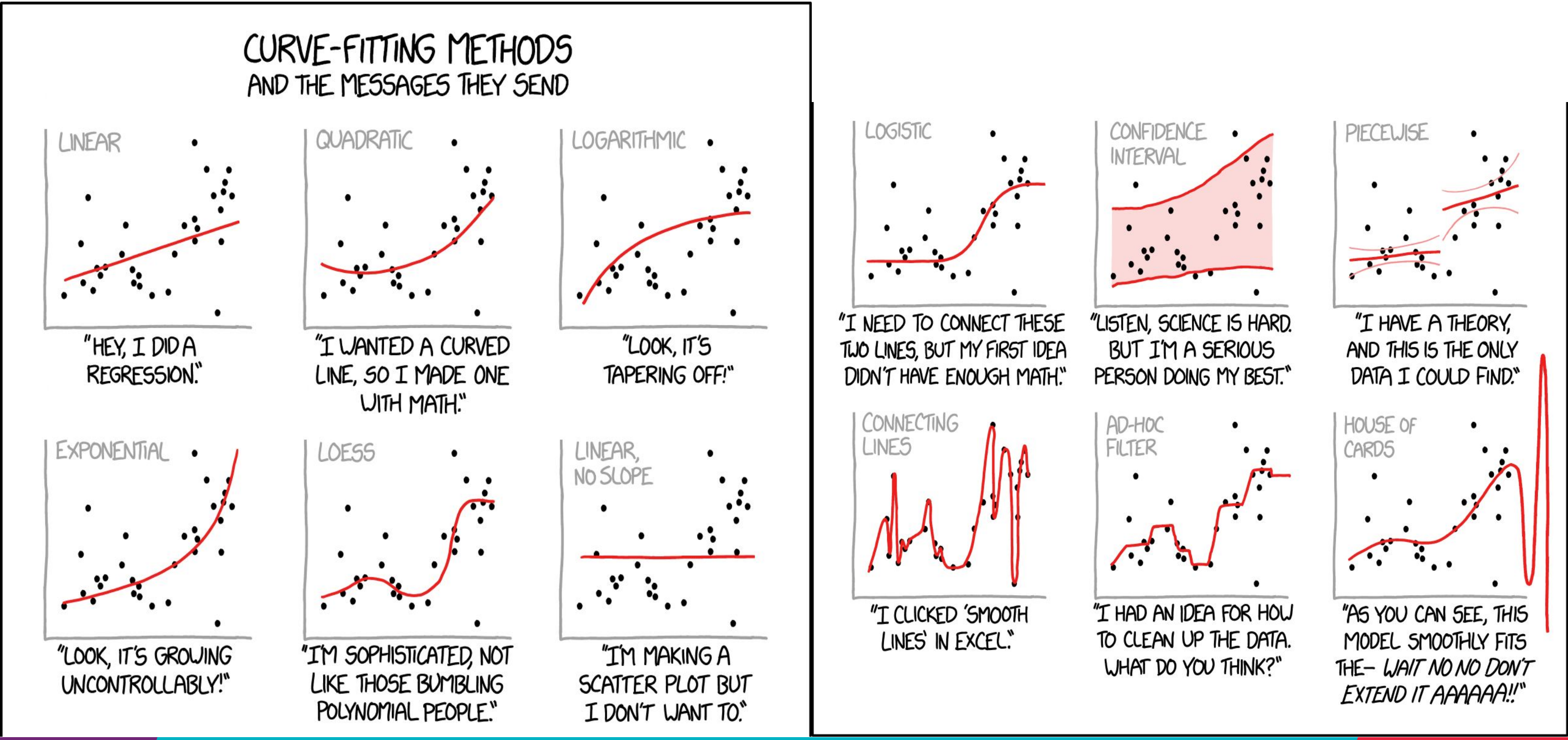
Goodness of fit

Badness of fit?

Lydia Brenner


CURVE-FITTING METHODS
AND THE MESSAGES THEY SEND
LINEAR
"HEY, I DID A REGRESSION."
QUADRATIC
"I WANTED A CURVED LINE, SO I MADE ONE WITH MATH."
LOGARITHMIC
"LOOK, IT'S TAPERING OFF!"
EXPONENTIAL
"LOOK, IT'S GROWING UNCONTROLLABLY!"
LOESS
"I'M SOPHISTICATED, NOT LIKE THOSE BUMBLING POLYNOMIAL PEOPLE."
LINEAR, NO SLOPE
"I'M MAKING A SCATTER PLOT BUT I DON'T WANT TO."
LOGISTIC
"I NEED TO CONNECT THESE TWO LINES, BUT MY FIRST IDEA DIDN'T HAVE ENOUGH MATH."
CONFIDENCE INTERVAL
"LISTEN, SCIENCE IS HARD. BUT I'M A SERIOUS PERSON DOING MY BEST."
PIECEWISE
"I HAVE A THEORY, AND THIS IS THE ONLY DATA I COULD FIND."
CONNECTING LINES
"I CLICKED 'SMOOTH LINES' IN EXCEL."
AD-HOC FILTER
"I HAD AN IDEA FOR HOW TO CLEAN UP THE DATA. WHAT DO YOU THINK?"
HOUSE OF CARDS
"AS YOU CAN SEE, THIS MODEL SMOOTHLY FITS THE- WAIT NO NO DON'T EXTEND IT AAAAAA!!"

The question: How well can I fit?

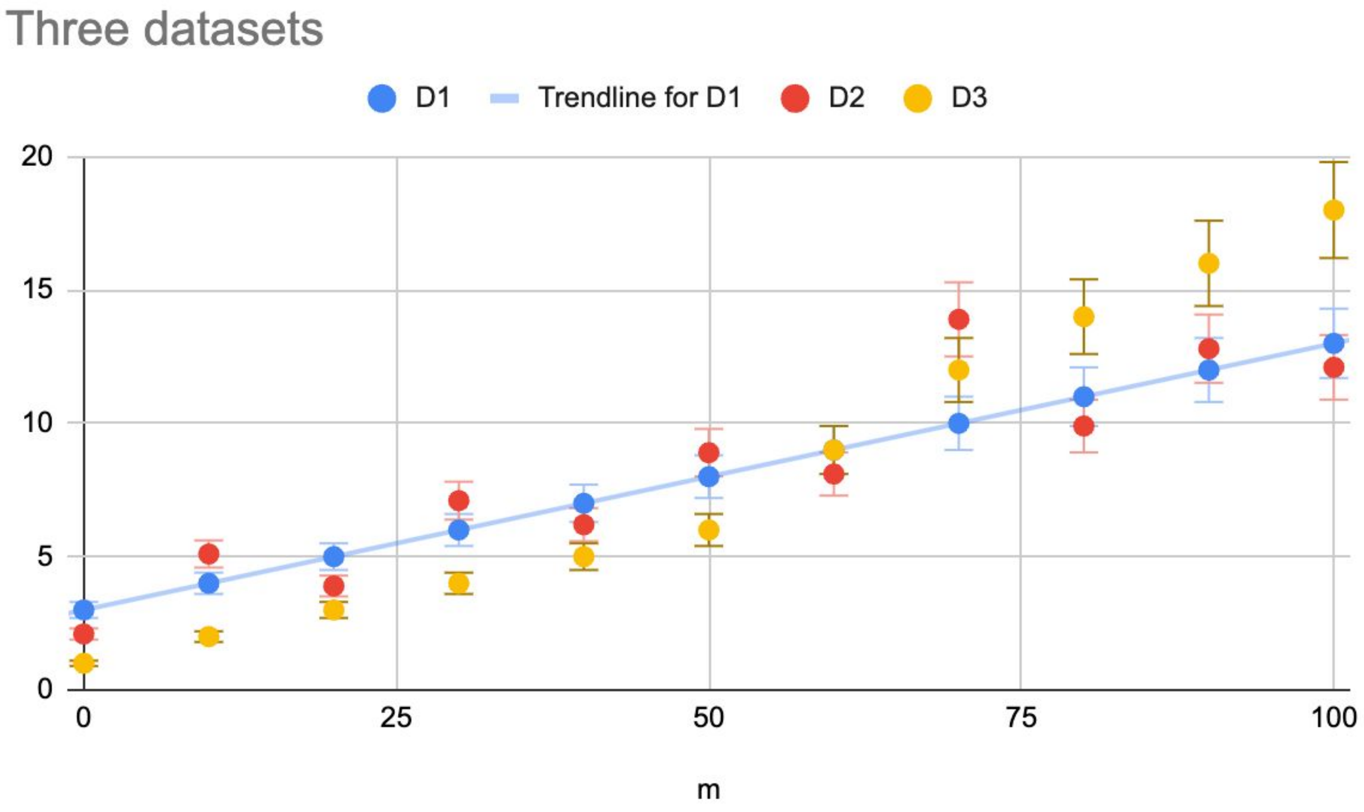
Basic question: how well does my hypothesis describe the data?

