

Statistics in FDM

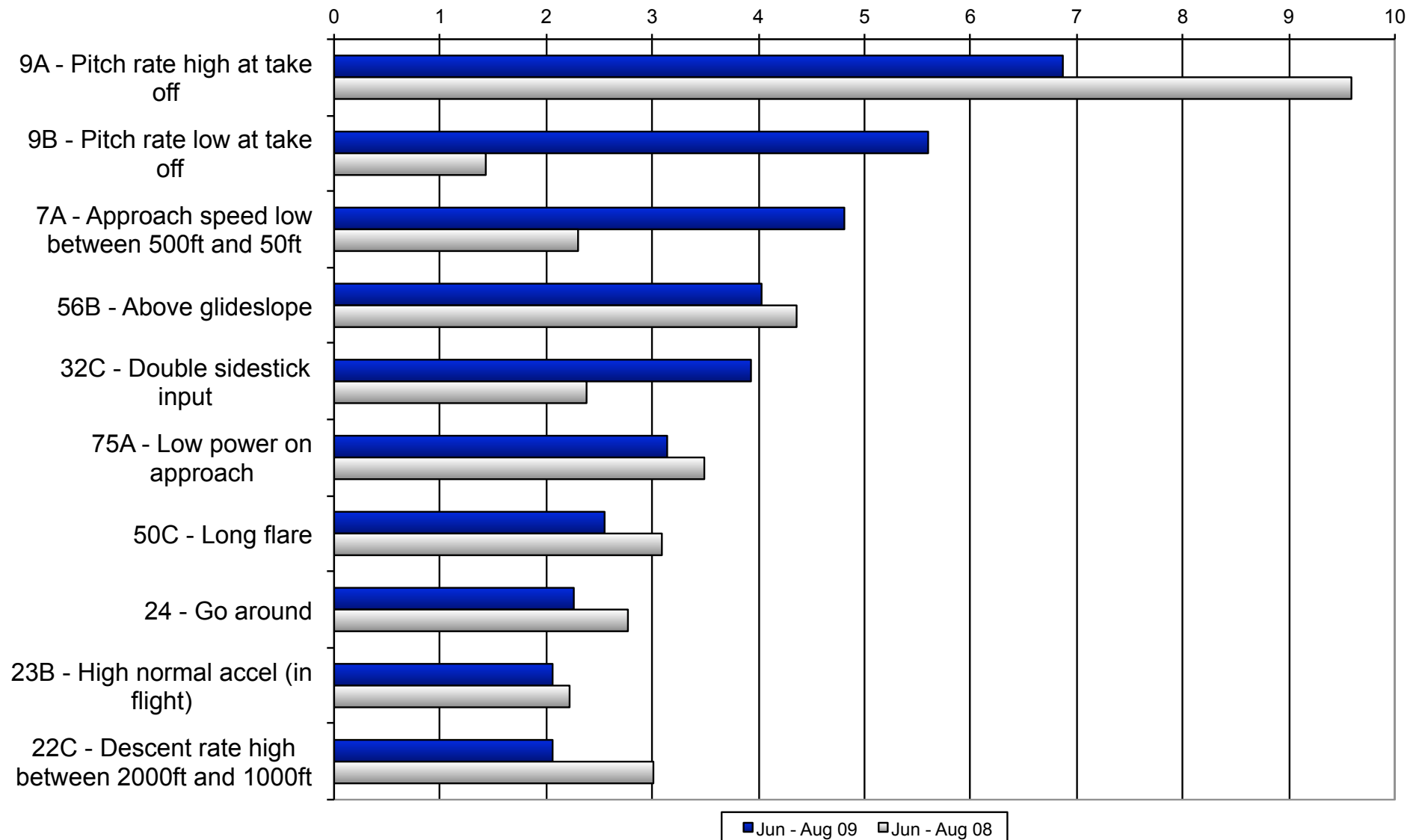
Dr Matthew Greaves &
David Barry

23rd January 2013
EOFDM Conference

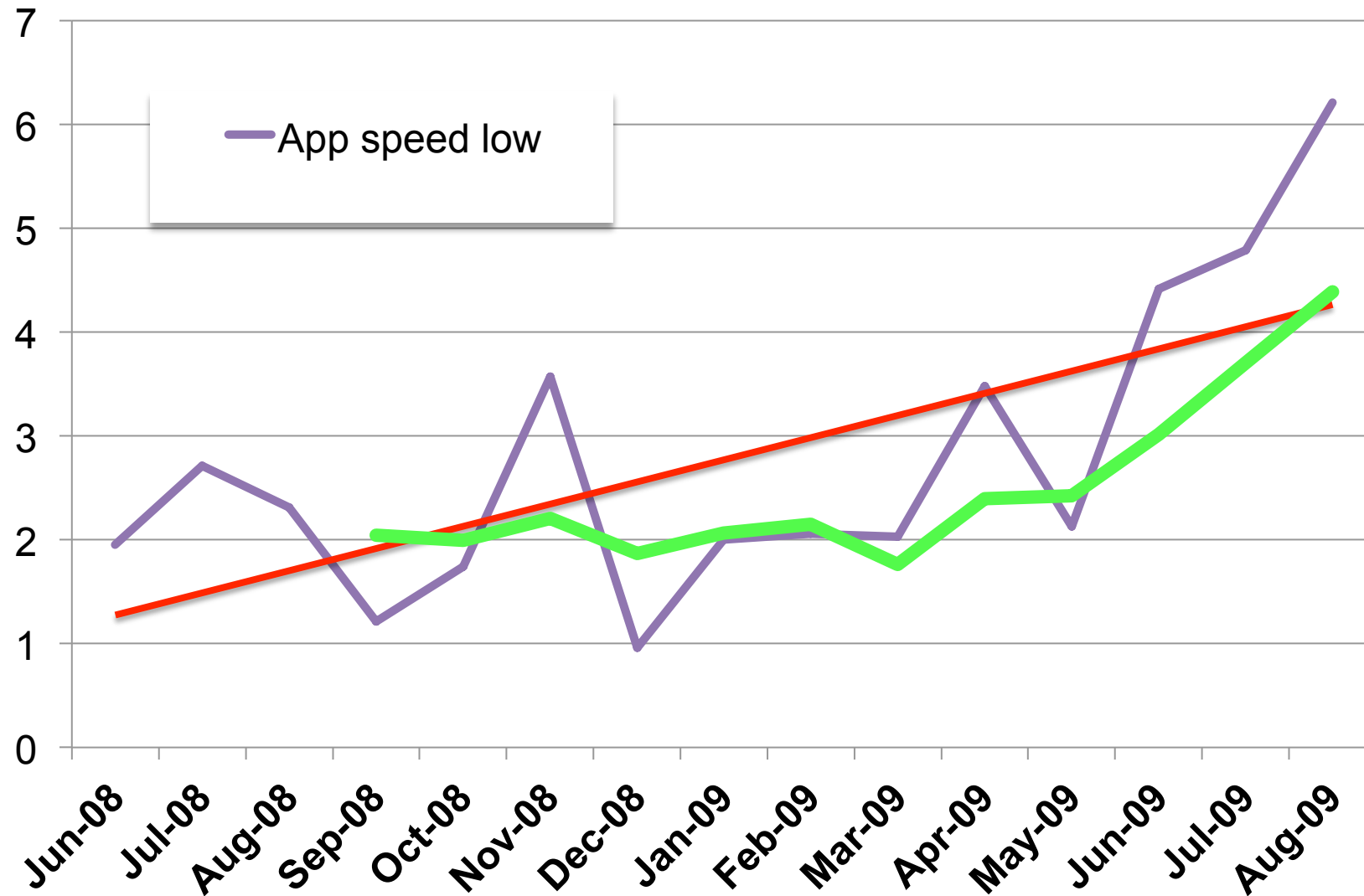
What are we trying to achieve?

- The primary purpose of FDM is to provide safety information to the SMS, allowing informed decision making and robust risk management
- Safety information must have integrity

Top 10 Events - Rate per 1000 sectors analysed



App speed low



Why do we need statistics in FDM?

