Lecture 3:

- Uncertainties & Error Propagation
- Testing Models: chi-squared
- “Scientific Method”

Uncertainty ~~Error~~ Propagation

The thing you want to measure → $$f(\mathbf{q}) = f(q_1, q_2, \ldots q_n)$$ ← Dependent parameters (*e.g.* temperature, position, time, pressure...)

Want to use the distribution f to propagate uncertainties in **q**, but

1) We don't necessarily know the full joint distribution of **q** (*i.e.* the probability distribution for all possible sets of values)
2) Even if we did, it's cumbersome to deal with!

So, instead, let's approximate things to first order and then estimate the variance of f

$$f(\mathbf{q}) \sim f(\boldsymbol{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial f}{\partial q_i} \right]_{\mathbf{q}=\boldsymbol{\mu}} (q_i - \mu_i)$$

Taylor expansion about the mean values for $\boldsymbol{q}$

where $$f(\boldsymbol{\mu}) = f(\mu_1, \mu_2, \ldots \mu_n)$$

$$f(\mathbf{q}) \sim f(\boldsymbol{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial f}{\partial q_i}\right]_{\mathbf{q}=\boldsymbol{\mu}} (q_i - \mu_i)$$

zero by definition

$$\langle f(\mathbf{q}) \rangle \sim f(\boldsymbol{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial f}{\partial q_i}\right]_{\mathbf{q}=\boldsymbol{\mu}} \overbrace{\langle q_i - \mu_i \rangle} = f(\boldsymbol{\mu})$$

$$\langle f^2(\mathbf{q}) \rangle \sim \left\langle \left[f(\boldsymbol{\mu}) + \sum_{i=1}^{n} \left[\frac{\partial f}{\partial q_i}\right]_{\mathbf{q}\,\boldsymbol{\mu}} (q_i - \mu_i) \right]^2 \right\rangle$$

$$= f^2(\boldsymbol{\mu}) + \left\langle \left[\sum_{i=1}^{n} \left[\frac{\partial f}{\partial q_i}\right]_{\mathbf{q}=\boldsymbol{\mu}} (q_i - \mu_i) \right] \left[\sum_{j=1}^{n} \left[\frac{\partial f}{\partial q_j}\right]_{\mathbf{q}=\boldsymbol{\mu}} (q_j - \mu_j) \right] \right\rangle$$

zero by definition

$$+ 2f(\boldsymbol{\mu}) \sum_{i=1}^{n} \left[\frac{\partial f}{\partial q_i}\right]_{\mathbf{q}=\boldsymbol{\mu}} \overbrace{\langle q_i - \mu_i \rangle}$$

$$= f^2(\boldsymbol{\mu}) + \sum_{i,j=1}^{n} \left[\frac{\partial f}{\partial q_i}\frac{\partial f}{\partial q_j}\right]_{\mathbf{q}=\boldsymbol{\mu}} \left\langle (q_i - \mu_i)(q_j - \mu_j) \right\rangle$$

So we get:

$$\sigma_f^2 = \left\langle f^2(\mathbf{q}) \right\rangle - \langle f(\mathbf{q}) \rangle^2$$

$$\simeq \sum_{i,j=1}^{n} \left[\frac{\partial f}{\partial q_i} \frac{\partial f}{\partial q_j} \right]_{\mathbf{q}=\boldsymbol{\mu}} \left\langle (q_i - \mu_i)(q_j - \mu_j) \right\rangle$$

"covariance matrix"

$$= \sum_{i,j=1}^{n} \left[\frac{\partial f}{\partial q_i} \frac{\partial f}{\partial q_j} \right]_{\mathbf{q}=\boldsymbol{\mu}} V_{ij}$$

If the q parameters are uncorrelated,

$$V_{ij} = \left\langle (q_i - \mu_i)(q_j - \mu_j) \right\rangle = 0 \quad for \quad i \neq j$$

$$V_{ii} = \left\langle (q_i - \mu_i)^2 \right\rangle = \sigma_i^2$$

$$\sigma_f^2 \simeq \sum_{i=1}^{n} \left[\frac{\partial f}{\partial q_i} \right]_{\mathbf{q}=\boldsymbol{\mu}}^2 \sigma_i^2$$

for independent parameter uncertainties

Some simple examples:

$$T_{tot} = t_1 + t_2$$

$$\sigma_T^2 \simeq \left[\frac{\partial T}{\partial t_1}\right]^2 \sigma_1^2 + \left[\frac{\partial T}{\partial t_2}\right]^2 \sigma_2^2 = \sigma_1^2 + \sigma_2^2$$

$$s = vt$$

$$\sigma_s^2 \simeq \left[\frac{\partial s}{\partial v}\right]^2 \sigma_v^2 + \left[\frac{\partial s}{\partial t}\right]^2 \sigma_t^2 = t^2\sigma_v^2 + v^2\sigma_t^2$$

$$\left(\frac{\sigma_s}{s}\right)^2 \simeq \left(\frac{\sigma_v}{v}\right)^2 + \left(\frac{\sigma_t}{t}\right)^2$$

For a quadrature addition of uncertainties, uncertainties that are half as big only carry 1/4 of the weight, and uncertainties that are 1/4 as big only carry 1/16 of the weight... **Only the dominant uncertainties matter!**