→ Would like a clearly understandable number

→ Would like it to match with visual input

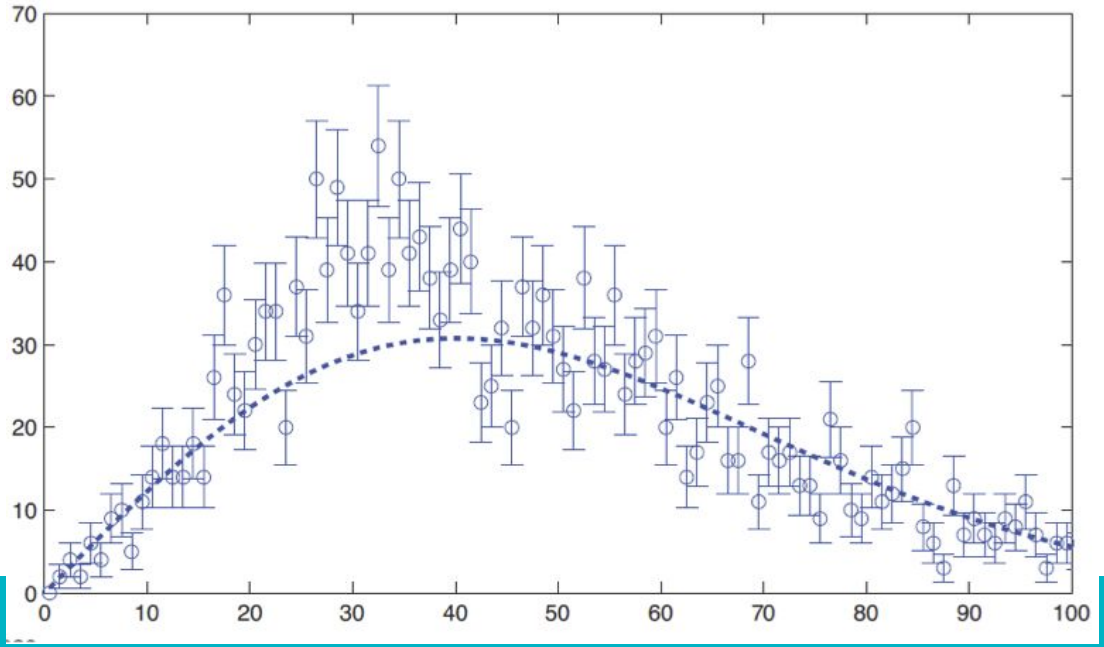
→ Would like it to have a meaningful interpretation in terms of the likelihood

The question: How well can I fit?



Is there a signal?

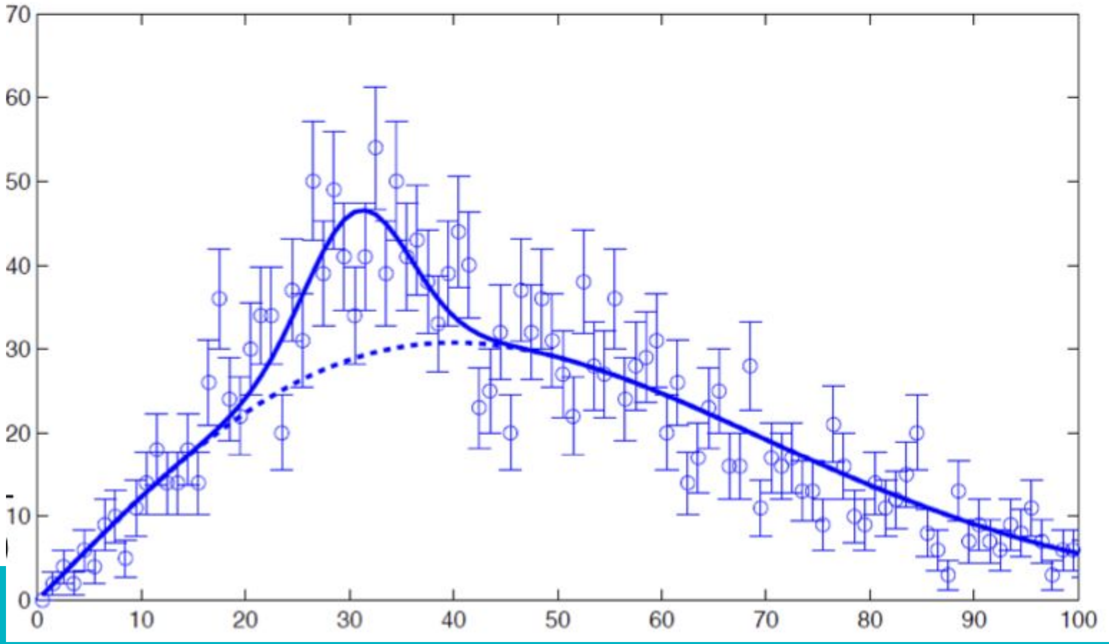
Try to distinguish background fluctuations from signals.



Nikhef

Is there a signal?

Looks like a signal around m=30 maybe



Nikhef

Is there a signal?

For GOF tests with binned data:

➔ compare observed event numbers n_i with expectation values f_i

Since no H1 specified there are many different GOF tests possible

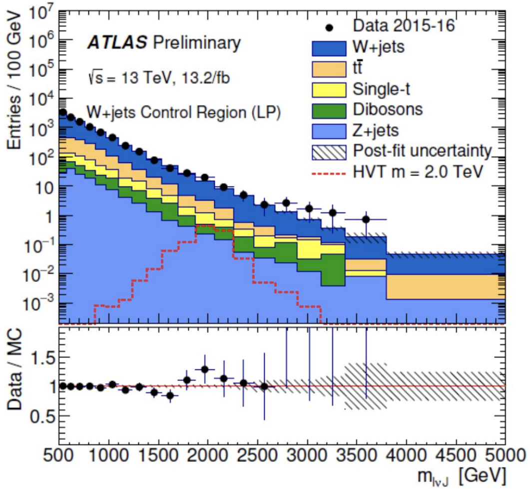
$$\chi^2 = \Sigma_i (f_i-n_i)^2/\sigma^2$$

➔ Basically does what you do by eye; Minimise distance from hypothesis to the data points

Do we use by eye?

