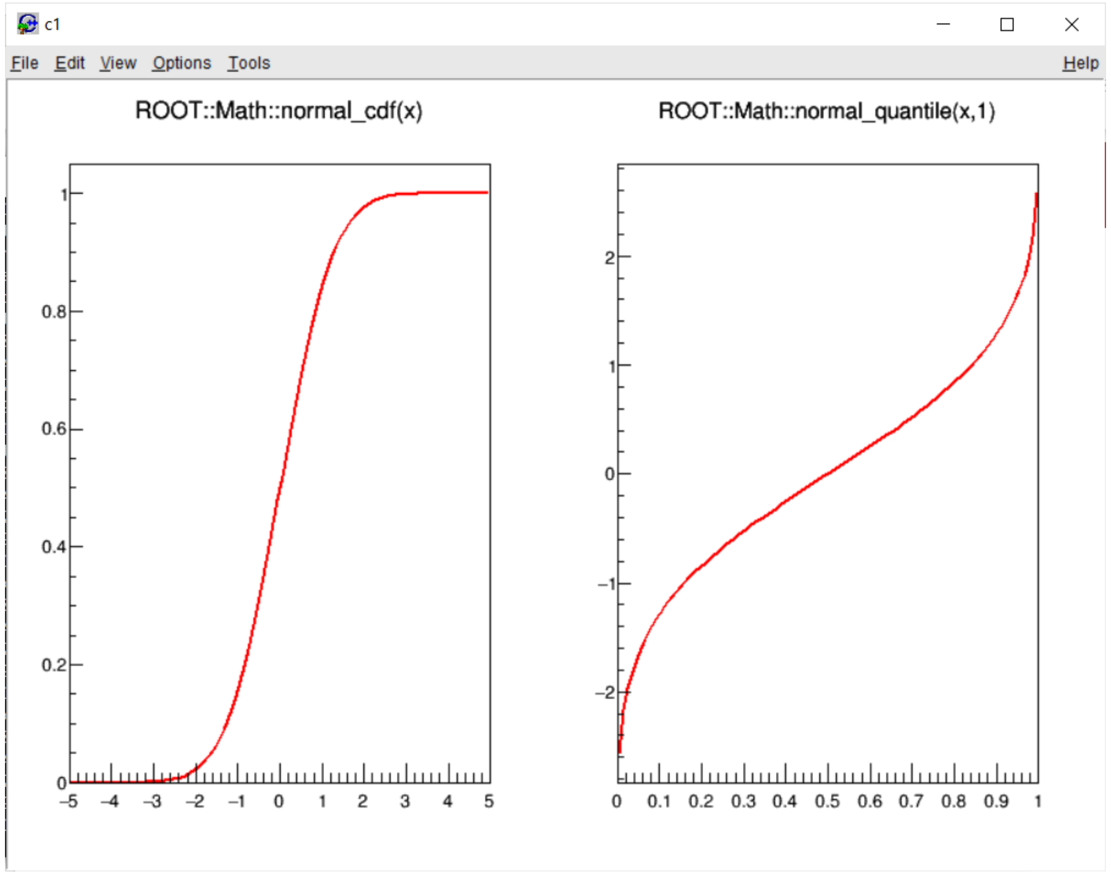
Significance: $$Z = \Phi^{-1}(1-p)$$

$$\Phi(x) = \texttt{ROOT::Math::normal\_cdf(x)}$$

$$\Phi^{-1}(x) = \texttt{ROOT::Math::normal\_quantile(x,1)}$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$





Example: Upper Limit of Neutrino Mass

In a neutrino mass measurement experiment, 500 events from $\pi^+ \rightarrow \mu^+ + \nu_\mu$ decays were examined and the neutrino mass-squared distribution is obtained as shown in Figure. We calculate mass-square of the neutrino from the kinematics of the decay using $m_\nu^2 = E_\nu^2 - p_\nu^2 = (E_\pi - E_\mu)^2 - (\mathbf{p}_\pi - \mathbf{p}_\mu)^2$. The spread of the mass is due to the detector resolution. Determine the upper limit of the neutrino mass at the 95% Confidence Level.

