Statistics and Modeling

Statistics is the grammar of science.
- Karl Pearson

There are three types of lies -- lies, damn lies, and statistics.
- Benjamin Disraeli? Mark Twain?

It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so.
- Mark Twain? Yogi Berra?

All models are wrong, but some models are useful.
- George E.P. Box

Data do not give up their secrets easily. They must be tortured to confess.
- Jeff Hopper, Bell Labs

For a series of discrete random events (photons hitting a detector), the probability of ***x*** events given an expectation of ***m*** is given by the **Poisson distribution *P_x***:

$$P_x = \frac{m^x e^{-m}}{x!}$$

Table 6.1. *Sample values of Poisson function P_x*

x:	0	1	2	3	4	5	6	7[a]	8	9
$m = 1$	0.368	0.368	0.184	0.061	0.015	0.003	0.001	7E–5	9E–6	1E–6
$m = 2$	0.135	0.271	0.271	0.180	0.090	0.036	0.012	0.003	0.001	2E–4
$m = 3$	0.050	0.149	0.224	0.224	0.168	0.101	0.050	0.022	0.008	0.003
$m = 4$[b]	0.018	0.073	0.147	0.195	0.195	0.156	0.104	0.060	0.030	0.013
$m = 6$[c]	0.002	0.015	0.045	0.089	0.134	0.161	0.161	0.138	0.103	0.069
$m = 10$[d]	5E–5	5E–4	0.002	0.008	0.019	0.038	0.063	0.090	0.113	0.125

[a] The notation 7E-5 indicates 7×10^{-5}.
[b] The values of P_x for $m = 4$ at $x = 10$ and 11 are 0.005 and 0.002 respectively.
[c] The values of P_x for $m = 6$ at $x = 10$–14 are 0.041, 0.023, 0.011, 0.005, 0.002.
[d] The values of P_x for $m = 10$ at $x = 10$–18 are: 0.125, 0.114, 0.095, 0.073, 0.052, 0.035, 0.022, 0.013, 0.007.

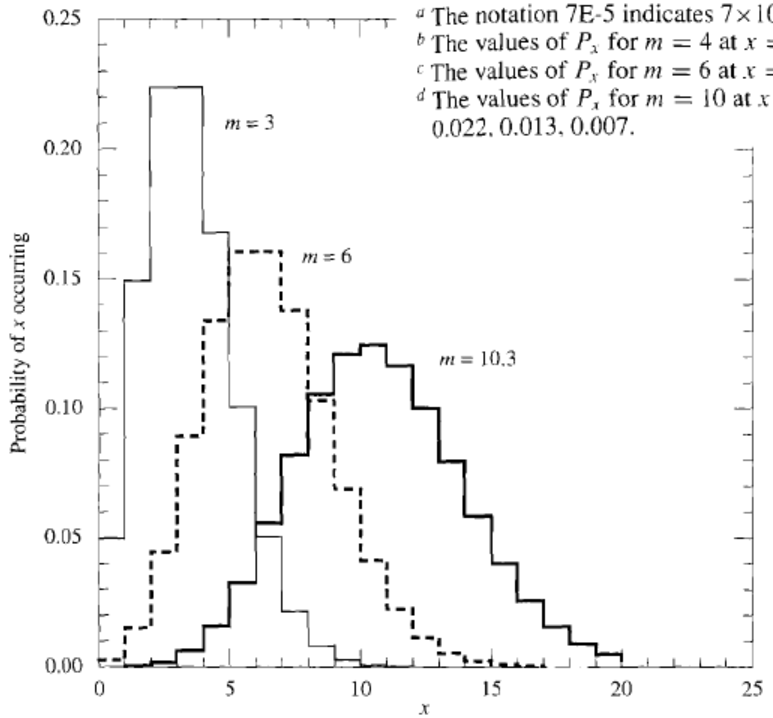


Figure 6.7. The Poisson distribution for small mean numbers, $m = 3.0$, 6.0 and 10.3. The ordinate gives the probability of the value x occurring, for the given mean value. Note the asymmetry of the histograms.

As m (the expectation value) gets large, the distribution resembles a **gaussian** or **normal** distribution.

$$dP = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-m)^2/2\sigma^2}$$

The **variance** of any distribution is defined as

$$\sigma^2 \equiv \frac{1}{n}\sum(x_i - m)^2$$

In the normal distribution, σ is independent from m. But for the Poisson distribution, **σ^2=m**.