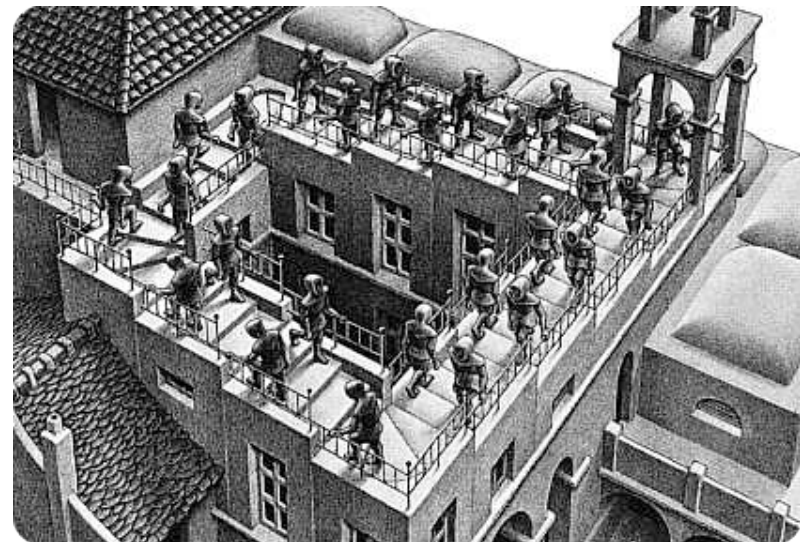
- For easier understanding
- To allow comparison with other data sets or earlier figures
- To give repeatability
- To support benchmarking
- To remove or compensate for bias