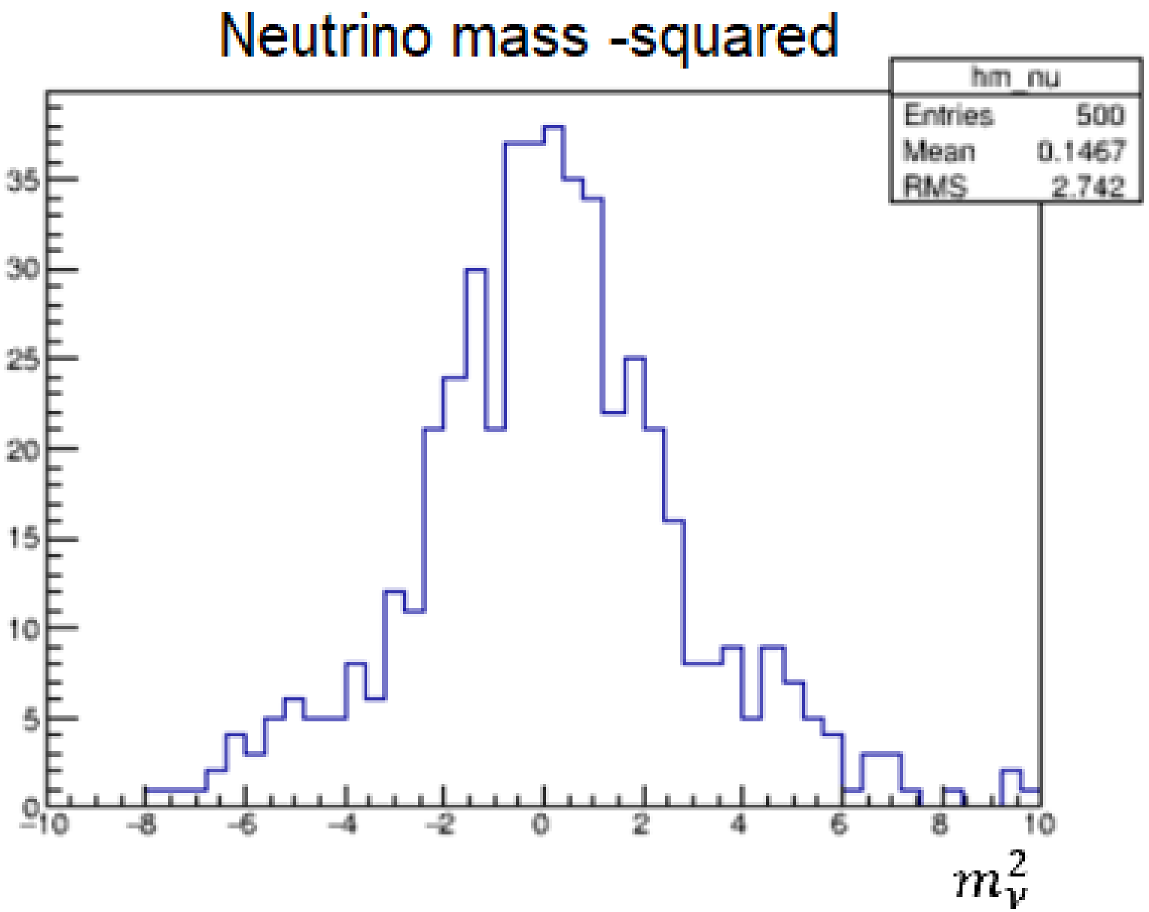
Solution

Upper limit of the mass can be found by:

$$m_{\text{upper}}^2 = \mu + z_c\sigma/\sqrt{n}$$
$$= 0.147 + (1.645)\left(2.742/\sqrt{500}\right)$$
$$= 0.349 \text{ MeV}^2$$
$$m_{\text{upper}} = \sqrt{0.349} = 0.59 \text{ MeV}$$
$$\boldsymbol{m_\nu < 0.59\ \mathbf{MeV}\ \ \mathbf{at\ 95\%\ CL}}$$

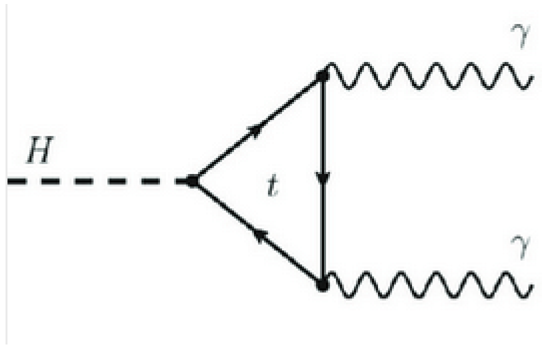
We are 95% confident that the neutrino mass is less than 0.59 MeV.

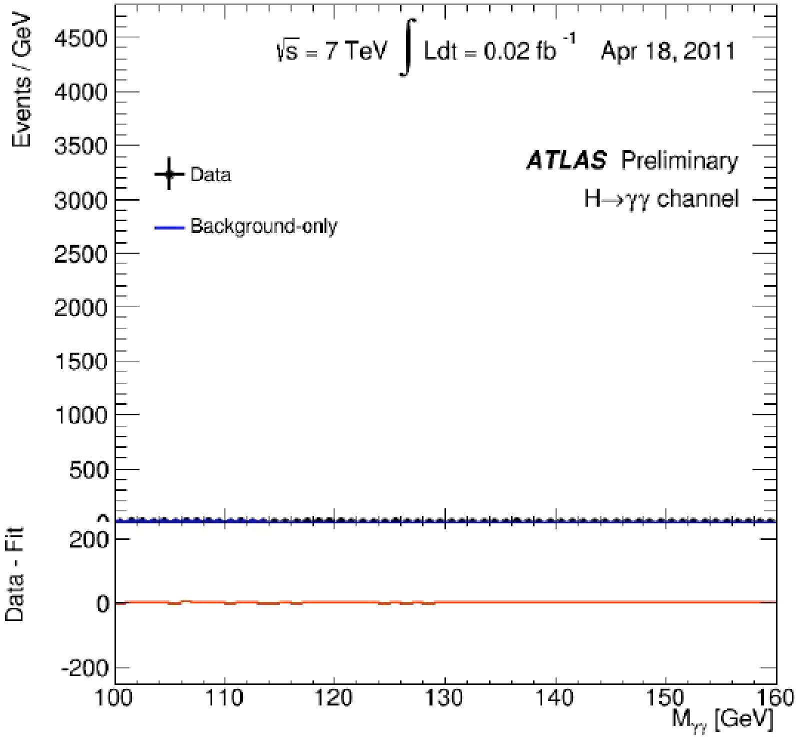
See also Exercise 6.