Terminology:
σ^2 = "variance"
σ = "standard deviation"

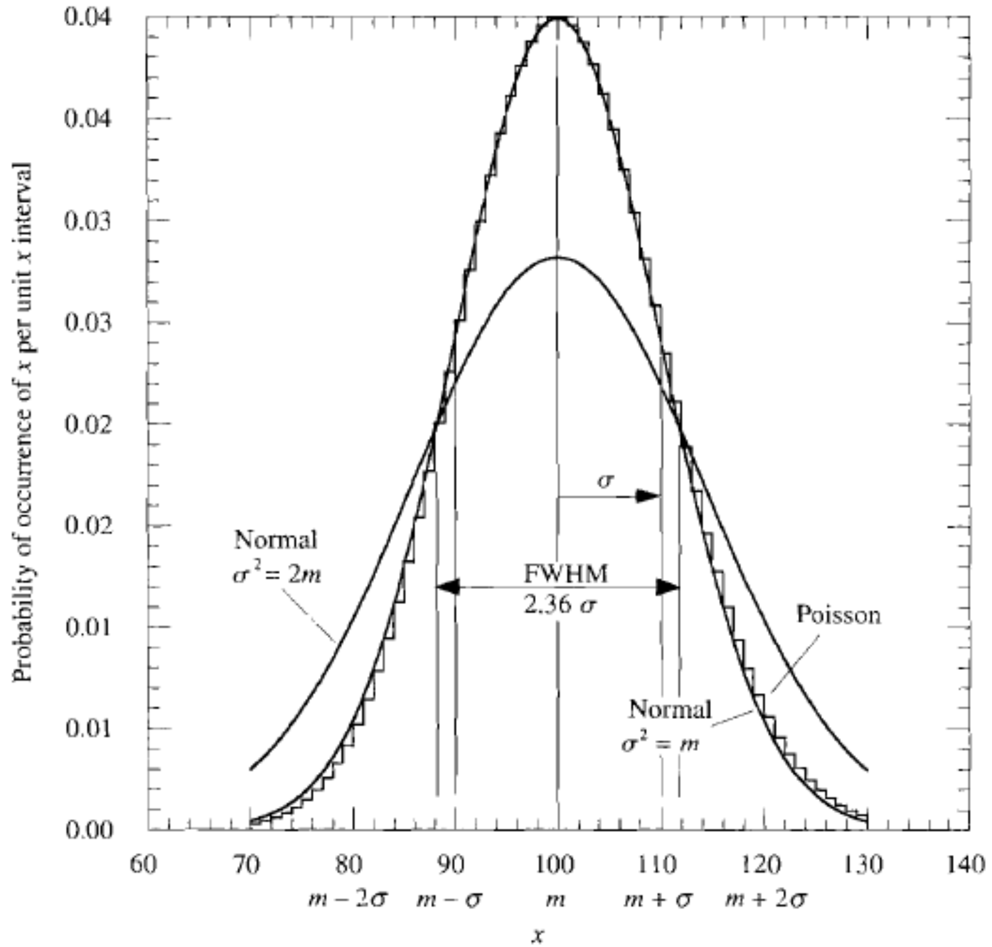


Figure 6.8. The Poisson (step curve) and normal distributions (smooth curves) for the mean value $m = 100$. The normal distribution is given for two values of the width parameter σ_w which is shown in the text to be equal to the standard deviation σ. The Poisson distribution approximates well the normal distribution if the latter has $\sigma = \sqrt{m}$. Note the slight asymmetry of the Poisson distribution relative to the normal distribution. The standard deviation and full width half maximum widths are shown for the higher normal peak; the two normal curves happen to cross at the FWHM point.

Detection significance

Say the background sky gives m=100 ***photons*** per pixel. By Poisson stats, the uncertainty in the sky level is then $\sigma=\sqrt{m} = \sqrt{100} = 10$ photons.

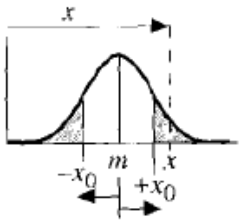
So the sky level is 100 ± 10 photons/pixel.

How faint of a (one pixel) star could you detect?

$N_* = 10$ photons ➔ 1σ detection, very likely to be just a sky fluctuation. A poor detection.

$N_* = 30$ photons ➔ 3σ detection, likelihood of a sky fluctuation is small. A good detection!

