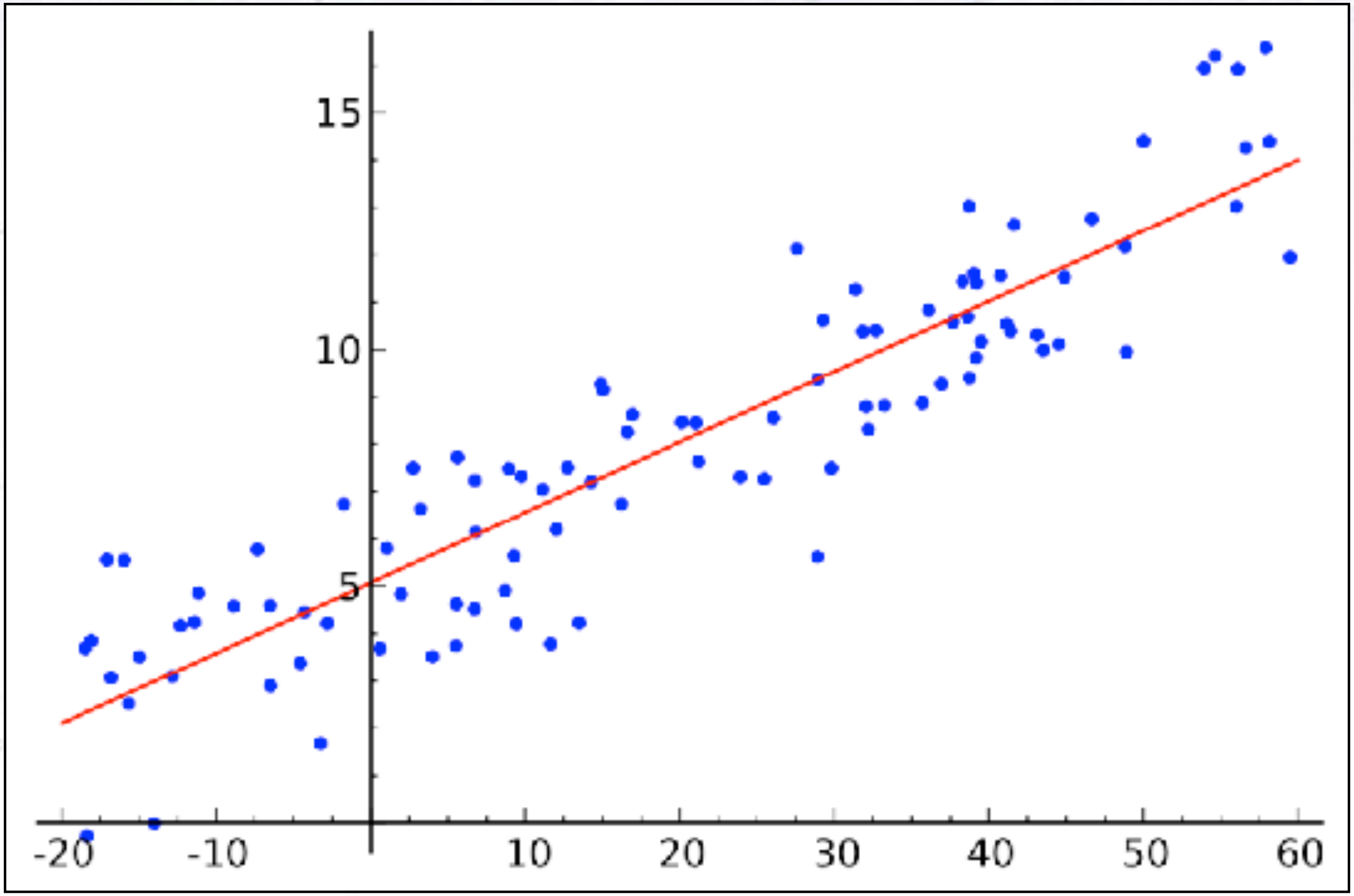
Method of Least Squares

Describe this method formally!!!! for 2024…



Applied Statistics

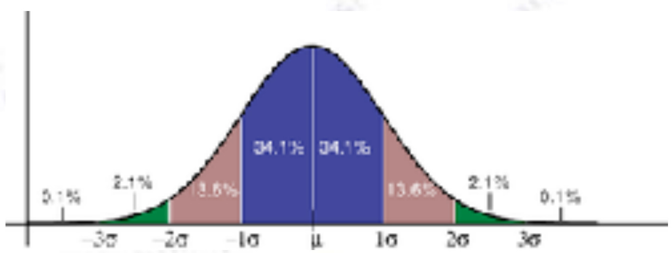
The Chi-Square Distribution, Fit & Test

The Chi-Square fit is also (originally) known as Method of Least Squares,
though this method does not include uncertainties on the data points involved.



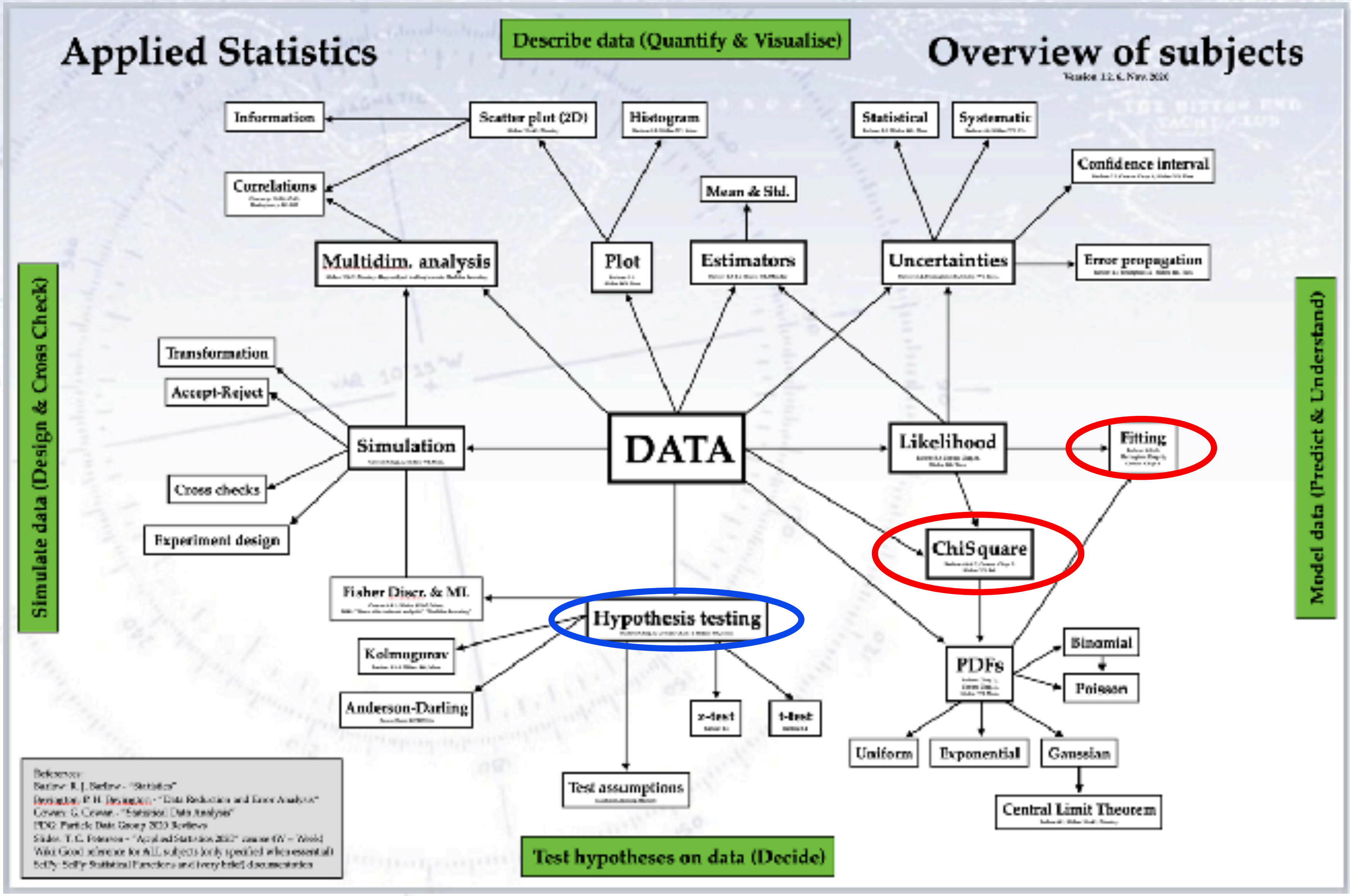



Troels C. Petersen (NBI)



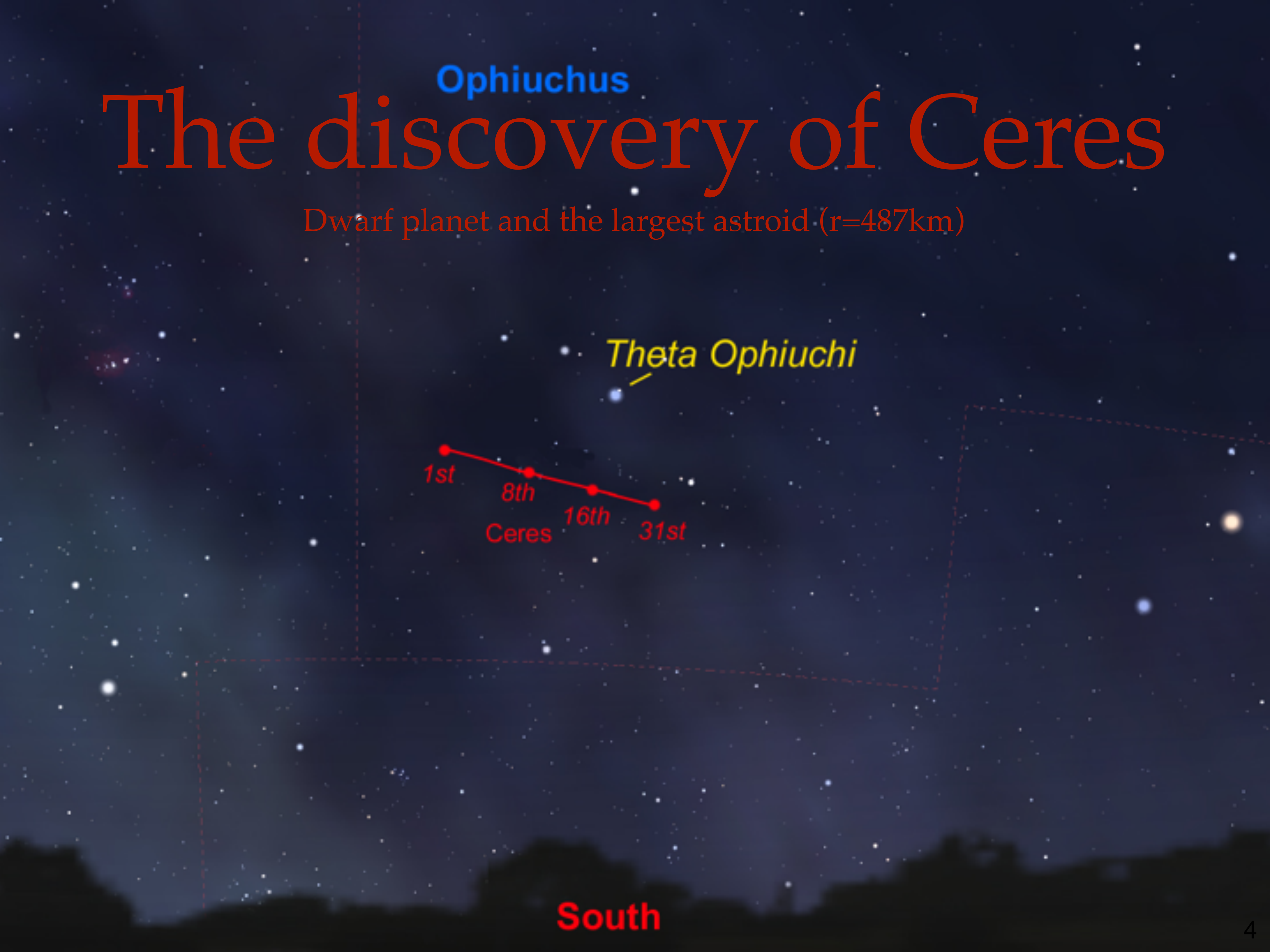
"Statistics is merely a quantisation of common sense"

ChiSquare


Applied Statistics
Describe data (Quantify & Visualise)
Overview of subjects
Information
Scatter plot (2D)
Histogram
Statistical
Systematic
Confidence interval
Correlations
Mean & Std.
Multidim. analysis
Plot
Estimators
Uncertainties
Error propagation
Simulate data (Design & Cross Check)
Model data (Predict & Understand)
Transformation
Accept-Reject
Simulation
DATA
Likelihood
Fitting
Cross checks
Experiment design
ChiSquare
Fisher Discr. & ML
Hypothesis testing
Kolmogorov
Anderson-Darling
z-test
t-test
PDFs
Binomial
Poisson
Uniform
Exponential
Gaussian
Central Limit Theorem
Test assumptions
Test hypotheses on data (Decide)

The discovery of Ceres

Dwarf planet and the largest astroid (r=487km)



The discovery of Ceres

Dwarf planet and the largest astroid (r=487km)

On the 1st of January 1801 Giuseppe Piazzi discovered "new light" and could follow this comet/planet until 11th of February. He published the positions, but due to Ceres being behind the sun, it would be out of sight until the following winter. Following the calculations of a 24 year old mathematician/physicist, it was recovered on the 31st of December 1801 by von Zach and H. Olbers.
The young man's name was Carl Friedrich Gauss, and the method he used/invented for this was…

Ophiuchus

The discovery of Ceres

Dwarf planet and the largest astroid (r=487km)

On the 1st of January 1801 Giuseppe Piazzi discovered “new light” and could follow this comet/planet until 11th of February. He published the positions, but due to Ceres being behind the sun, it would be out of sight until the following winter. Following the calculations of a 24 year old mathematician/physicist, it was recovered on the 31st of December 1801 by von Zach and H. Olbers.

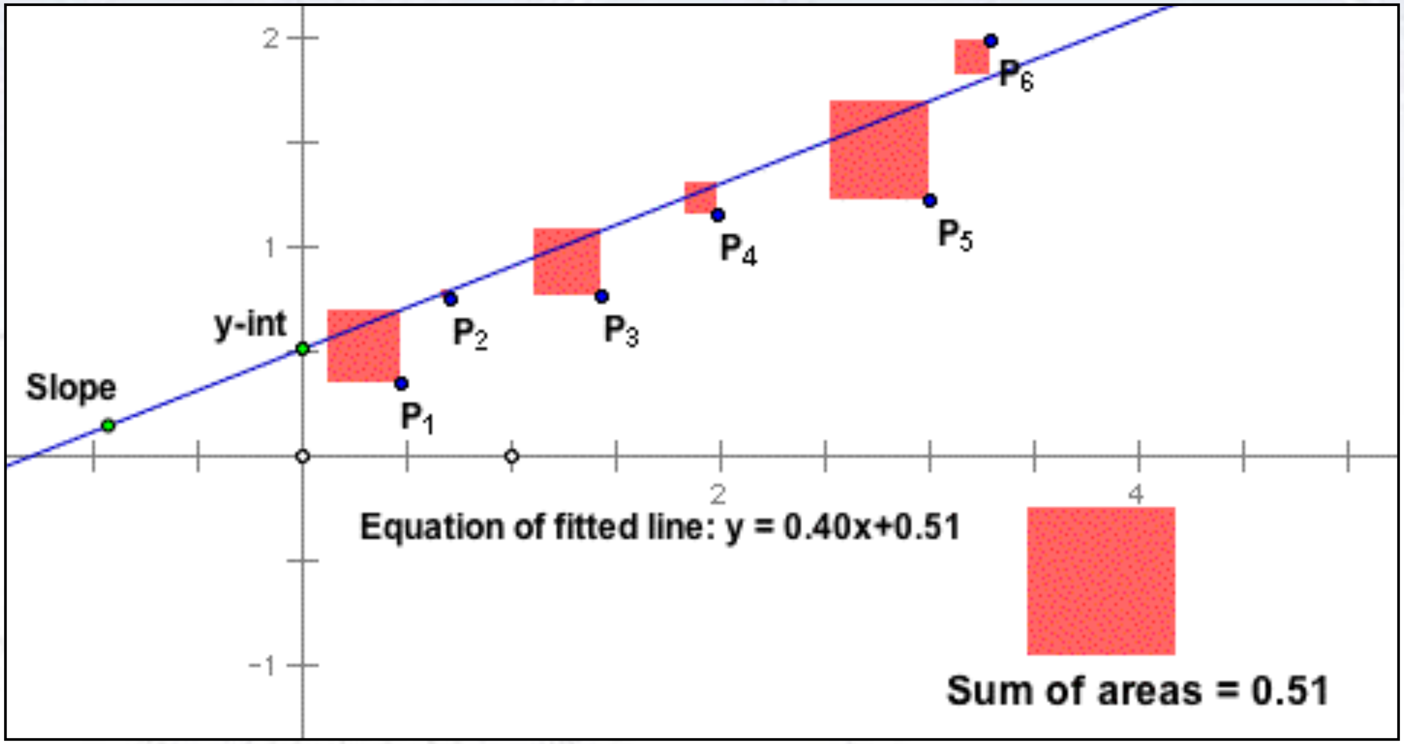
The young man’s name was Carl Friedrich Gauss, and the method he used/invented for this was…

…method of least squares!