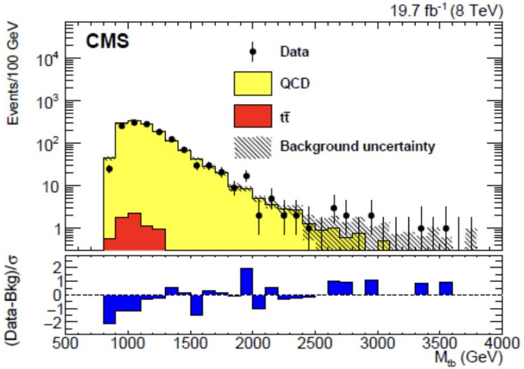
Yes!

ATLAS-CONF-2016-062 Diboson resonance

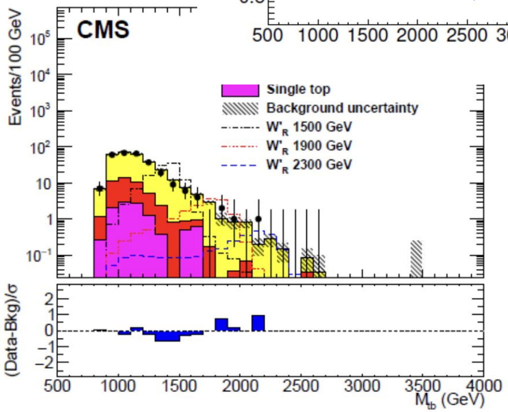


"Good agreement is observed (optical inspection of pulls) between the data and the background prediction"

arXiv:1509.06051v2 W′ → tb



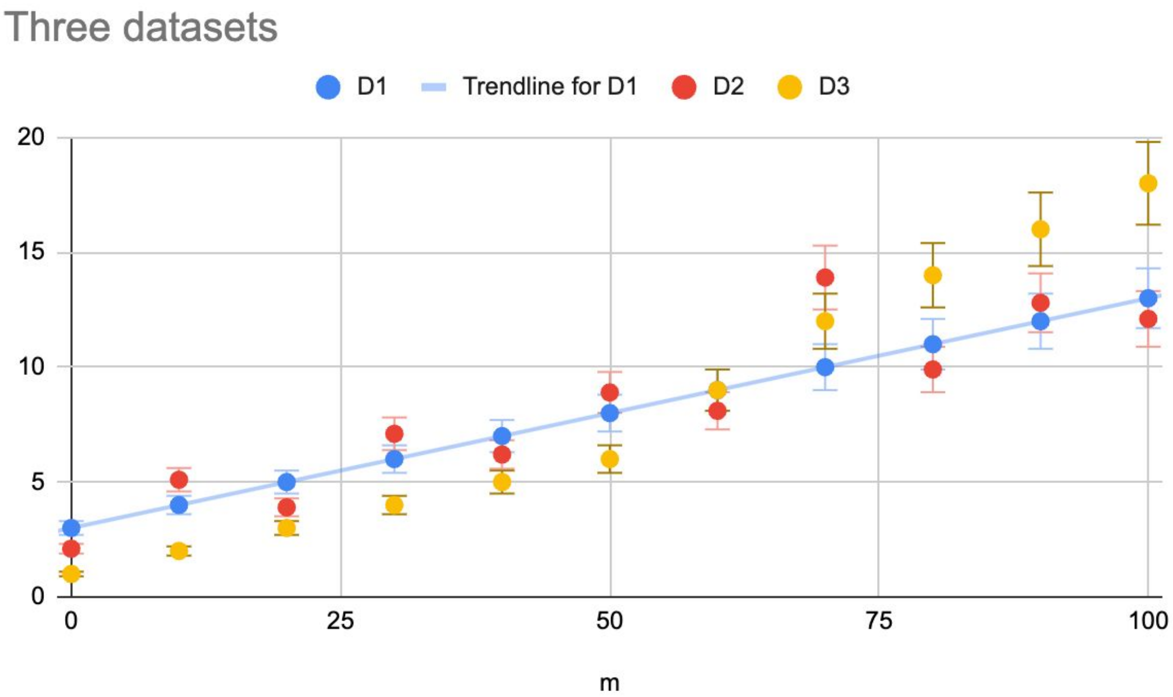
"This test (optical inspection) shows good agreement "



"Good agreement (optical inspection) between data and expectation from SM"

Nikhef

Is there a signal?



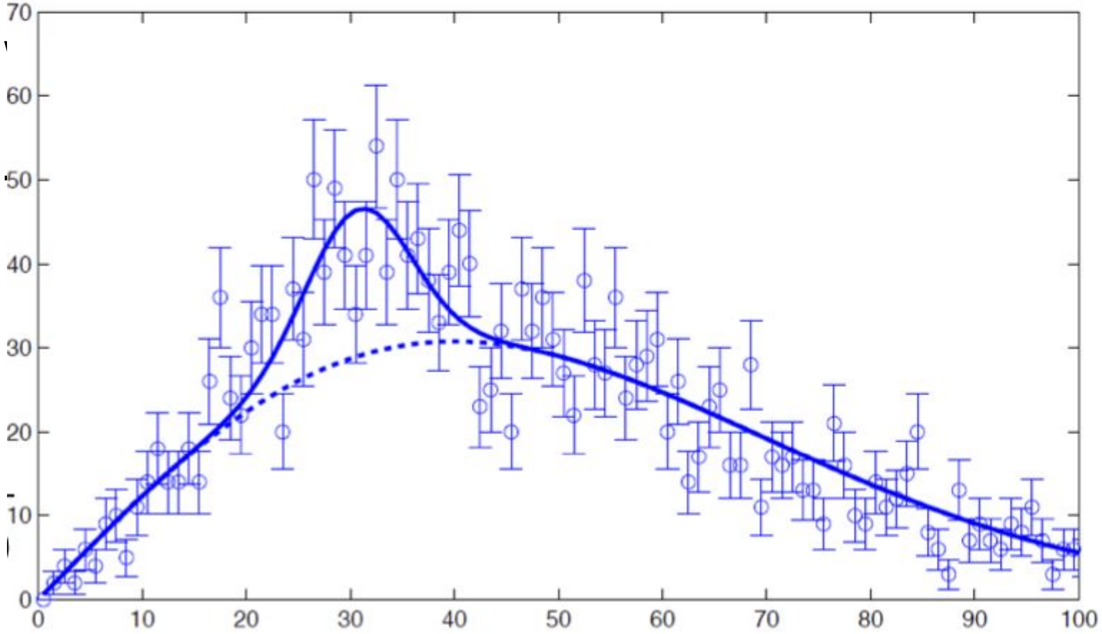
$$\chi^2 = \Sigma_i (f_i - n_i)^2/\sigma^2$$

χ^2 throws away all sign and order info

- ➔ Not very sensitive to correlated shifts in a certain region.
- ➔ Apply further GOF tests to check all data/model facets!

Nikhef

Is there a signal?

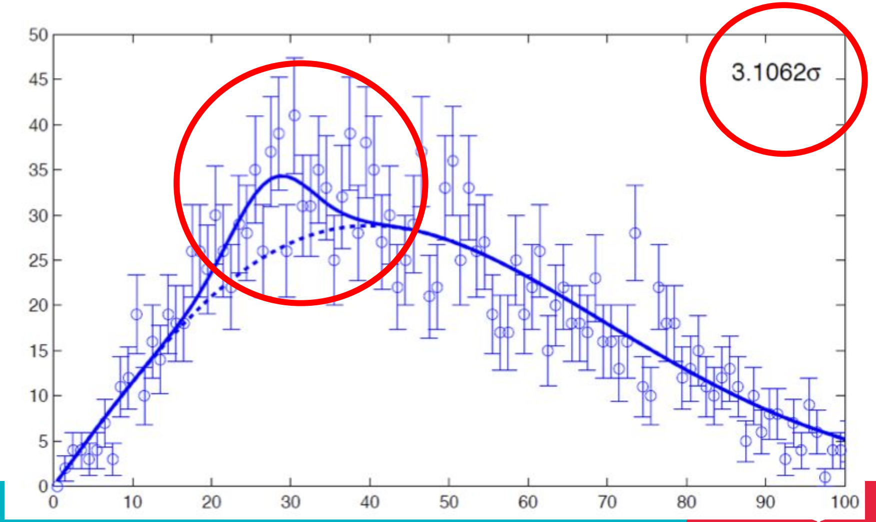


$$q_{fix,obs} = -2\ln\frac{L(b)}{L(\hat{\mu}s(m=30)+b)}$$

What about these signals?