More General Example: Measurement of Linear Thermal Expansion Coefficient

Measure h by timing the drop of snowballs on one particularly cold day, then compare with h_0 to determine α

data point	Time (s)	Temp (°C)
1	2.02	-5.6
2	1.99	-4.8
3	2.05	-4.4
.	.	.
.	.	.
n	2.01	-5.3

In this simple analysis, we're interested in determining the average values of drop time and temperature for the day:

$$t_\mu \sim \tilde{t}_\mu = \frac{1}{n}\sum_{i=1}^{n} t_i \qquad T_\mu \sim \tilde{T}_\mu = \frac{1}{n}\sum_{i=1}^{n} T_i$$

Then, from the relation $h = h_0[1 - \alpha(T - T_0)]$

estimate the expansion coefficient:

$$\tilde{\alpha} = \frac{\frac{\tilde{h}_\mu}{h_0} - 1}{\tilde{T}_\mu - T_0} = \frac{\frac{g\tilde{t}_\mu^2}{2h_0} - 1}{\tilde{T}_\mu - T_0}$$

Now we want to find the uncertainty in $\tilde{\alpha}$ by propagating the uncertainties in $\tilde{t}_\mu$ and $\tilde{T}_\mu$

data point	Time (s)	Temp (°C)
1	2.02	-5.6
2	1.99	-4.8
3	2.05	-4.4
.	.	.
.	.	.
n	2.01	-5.3

$$V_{ij} =$$

	time	Temp
time	σ_t^2	$\text{cov}(t, T)$
Temp	$\text{cov}(t, T)$	σ_T^2

Approach 1: Evaluate Approximately Using the Data

$$\sigma_t^2 = \left\langle (t - t_\mu)^2 \right\rangle \sim \frac{1}{n-1} \sum_{i=1}^{n} (t_i - \tilde{t}_\mu)^2$$

$$\sigma_T^2 = \left\langle (T - T_\mu)^2 \right\rangle \sim \frac{1}{n-1} \sum_{i=1}^{n} (T_i - \tilde{T}_\mu)^2$$

$$\text{cov}(t, T) = \left\langle (t - t_\mu)(T - T_\mu) \right\rangle \sim \frac{1}{n-1} \sum_{i=1}^{n} (t_i - \tilde{t}_\mu)(T_i - \tilde{T}_\mu)$$

Drawback: Requires a large enough data set so that estimates are well determined

data point	Time (s)	Temp (ºC)
1	2.02	-5.6
2	1.99	-4.8
3	2.05	-4.4
.	.	.
.	.	.
n	2.01	-5.3

$V_{ij} =$

	time	Temp
time	σ_t^2	$\text{cov}(t, T)$
Temp	$\text{cov}(t, T)$	σ_T^2

Approach 2: Use Calibration Measurements and/or Physical Models

$\sigma_t^2 = \left\langle (t - t_\mu)^2 \right\rangle$ from calibration of timing accuracy

$\sigma_T^2 = \left\langle (T - T_\mu)^2 \right\rangle$ from calibration of temperature reading accuracy

Temperature variations during the day are sufficiently small that the correlation with time measurements is very weak, so $\text{cov}(t, T) \sim 0$

Drawback: Model could be wrong

The temperature variations might influence reaction times and this might have a noticeable systematic impact on the stopwatch measurements

data point	Time (s)	Temp (°C)
1	2.02	-5.6
2	1.99	-4.8
3	2.05	-4.4
.	.	.
.	.	.
n	2.01	-5.3

$V_{ij} =$

	time	Temp
time	σ_t^2	$\text{cov}(t, T)$
Temp	$\text{cov}(t, T)$	σ_T^2

Best Approach: **Do both!**

Check the consistency or your model and calibrations with the data
(If things don't add up, dig around to understand it!)

$$\alpha = \frac{\frac{gt^2}{2h_0} - 1}{T - T_0}$$

$$\left[\frac{\partial \alpha}{\partial t}\right]_{t_\mu, T_\mu} = \frac{\frac{gt_\mu}{h_0} - 1}{T_\mu - T_0}$$

$$\left[\frac{\partial \alpha}{\partial T}\right]_{t_\mu, T_\mu} = -\frac{\frac{gt_\mu^2}{2h_0} - 1}{(T_\mu - T_0)^2}$$

$$\sigma_\alpha^2 = \left(\left[\frac{\partial \alpha}{\partial t}\right]_{t_\mu, T_\mu}, \left[\frac{\partial \alpha}{\partial T}\right]_{t_\mu, T_\mu} \right) \begin{pmatrix} \sigma_t^2 & \mathrm{cov}(t, T) \\ \mathrm{cov}(t, T) & \sigma_T^2 \end{pmatrix} \begin{pmatrix} \left[\frac{\partial \alpha}{\partial t}\right]_{t_\mu, T_\mu} \\ \left[\frac{\partial \alpha}{\partial T}\right]_{t_\mu, T_\mu} \end{pmatrix}$$

$$\sigma_\alpha^2 = \left[\frac{\partial \alpha}{\partial t}\right]_{\boldsymbol{\mu}}^2 \sigma_t^2 + \left[\frac{\partial \alpha}{\partial T}\right]_{\boldsymbol{\mu}}^2 \sigma_T^2 + 2\left[\frac{\partial \alpha}{\partial t}\frac{\partial \alpha}{\partial T}\right]_{\boldsymbol{\mu}} \mathrm{cov}(t, T)$$

The Statistical Calculation That You Should Have Done at the Start!