South

Method of Least Squares

The problem at hand is determining the curve that best fitted data:

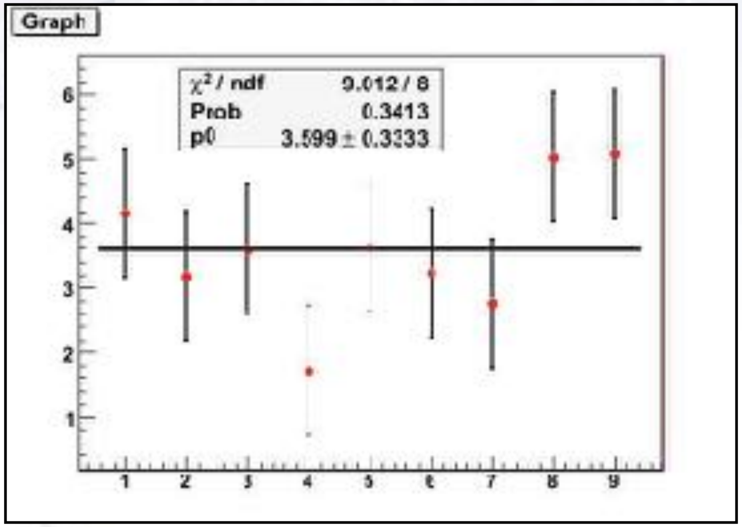


The "best fit" is found by minimising the sum of the squares…

Originally, uncertainties were not included (not "invented" yet!)

Method of Least Squares

The method of least squares is a standard approach to the approximate solution of **overdetermined systems**, i.e. sets of equations in which there are **more equations than unknowns**.

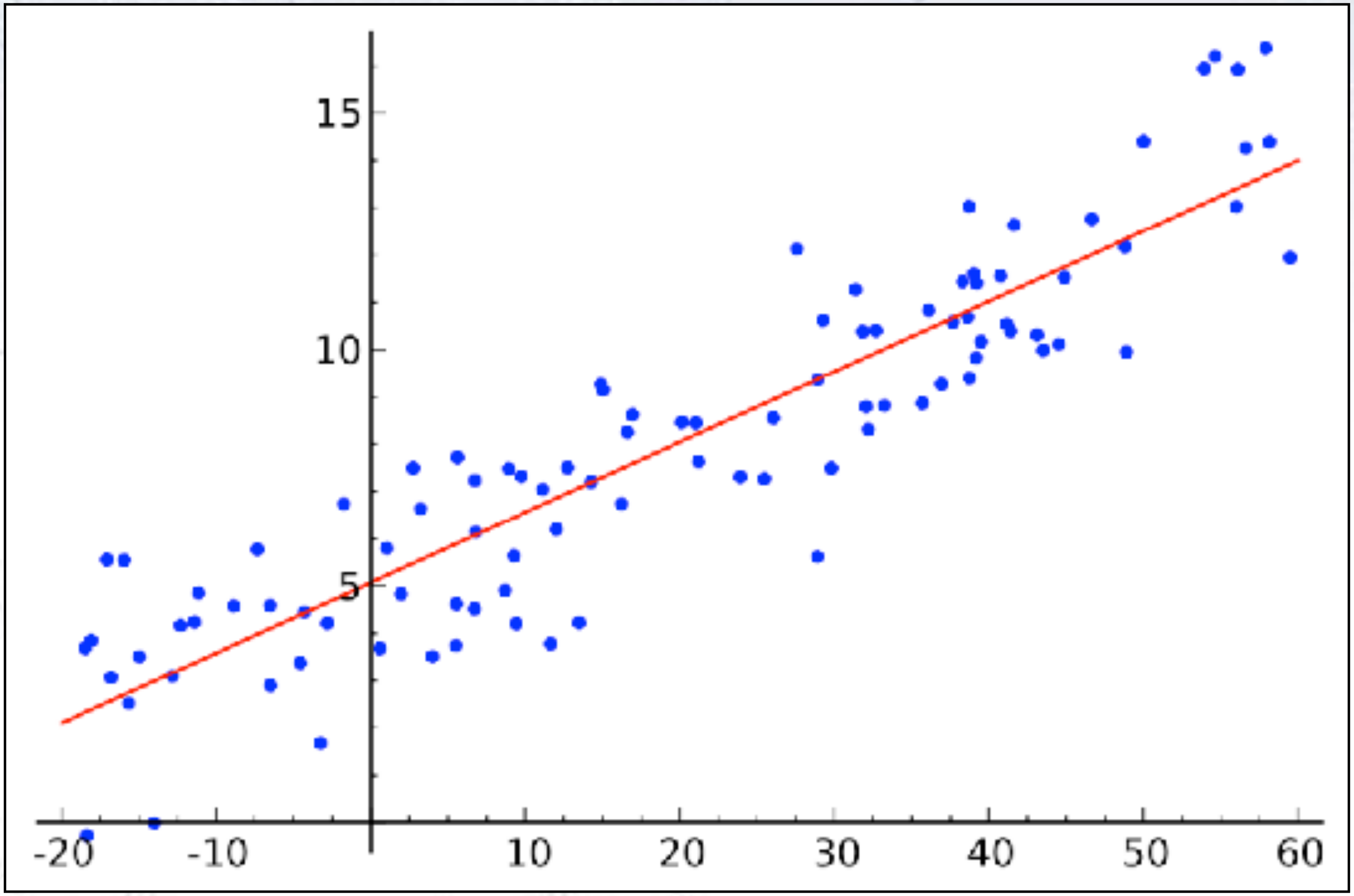


"Least squares" means that the overall solution minimises the **sum of the squares** of the errors made in solving every single equation.

The most important application is in **data fitting**. The best fit in the least-squares sense minimises the sum of squared residuals, a residual being the difference between an observed value and the fitted value provided by a model.

Method of Least Squares

The problem at hand is determining the curve that best fitted data:



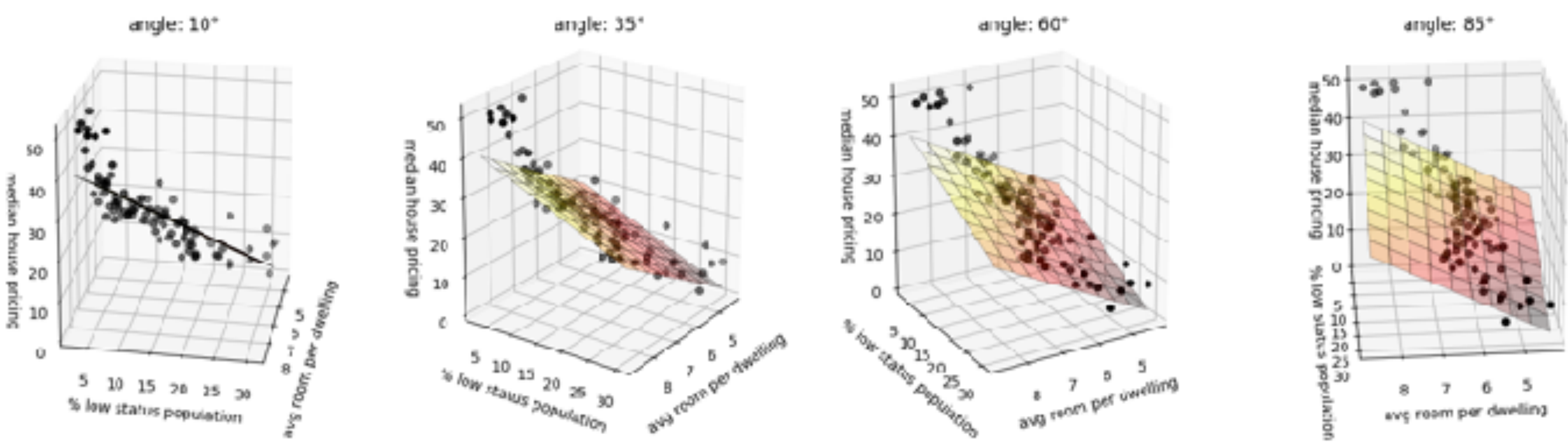
Originally, uncertainties were not included (not “invented” yet!)

MultiDim Linear Regression

The advantage of the linear regression (without uncertainties) is that it can be done “easily” in multiple dimensions.

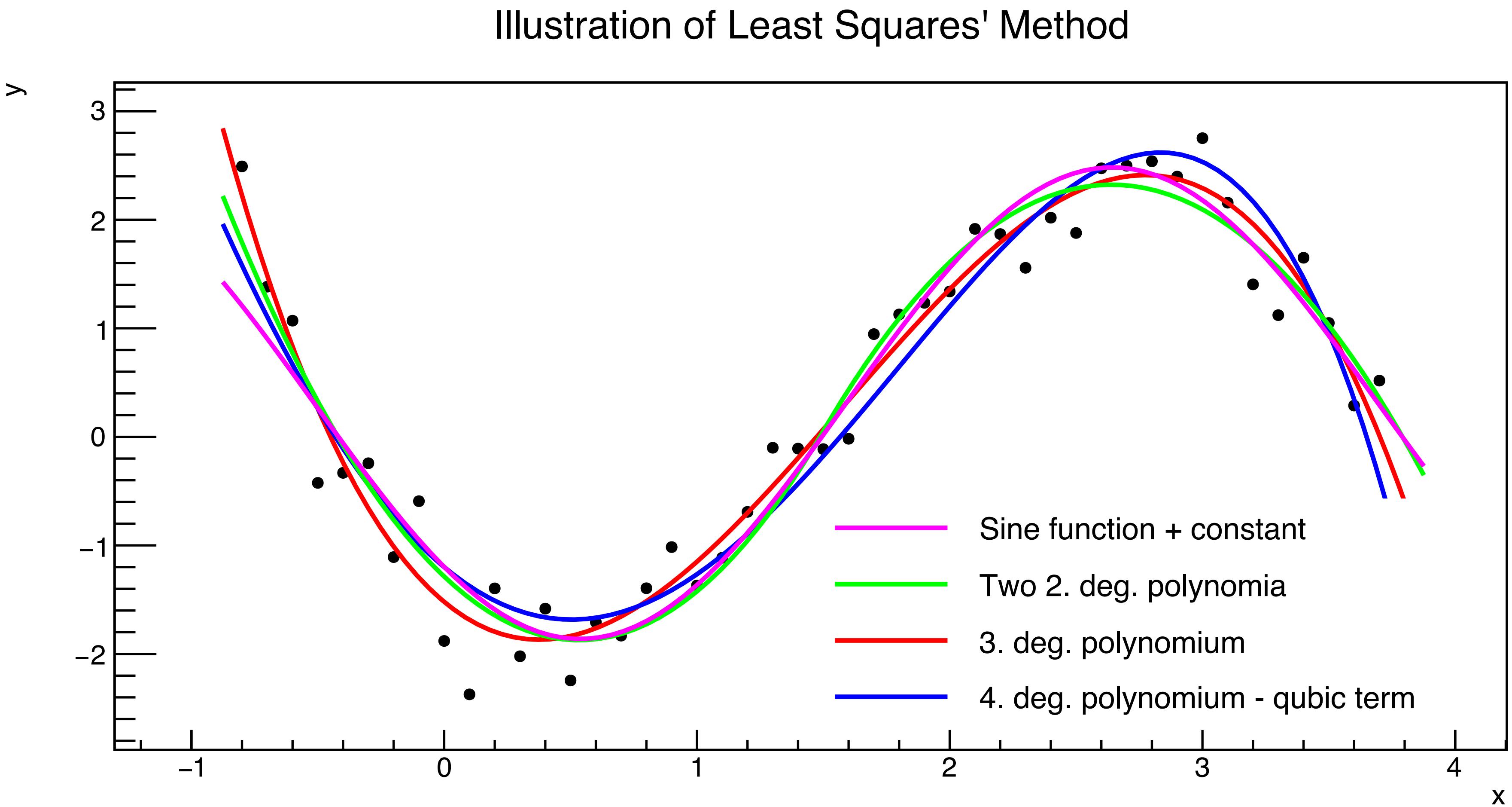
This has been used in several fields of science, in particular economy, but also medicin.
The model typically used is given below, and the analysis called “ANOVA”.

$$f(x) = ax_1 + bx_2 + c$$



Method of Least Squares

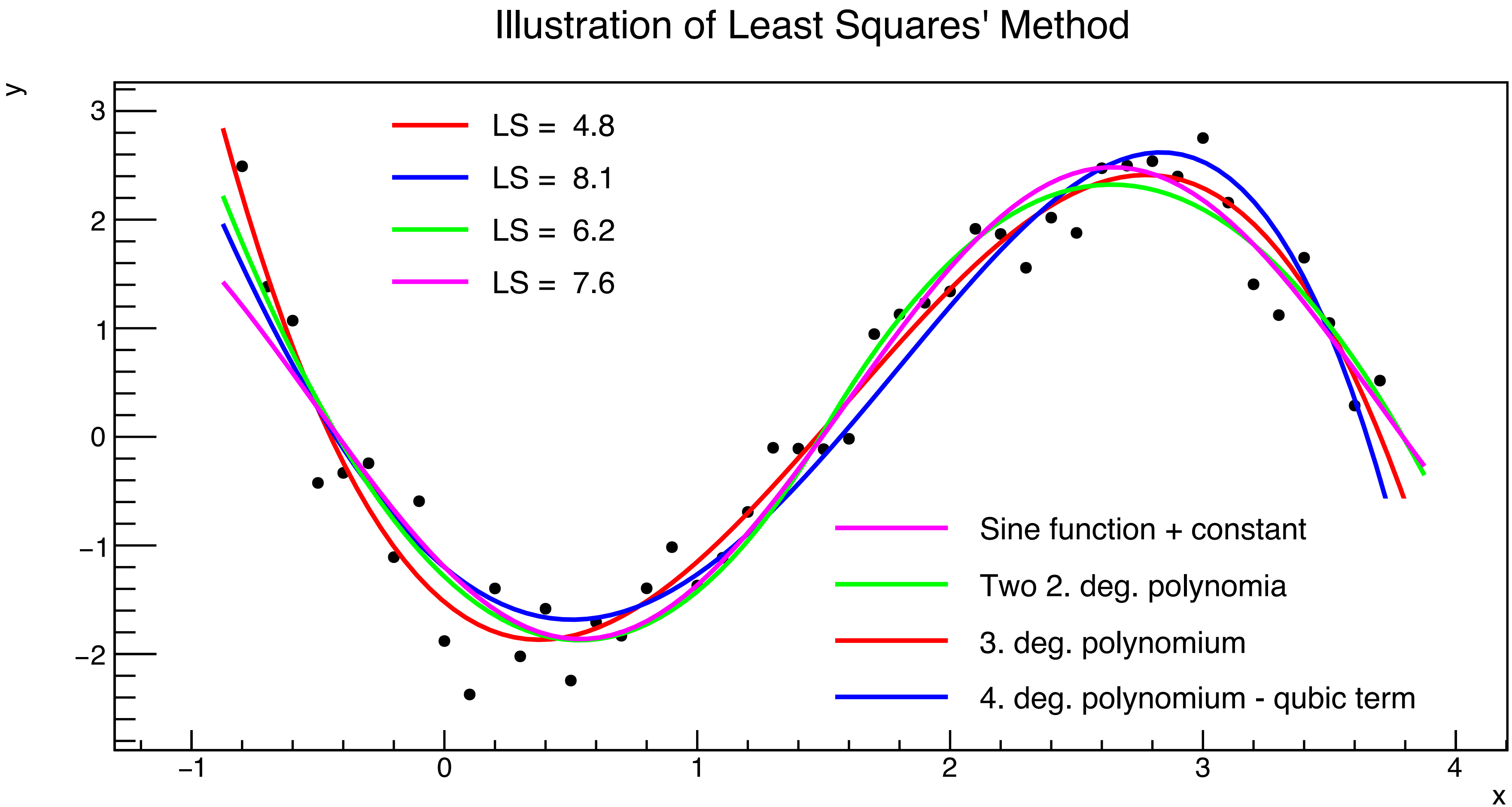
Look at the figure below, and determine which curve fits best...



Well, what do you define as "best"?