Example: Discovery of Higgs Boson ($H \rightarrow \gamma\gamma$)

- When did a peak become a discovery? Or
- When did we consider it as incompatible with the background hypothesis (SM without Higgs) ?
- A method can be as follow. Estimate
  N_B (# of background) and
  N_D (# number of all data)
  under the peak, then calculate the significance
  $Z = \Phi^{-1}(1-p)$
  `sigma = ROOT.Math.normal_quantile_c(p,1)`
  $p = 2.87 \times 10^{-7}$ corresponds to sigma = 5





Note:

Invariant mass of two photons can be calcuated by:

$$M_{\gamma\gamma}^2 = (E_1 + E_2)^2 - (\mathbf{p}_1 + \mathbf{p}_2)^2 = 2E_1E_2(1 - \cos(\theta))$$

θ is the angle between photons and energies of photons are measured in electromagnetic calorimeter (ECAL)

Optimal test statistic

According to the **Neymann-Pearson lemma** the likelihood ratio of two alternative hypotheses H_1 and H_0 is the best test statistic, a scalar function with the maximum power, i.e. highest probability to reject H_0 if H_1 is true.

$$\lambda(\boldsymbol{x}) = \frac{f(\boldsymbol{x}|H_1)}{f(\boldsymbol{x}|H_0)}$$

Define

H_0 (background only)

H_1 (signal+background) with signal peak

In particle physics we use likelihood ratio (often Q is used as symbol)

$$Q = L_b/L_{sb}$$

or use log-likelihood

$$q = -2\ln Q = -2(L_b^*/L_{sb}^*)$$

where -2 makes it equal to the $\chi 2$ distribution for large counts (remember $\chi 2 = -2L^* + \text{const}$)

Algorithm

1. Define analitical form of the background function (f_b) and signal+background function (f_{sb})
2. For each test mass m_H (say $m_H = [100\ \mathrm{GeV}, 200\ \mathrm{GeV}]$ step 1 GeV)

   Fit f_b to full data and calculate $\chi_b^2 = -2L_b^*$

   Fit f_{sb} to full data and calculate $\chi_{sb}^2 = -2L_{sb}^*$

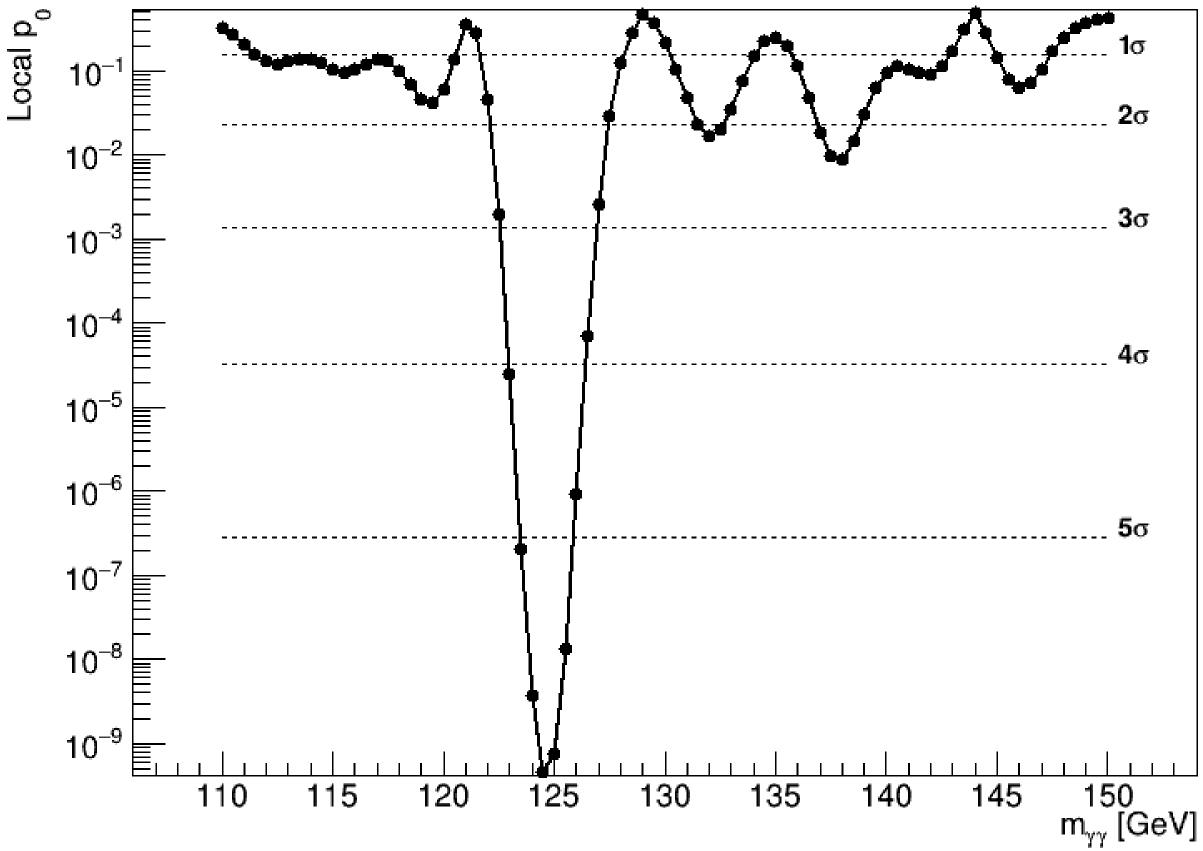
   Calculate test statistic: $q = -2\ln\left(\frac{L_b}{L_{sb}}\right) = -2\ln(L_b) + 2\ln(L_{sb}) = \chi_b^2 - \chi_{sb}^2$