Table 6.2. *Normal distribution probabilities*



$\left(\frac{x_0}{\sigma}\right)^a$	Area (shaded) at $\lvert x - m\rvert > x_0$ [b]	$\left(\frac{x_0}{\sigma}\right)^a$	Area (shaded) at $\lvert x - m\rvert > x_0$ [b]
0	1.00	2.5	0.0124
0.5	0.617	3.0	0.00270
1.0	0.317	3.5	4.65×10^{-4}
1.2	0.230	4.0	6.34×10^{-5}
1.4	0.162	5.0	5.73×10^{-7}
1.6	0.110	6.0	2.0×10^{-9}
1.8	0.0719	7.0	2.6×10^{-12}
2.0	0.0455		

[a] Ratio of deviation x_0 to standard deviation σ. The standard deviation σ is equal to σ_w, the width parameter of the distribution.
[b] Probability of occurrence of deviation greater than $\pm x_0$.

A more rigorous signal-to-noise calculation

Consider measuring the flux from a star in an aperture that includes n_{pix} pixels.

Signal:

- N_*, the total number of photons from the star.

Noise:

- Total Poisson noise from the star: $\sigma = \sqrt{N_*}$
- Per-pixel Poisson noise from the sky: $\sigma = \sqrt{N_S}$
- Per-pixel Poisson noise from dark current: $\sigma = \sqrt{N_D}$
- Per-pixel CCD read noise $\sigma = N_R$

These N's all refer to photons or electrons, not counts!

These noise contributions add in quadrature, so we get

$$\frac{S}{N} = \frac{N_*}{\sqrt{N_* + n_{pix}(N_S + N_D + N_R^2)}}$$

"The CCD Equation"

see Howell, Chapter 4.4

Example: Schmidt Telescope + CCD

- gain = 2.5 e^-/ADU
- read noise = 3.6 e^-
- N_D = 0 ADU

$$\frac{S}{N} = \frac{N_*}{\sqrt{N_* + n_{pix}(N_S + N_D + N_R^2)}}$$

In a 60s exposure in the M filter, we get

- Sky = 80 ADU = 200 photons (±14) per pixel
- A star with a peak of 20,000 ADU has 136,000 ADU (340,000 photons) inside a circular aperture of r=5 pixels (so $n_{pix} = \pi 5^2 \approx 80$).

$$\frac{S}{N} = \frac{340,000}{\sqrt{340,000 + 80(200 + 0 + 3.6^2)}} = 570$$

For a star that peaks at 100 ADU, the same calculation gives S/N = 12.

Magnitude error:

- $\sigma_{mag} = 1.0857\ (\sigma_{flux}/\text{flux}) = 1.0857/(S/N)$.
- Star 1: S/N = 570, so σ_{mag}=0.002 mag
- Star 2: S/N = 12, so σ_{mag}=0.09 mag

S/N scaling with exposure time

$$\frac{S}{N} = \frac{N_*}{\sqrt{N_* + n_{pix}(N_S + N_D + N_R^2)}}$$

Case 1: Bright objects

N_* dominates, so $S/N \approx N_*/\sqrt{N_*} \approx \sqrt{N_*}$
Since N_* scales with exposure time, $S/N \sim \sqrt{t_{exp}}$

Case 2: Detector limited

N_R dominates, so $S/N \approx N_*/N_R$
N_R is independent of exposure time, so $S/N \sim t_{exp}$

Random vs Systematic Error

Precision: How well can you measure a quantity? How repeatable is your measurement? Usually captured by “random errors.”

Accuracy: How well does your measurement actually recover the value you are trying to measure? Source of “systematic errors.”

Precision vs Accuracy / Random vs Systematic is critical to understand, extremely hard to quantify in practice.

If you measure a value and do not give some estimate of uncertainty or some discussion of systematic errors, your measurement is useless.