Method of Least Squares

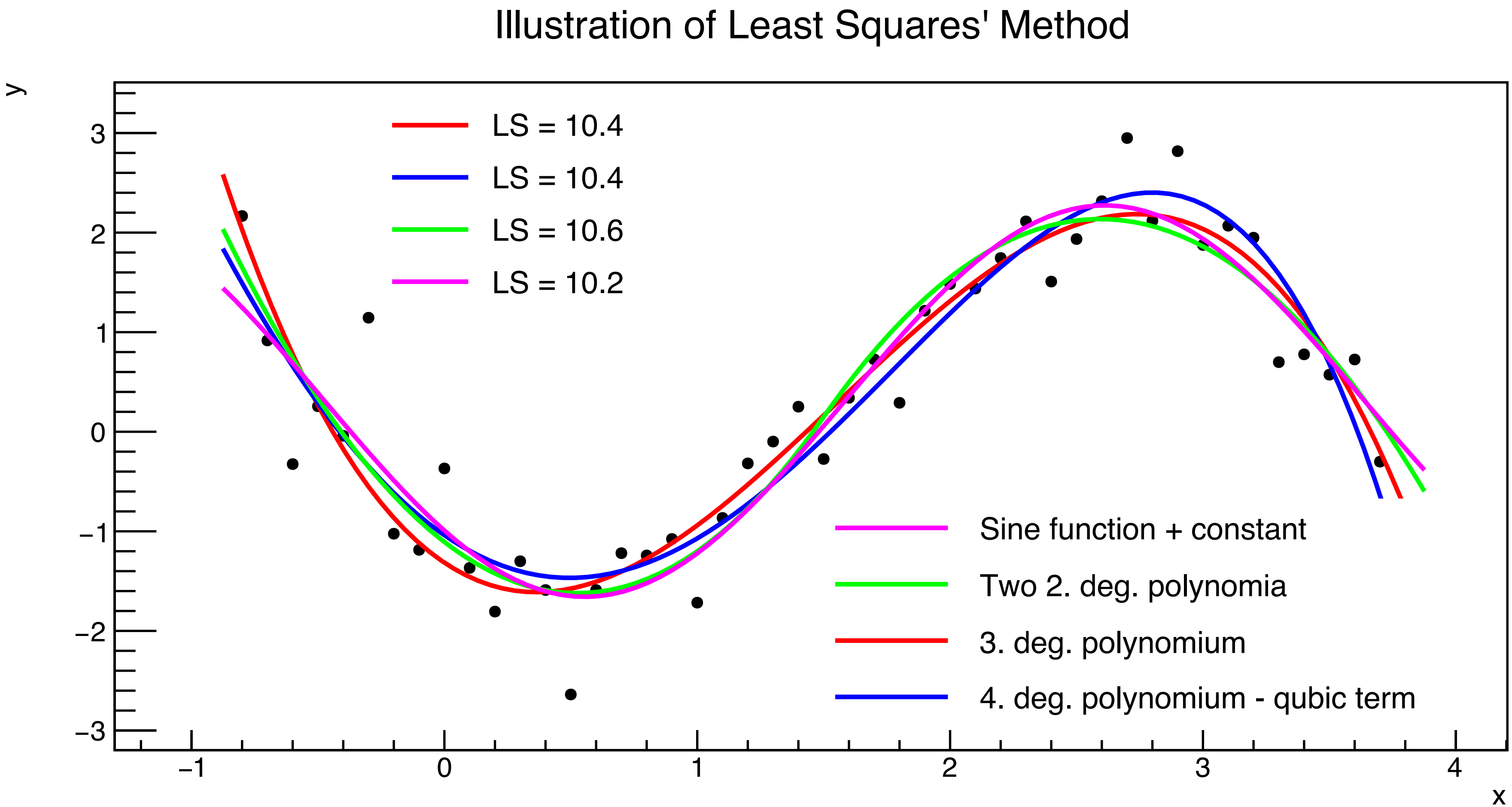
Look at the figure below, and determine which curve fits best...



Well, what do you define as "best"? And how good is it?!?

Method of Least Squares

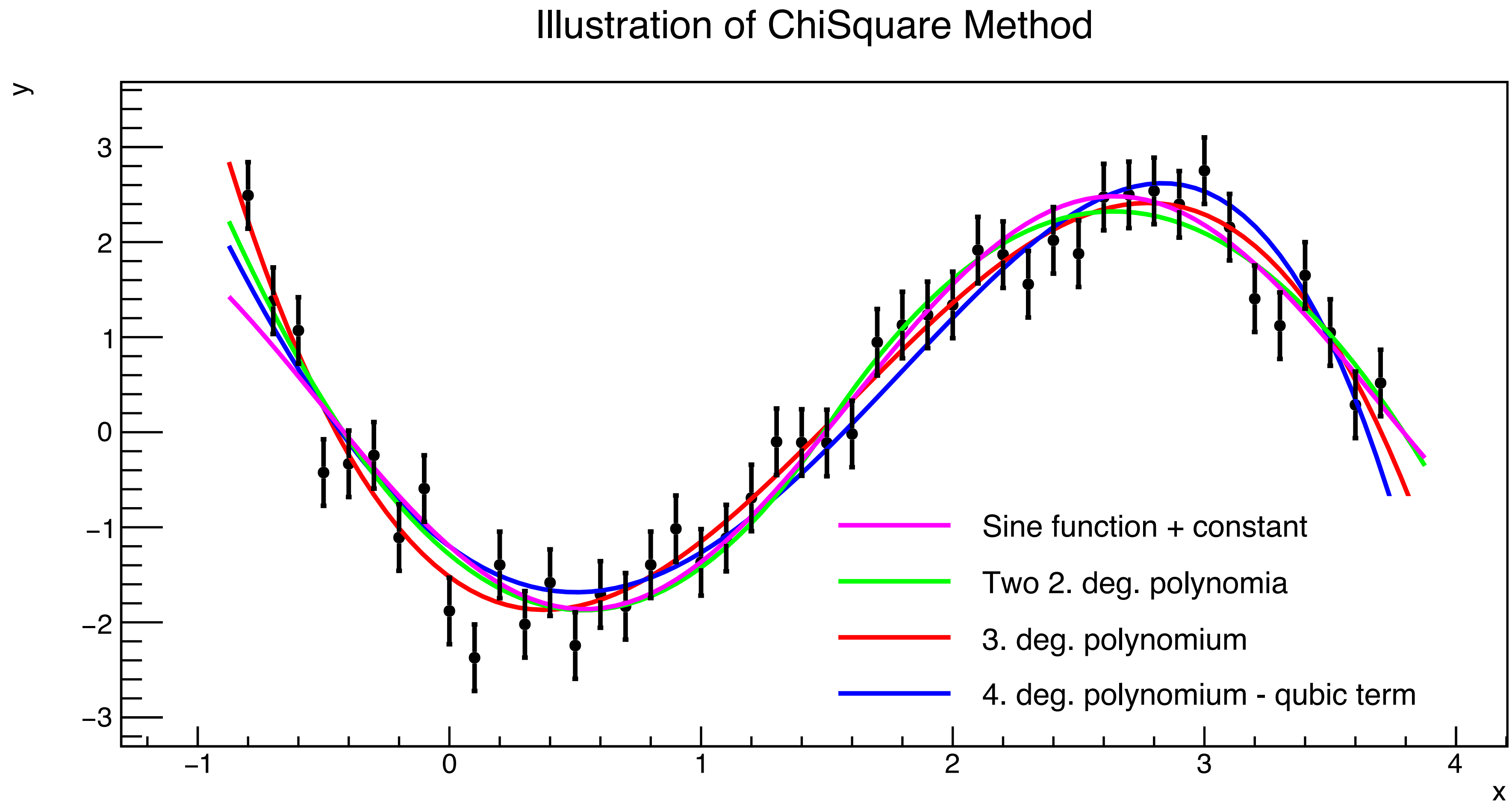
Look at the figure below, and determine which curve fits best...



Well, what do you define as "best"? And how good is it?!?

Chi-Square method

Look at the figure below, and determine which curve fits best...



Well, what do you define as "best"?

Defining the Chi-Square

<u>Problem Statement:</u> Given N data points (x, y, σ_y), adjust the parameter(s) θ of a model, such that it fits data best.

The best way to do this, given uncertainties σ_i on y_i is by minimising:

$$\chi^2(\theta) = \sum_i^N \frac{(y_i - f(x_i, \theta))^2}{\sigma_i^2}$$

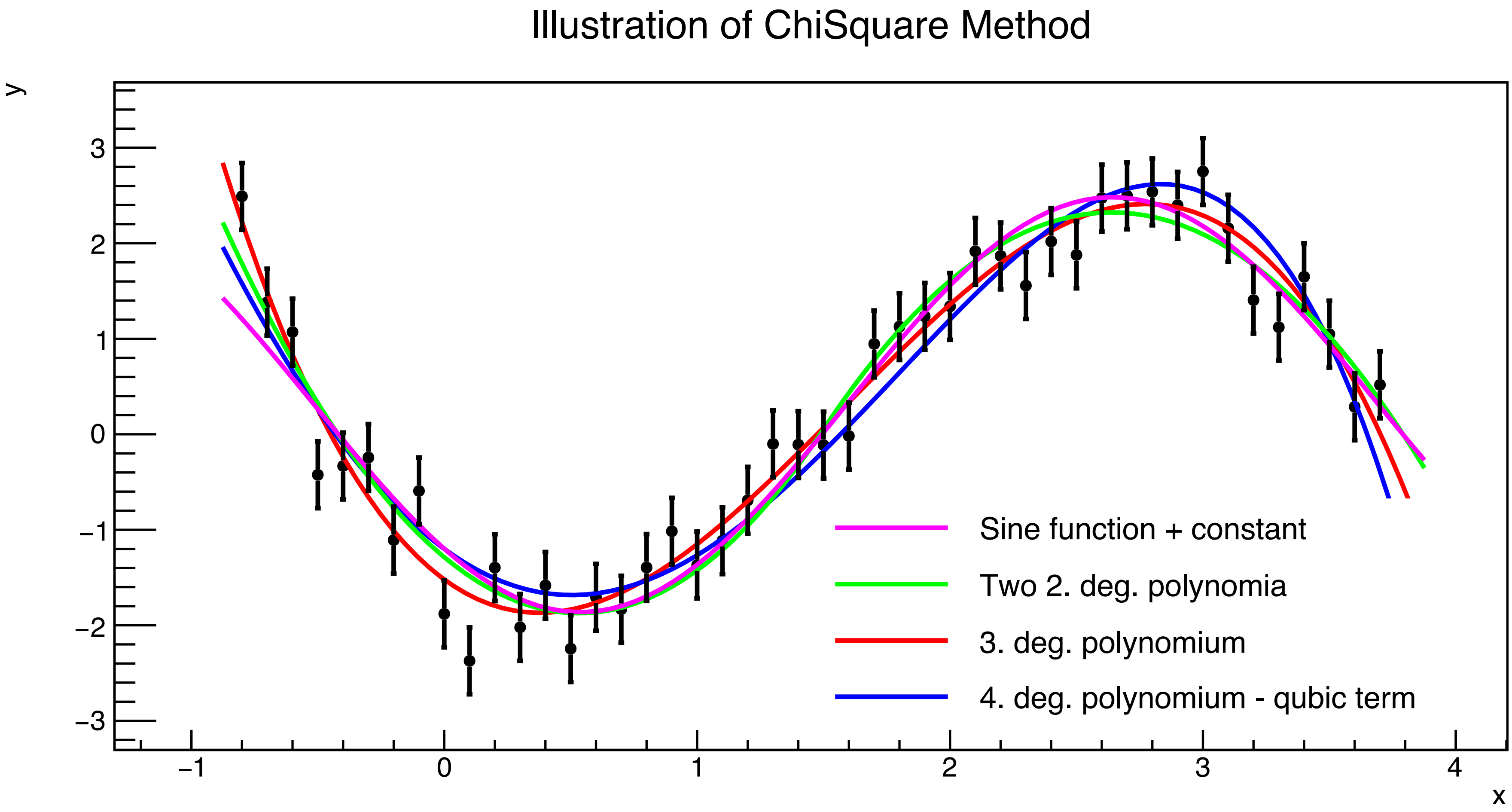
The power of this method is hard to overstate!

Not only does it provide a simple, elegant and unique way of fitting data, but more importantly it provides a **goodness-of-fit measure**.

This is the Chi-Square test!

Chi-Square method

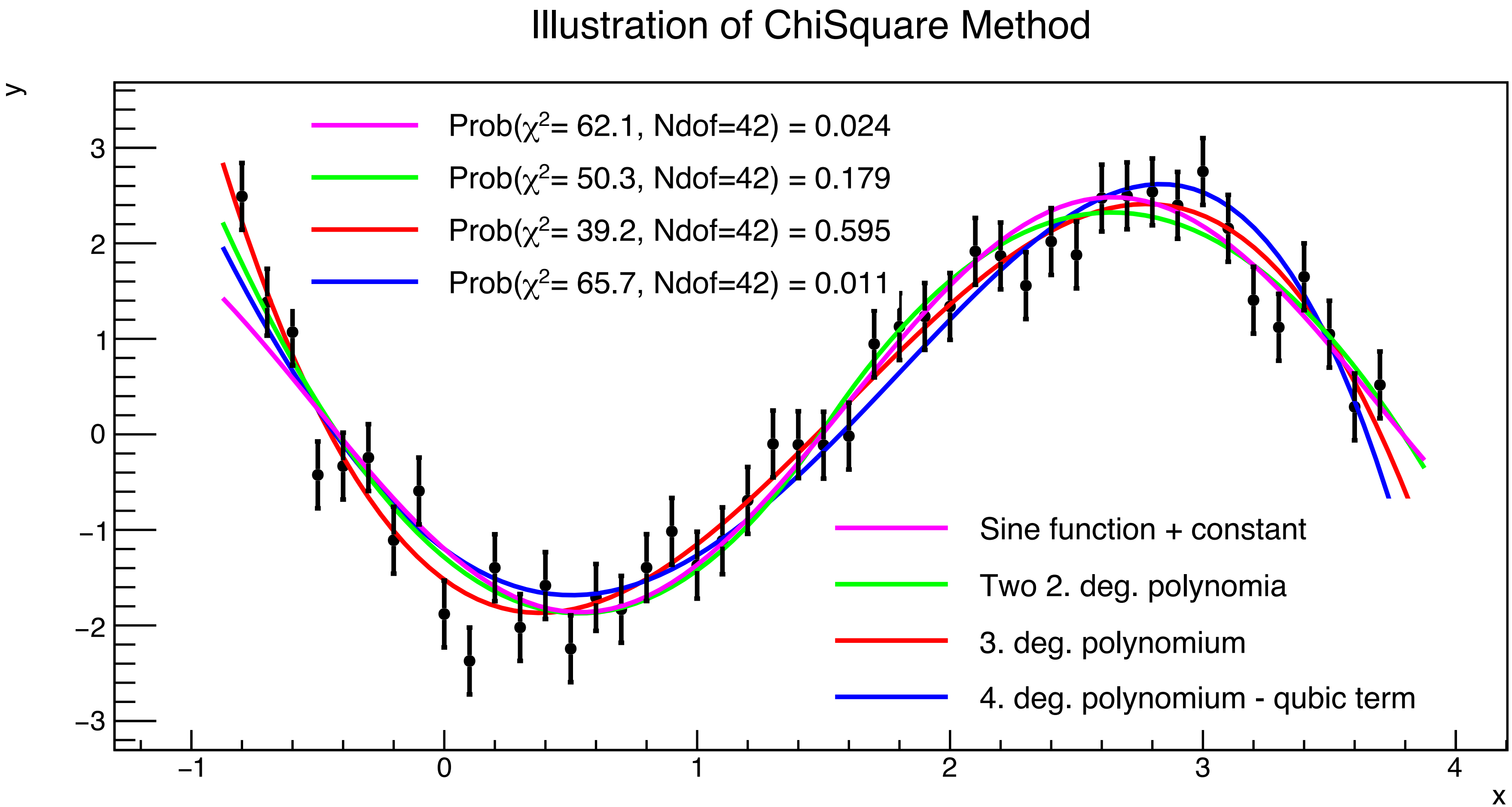
Look at the figure below, and determine which curve fits best...



Well, what do you define as "best"?