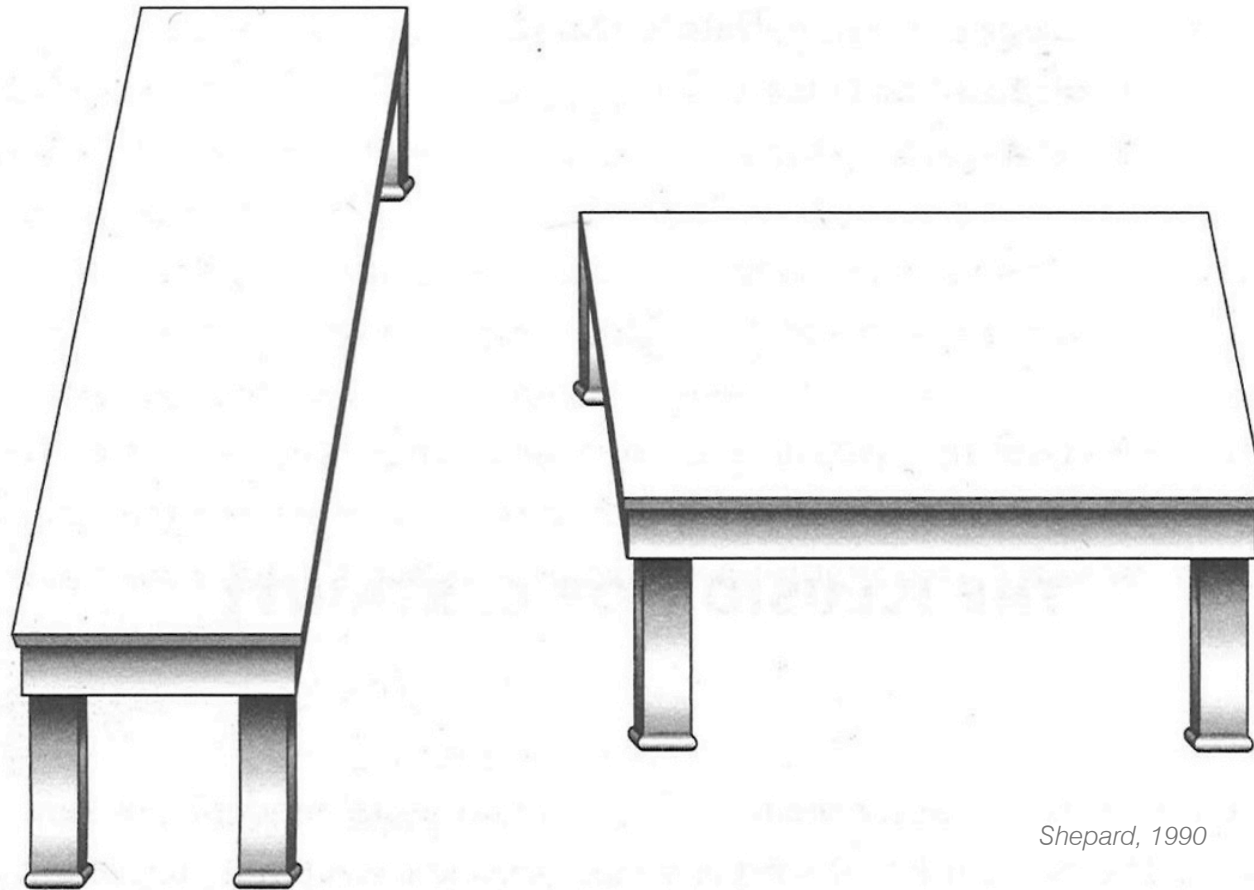
Heuristics



- Informal reasoning
- Rules of thumb
- Intuition
- Efficient ways of knowing in many situations which help humans to survive
- However, they can introduce cognitive illusions



Heuristics



Shepard, 1990

Biases

- Confirmation bias – we search out evidence which supports our hypothesis
- Availability bias - our attention is drawn to the exceptional and dramatic
- Social bias - we are in situations that reinforce our beliefs, and our beliefs were influenced by that situation
- Attribution bias – our successes are earned, others' are luck
- Post hoc ergo propter hoc



Randomness

- Does this pattern represent ‘streak shooting’?