The 10/90 rule: you spend 10% of your time getting “the answer”. You spend the other 90% understanding your uncertainties.

courtesy Matt Bershady, University of Wisconsin

Characterizing distributions

- Moments
- 1st: Mean, $\bar{x}$ (location)
  - Other 1st-moment indicators:
    - *median* (robust estimator)
    - *mode*
- 2nd: Standard deviation, σ (width)
  - Other 2nd-moment indicators:
    - *Average deviation* (robust estimator): $AD = \frac{1}{N}\sum_{i=1}^{N}|x_i - \bar{x}|$
    - *full-width half-maximum (FWHM)*
- 3rd: Skew, s (symmetry)
- 4th: Kurtosis, k (shape)

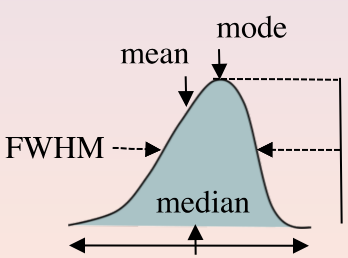
$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

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

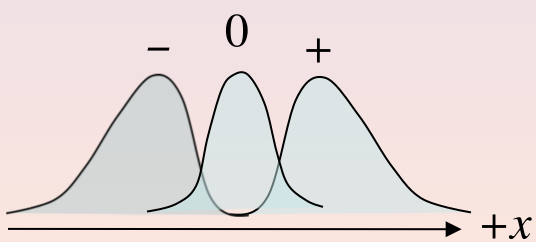
$$s = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma}\right)^3$$

$$k = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma}\right)^4 - 3$$

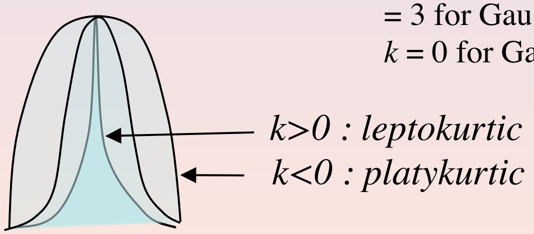
= 3 for Gaussian
k = 0 for Gaussian



1st & 2nd moments



skew



kurtosis

Error Propagation

If errors are **gaussian** and **uncorrelated**, we can add each error source in quadrature.
(But uncorrelated gaussian errors are often a bad assumption!)

Error in mean: $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$$

For propagating small errors, we can use a **Taylor expansion**. If you are calculating some property C from measurements of x, y, and z:

C = f(x,y,z) then $\sigma_C^2 = [(df/dx)*\sigma_x]^2 + [(df/dy)*\sigma_y]^2 + [(df/dz)*\sigma_z]^2$

Example, the absolute magnitude of M87
m = 8.63 ± 0.04, d=16.0 ± 1.1 Mpc

since m-M = 5log(d)-5, M=-5log(d)+5+m = -22.4

- dM/dm = 1
- dM/dd = -5/(d ln(10)) ≈ -2.17/d
- then $\sigma_M^2 = [1*0.04]^2 + [(-2.17/16)*1.1]^2$
- so σ_M = 0.15 mags

but this is random error, not systematic!

Correlations

Linear:

- $y = mx + b$
- multidimensional: $z = mx + ny + b$

Nonlinear: *try to linearize them!*

Example #1: Exponential surface brightness of a disk

Raw form: $I(r)=I_0 e^{-r/h}$
Linearized form: $\ln(I) = \ln(I_0) - r/h$

In surface brightness:

remember $\log(x)=\ln(x)/\log(10)$

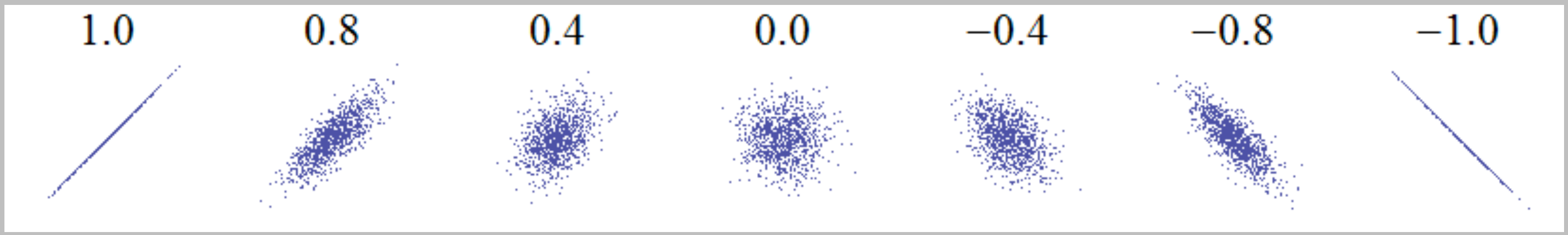
$$\mu(r) = \mu_0 + \frac{2.5}{\ln 10}\frac{r}{h}$$

Example #2: Power law form form of Tully-Fisher

Raw form: $L \sim V_{circ}^{\alpha}$
Linearized form: $\log(L) = \alpha\log(V_{circ}) + C$

Correlations

Pearson's correlation coefficient, r, measures linear correlation between two variables.