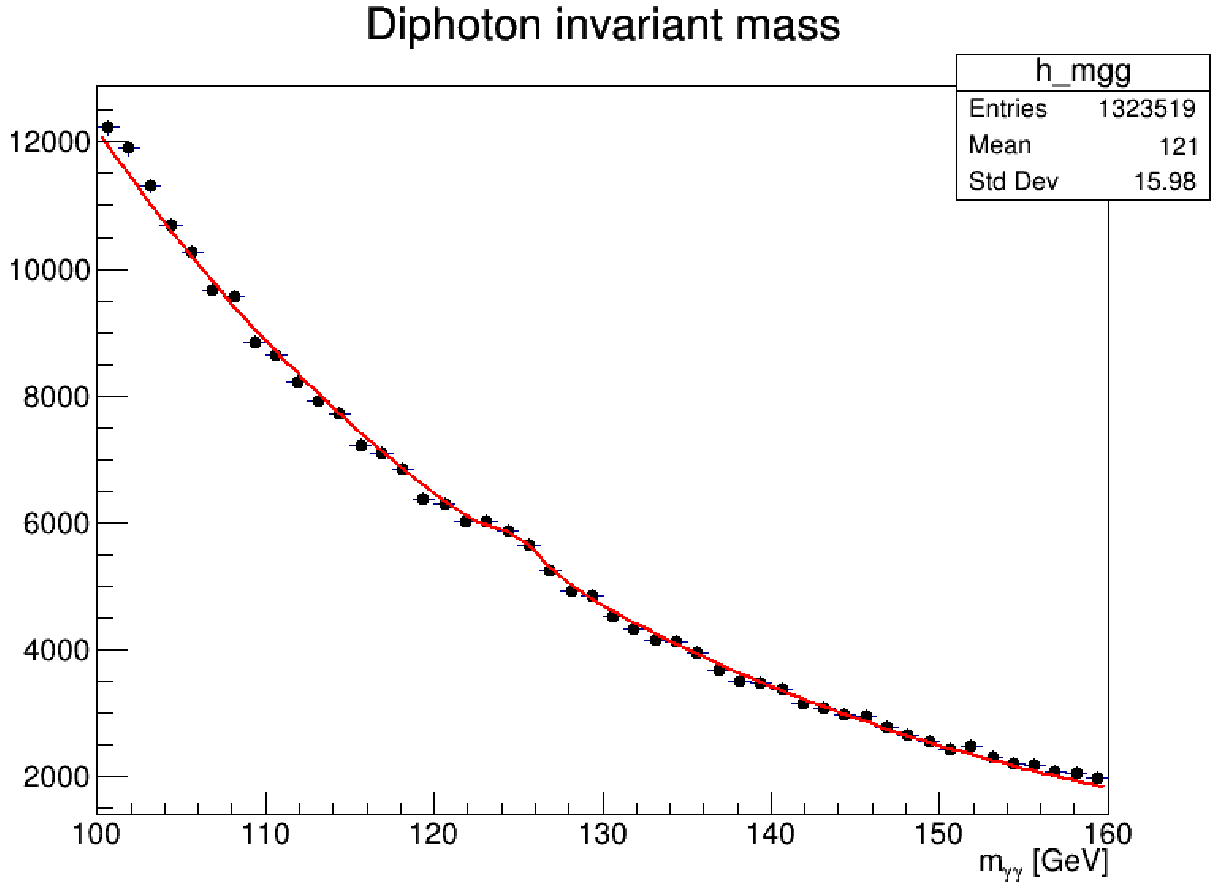
   Calcuate significance: $Z = \sqrt{q}$