0XXX0XXX0XX000X00XX00

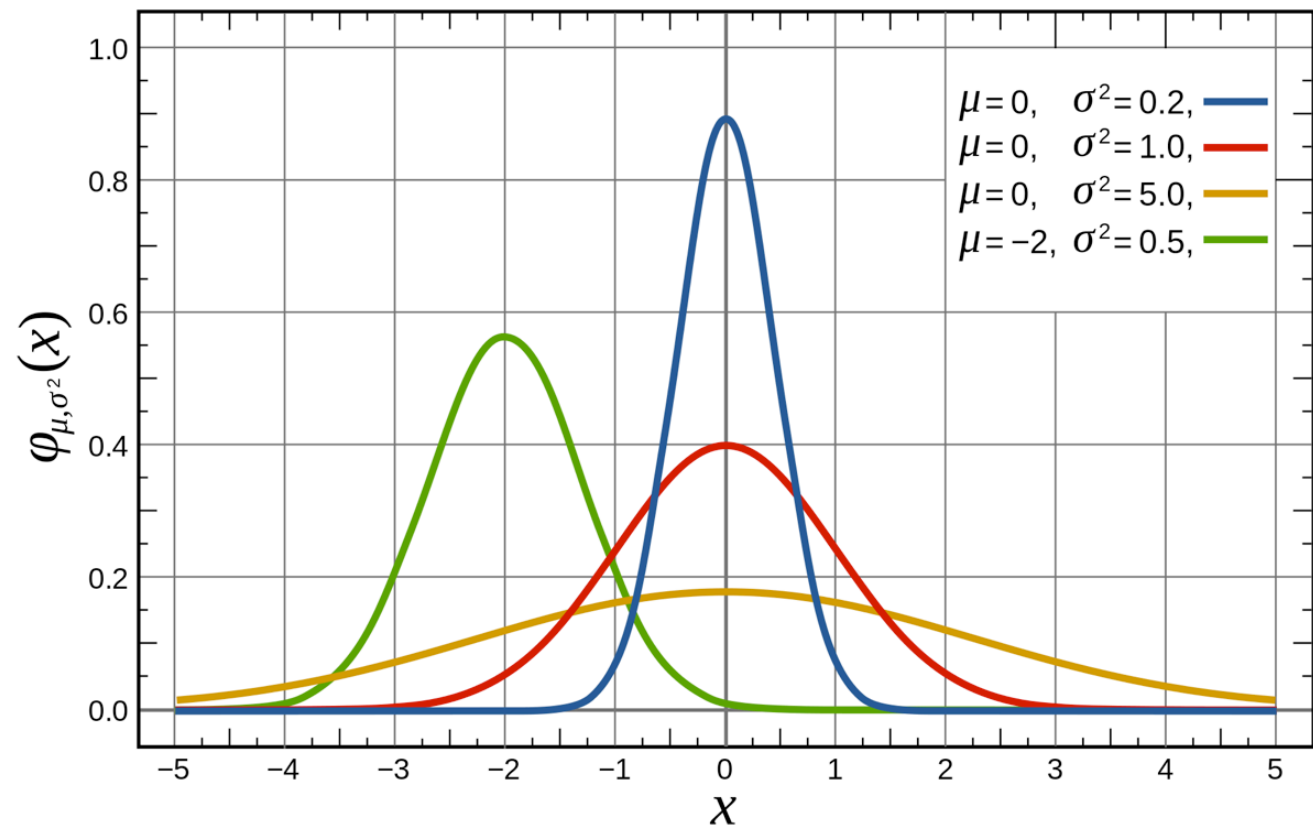
- 6 of first 8
- 8 of first 11

Gilovich T, Vallone R, Tversky A. The hot hand in basketball: on the misperception of random sequences.
Cog Psych (1985); 17: 295-314

- All samples fluctuate (sample error)
- Statistics are about making sense of this
- We can expect extremes by chance
- “Regression toward the mean”
- Need to avoid chasing chance – it can give an Illusion of system management
- Causation is attractive

Normal or Gaussian distribution

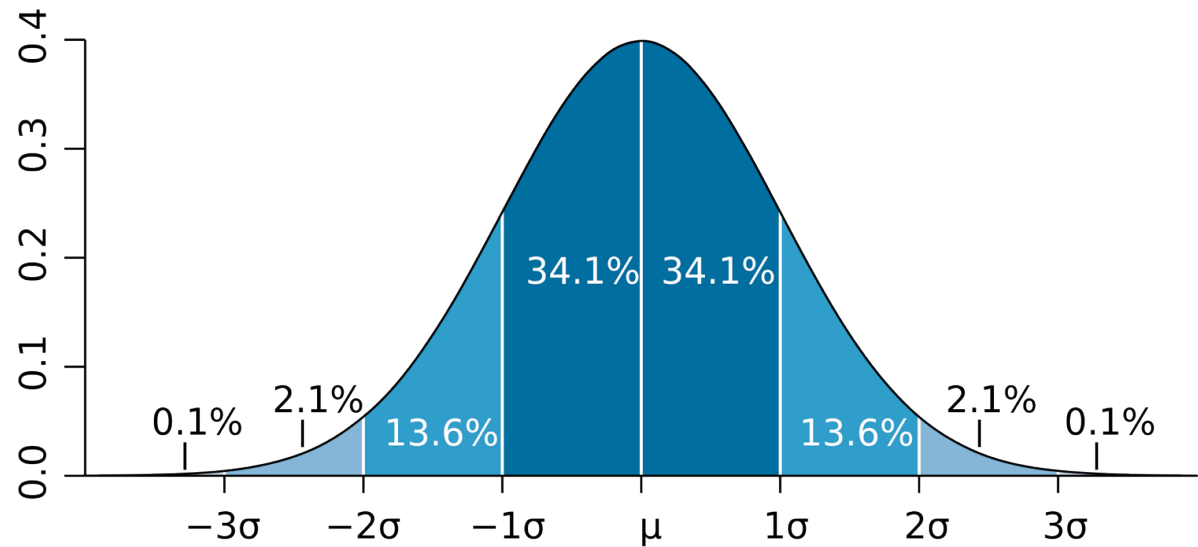
- Mean is μ
- Standard deviation is σ
- Symmetrical
- Mean, median and mode are all the same
- Tests for normal distribution