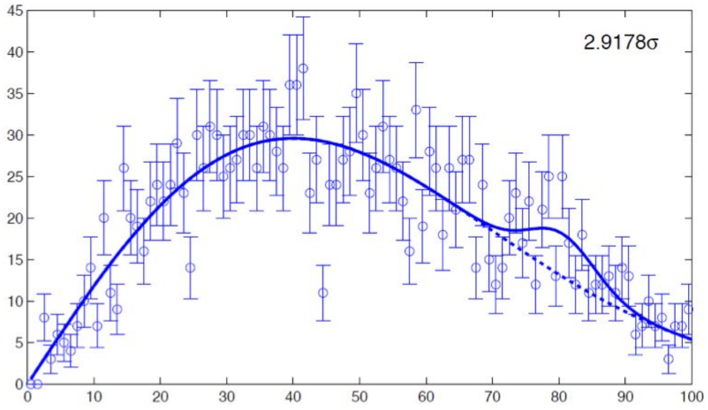
By eye; does not really look like a signal but: >3 sigma!

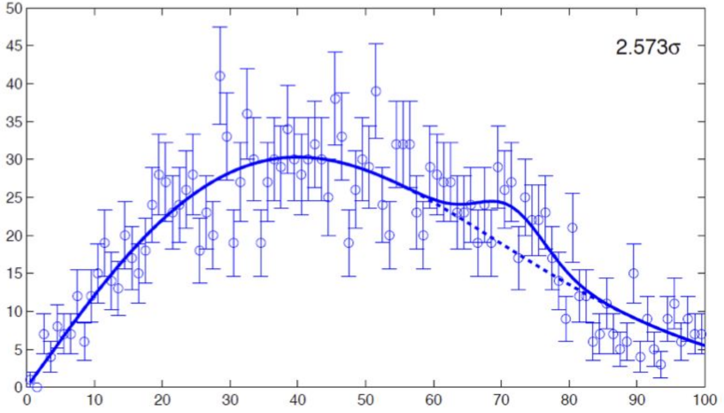
$$q_{fix,obs} = -2\ln\frac{L(b)}{L(\hat{\mu}s(m=30)+b)}$$

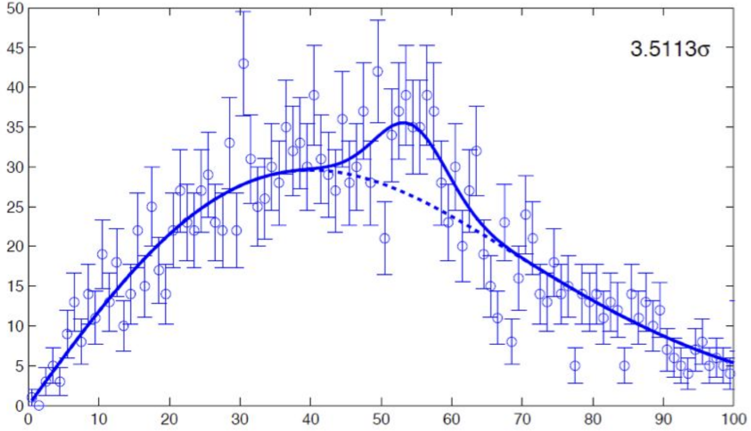


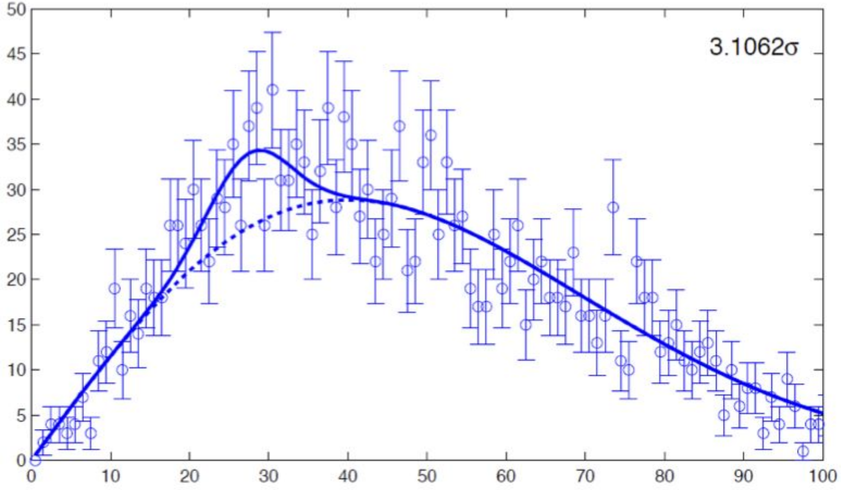
Nikhef

What about these signals?