Slope(s), intercept, and their uncertainties: $m \pm \sigma_m$, $b \pm \sigma_b$

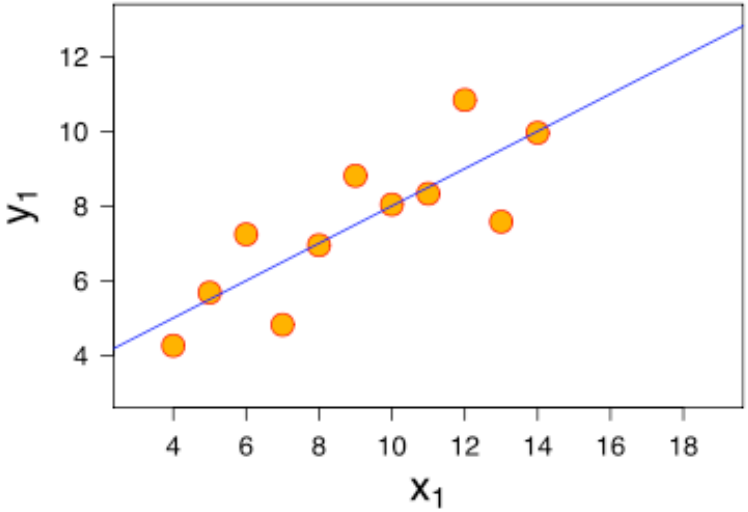
RMS scatter around the fit:

$$\sigma_{RMS} \equiv \frac{1}{N} \sum (y_i - y_{fit})^2$$

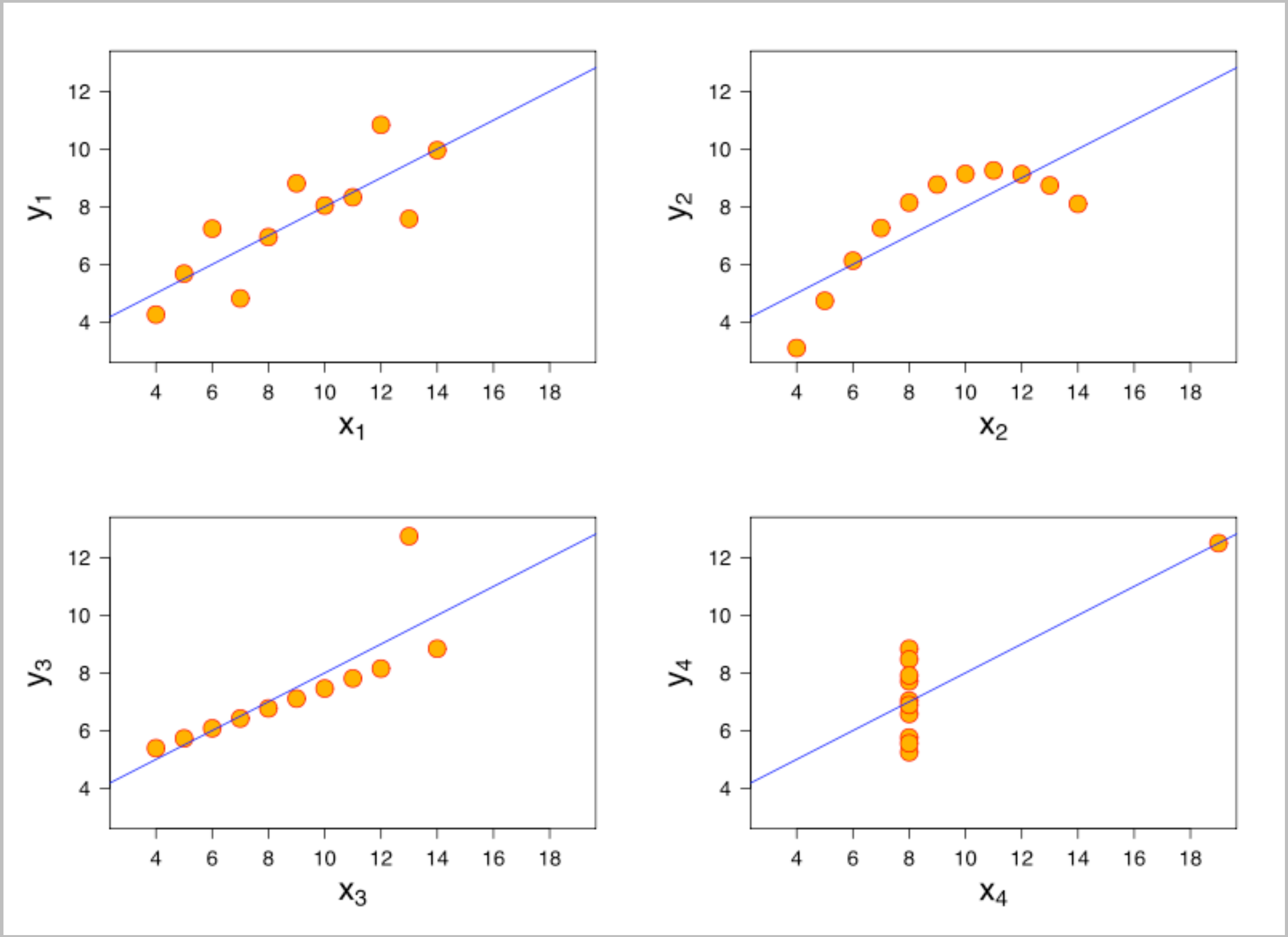
The importance of scatter:

- Measure of other dependencies (data noise, additional physical parameters)
- Determines the "accuracy" to which an individual object obeys the relationship.

But, beware……..

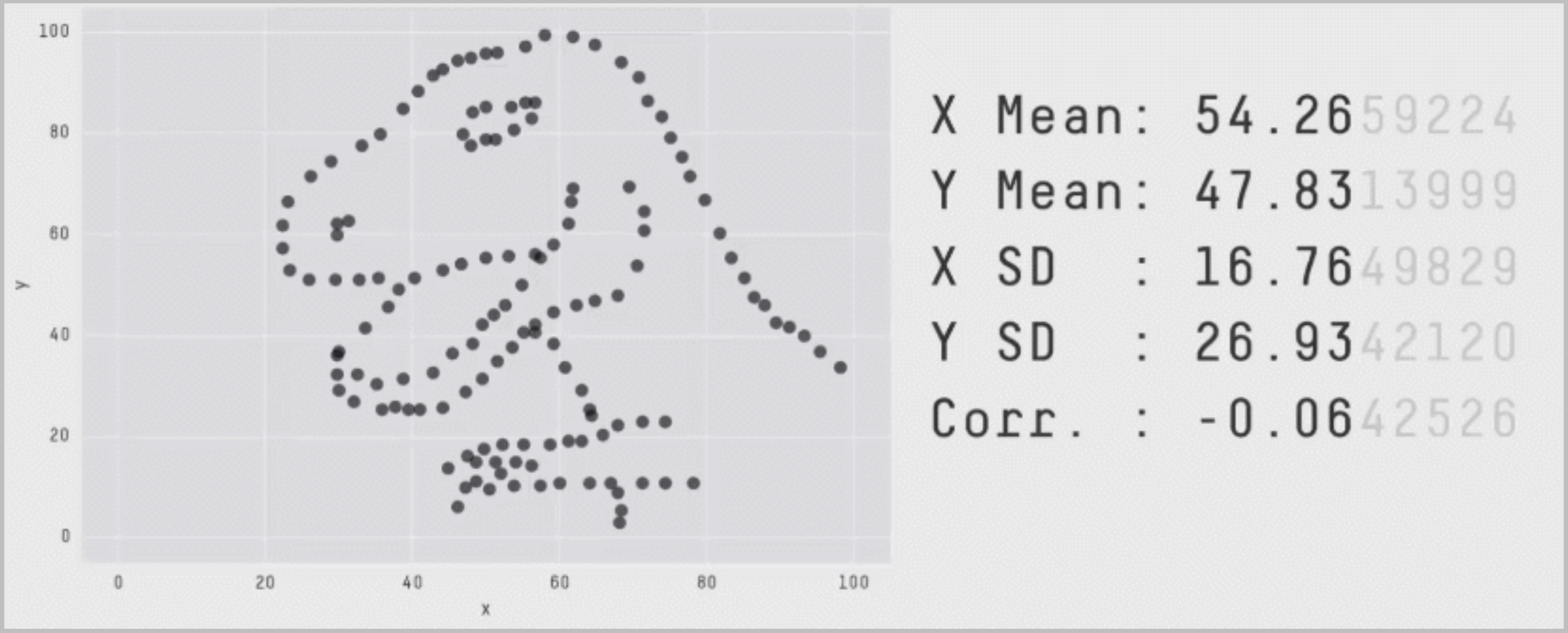


Anscombe's quartet: Fit y=mx+b and get the same r, m, b, σ_m, σ_b, σ_{RMS}



Anscombe’s quartet: Fit y=mx+b and get the same r, m, b, σ_m, σ_b, σ_{RMS}

Beware the datasaurus!