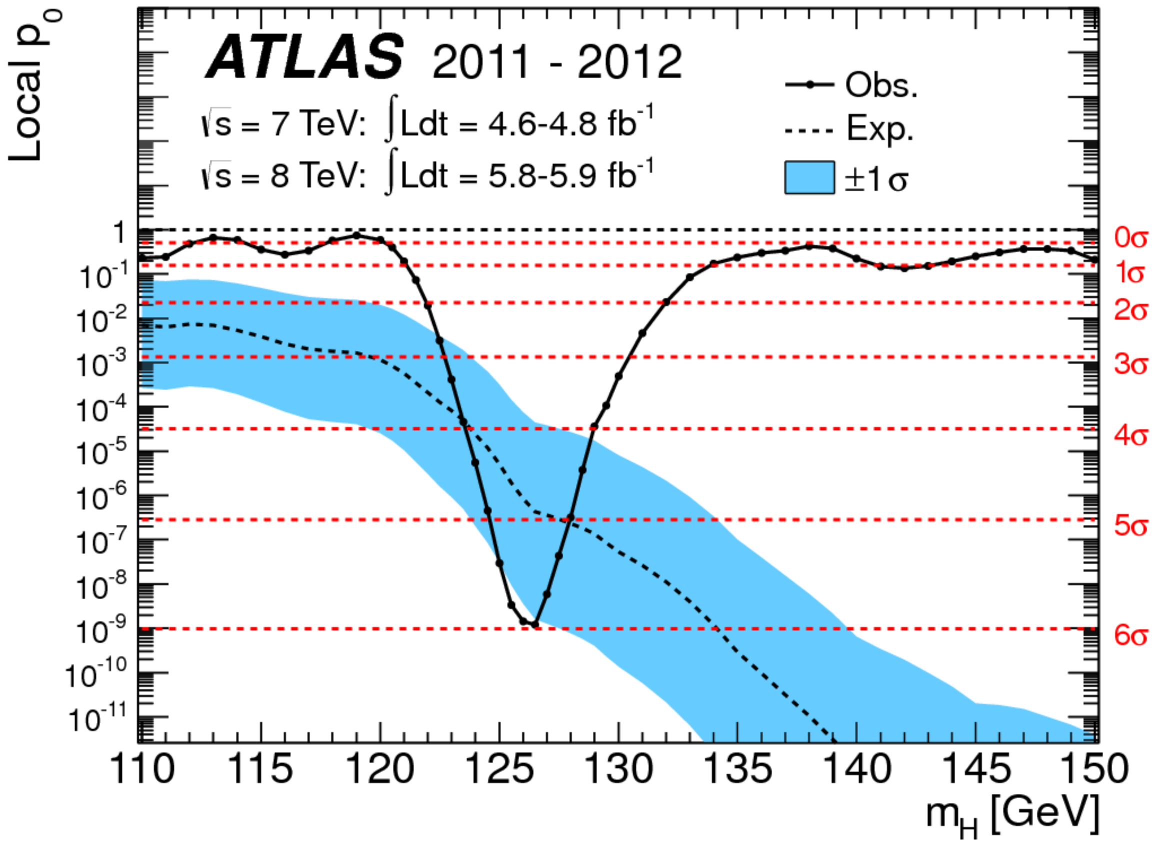
   Calculate p-value: $p = 1 - \Phi(Z)$

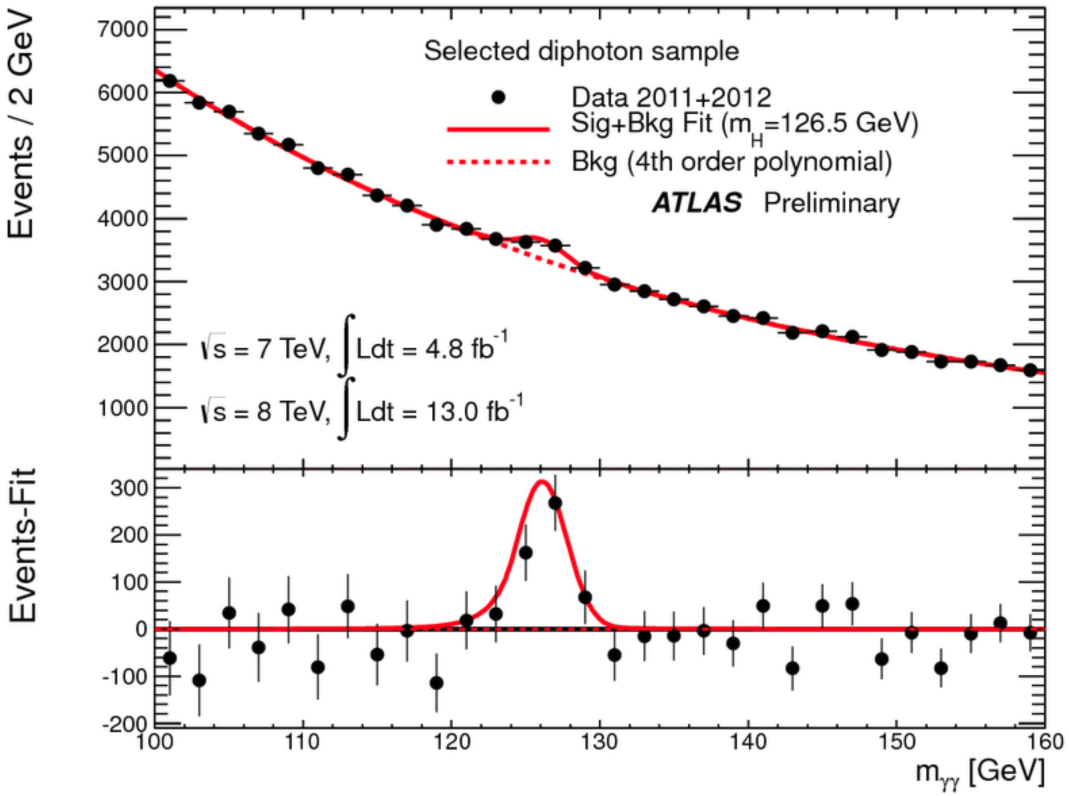
   Output m_H, Z, p
3. Plot m_H vs p or m_H vs Z
4. If $Z \geq 5$, we have discovered a new physics!