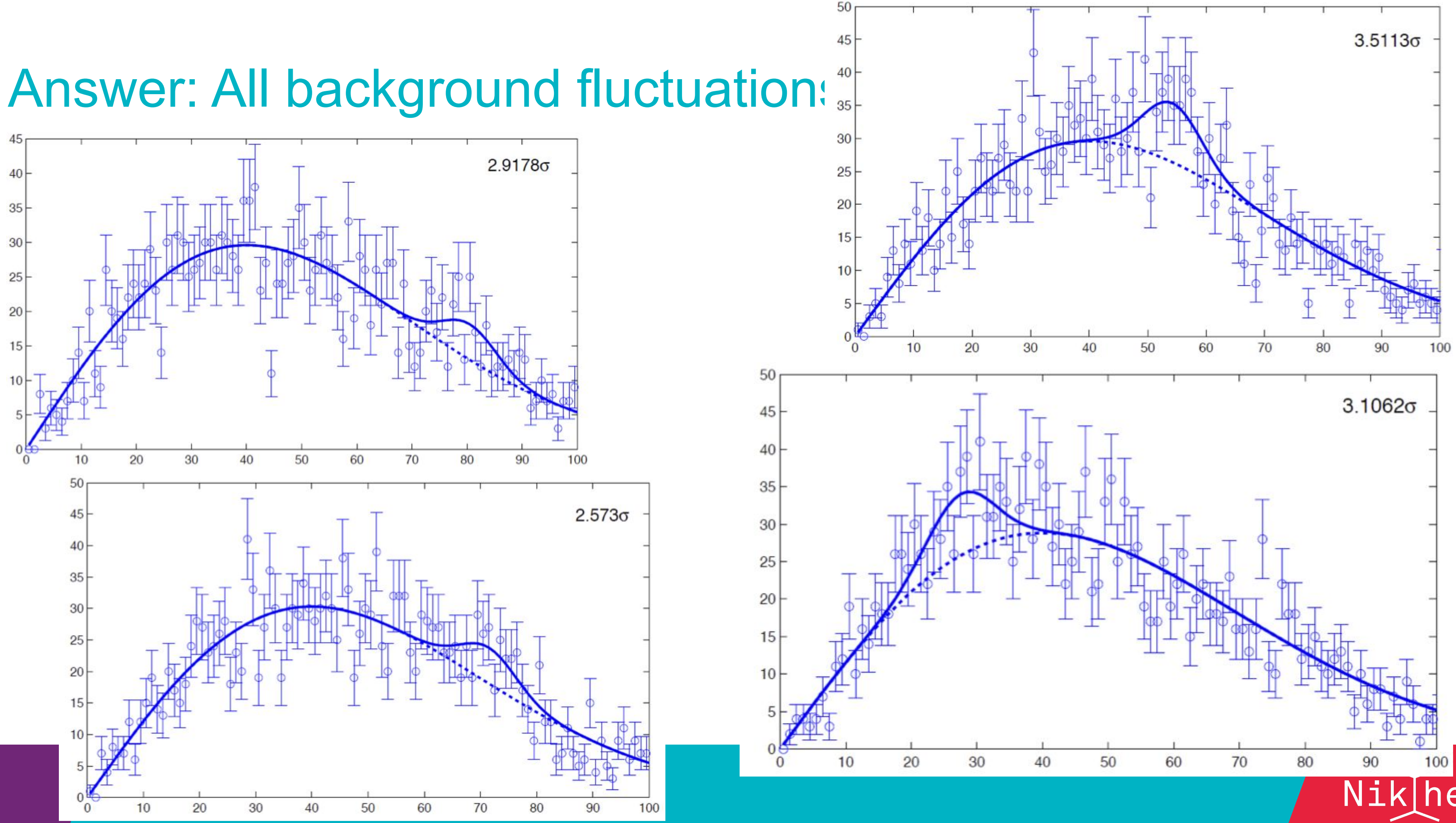






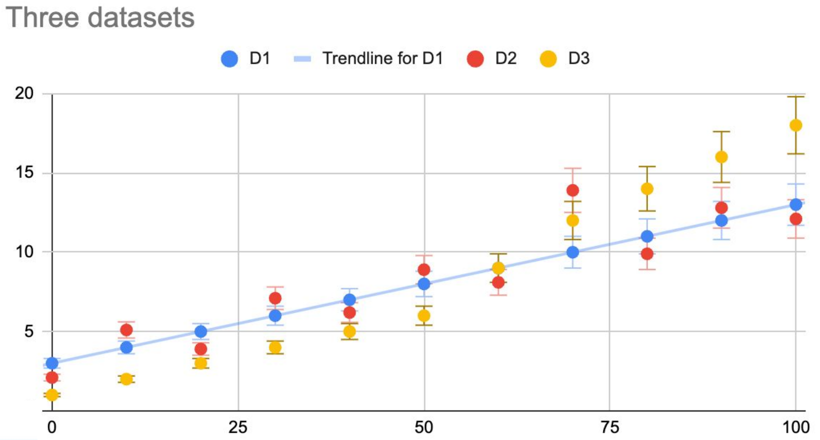


Answer: All background fluctuations



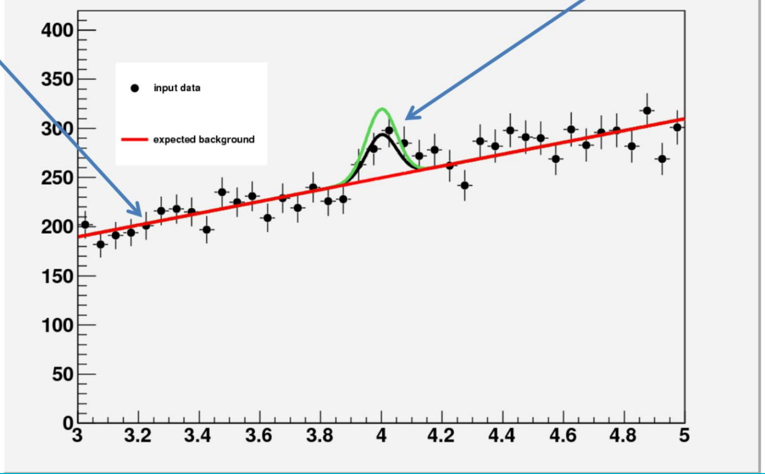
How many parameters do I need?