Modeling Uncertainty

Let’s say you have used your data to estimate some parameter. For example, you have a set of x,y data points and you’ve fit a line to the data and estimated a slope and intercept. How do we estimate our uncertainties on these values:

Least squares fitting: what usually comes out of your computer. Implicitly assumes uncorrelated Gaussian statistics. Different algorithms can give different estimates, particularly in low-N or presence of outliers.

Resampling (non-parametric):

- **Jack-knife:** go through your data i=1,N times, tossing out data point i and redoing you estimate. Look at variation.

- **Bootstrap:** go through your dataset picking out N data points at random. Do this as many times as you can stand, look at variation.

Bayesian Estimation

We speak in terms of probabilities. What is the probability you'd get the data you measure given some underlying model?

Bayes' theorem:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

A = your dataset
B = the parameter you're trying to measure

P(B|A): ***The posterior probability***. What is the probability of B, given that you've measured A? Your best estimate is the B that is most-likely.

P(A|B): ***The likelihood function***. What is the probability of measuring A, given that model B is true?

P(B): ***The prior***. What is the probability of B?

P(A): Normalizing factor. What is the probability you could measure A to begin with?

Bayesian Estimation

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Let's get specific, but simple.

The luminosity function of a galaxy cluster – the number of galaxies as a function of their luminosity, N(L).

Adopt the Schecter function: $N(L) = \Phi_0 \, L^{\alpha} e^{-L/L_*}$

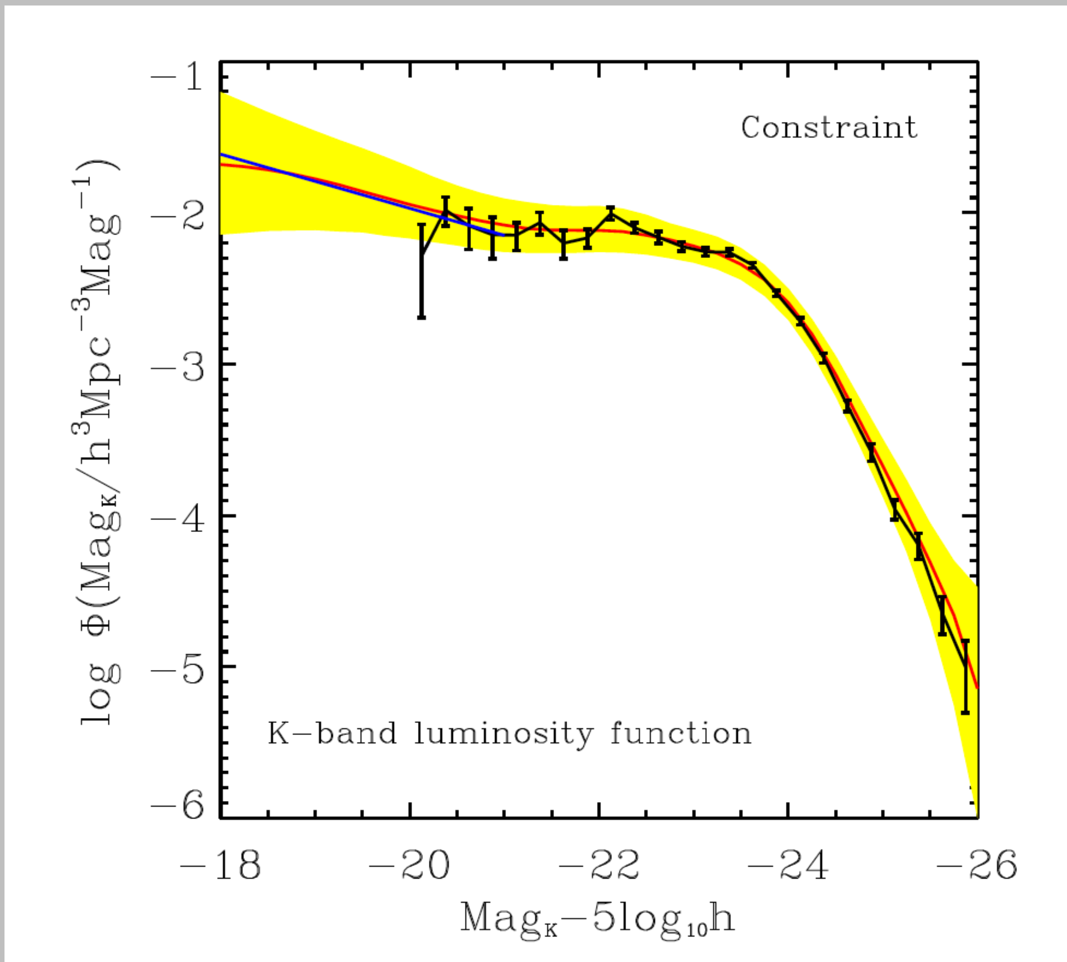
You've measured a bunch of galaxy luminosities – how do you estimate Φ_0, α, and L_*?