Chi-Square method

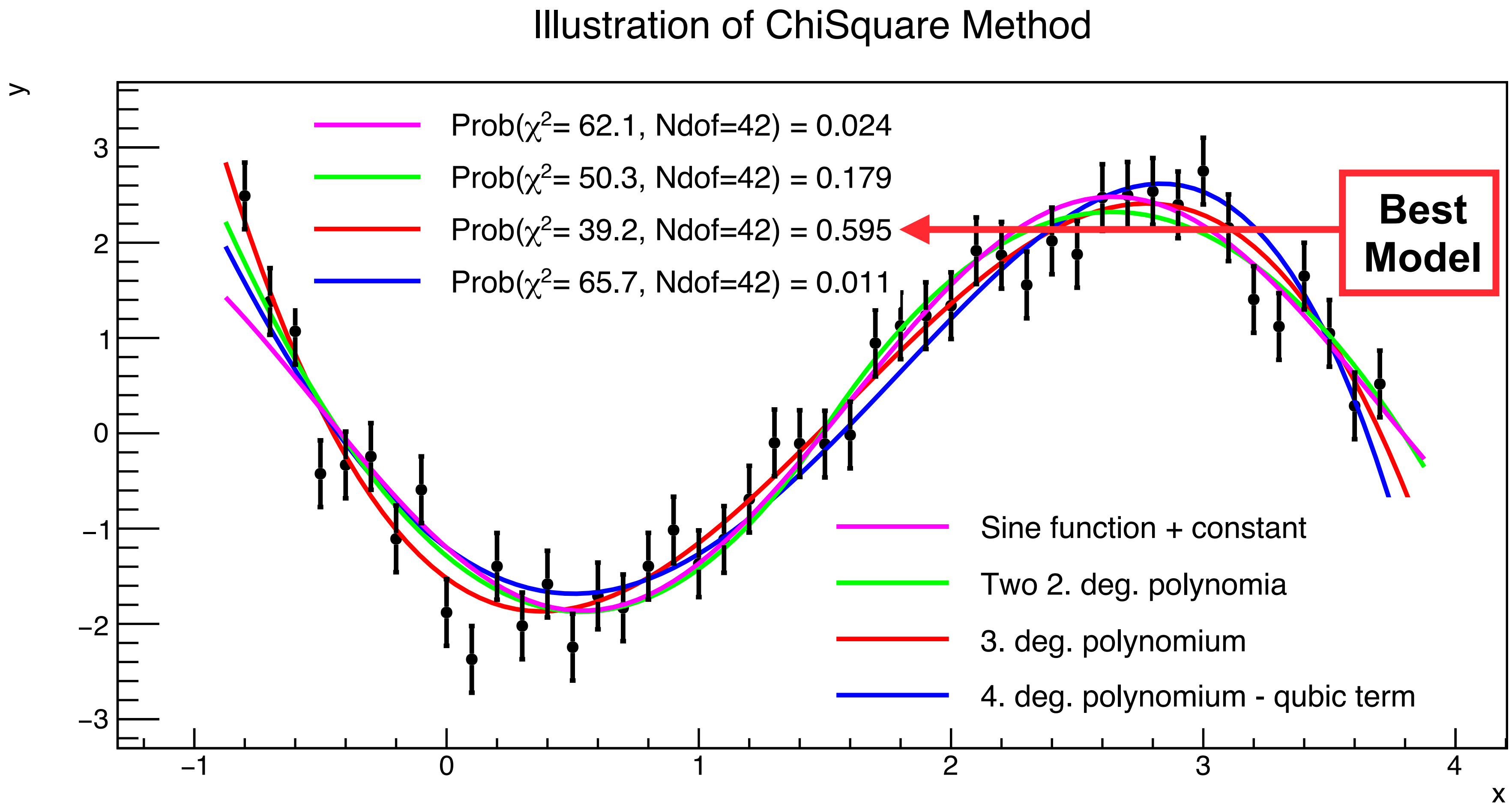
Look at the figure below, and determine which curve fits best…



Well, what do you define as "best"? The Chi2 quantifies this!

Chi-Square method

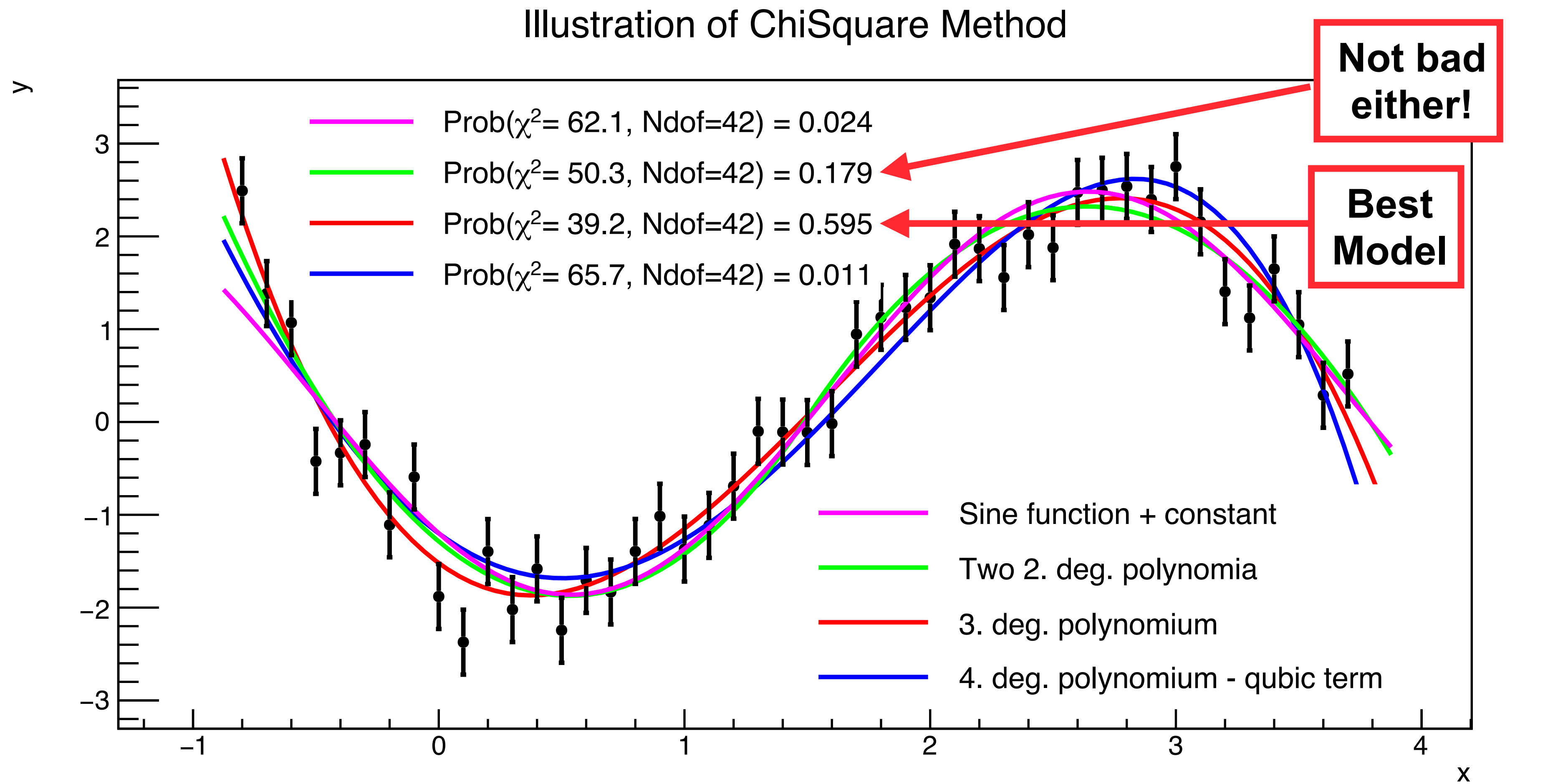
Look at the figure below, and determine which curve fits best...



Well, what do you define as "best"? The Chi2 quantifies this!

Chi-Square method

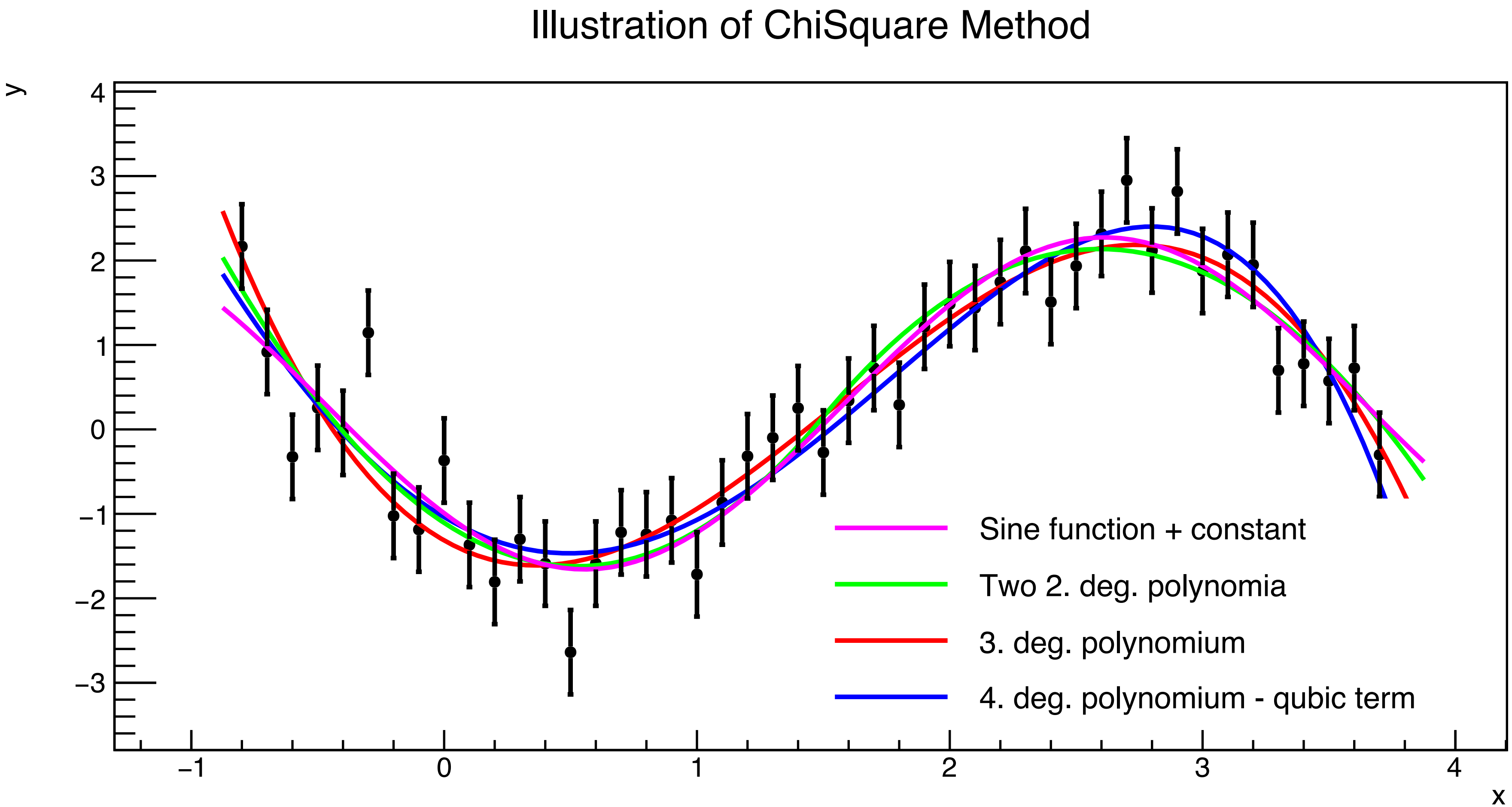
Look at the figure below, and determine which curve fits best...



Well, what do you define as "best"? The Chi2 quantifies this!

Chi-Square method

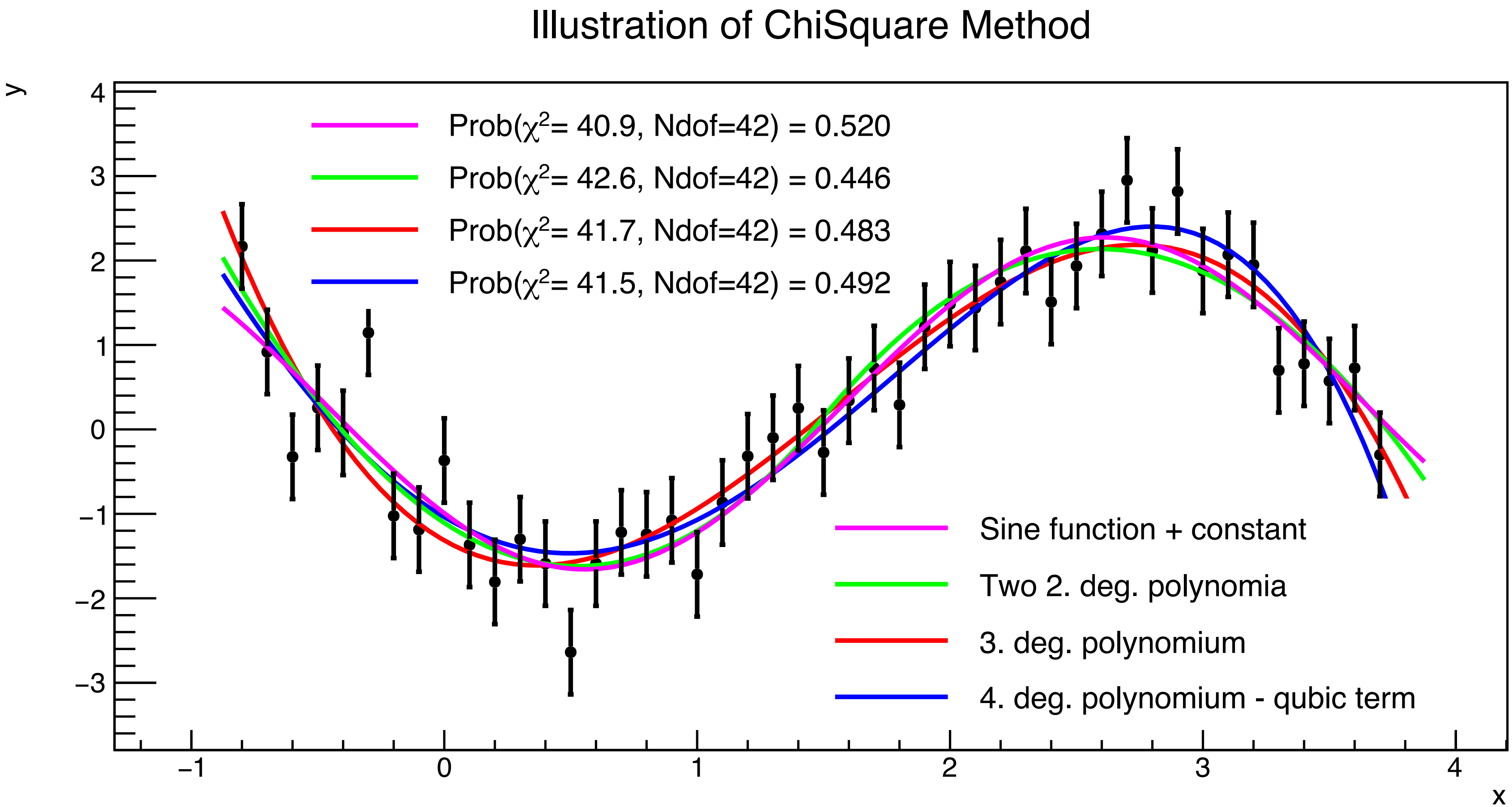
Look at the figure below, and determine which curve fits best...



What about now with **larger** errors?

Chi-Square method

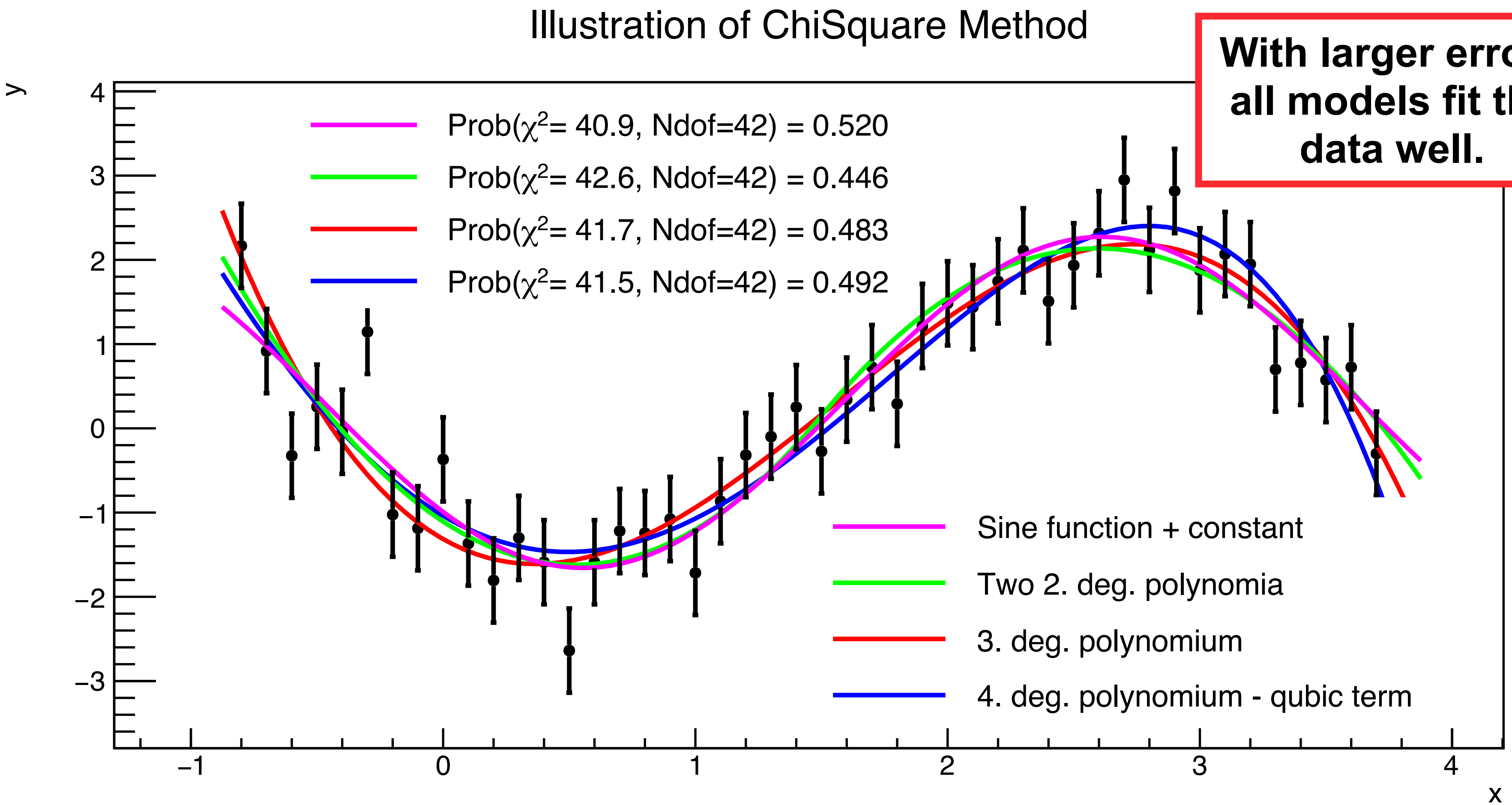
Look at the figure below, and determine which curve fits best...