Typical linear expansion coefficients for building materials ~5×10^{-6} per ^{o}C

Take $(T-T_o) \sim 20^{o}C$

$h_0 - h \sim (20m)(5 \times 10^{-6})(20^{o}C) = 0.002m$

velocity at impact = $\sqrt{2gh_0} = \sqrt{2(9.8m/s^2)(20m)} \sim 20m/s$

So, timing must be known to an accuracy of (0.002/20) = 0.0001s

Accuracy of any one timing timing measurement ~ 0.1s

But we improve by averaging lots of measurements according to $\sigma_m = \frac{\sigma}{\sqrt{n}}$

How many measurements do we need?

$$n = \frac{\sigma^2}{\sigma_m^2} \sim \left(\frac{0.1}{0.0001}\right)^2 = 10^6$$

(ignoring systematic uncertainties!)

Statistical Uncertainties

Fundamental, calculable, random variations due to an inherent limited sampling of the underlying distribution (i.e. counting statistics).

Systematic Uncertainties

Incidental, estimated (bounded), systematic biases incurred as a result of limited measurement precision (also always present).

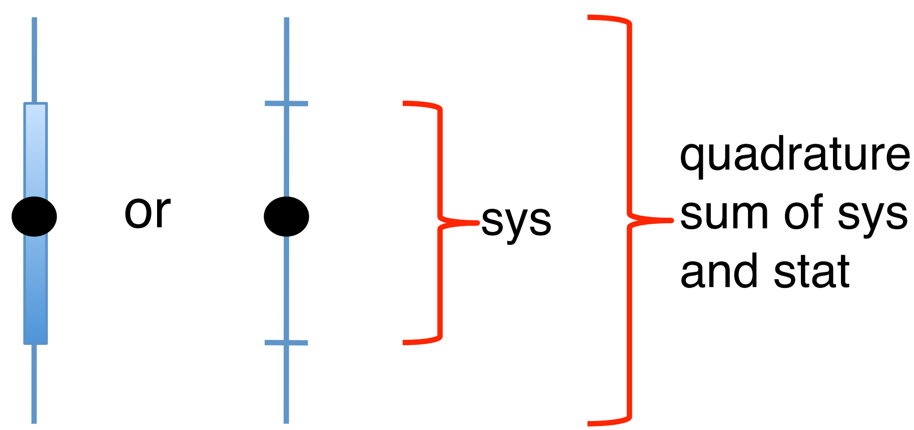
There is no universally applicable method for estimating/bounding* systematic uncertainties. A typical approach often relies on independent cross-checks, accounting for possible statistical limitations of calibration procedures, knowledge about the experimental design and general consistency arguments.

* Systematic errors that are "determined" become corrections!

Because of their very different nature, there is no standard, mathematically rigorous way to combine the 2 types of uncertainties. The convention is thus to quote results in the form:

Result ± Uncertainty (stat) ± Uncertainty (sys)

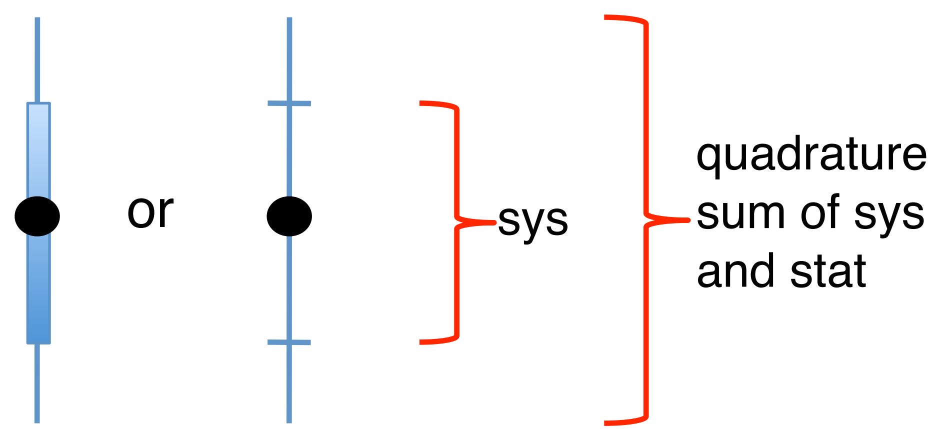
And error bars such as:



How do you then make use of such data points to fit a model?

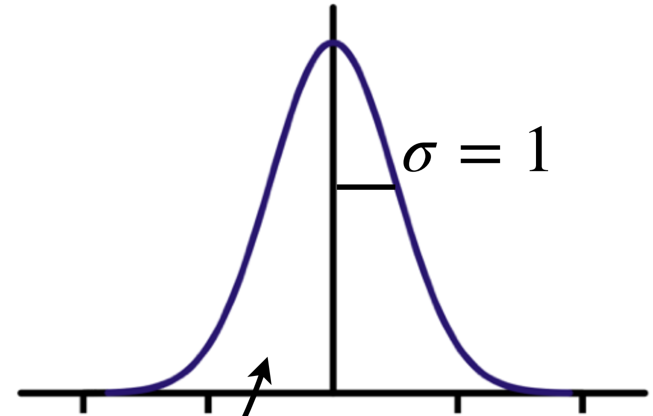
It is often generally assumed that systematic uncertainties can be treated in a similar way to statistical uncertainties, with careful attention to correlations.