Standard deviation

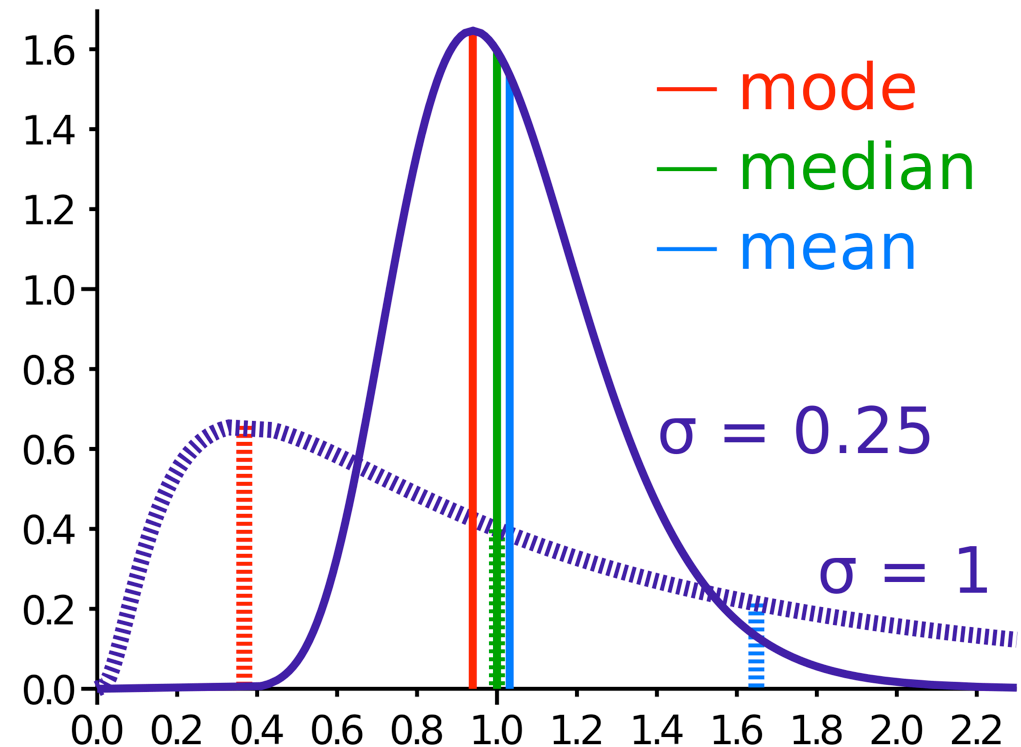
σ

- $\pm 1 \sigma$ gives 68%
- $\pm 2 \sigma$ gives 95%
- $\pm 3 \sigma$ gives 99.7%
- Variance:
 $\text{var}(x) = \sigma^2$



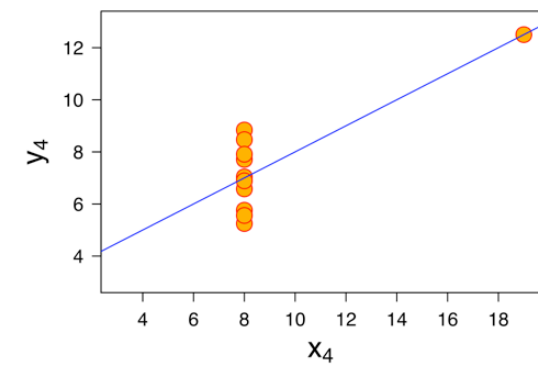
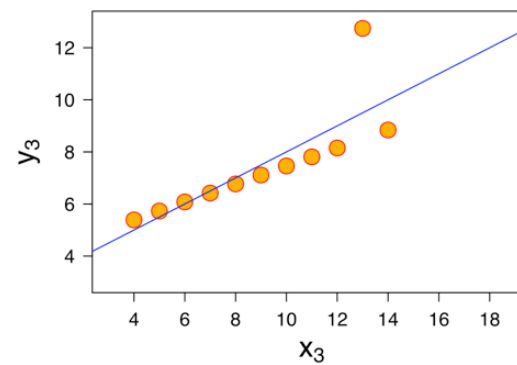
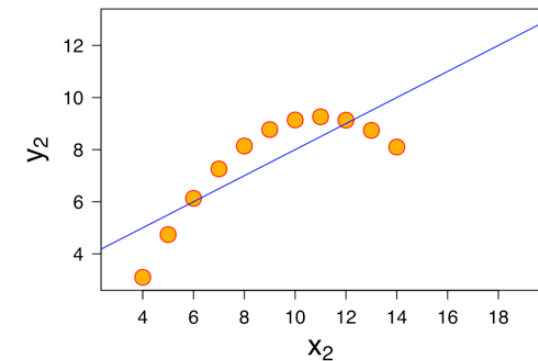
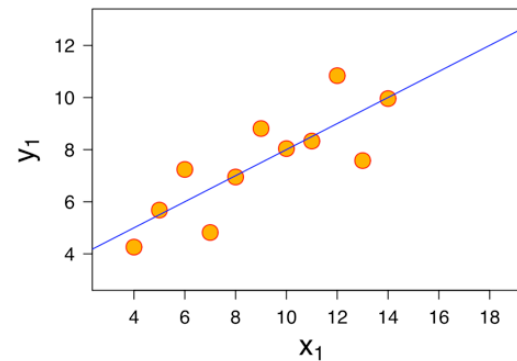
Mean, median or mode?