What about now with **larger** errors?

Chi-Square method

Look at the figure below, and determine which curve fits best...

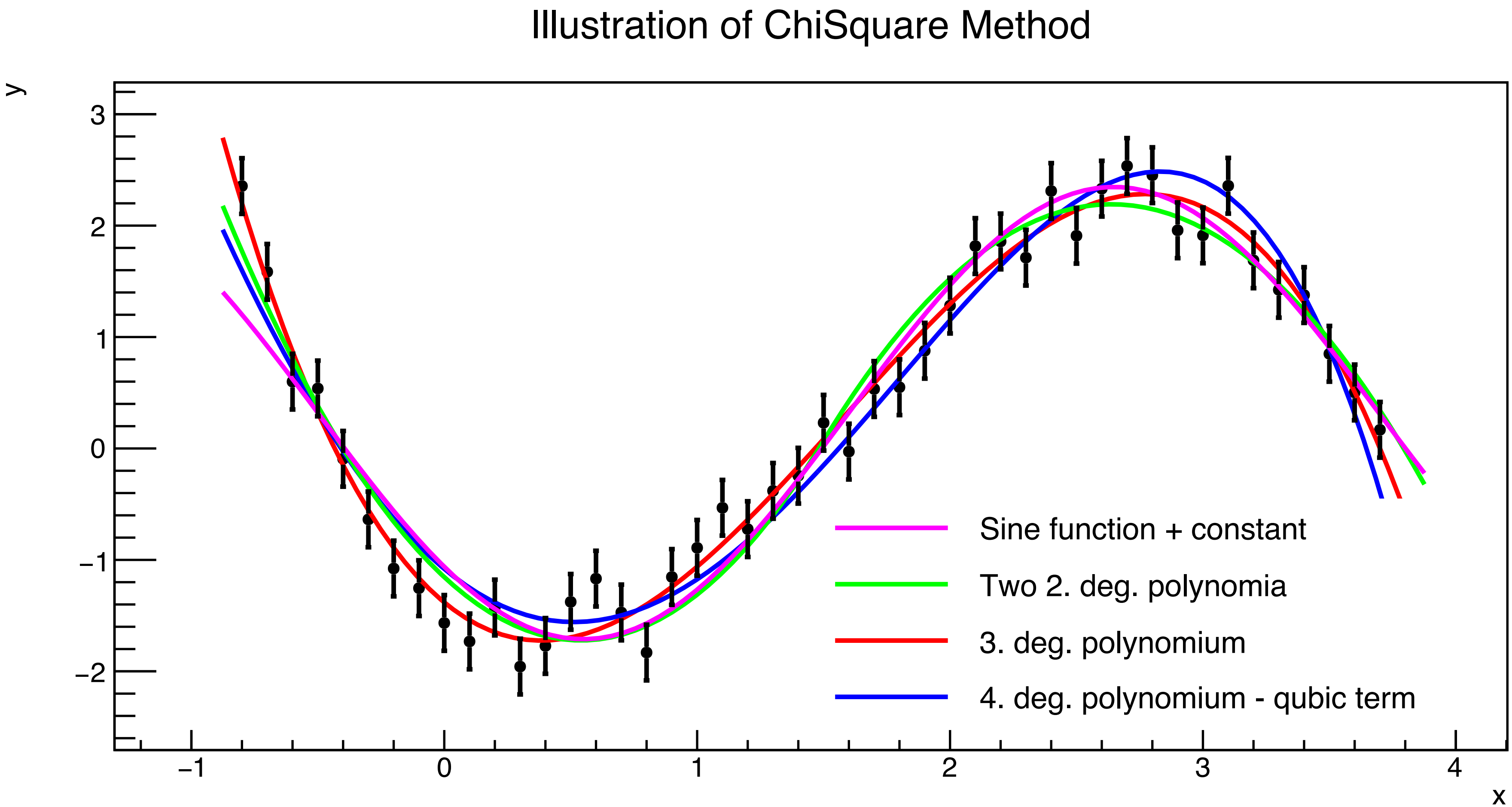


With larger errors all models fit the data well.

What about now with **larger** errors?

Chi-Square method

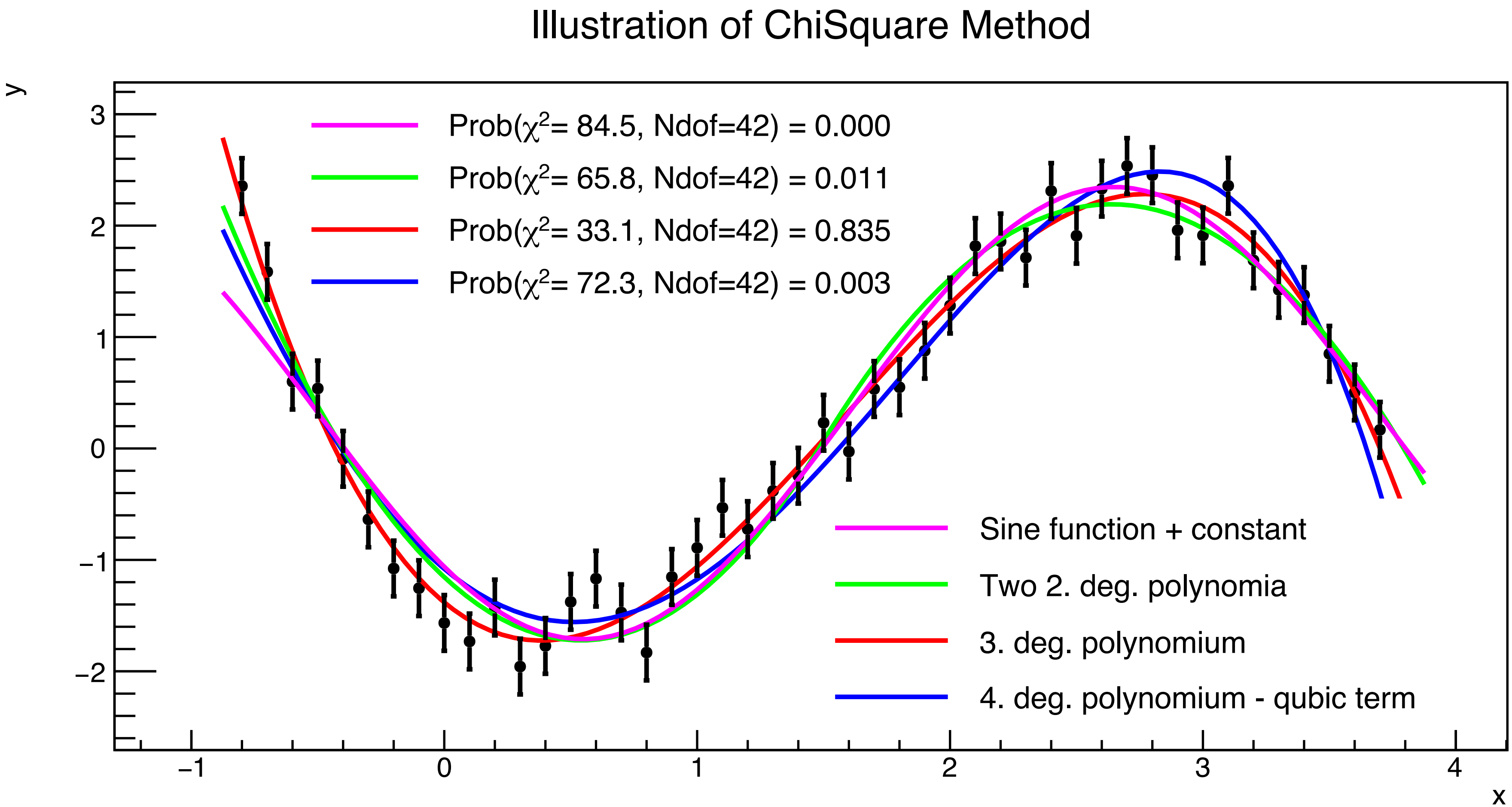
Look at the figure below, and determine which curve fits best...



What does **smaller** errors do?

Chi-Square method

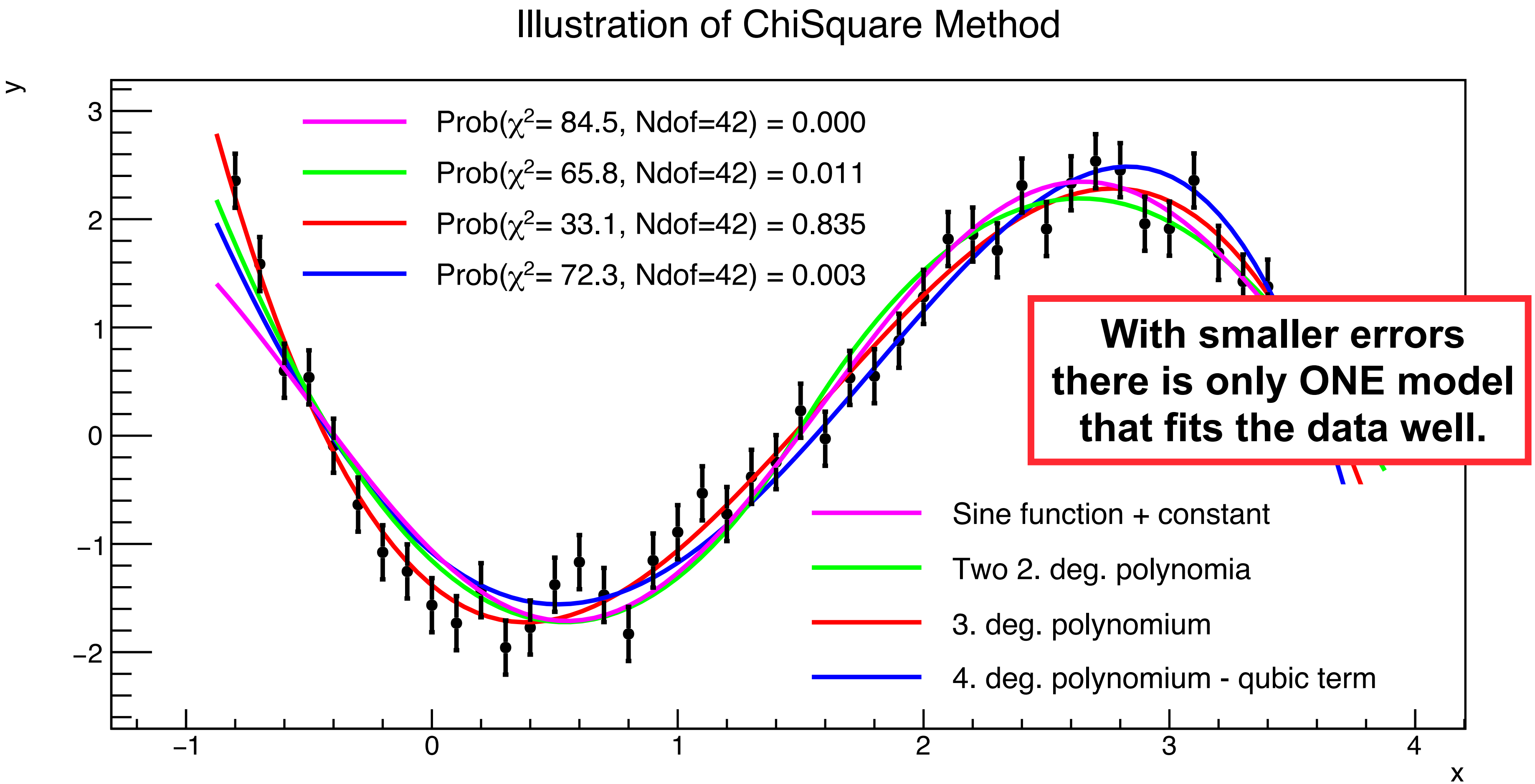
Look at the figure below, and determine which curve fits best...



What does **smaller** errors do?

Chi-Square method

Look at the figure below, and determine which curve fits best...



What does **smaller** errors do?

Defining the Chi-Square

<u>Problem Statement:</u> Given N data points (x,y,σ_y), adjust the parameter(s) θ of a model, such that it fits data best.

The best way to do this, given uncertainties σ_i on y_i is by minimising:

$$\chi^2(\theta) = \sum_i^N \frac{(y_i - f(x_i, \theta))^2}{\sigma_i^2}$$

The power of this method is hard to overstate!

Not only does it provide a simple, elegant and unique way of fitting data, but more importantly it provides a **goodness-of-fit measure**.

This is the Chi-Square test!

Defining the Chi-Square

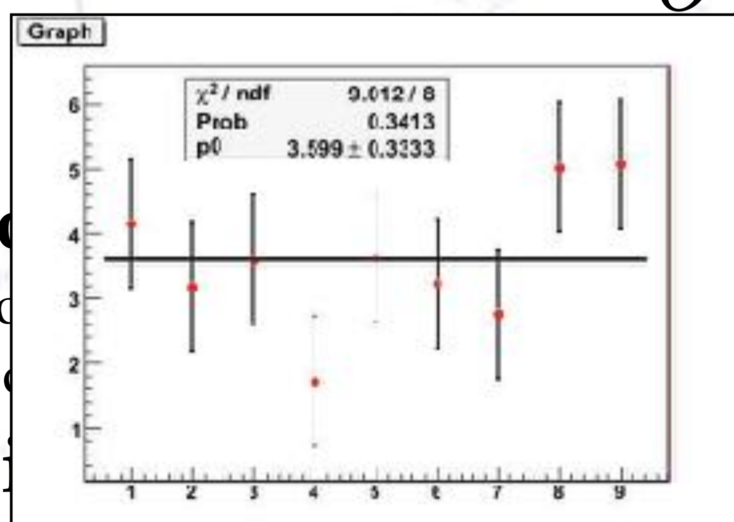
Problem St

Note that when doing a weighted mean, one should check if the measurements agree with each other!
This can be done with a ChiSquare test.

arameter(s)

The best w ... imising:

$$\chi^2(\theta) = \sum^{N} \frac{(y_i - f(x_i, \theta))^2}{\sigma_i^2}$$



The power c ... to overstate!

Not only does it pro ... que way of fitting data, but more imp ... **-of-fit measure**.

Thi ... test!

Weighted mean & ChiSquare

The weighted mean is actually an **analytical ChiSquare minimisation to a constant**. The result is the same, and one can then calculate Prob(χ^2, Ndof).

<u>Example:</u>

Data (from pendulum experiment) could be four length measurement (in mm):

d : [17.8 ± 0.5, 18.1 ± 0.3, 17.7 ± 0.5, 17.7 ± 0.2]

The output from the above data is (many digits for *checks only*):

Mean	= 17.8098 mm
Error on mean	= 0.15057 mm
ChiSquare	= 1.28574
Ndof	= 3
Probability	= 0.7325213

NOTE: This seems a very nice (and precise) result, and it may very well be. BUT, it might also be, that we all four estimated it from the same photo or similarly, which could be biased by an angled view. Then we would be fooling ourselves. We will discuss such **"systematic uncertainties"** more!

Weighted mean & ChiSquare

The weighted mean is actually an **analytical ChiSquare minimisation to a constant**. The result is the same, and one can then calculate Prob(χ^2, Ndof).

<u>Example:</u>
Data (from pendulum experiment) could be four length measurement (in mm):

d : [17.8 ± 0.5, 18.1 ± 0.3, 17.7 ± 0.5, 17.7 ± 0.2]

The output from the above data is (many digits for *checks only*):

Mean	=	17.8098 mm
Error on mean	=	0.15057 mm
ChiSquare	=	1.28574
Ndof	=	3
Probability	=	0.7325213

d = (17.81 ± 0.15) mm
p(χ^2=1.3, N_{dof}=3) = 0.73

NOTE: This seems a very nice (and precise) result, and it may very well be. BUT, it might also be, that we all four estimated it from the same photo or similarly, which could be biased by an angled view. Then we would be fooling ourselves. We will discuss such **"systematic uncertainties"** more!