Ideally, the best way to treat systematic uncertainties are as free parameters in the model fit, constrained by the separately determined bounds on their values.



Consider:

$$\chi^2 \equiv \sum_{i=1}^{n} g_i^2$$



where g_i are samples drawn from a normal (*i.e.* Gaussian) distribution of unit variance

Then the distribution of this quantity defines a χ^2 ("chi-squared") distribution with **n** *degrees of freedom*

effective number of independent samples contributing to the variance

The χ^2 probability density function for $\mathbf{n}$ degrees of freedom has the form:

$$P(\chi^2, n) = \frac{(\chi^2)^{\frac{n}{2}-1} e^{-\frac{\chi^2}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}$$

Where $\Gamma(k) = (k-1)!$ **or** $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$

if k = positive integer

for any real value z

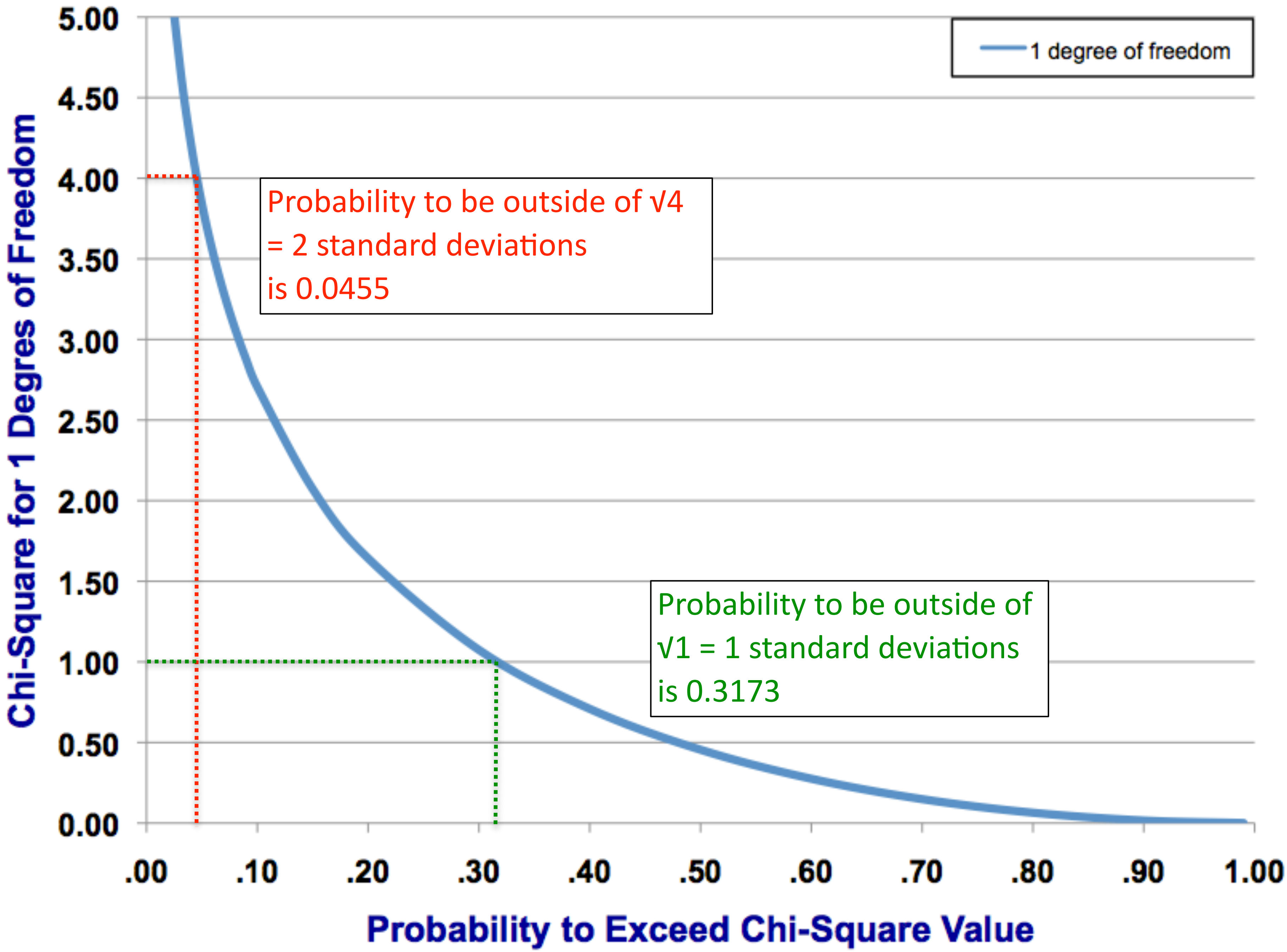
Note that: $P(\chi^2, 2) = \frac{1}{2} e^{-\frac{\chi^2}{2}}$

(will come back to this)