$\chi^2 = \Sigma_i (f_i - n_i)^2/\sigma^2$

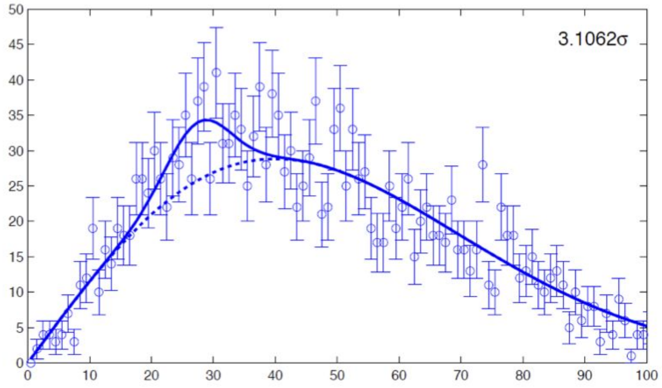


1) H_1: add more background pars

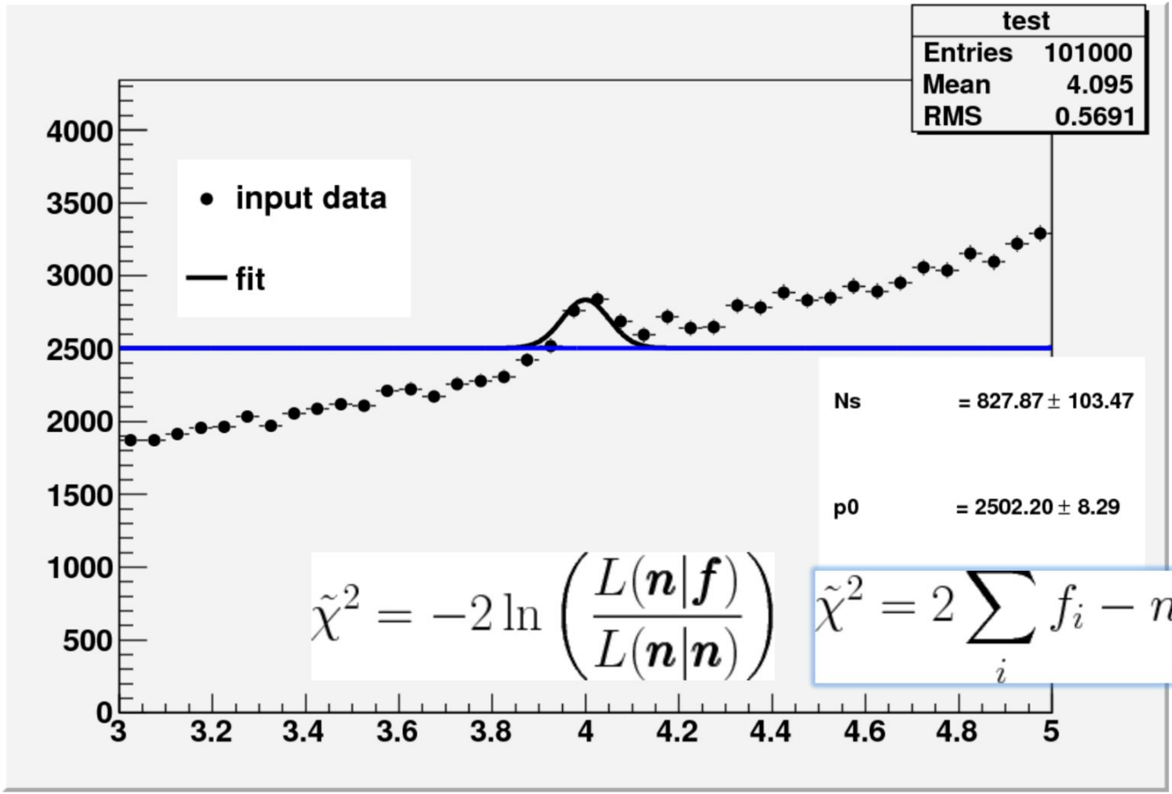
2) H_1: there is a signal



$$q_{fix,obs} = -2\ln\frac{L(b)}{L(\hat{\mu}s(m=30)+b)}$$



How many parameters do I need? – Example



Fit function:
gauss+p0

$$\chi^2 = \Sigma_i (f_i - n_i)^2/\sigma^2$$

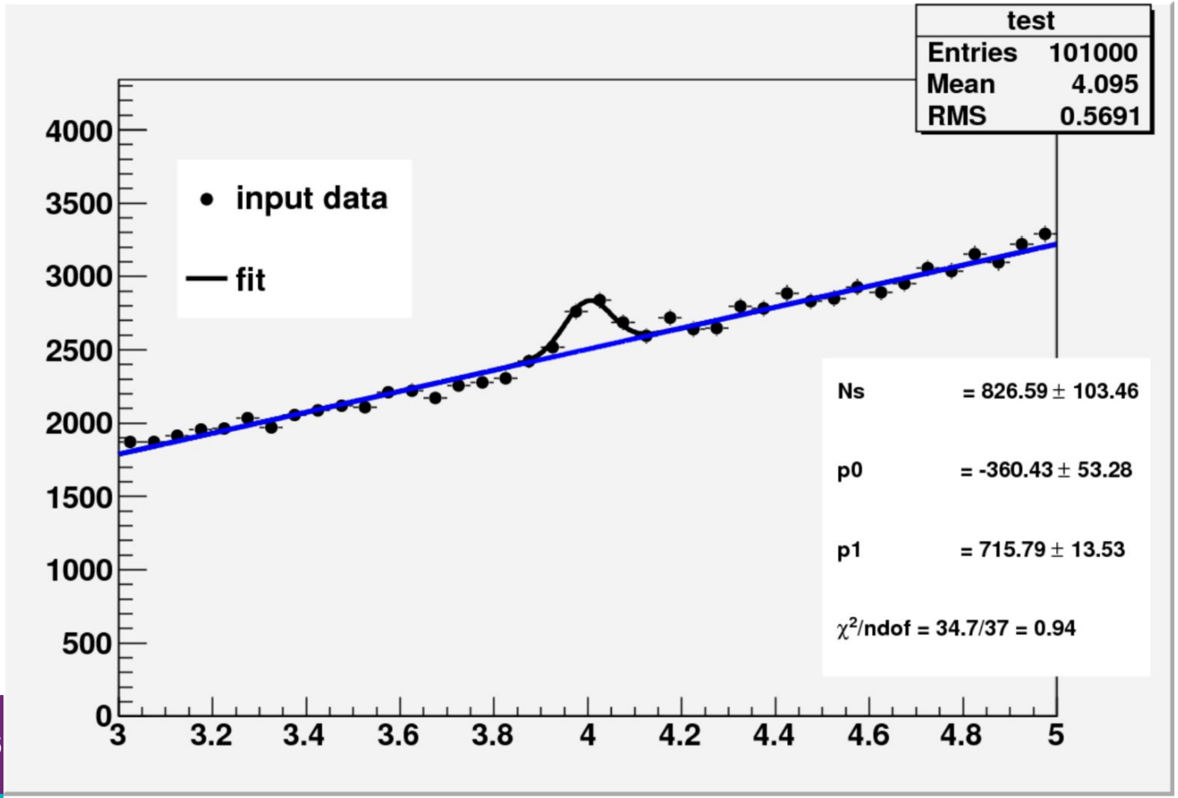
$$\tilde{\chi}^2 = -2\ln\left(\frac{L(\boldsymbol{n}|\boldsymbol{f})}{L(\boldsymbol{n}|\boldsymbol{n})}\right)$$

$$\tilde{\chi}^2 = 2\sum_i f_i - n_i + n_i \ln(n_i/f_i)$$

$$\tilde{\chi}^2 = 2880$$

Nikhef

How many parameters do I need? – Example

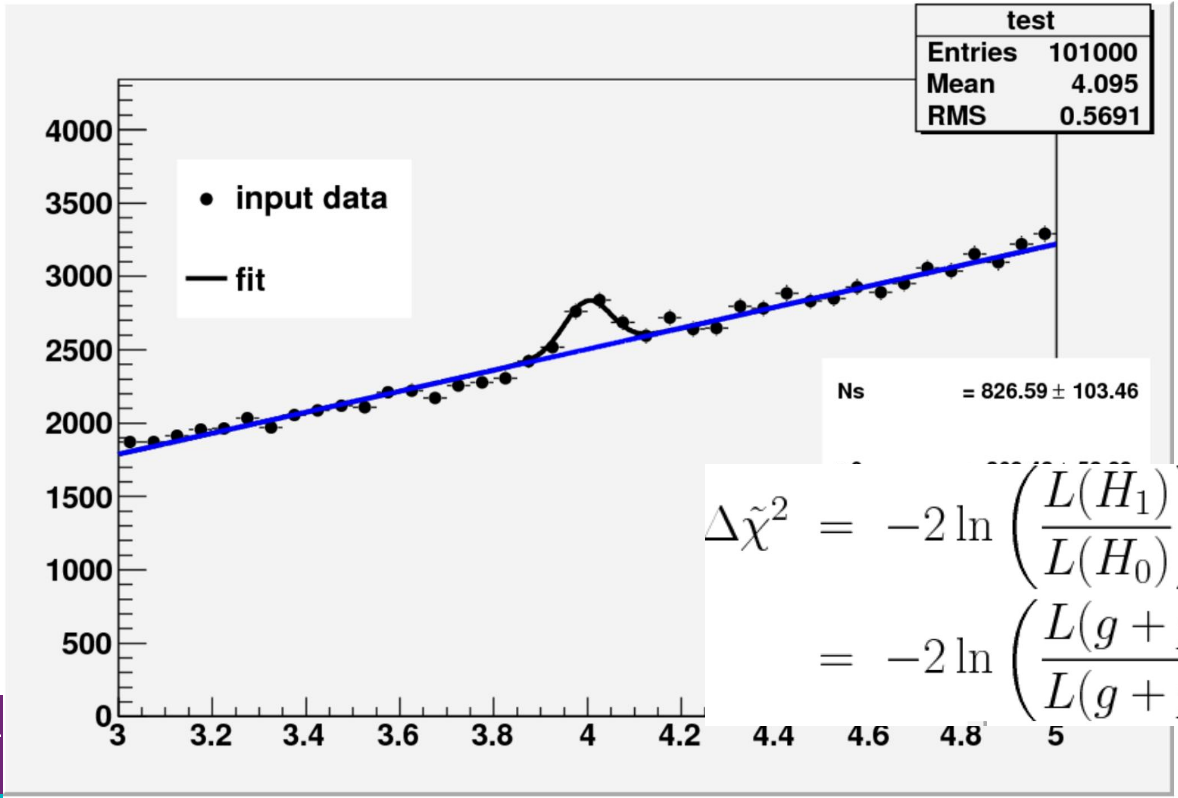


Fit function: gauss+p1

$$\tilde{\chi}^2 = 34.7$$

Should we stop here?