Why the ChiSquare is great

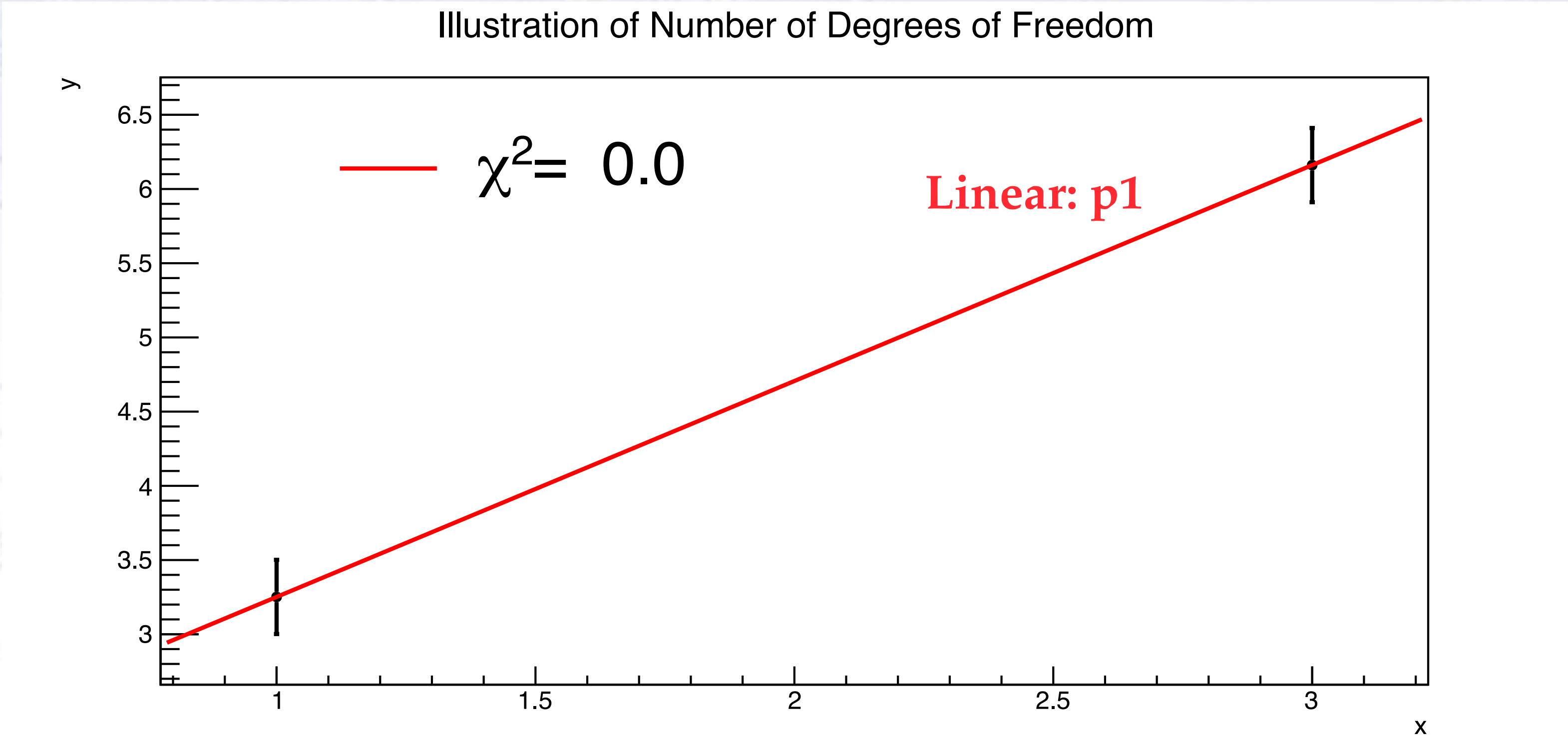
…but not its magic

Number of degrees-of-freedom

How to find/calculate the Number of degrees-of-freedom (Ndof) in a fit?

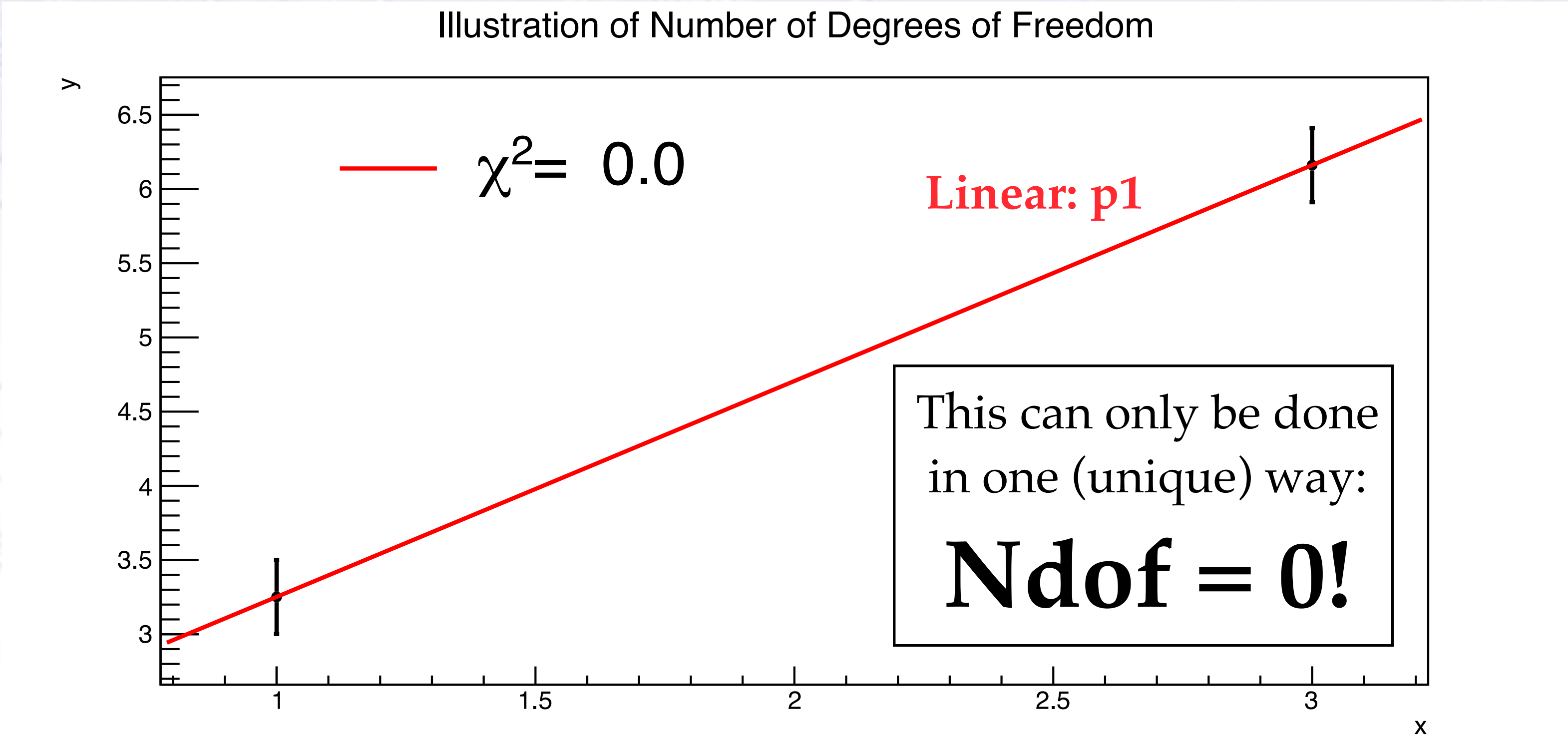
Number of degrees-of-freedom

How to find / calculate the Number of degrees-of-freedom (Ndof) in a fit?



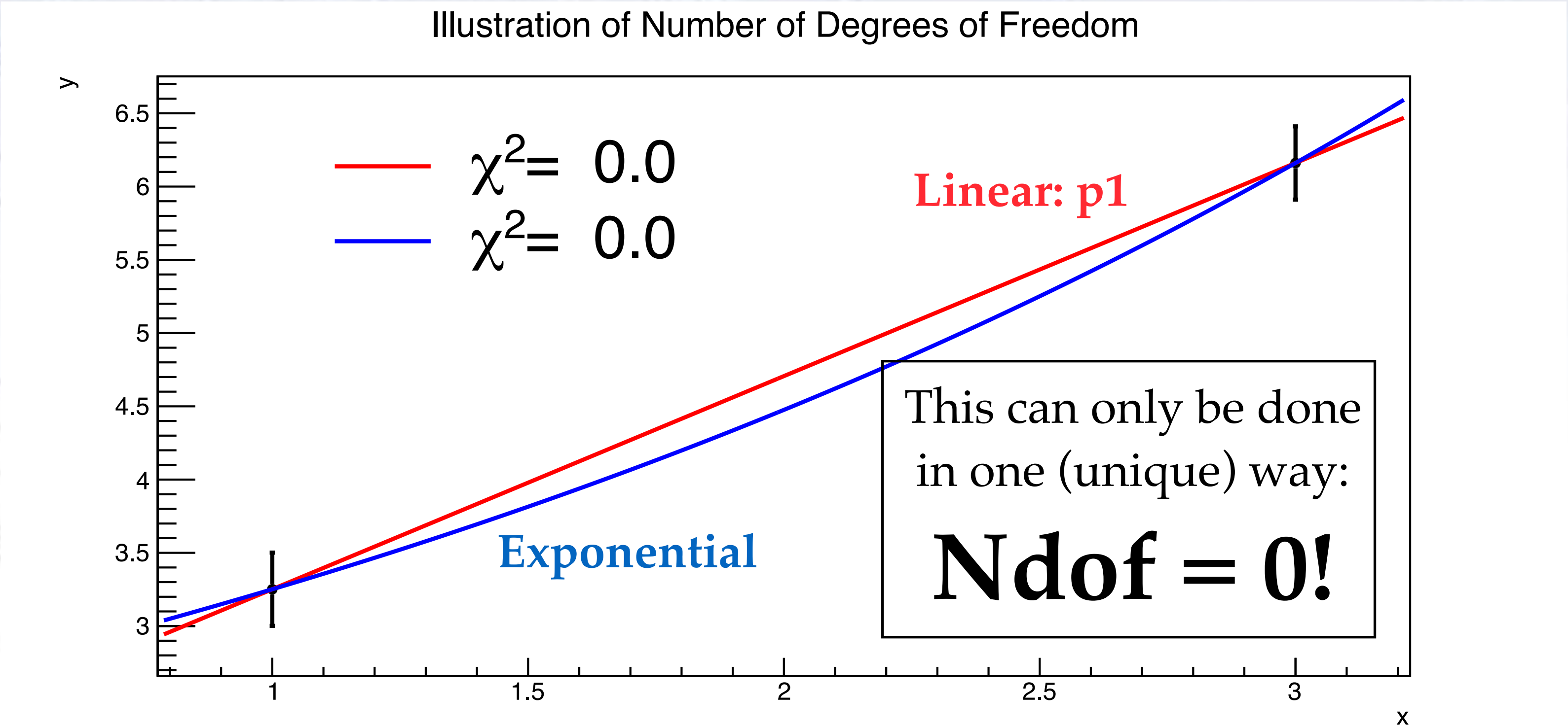
Number of degrees-of-freedom

How to find/calculate the Number of degrees-of-freedom (Ndof) in a fit?



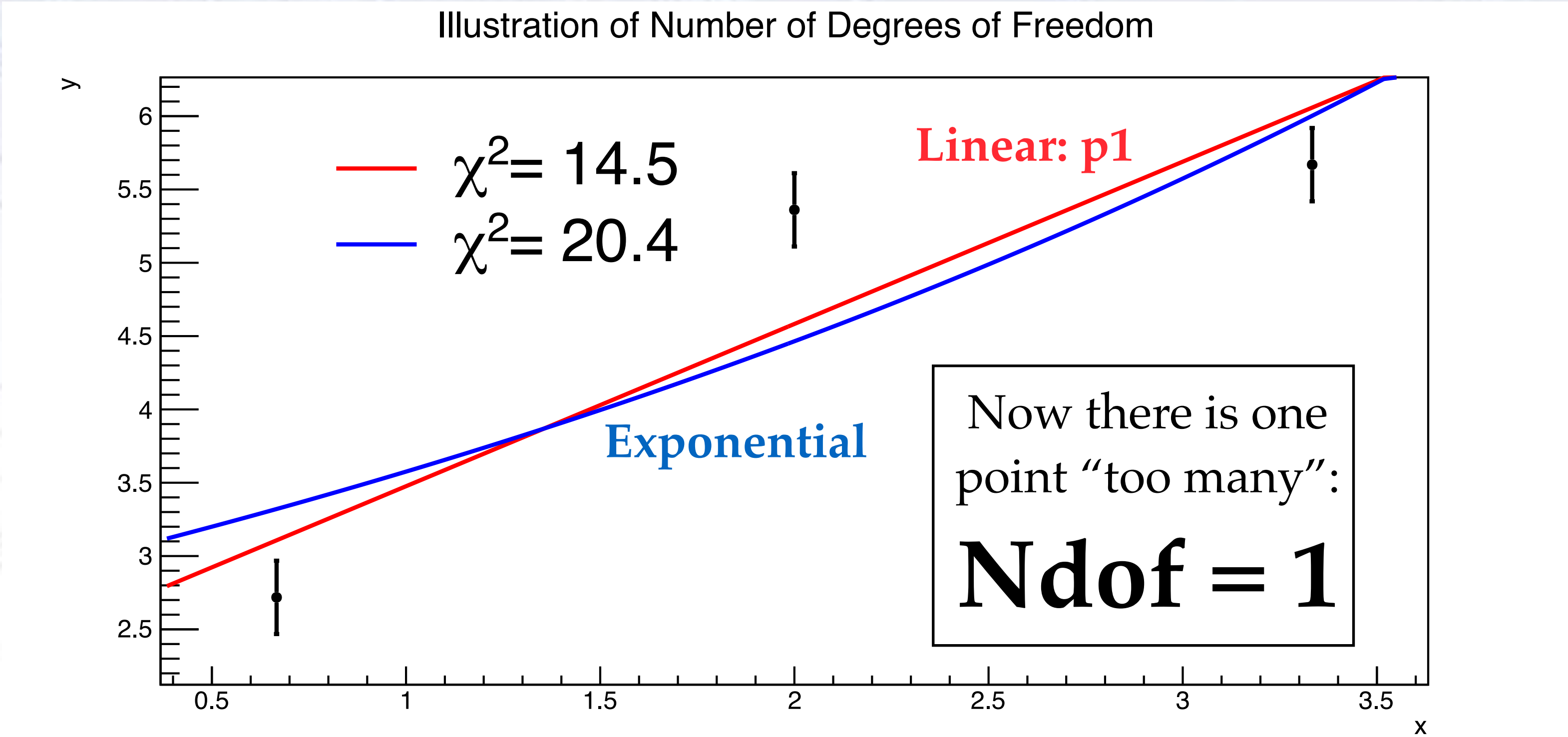
Number of degrees-of-freedom

How to find/calculate the Number of degrees-of-freedom (Ndof) in a fit?



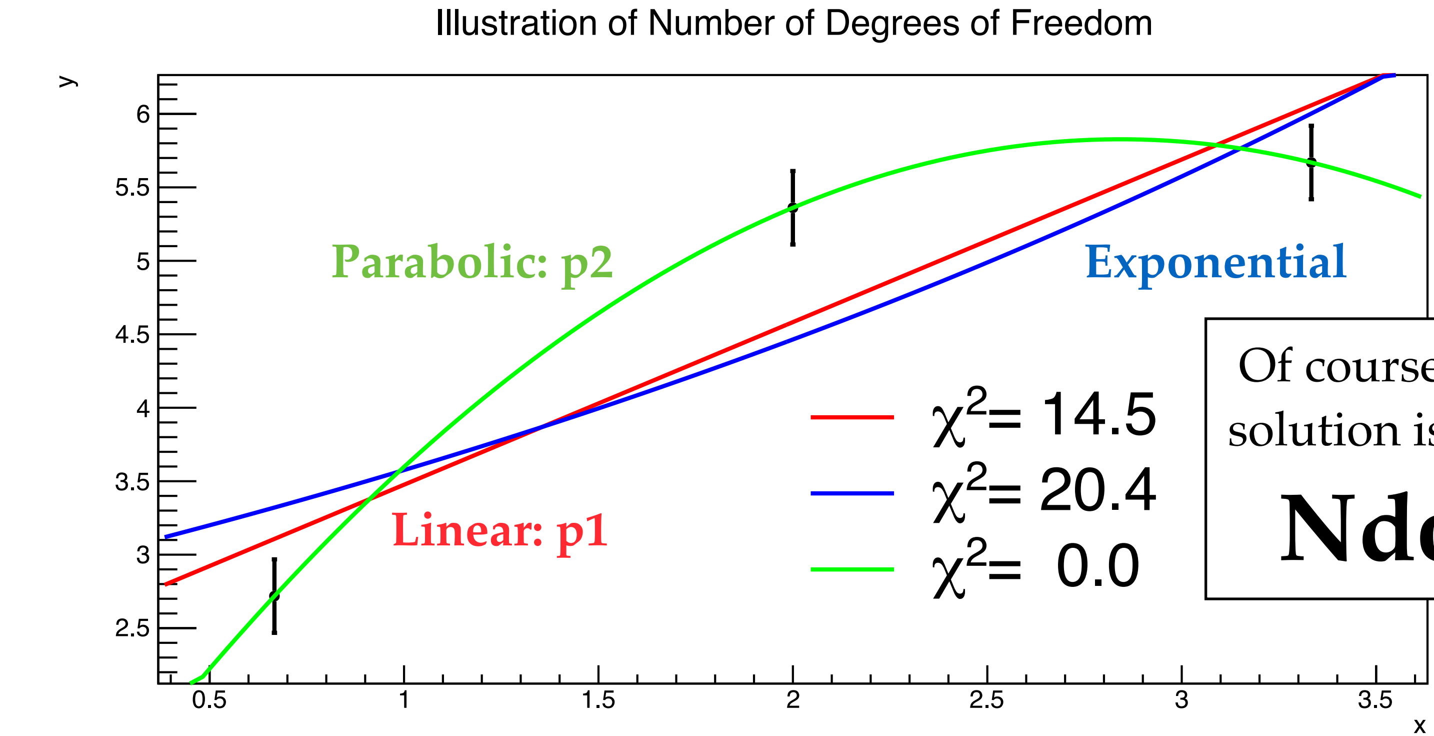
Number of degrees-of-freedom

How to find/calculate the Number of degrees-of-freedom (Ndof) in a fit?