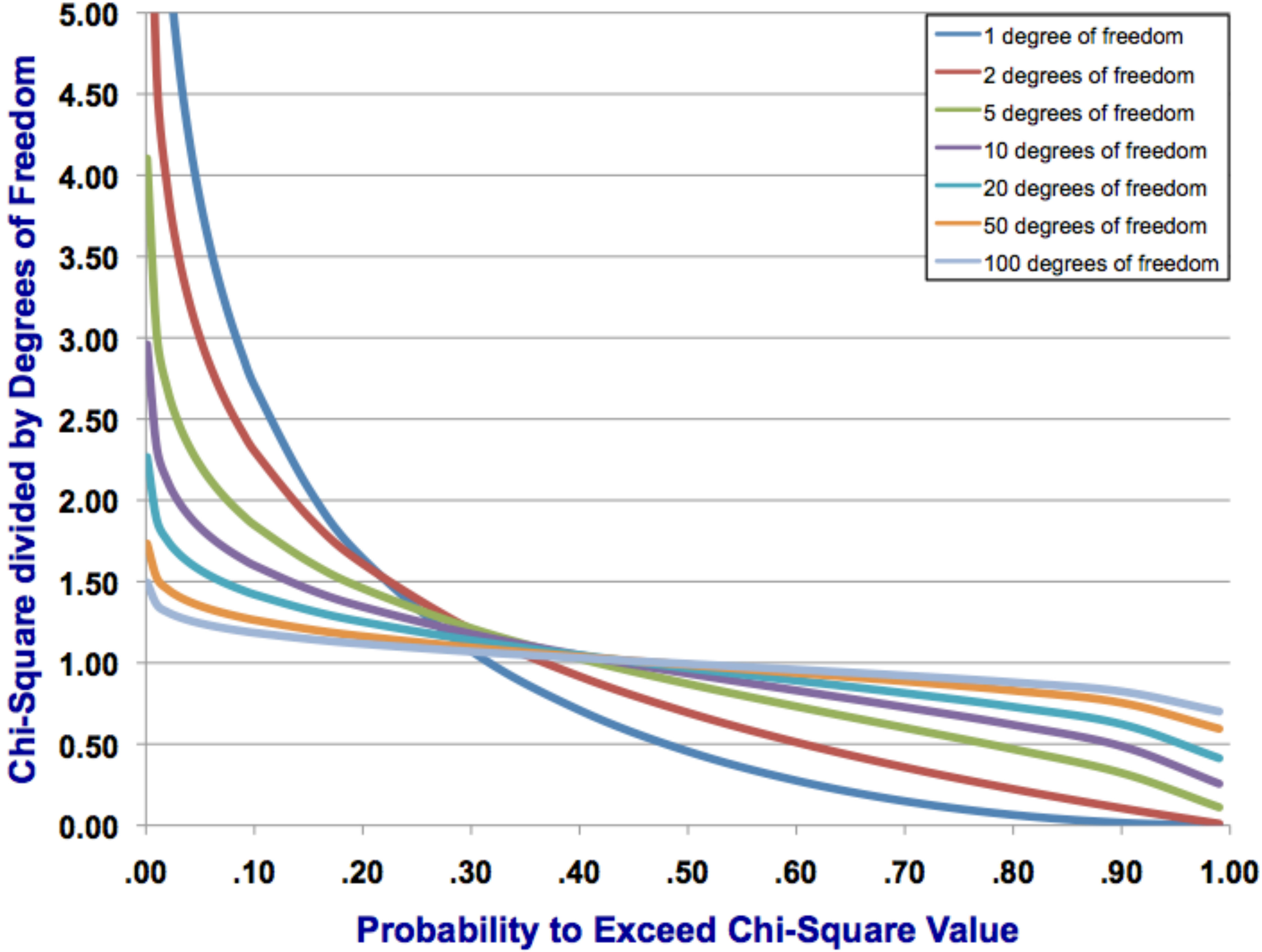
And the integral probability is given by:

$$P(>\chi^2, n) = 1 - \frac{\gamma\left(\frac{n}{2}, \frac{\chi^2}{2}\right)}{\Gamma(\frac{n}{2})}$$

where $\gamma(z, \alpha) = \int_0^\alpha x^{z-1} e^{-x} dx$



1 degree of freedom
Probability to be outside of √4
= 2 standard deviations
is 0.0455
Probability to be outside of
√1 = 1 standard deviations
is 0.3173
Chi-Square for 1 Degres of Freedom
Probability to Exceed Chi-Square Value
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
.00
.10
.20
.30
.40
.50
.60
.70
.80
.90
1.00



1 degree of freedom
2 degrees of freedom
5 degrees of freedom
10 degrees of freedom
20 degrees of freedom
50 degrees of freedom
100 degrees of freedom
Chi-Square divided by Degrees of Freedom
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
.00
.10
.20
.30
.40
.50
.60
.70
.80
.90
1.00
Probability to Exceed Chi-Square Value

Pearson's χ^2 Test

So, for example, if we have a model, **m**, involving **k** free parameters (determined by a fit to the data) that seeks to predict the values, **x**, of **n** data points, each with *normally distributed uncertainties*, we can construct the sum:

$$S \equiv \sum_{i=1}^{n} \frac{(x_i - m_i)^2}{\sigma_{m_i}^2}$$

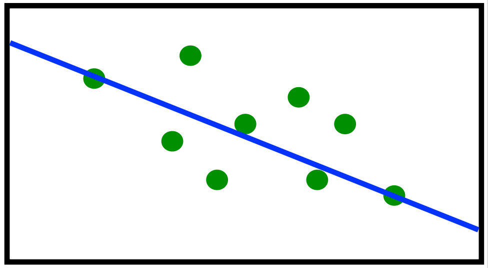
normalises things to give a Gaussian distribution with unit variance

for binned data (Poisson statistics)

$$\sigma_{m_i}^2 \cong m_i$$

S will then be distributed as a χ^2 distribution with **n-k** degrees of freedom, and can thus be used as a statistic to determine how well the model matches the data.