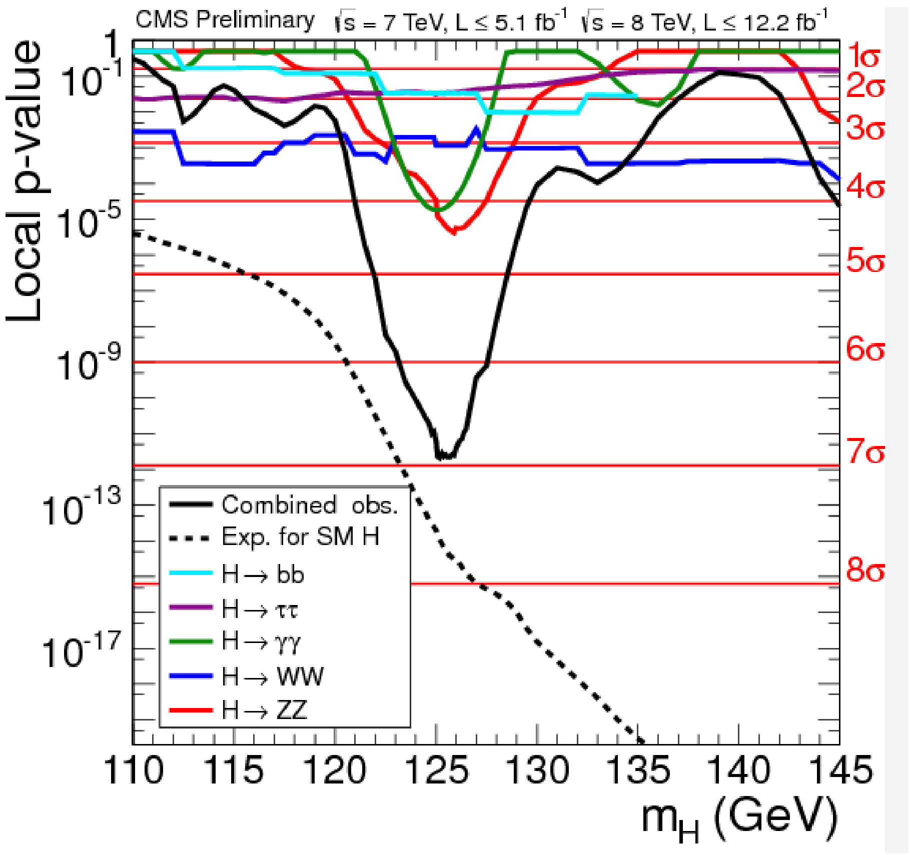
See course web page for the ROOT implementation of this algorithm.