Number of degrees-of-freedom

How to find/calculate the Number of degrees-of-freedom (Ndof) in a fit?



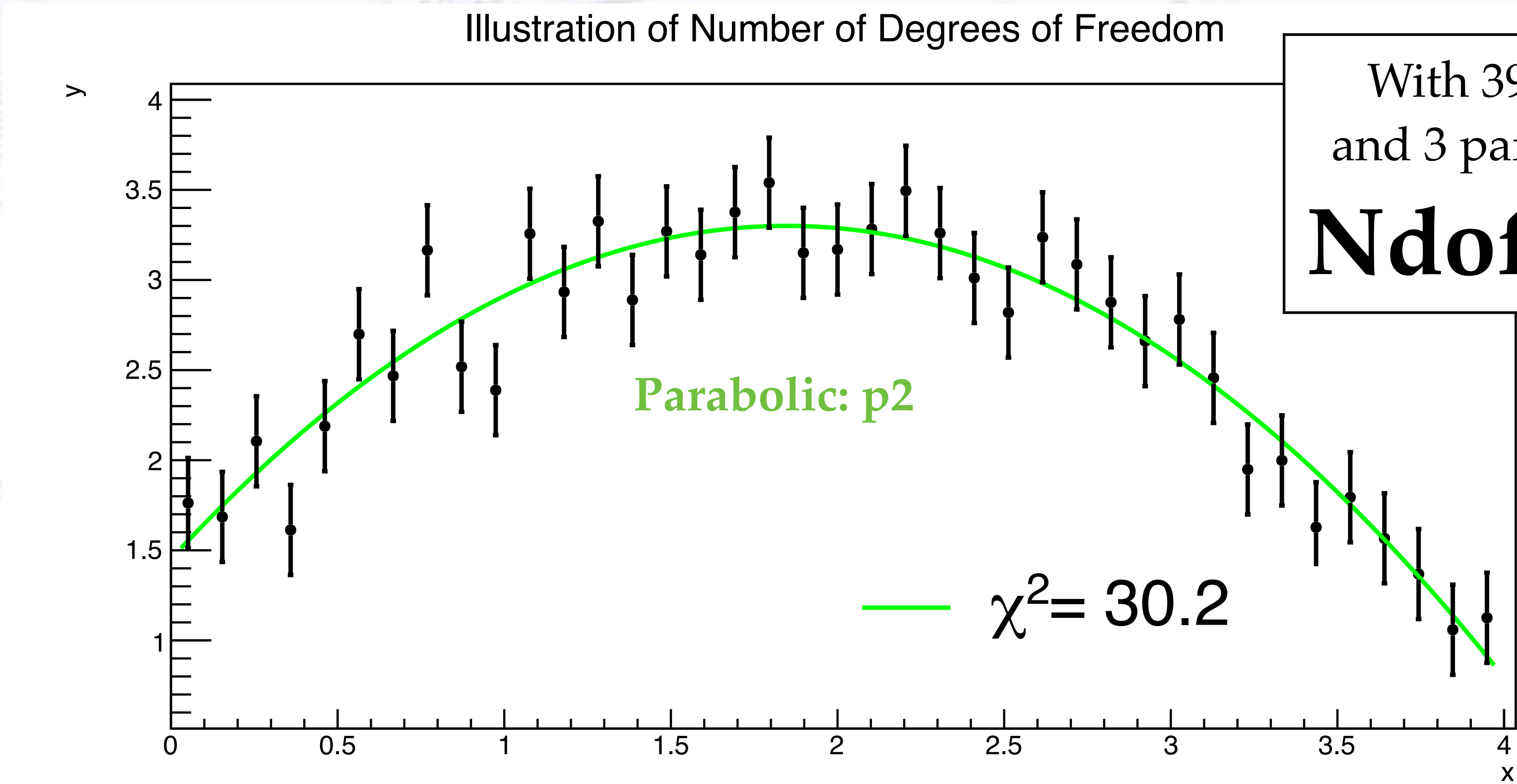
Of course for pol2 the solution is still unique:

Ndof = 0

Number of degrees-of-freedom

The number of degrees-of-freedom, Ndof, can be calculated as the number of points in the fit minus the number of parameters in the fit function:

$$N_{\text{dof}} = N_{\text{data points}} - N_{\text{fit variables}}$$

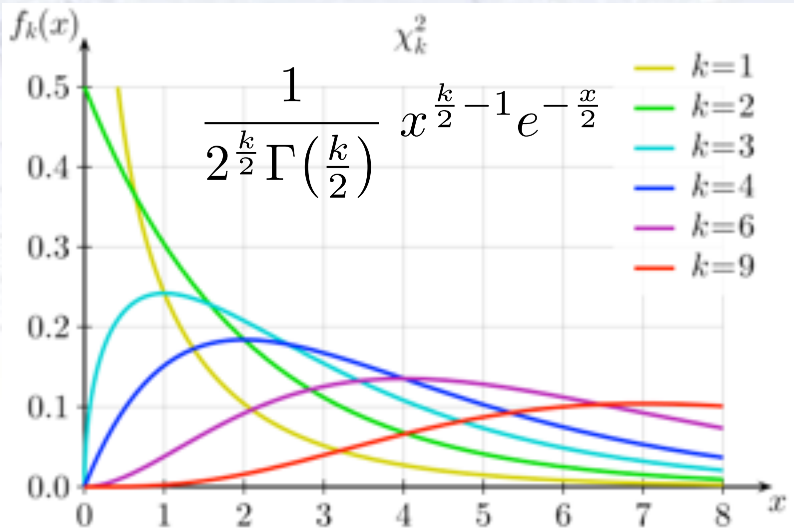


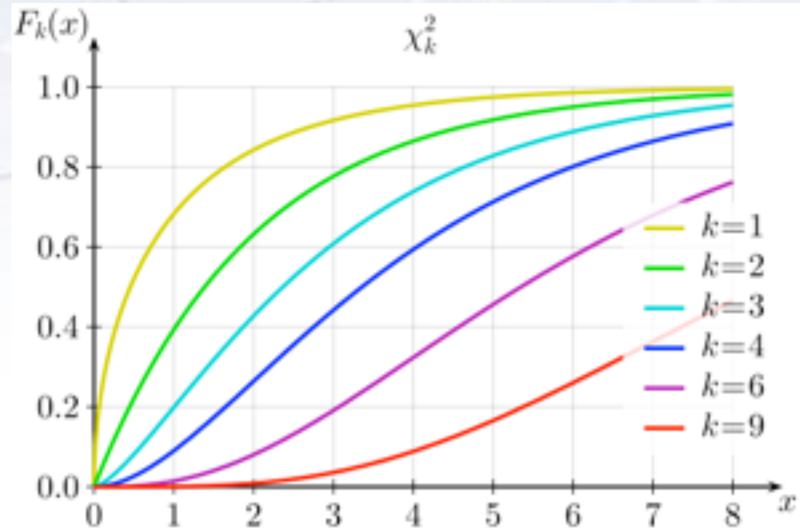
With 39 points and 3 parameters:

Ndof = 36

The Chi-Square distribution and test

The **Chi-Square distribution** for N_{dof} degrees of freedom is the distribution of the sum of the squares of N_{dof} normally distributed random variables.





The **Chi-Square test** consists of comparing the Chi-Square value obtained from a fit with the PDF of expected Chi-Square values. This allows the calculation of the *probability* of observing something with the same Chi-Square value or higher...

Rule of thumb: Chi-Square should roughly match N_{dof}

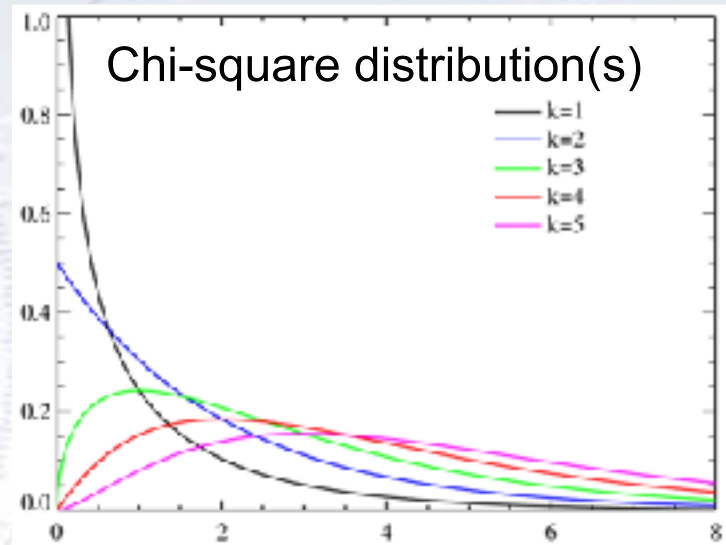
Chi-Square probability calculation

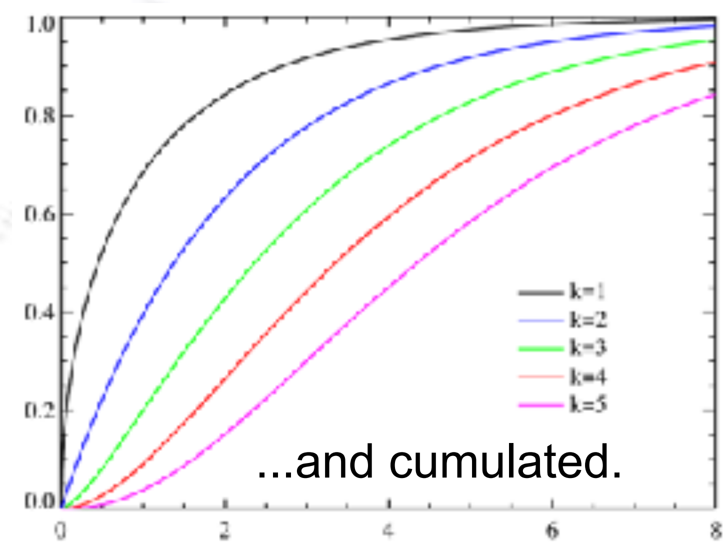
Given a **Chi-square value** and a **number of degrees of freedom** (Ndof), one can obtain a **"goodness-of-fit"**.

It is known, what Chi-square values to expect given the Ndof. One can therefore compare to this (Chi-square) distribution, and see...

what is the probability of getting this Chi-square value or something worse, assuming this is the correct fit function!

<u>Example:</u>
A fit gave the Chi-square 7.1 with 5 dof. The chance of getting this Chi-square or worse is... (reading the pink bottom curve (Ndof = k = 5) at 7.1)...



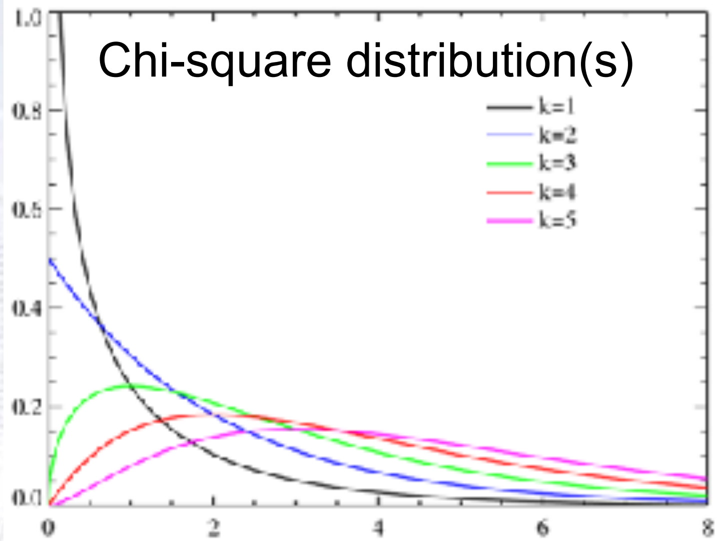


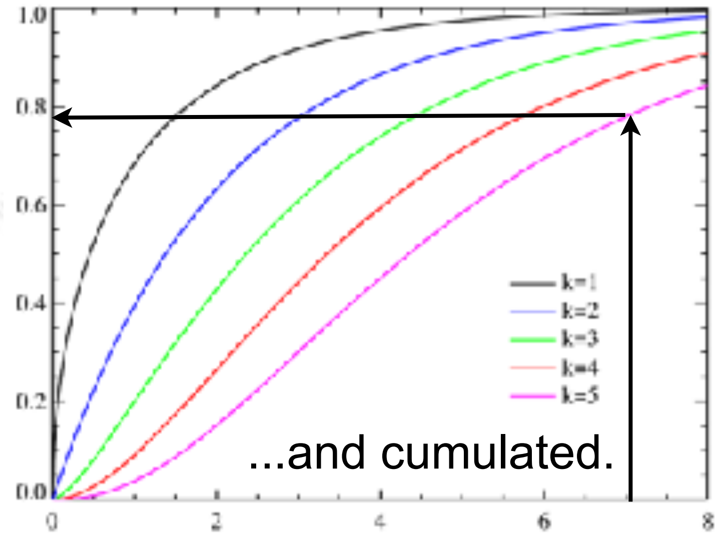
Chi-Square probability calculation

Given a **Chi-square value** and a **number of degrees of freedom** (Ndof), one can obtain a **"goodness-of-fit"**.

It is known, what Chi-square values to expect given the Ndof. One can therefore compare to this (Chi-square) distribution, and see...

what is the probability of getting this Chi-square value or something worse, assuming this is the correct fit function!





Example:

A fit gave the Chi-square 7.1 with 5 dof. The chance of getting this Chi-square or worse is... (reading the pink bottom curve (Ndof = k = 5) at 7.1)... 1 - 0.78 = **22%**

Chi-Square probability calculation

In the table below, one can get a quick estimate for low N_{dof}.

Degrees of freedom (df)	χ^2 value [16]										
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
P value (Probability)	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
	Non-significant								**Significant**		

Chi-Square probability calculation

In the table below, one can get a quick estimate for low N_{dof}.

Degrees of freedom (df)	χ^2 value [16]										
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9.49	13.28	18.47
5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
P value (Probability)	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
	Non-significant								**Significant**		

Python:

chi2_prob = stats.chi2.sf(chi2_value, N_{DOF})

sf (survival function) = 1 - CDF

Chi-Square probability interpretation

The Chi-Square probability can **roughly** be interpreted as follows:

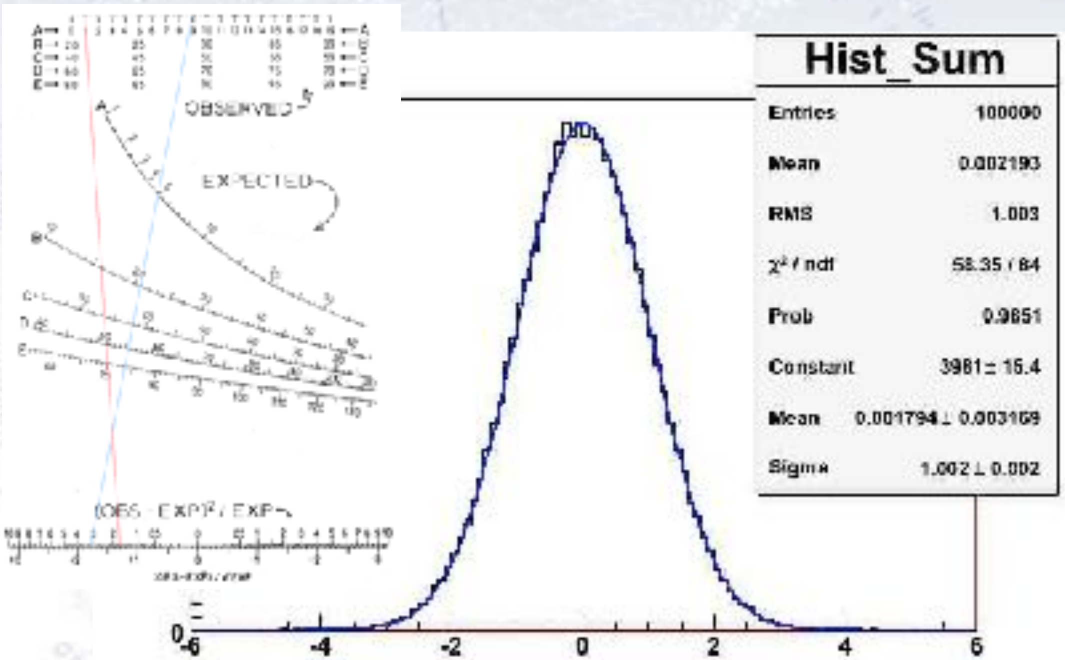
- If χ^2 / Ndof $\simeq$ 1 or more precisely if $0.01 < p(\chi^2, Ndof) < 0.99$, then all is good.
- If χ^2 / Ndof $\gg$ 1 or more precisely if $p(\chi^2, Ndof) < 0.01$, then your fit is bad, and your hypothesis is probably not correct.
- If χ^2 / Ndof $\ll$ 1 or more precisely if $0.99 < p(\chi^2, Ndof)$, then your fit is TOO good and you probably overestimated the errors.