For example, imagine fitting a straight line (2 parameters: slope and intercept) to a set of data. You can always force the line to go through 2 of the data points exactly, so only n-2 of the data points will contribute to the variance around the model



S will then be distributed as a χ^2 distribution with **n-k** degrees of freedom, and can thus be used as a statistic to determine how well the model matches the data.

"If my model is correct, how often would a randomly drawn sample of data yield a value of χ^2 at least as large as this?"

Determining the best values for the model parameters by choosing them so as to minimise χ^2 is called the "**Method of Least Squares**."

Note that, if you vary one of the model parameters from its best fit value until χ^2 increases by 1, this therefore represents the change in the model parameter associated with 1 unit of variance in the fit quality (*i.e.* the "1σ uncertainty" in the model parameter).

Example:

A newly commissioned underground neutrino detector sees a rate of internal radioactive contamination decreasing as a function of time. Measurements of the number of such events observed are taken on 10 consecutive days. Determine the best fit mean decay time in order to determine the source of the contamination.

decay probability:

$$P(t) = \frac{1}{t_o} e^{-\frac{t}{t_o}}$$

t_o = mean decay lifetime

Centre of Time Bin (days)	Measured # counts
0.5	115
1.5	76
2.5	60
3.5	66
4.5	56
5.5	35
6.5	25
7.5	32
8.5	22
9.5	25

Table of model predictions for different values of t0

m = (Ntot/(1-exp(-10/t0))*[exp(-(t-0.5)/t0) - exp(-(t+0.5)/t0)] ~ (Ntot/(1-exp(-10/t0))*(1/t0)*exp(-t/t0)

t0:	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12
	151	135	123	114.4	107.3	101.6	96.9						80.3	[illegible]					03838
	108	101	96.1	91.62	87.88	84.72	82						71.8	[illegible]					14747
	77.3	76.2	74.8	73.37	71.95	70.64	69.4						64.3	[illegible]					88526
	55.4	57.3	58.3	58.75	58.91	58.89	58.8						57.5	[illegible]					88166
	39.7	43.1	45.4	47.04	48.23	49.1	49.8						51.5	[illegible]					79617
	28.4	32.4	35.3	37.67	39.49	40.94	42.1	43.1	43.9	44.6	45.1	45.6	46.1	46.4	46.8	47	47.3	47.5	47.731552
	20.4	24.3	27.5	30.16	32.33	34.13	35.6	36.9	38	39	39.8	40.6	41.2	41.8	42.3	42.8	43.2	43.6	43.915148
	14.6	18.3	21.4	24.15	26.47	28.46	30.2	31.7	33	34.1	35.1	36.1	36.9	37.6	38.3	38.9	39.4	39.9	40.403886
	10.5	13.7	16.7	19.34	21.67	23.73	25.5	27.2	28.6	29.9	31	32.1	33	33.9	34.6	35.3	36	36.6	37.17337
	7.49	10.3	13	15.49	17.74	19.78	21.6	23.3	24.8	26.1	27.4	28.5	29.5	30.5	31.3	32.1	32.9	33.6	34.201151

normalised to total number of events observed in 10 days

often approximate integral over bin with the average

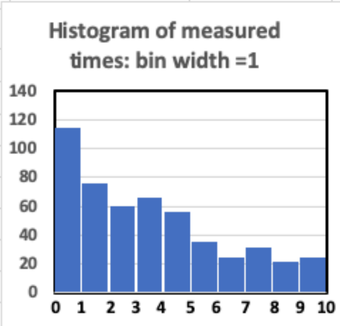


Table of ((n-m)^2)/m for different values of t0

t0:	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	12
	8.38	2.96	0.57	0.003	0.547	1.764	3.38	5.22	7.17	9.17	11.2	901
	9.4	6.38	4.2	2.664	1.606	0.897	0.44	0.17	0.04	0	0.03	745
	3.86	3.46	2.94	2.435	1.985	1.601	1.28	1.02	0.8	0.63	0.49	399
	2.04	1.32	1.02	0.896	0.854	0.858	0.89	0.93	0.99	1.05	1.12	175
	6.72	3.89	2.48	1.707	1.252	0.969	0.78	0.66	0.57	0.51	0.46	491
	1.52	0.22	0	0.189	0.51	0.861	1.2	1.52	1.8	2.05	2.28	175
	1.05	0.02	0.23	0.883	1.662	2.444	3.18	3.86	4.47	5.02	5.52	389
	20.8	10.3	5.2	2.551	1.156	0.441	0.11	0	0.03	0.13	0.28	983
	12.7	4.98	1.68	0.366	0.005	0.126	0.49	0.98	1.52	2.07	2.62	432
	40.9	20.9	11.1	5.846	2.969	1.376	0.53	0.13	0	0.05	0.21	399
Chi-squared sum:	107	54.5	29.4	17.54	12.54	11.34	12.3	14.5	17.4	20.7	24.2	305