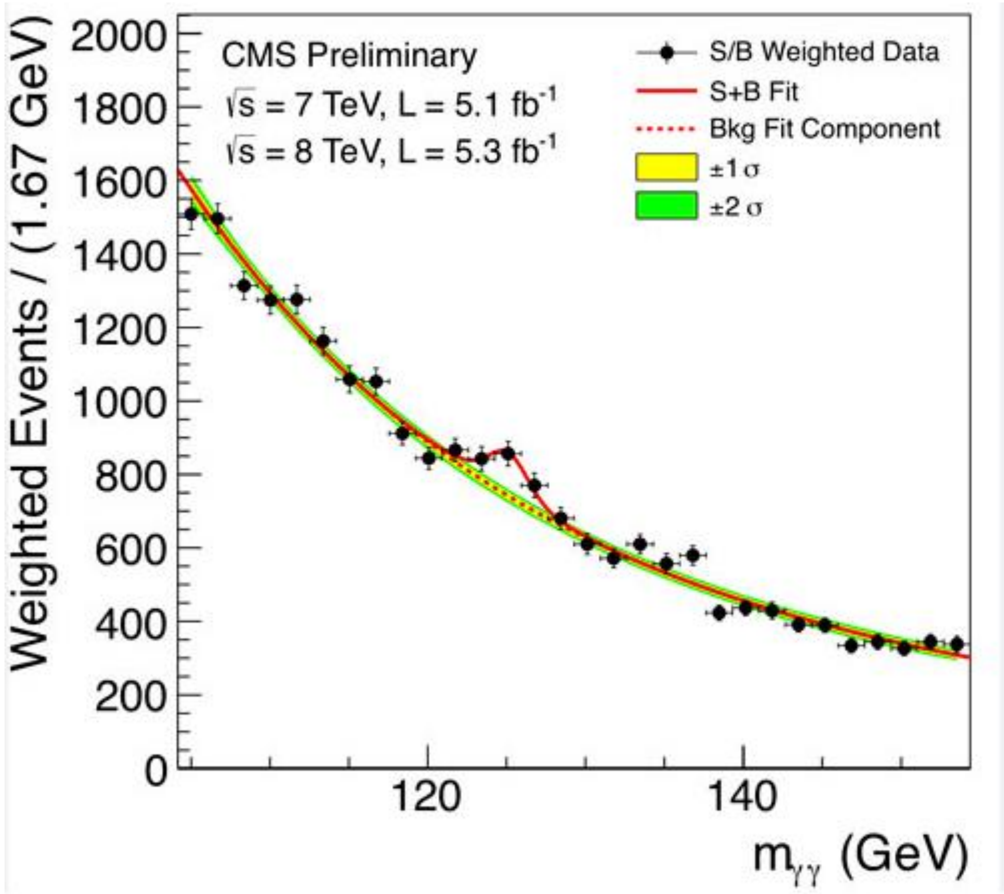

Local p_0
1σ
2σ
3σ
4σ
5σ
$m_{\gamma\gamma}$ [GeV]


Diphoton invariant mass
h_mgg
Entries 1323519
Mean 121
Std Dev 15.98
$m_{\gamma\gamma}$ [GeV]


ATLAS 2011 - 2012
√s = 7 TeV: ∫Ldt = 4.6-4.8 fb-1
√s = 8 TeV: ∫Ldt = 5.8-5.9 fb-1
Obs.
Exp.
±1 σ
Local p0
mH [GeV]
0σ
1σ
2σ
3σ
4σ
5σ
6σ


Selected diphoton sample
Data 2011+2012
Sig+Bkg Fit (mH=126.5 GeV)
Bkg (4th order polynomial)
ATLAS Preliminary
√s = 7 TeV, ∫Ldt = 4.8 fb-1
√s = 8 TeV, ∫Ldt = 13.0 fb-1
Events / 2 GeV
Events-Fit
mγγ [GeV]


CMS Preliminary √s = 7 TeV, L ≤ 5.1 fb-1 √s = 8 TeV, L ≤ 12.2 fb-1
Local p-value
1σ
2σ
3σ
4σ
5σ
6σ
7σ
8σ
Combined obs.
Exp. for SM H
H→ bb
H→ ττ
H→ γγ
H→ WW
H→ ZZ
mH (GeV)


CMS Preliminary
√s = 7 TeV, L = 5.1 fb-1
√s = 8 TeV, L = 5.3 fb-1
S/B Weighted Data
S+B Fit
Bkg Fit Component
±1 σ
±2 σ
Weighted Events / (1.67 GeV)
mγγ (GeV)

Exercises

1. The CO_2 level in the atmosphere is measured by two instruments. The first instrument has a known precision (σ) of 2.6 ppm while the precision of the second is 3.9 ppm. The first instrument takes 20 samples and determines the mean concentration to be 401.3 ppm. The second instrument takes 30 samples and determines the mean concentration to be 400.8 ppm. Calculate a 95% confidence interval for the difference between the two measurements. Comment on the result.

Ans:

We are 95% confident that the diff. between the population means is inside the interval $-1.3 < \bar{x}_1 - \bar{x}_2 < 2.3$

Since zero is contained well inside the limits, it looks like the result is consistent with $\bar{x}_1 = \bar{x}_2$ *the two instruments agree with each other.*

2. The mean lifetime of a sample of 100 fluorescent light bulbs produced by a company is computed to be 1570 hours with a standard deviation of 120 hours. If μ is the mean lifetime of all the bulbs produced by the company, test the hypothesis $\mu = 1600$ hours against the alternative hypothesis $\mu \neq 1600$ hours, using a level of significance of (a) 0.05 and (b) 0.01. (c) Find the P value of the test.

3. Radiation levels in mSv/a are sampled around a nuclear power station.

The result of the 35 sampling in mSv/a is

0.9437	0.7813	0.7836	0.7102	0.9503	0.9171	0.7853
0.6704	0.7660	0.7920	0.8308	0.7825	0.8459	0.8413
0.6934	0.7207	0.7393	0.7603	0.7771	0.8071	0.5334
0.7649	0.9178	0.7100	0.8900	0.8375	0.8034	0.8224
0.6651	0.7984	0.9504	0.8149	0.8097	0.7399	0.8032

Does this data suggest that the radiation levels are signficantly above the normal expected background radiation level of 0.806 mSv/a ? Apply a hypothesis test and calculate p-value.