- So a change in mean could be brought about by a change in distribution
- Or a change in distribution may not be reflected by the median value
- This affects not only your reporting but your trigger criteria



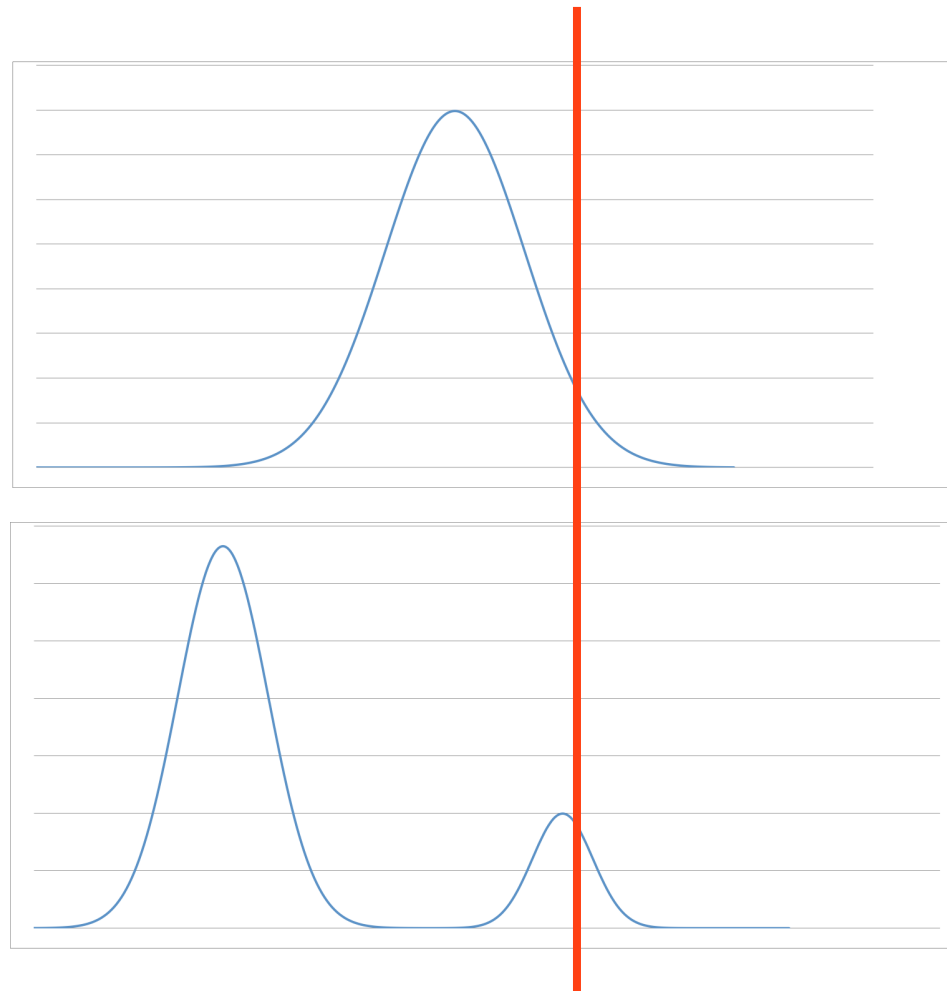
Lies, damn lies and statistics

- All have:
 - same mean
 - same standard deviation
 - same correlation
 - same regression line



So what?