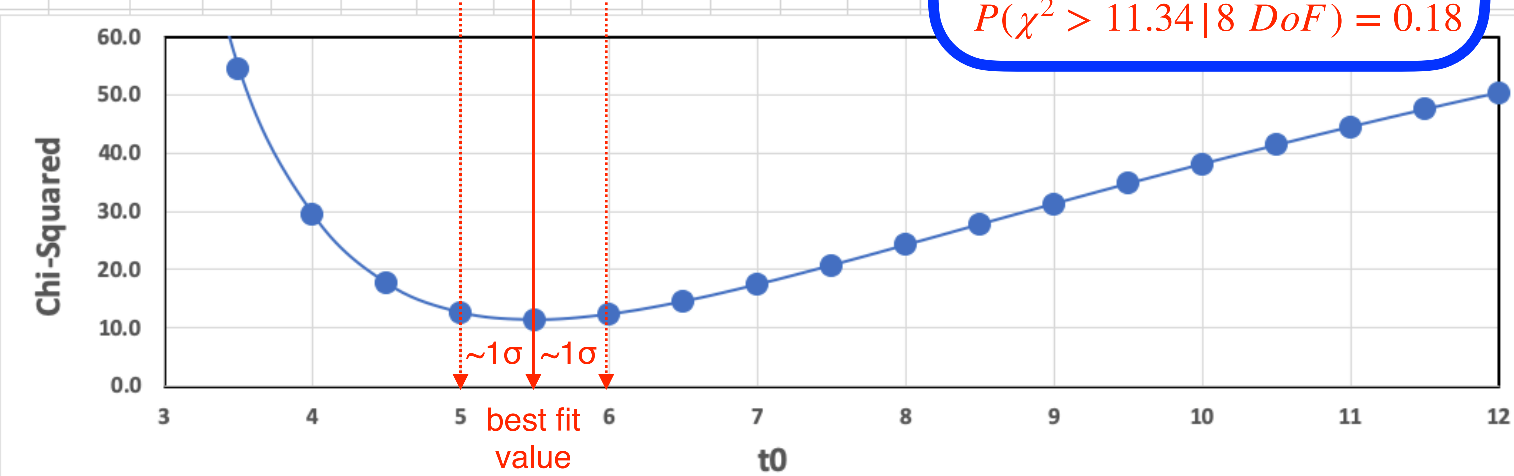
How many degrees of freedom?

- 10 independent data points
- fit parameter t_o
- but normalisation is also based on observed data (for single bin, variance would be zero)

DoF = 10 - 2 = 8

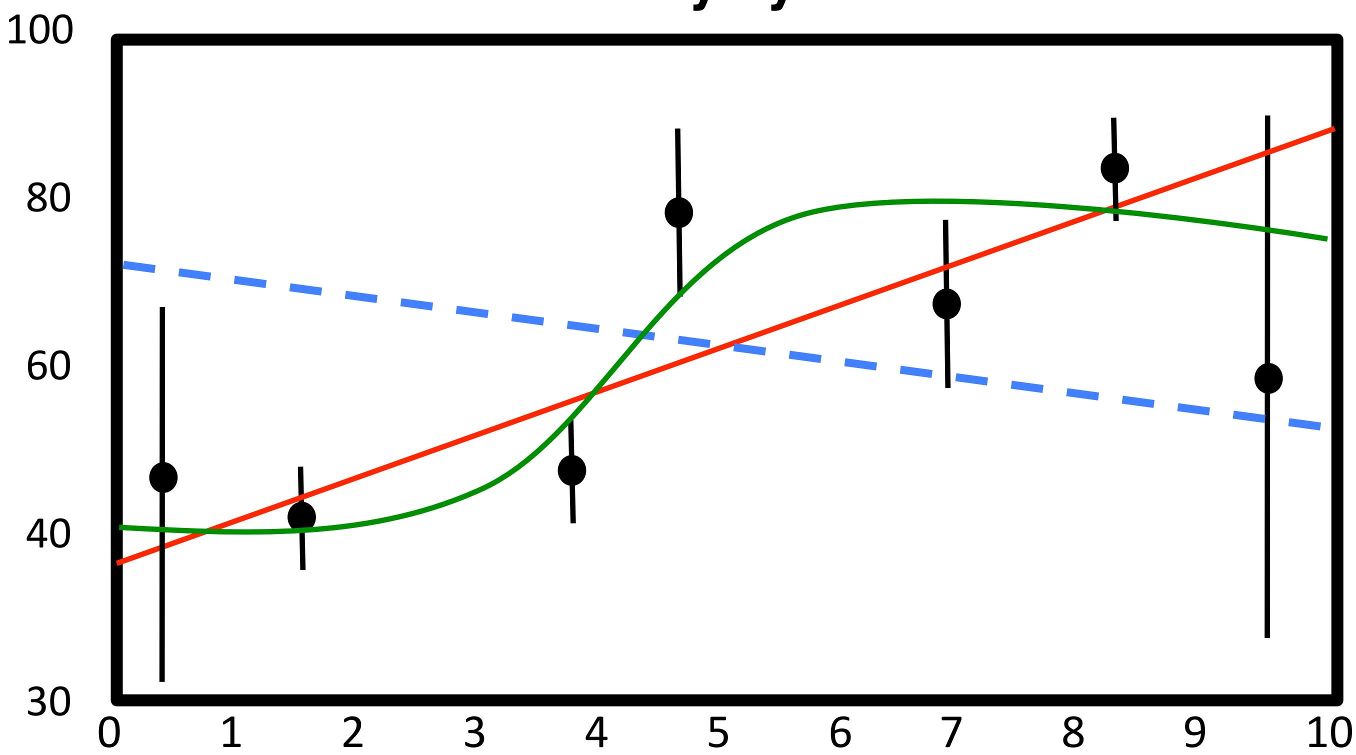
How good is the fit?