How many parameters do I need? – Example



Fit function: gauss+p1

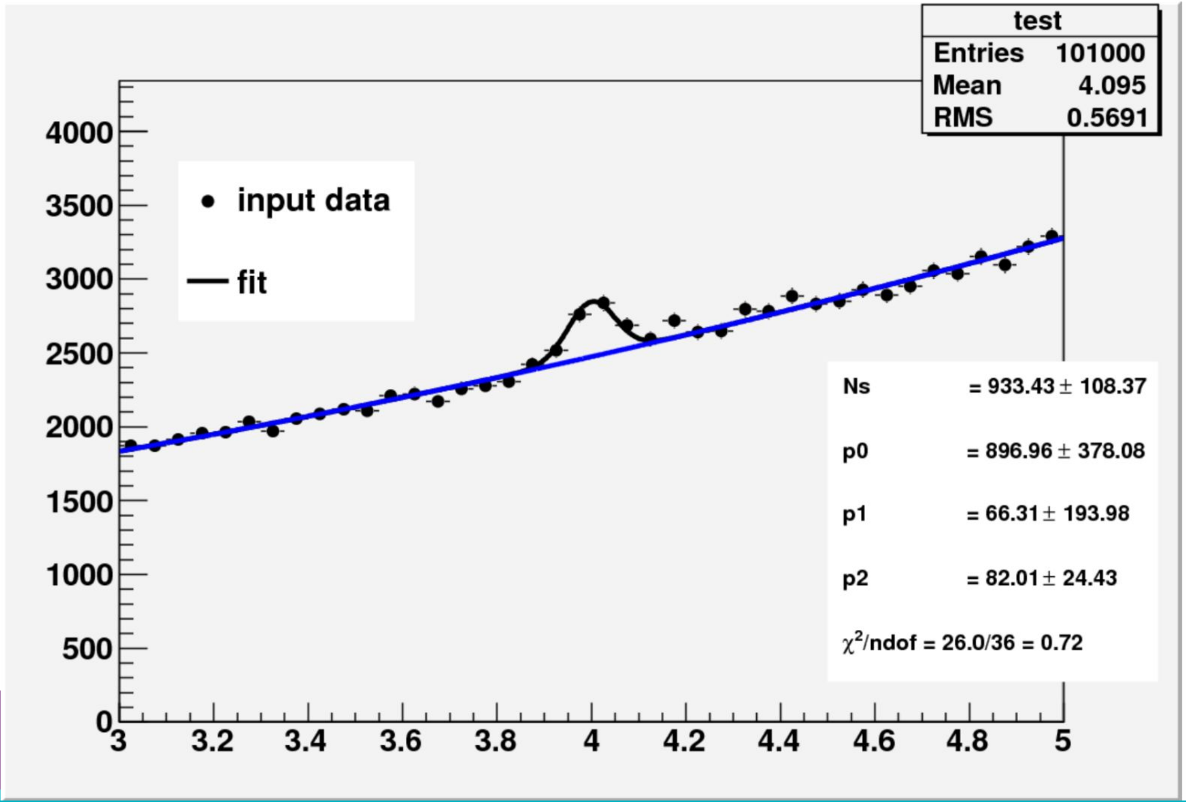
$$\tilde{\chi}^2 = 34.7$$

Should we stop here?

$$\Delta\tilde{\chi}^2 = -2\ln\left(\frac{L(H_1)}{L(H_0)}\right) = -2\ln\left(\frac{L(g+p_1)}{L(g+p_0)}\right) = -2845.3$$

Nikhef

How many parameters do I need? – Example



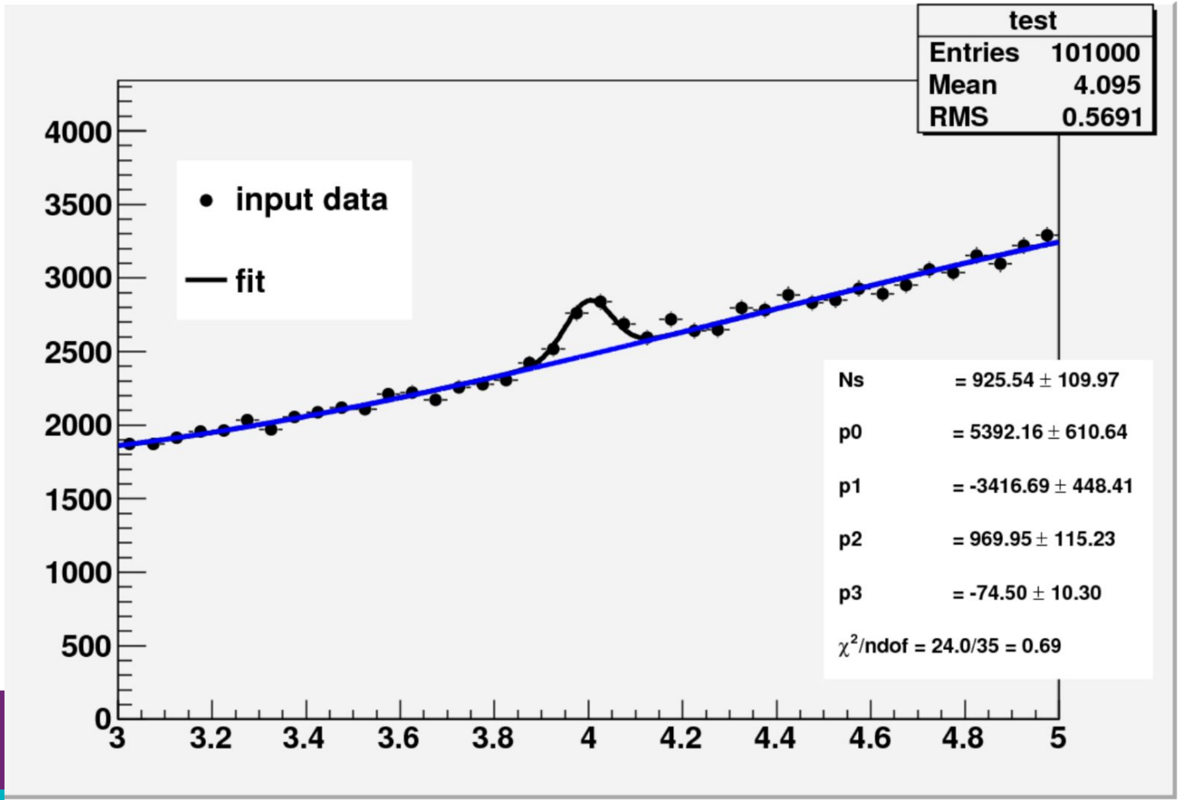
Fit function:
gauss+p2

$$\tilde{\chi}^2 = 26.0$$

$$\Delta\tilde{\chi}^2 = -8.7$$

Should we stop here?

How many parameters do I need? – Example



Fit function: gauss+p3

$$\tilde{\chi}^2 = 24.0$$

$$\Delta\tilde{\chi}^2 = -2.0$$

Please stop!