Ans:

null hypothesis H_0: $\mu = 0.806$ mSv/a

alternative hypothesis H_1: $\mu > 0.806$ mSv/a

p-value = 0.214. It is likely ($p \approx 21\%$) we cannot reject the null hypothesis.

4. Consider the data in the previous exercise.

Form a two-sided 95% confidence interval for the radiation level, and compare it to the expected background radiation level of 0.806 mSv/a.

Ans: $0.764 < \mu < 0.822$

Conclusion:
We are highly confident that the true mean radiation level is in this range which also contains the expected background level.

5. In a particle physics experiment, the lifetime of an unstable particle is studied using its decay events. From a sample of 360 decays, the average measured lifetime is 2.6 ps and associated std. dev. is 0.3 ps.

(a) Determine the 90% confidence interval including the true mean lifetime τ of the particle.

(b) Determine the 90% upper one-sided confidence bound for the true mean lifetime τ of the particle.

(c) Using the relation between lifetime and decay width given by $\tau = \hbar/\Gamma$, calculate the corresponding lower bound on the decay width Γ in part (b).

6. In this exercise, you will do some analysis. Consider the file **`NeutrinoMass.root`** stores the mometum components (in MeV) of the charged particles in the decay $\pi^+ \to \mu^+ + \nu_\mu$.

Tasks:

i. Download the root file: `wget http://www1.gantep.edu.tr/~bingul/ep228/NeutrinoMass.root`

ii. Calculate invariant mass-square of neutrinos from the decay kinematics using **`TLorentzVector`**.
(The spread of the mass-square is due to the detector resolutions).

iii. Fit the mass-square distribution to a Gaussian function and obtain the mean and std.

iv. Using fit parameters, determine the upper limit of the neutrino mass at the 99% Confidence Level.

7. Integral of Gaussian function and Histogram

Tasks:

i. Generate a Gaussian distribution of 5000 entries with mean 0 and standard deviation 1.

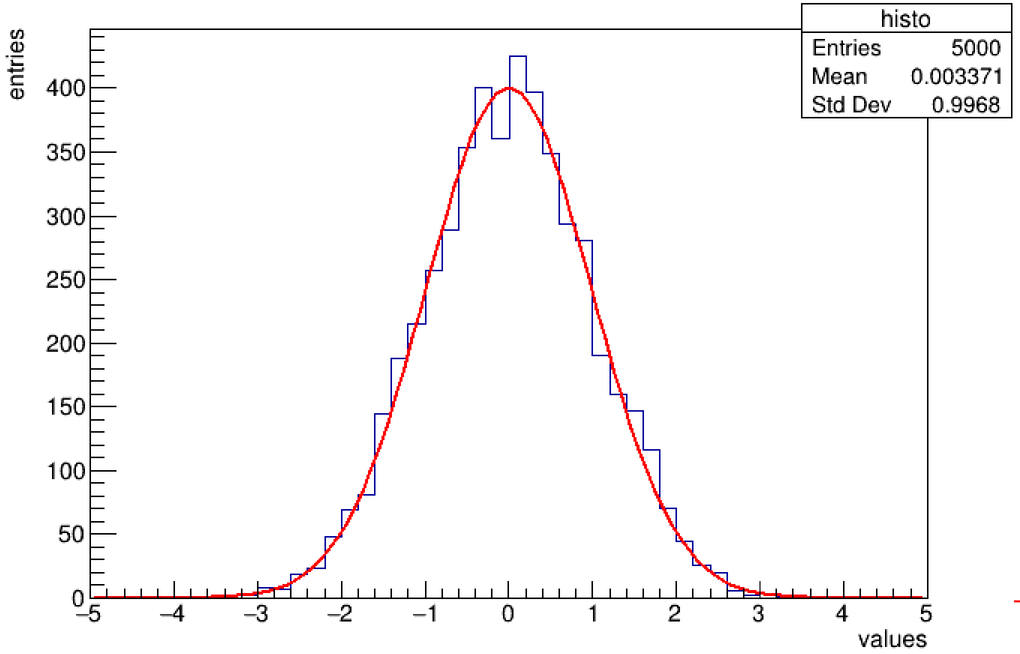
ii. Fit the distribution to Gaussian function

iii. Calculate integral and integral error of the fitted Gaussian function in part ii for the range [-4,4]

iv. Print the covariance matrix of the fit parameters.

v. Investigate the integral error based on the covariance matrix.

vi. Compare integral of the fit function and integral of histogram.



8. Given data:

x = [1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0]

y = [2.7, 3.9, 5.5, 5.8, 6.5, 6.3, 7.7, 8.5, 8.7]

σ = [0.3, 0.5, 0.7, 0.6, 0.4, 0.3, 0.7, 0.8, 0.5]

Fit (a) a linear function (b) a quadratic function to the data. Compare the goodness of the fits.

See also: https://www.pp.rhul.ac.uk/~cowan/ph3010/statistics/2020/ph3010_stat.pdf