If the statistics behind the plot is VERY high (great than 10^6), then you might have a hard time finding a model, which truly describes all the features in the plot (as now tiny effects become visible), and one hardly ever gets a good Chi-Square probability. However, in this case, one should not worry too much, unless very high precision is wanted.

Anyway, the Chi-Square still allows you to compare several models, and determine which one is the better.

Chi-Square for binned data

If the data is binned (i.e. put into a histogram), then Pearson's Chi-square applies:



The formula (based on Poisson statistics) is:

$$\chi^2 = \sum_{\mathrm{i} \,\in\, \mathrm{bin}} \frac{(O_i - E_i)^2}{E_i}$$

Chi-Square for binned data

While Pearson's Chi-square test is quite useful, it has some limitations, especially when some bins have low statistics.

The expected cell count (E_i) should not be too low. Some require 5 or more, and others require 10 or more. A common rule is 5 or more in 80% of bins, but no cells with zero expected count. When this assumption is not met, Yates's Correction can be applied.

One alternative is to divide by O_i when O_i is not 0.

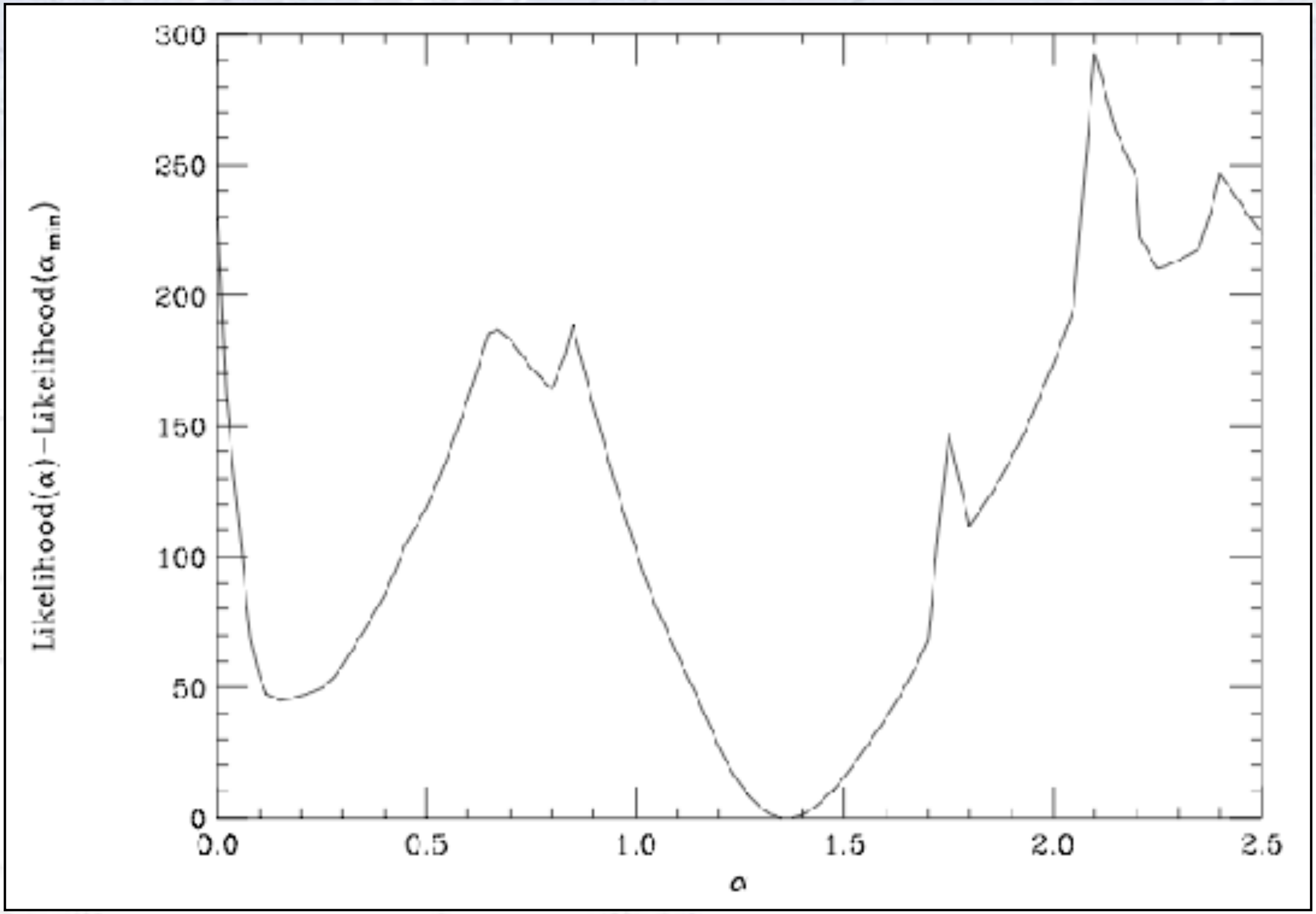
Another alternative is the likelihood fit, which does not suffer under low statistics.

$$\chi^2 = \sum_{\mathrm{i} \in \mathrm{bin}} \frac{(O_i - E_i)^2}{E_i}$$

Yet, another alternative is the G-test, which is more robust at low statistics. However, I've never seen it in use.

$$G = 2 \sum_{\mathrm{i} \in \mathrm{bin}} O_i \ \ln(O_i / E_i)$$

Example of Chi-Square



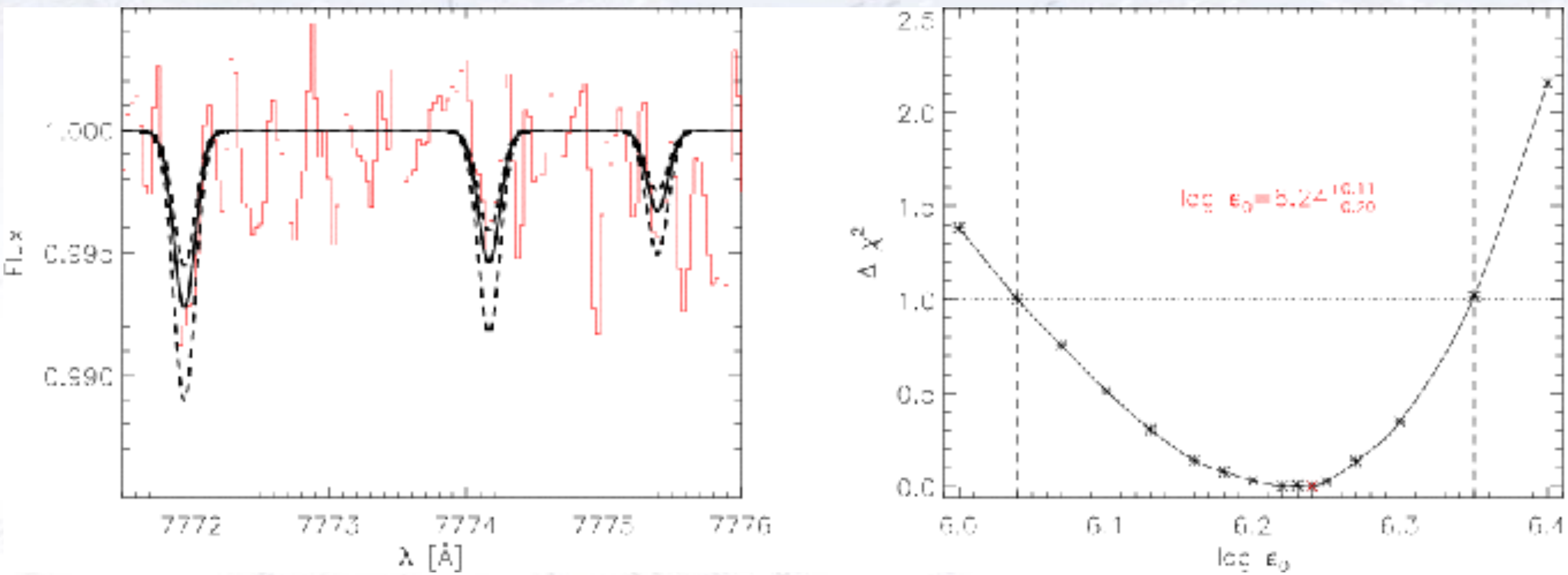
The fact that there are several minima makes fitting difficult/uncertain!

Always give good starting values!!!

Why the ChiSquare is (near) magic

Example of Chi-Square

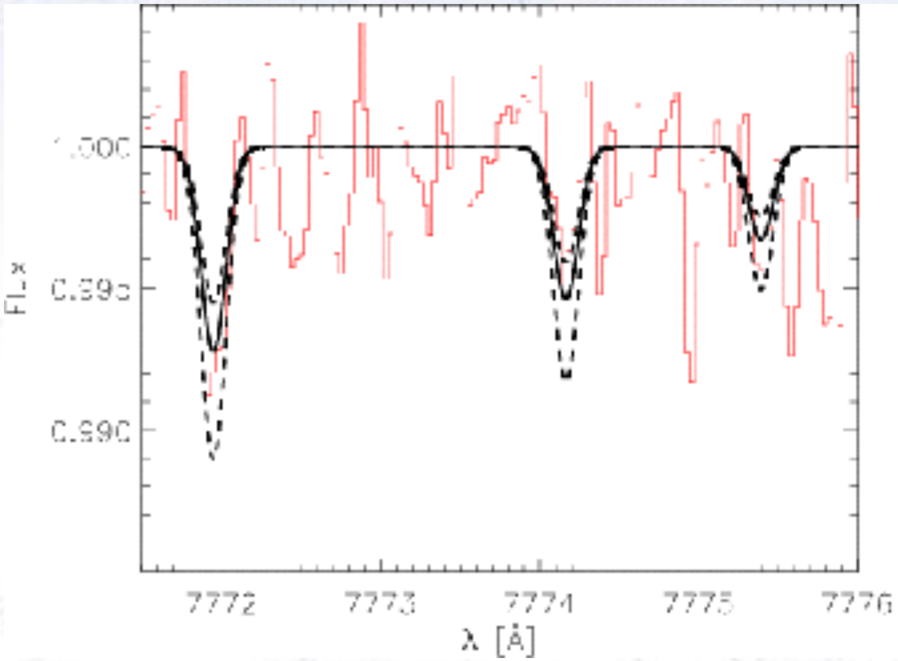
Uncertainties need not always be symmetric (though that is usually better!)

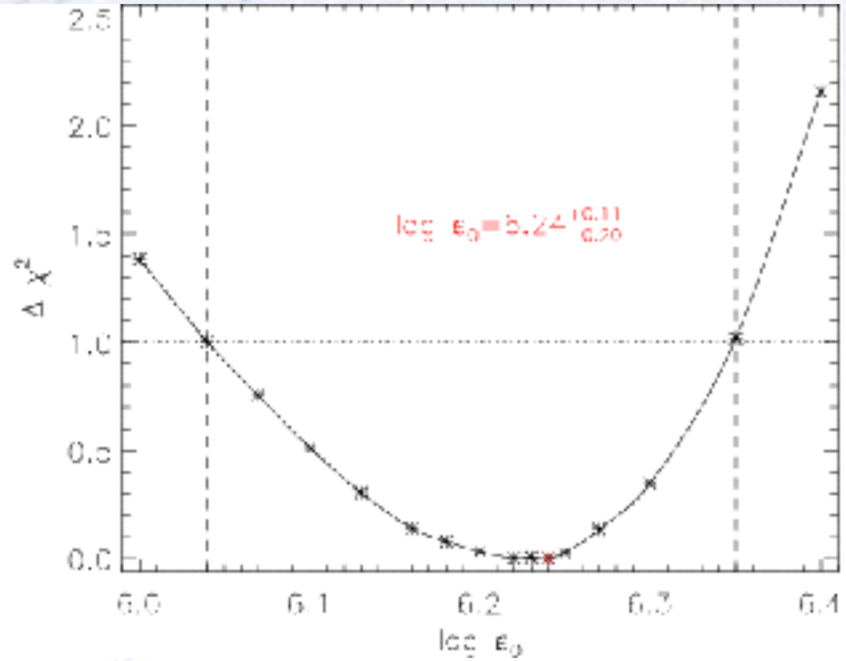


The uncertainty on a parameter is found where the Chi2 has increased by 1 from the minimum.

Example of Chi-Square

Uncertainties need not always be symmetric (though that is usually better!)

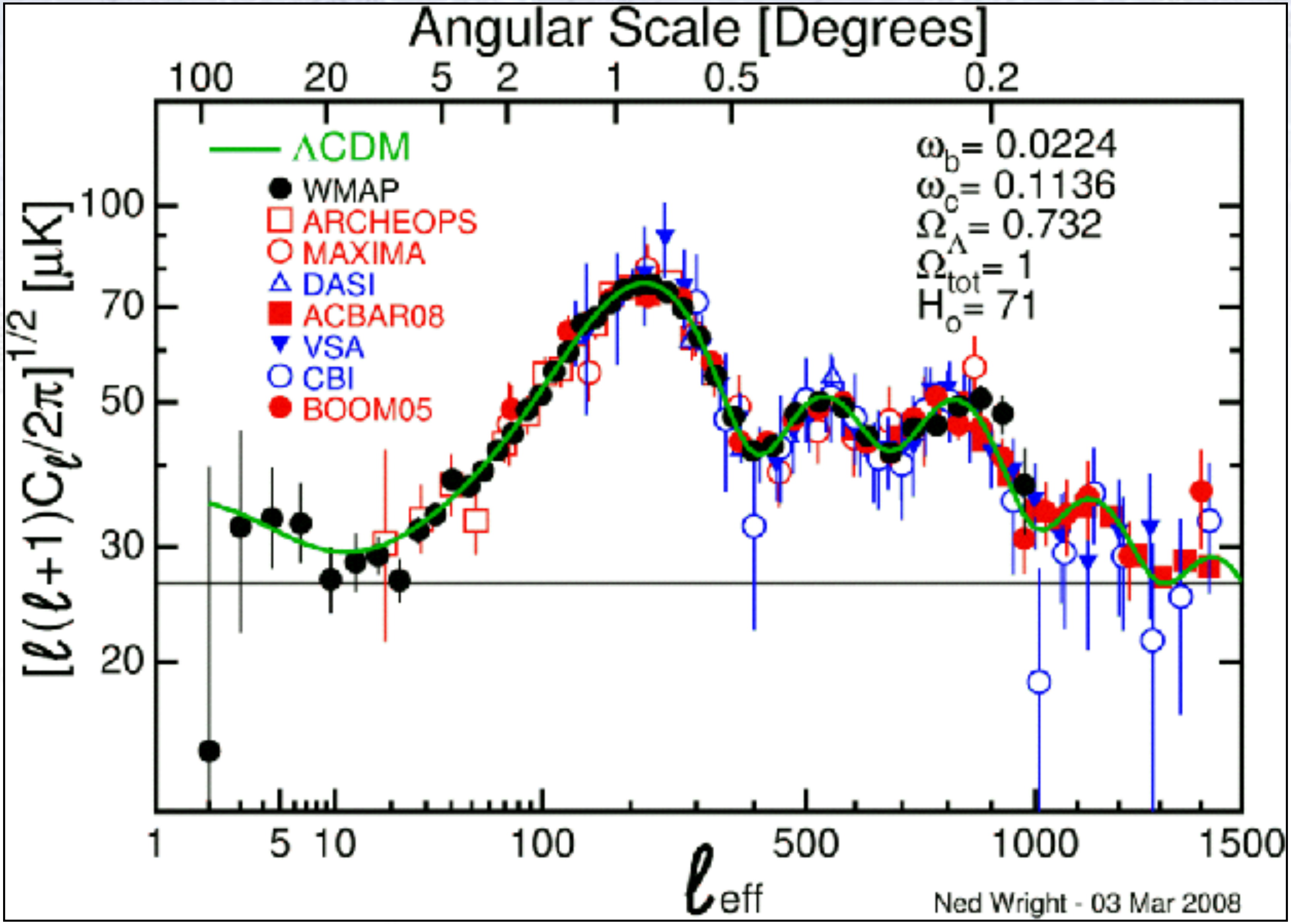




Please commit to memory!

The uncertainty on a parameter is found where the Chi2 has increased by 1 from the minimum.

Example of Chi-Square



Notes on the ChiSquare method

"It was formerly the custom, and is still so in works on the theory of observations, to derive the method of least squares from certain theoretical considerations, the assumed normality of the errors of the observations being one such.
It is however, more than doubtful whether the conditions for the theoretical validity of the method are realised in statistical practice, and the student would do well to regard the method as recommended chiefly by its comparative simplicity and by the fact that it has ***stood the test of experience****"*.

[G.U. Yule and M.G. Kendall 1958]