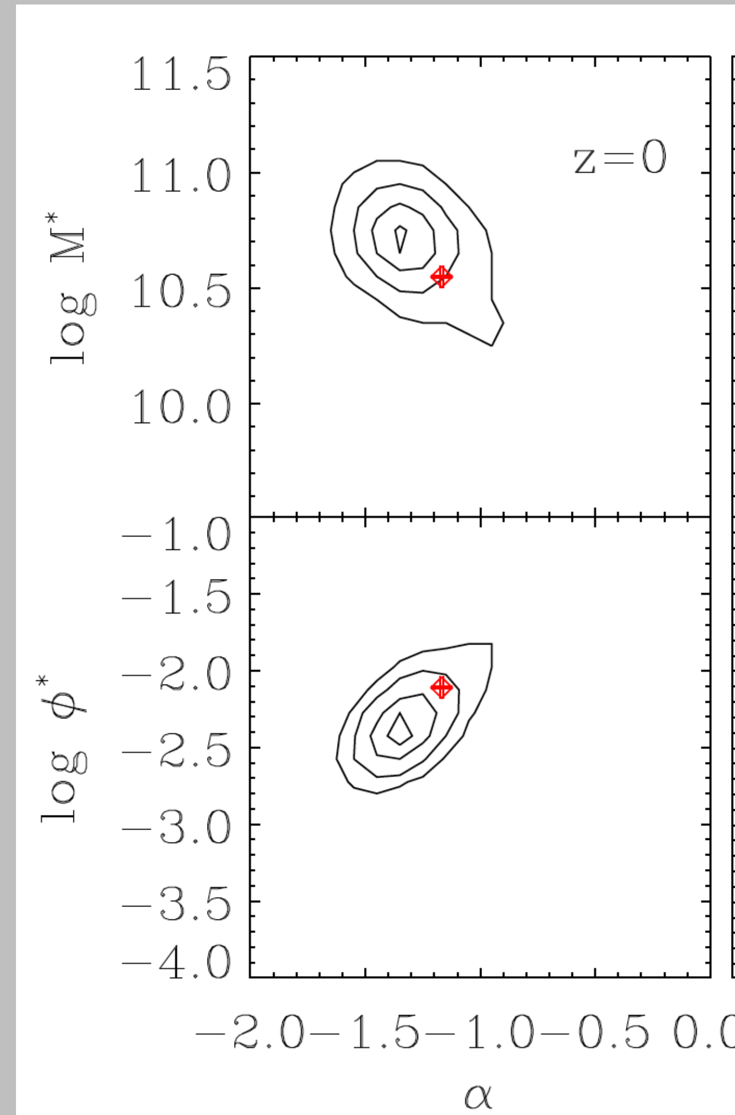
Classical: bin in L, plot N(L), do a (non-linear) chi-sq fit, solve for the parameters

Bayesian:

- A = your measurements
- B = the LF parameters Φ_0, α, and L_*
- P(A|B) = probability of measuring my dataset given some particular value of α and L_*, in other words the model for the luminosity function.
- P(B) = my prior beliefs about Φ_0, α, and L_*?
- P(B|A) = the probability that I'd get my data given some particular set of Φ_0, α, and L_*

Bayesian Estimation





from Lu+12