How many parameters do I need? – Example

If H0 correct then according to Wilks' theorem: $-\Delta\chi^2$ should follow a χ^2 function with ndf=1 (in asymptotic regime of large n)

Note: p-values for χ^2: TMath::Prob(χ2 obs,ndf)

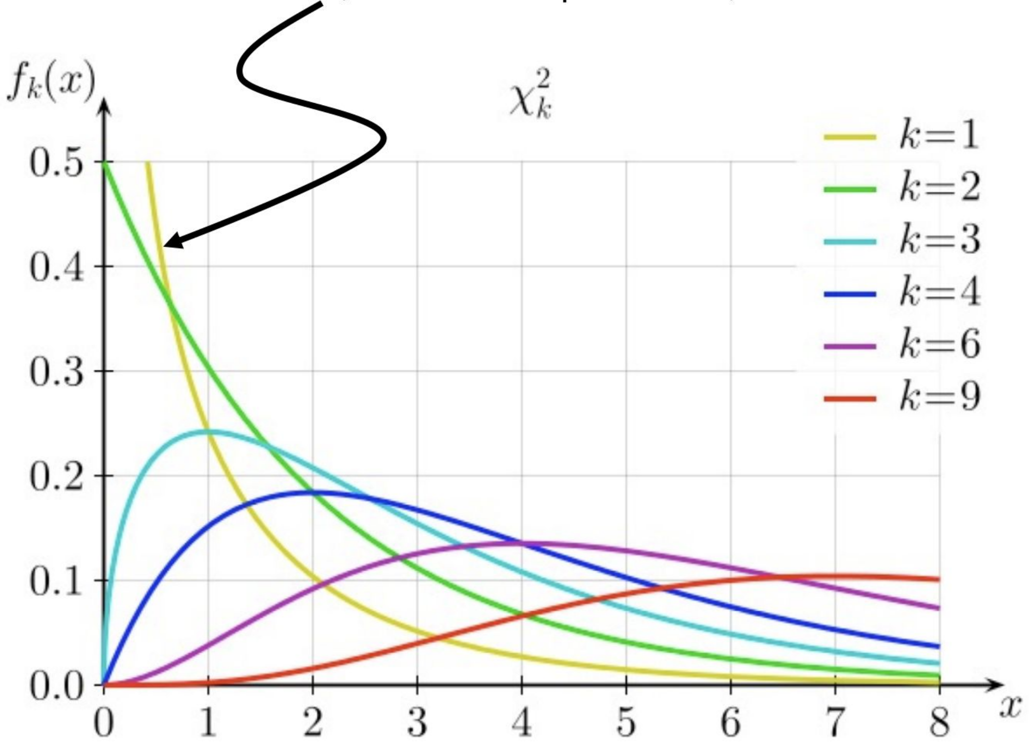
g+p0	$\tilde{\chi}^2 = 2880$		
g+p1	$\tilde{\chi}^2 = 34.7$	$\Delta\tilde{\chi}^2 = -2845.3$	
g+p2	$\tilde{\chi}^2 = 26.0$	$\Delta\tilde{\chi}^2 = -8.7$	P value = 0.003 Favoured over g+p1
g+p3	$\tilde{\chi}^2 = 24.0$	$\Delta\tilde{\chi}^2 = -2.0$	P value = 0.15 Not favoured over g+p2

χ^2

Remember:

What does a χ^2 distribution look like for n=1?

- Note that it for n=1, it does not peak at 1, but rather at 0…



Samuel S. Wilks
(1906-1964)

Wilks theorem

- H0: Additional parameters (as predicted by H1) not needed
- If H0 correct then according to Wilks' theorem: $-\Delta\chi^2 = -2\ln[L(H1)/L(H0)]$ should follow for $n \to \infty$ χ^2 function with ndf = #added parameters

Wilks' theorem only applies for nested hypotheses:

H0: 1st order polynomial H1: 2nd order polynomial ✓

H0: 1st order polynomial H1: a·exp(bx+cx2) ✕

Nikhef

Improving on χ^2

Likelihood ratio is an improved χ^2 – S. Baker & R.D. Cousins, NIM 221 (1984) 437

- ➔ Still a single number
- ➔ "Optimal estimator" e.g. parameter estimation
- ➔ Requires a second hypothesis
- ➔ Allows taking uncertainties into account systematically!!
  - ◆ What if there's a correlation?
  - ◆ What if there's a systematic uncertainty

See Wouter's slides from yesterday

Alternatives to Likelihoods and χ^2

Kolgomorov-Smirnov test

F_c: Cumulative distribution function

F_e: Empirical distribution function

$$q_{GoF,KS} = \sup \left| F_c(x) - F_e(x) \right|$$

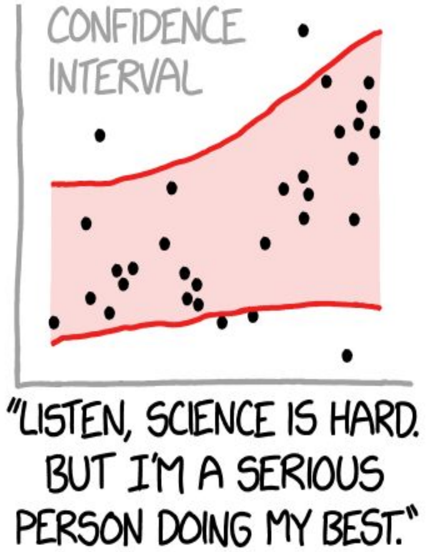
Anderson-Darling test

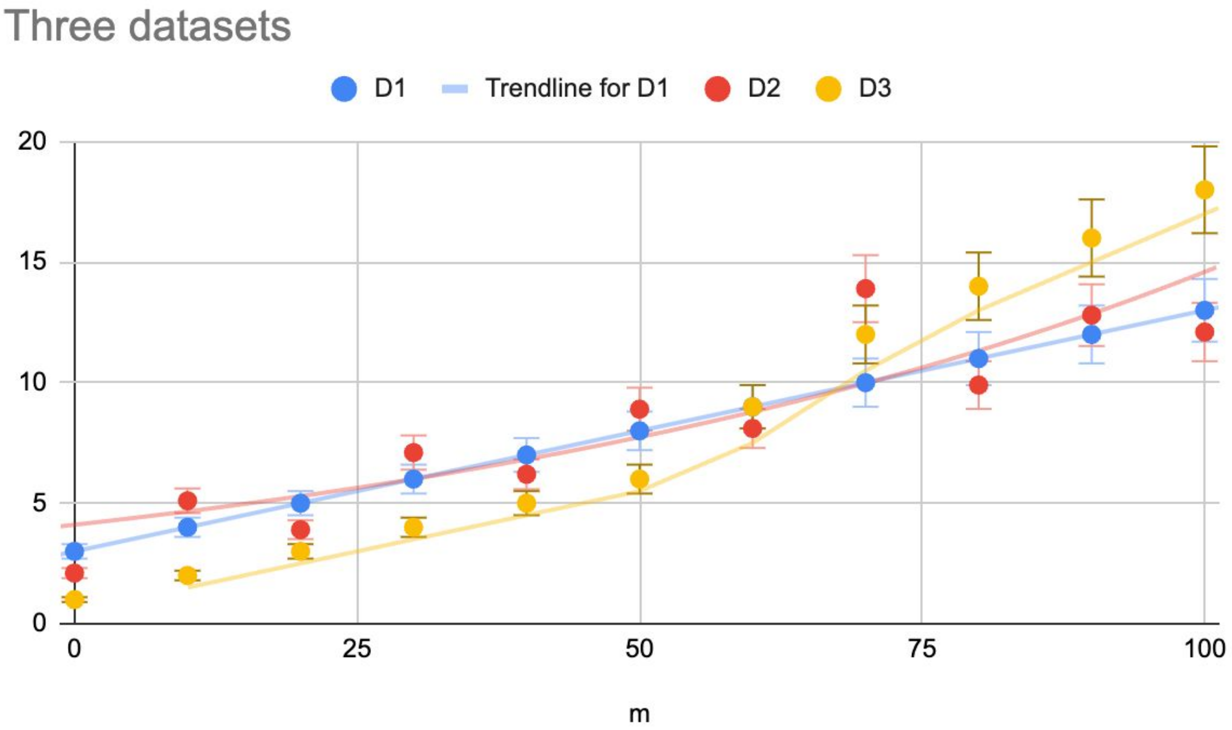
$$q_{GoF,AD} = n \cdot \int dF_e(x) \frac{(F_c(x) - F_e(x))^2}{F_e(x) \cdot (1 - F_e(x))}$$

Nikhef

Summary

Not too bad…





Nikhef