$$P(\chi^2 > 11.34 | 8\ DoF) = 0.18$$



^{222}Rn
mean lifetime
= 5.51 days

"Chi by eye"



$$\chi^2 \sim (0.3)^2 + (0.1)^2 + (1.2)^2 + (1.7)^2 + (0.3)^2 + (0.8)^2 + (0.8)^2 = 5.8$$

Degrees of Freedom = 7 – 2 = 5

NOTE: This doesn't tell you which model is correct,

but it can tell you which models don't fit well!

Quoted chance probability based on correlation analysis: **6 x 10^{-4}**

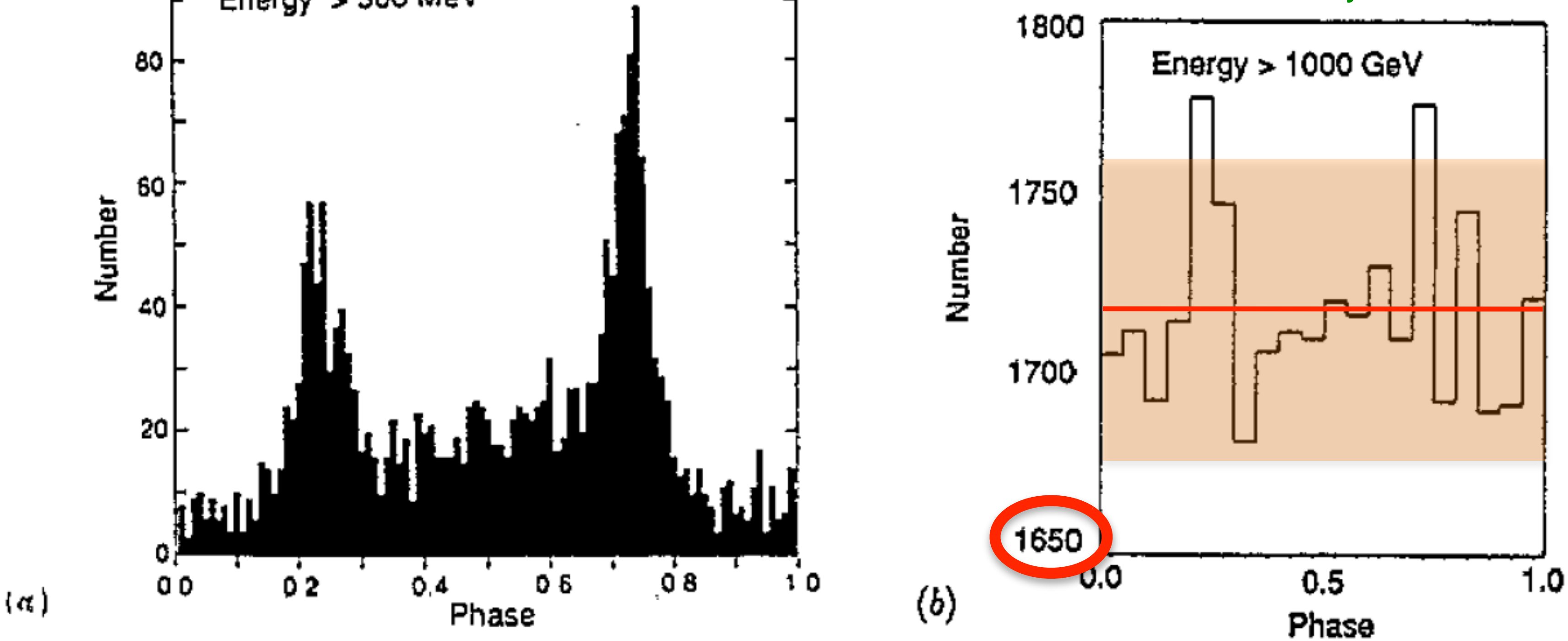


Figure 1. (*a*) Light curve for Geminga obtained with EGRET. (*b*) The VHE γ-ray light curve of Geminga plotted at the GeV γ-ray phase, as derived from the COS-B ephemeris.

x
1705
1712
1693
1715
1778
1756
1681
1707
1712
1710
1721
1717
1731
1710
1777
1693
1747
1690
1692
1722

Avg (m) = 1718.45

$$\chi^2 = \sum_{i=1}^{20} \frac{(x_i - 1718.45)^2}{1718.45} = 8.22$$

Wuant 'em Effect

DoF = 20 – 1 = 19 **(98.4% chance of getting something larger)**

Scientific Method:

ORDER!!

Model Model Model Model Model Model Model Model Model Model MODEL Model Model

Simplest and most predictive

Test for reproducible predictions to disprove

Rejected with high confidence

Model

Next simplest & most predictive

Not rejected with high confidence

Model

A theory is judged not on what it can explain, but on what it can reproducibly predict!

We don't prove models correct; we reject those models that are wrong!

Test for reproducible predicti...

Rejected with high confidence

Model

Don't state that data are "consistent" with a given model, but rather that they are **"not inconsistent."**