- Why are distributions important?
- They can tell you whether that was expected....
- ...and if not, how to address a problem



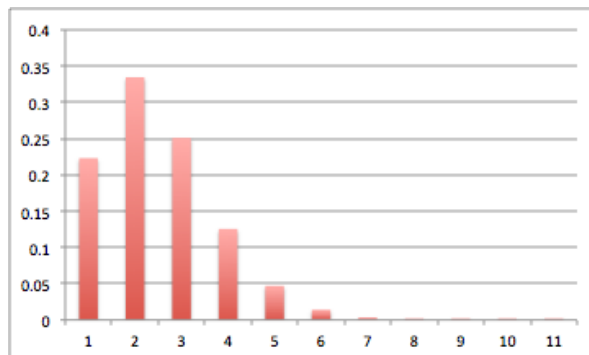
Poisson distribution

- The probability that a certain number of random events occur in a given unit of time or space can be calculated using a Poisson probability distribution

$$p(x) = \frac{e^{-\mu} \mu^x}{x!}$$

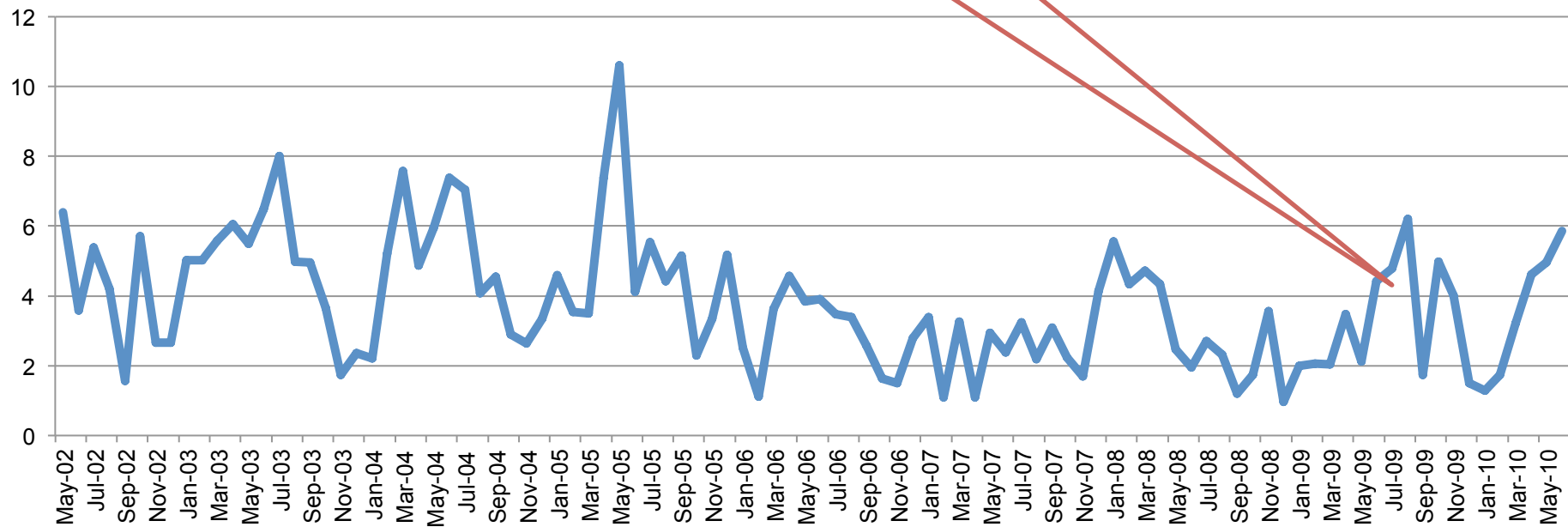
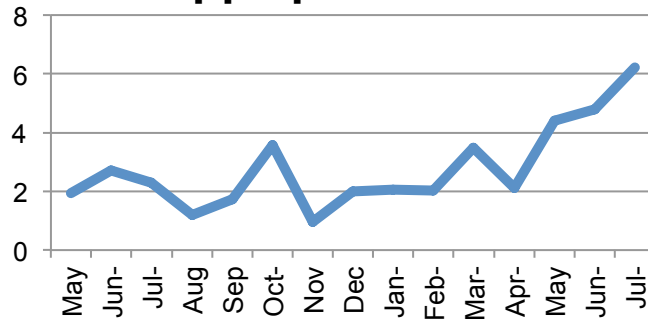
- Assuming 1500 flights per month and 1 unstabilised approach per 1000 flights gives a mean of 1.5 per month
- Therefore, the probability of, say, 0 unstabilised approaches in one month is

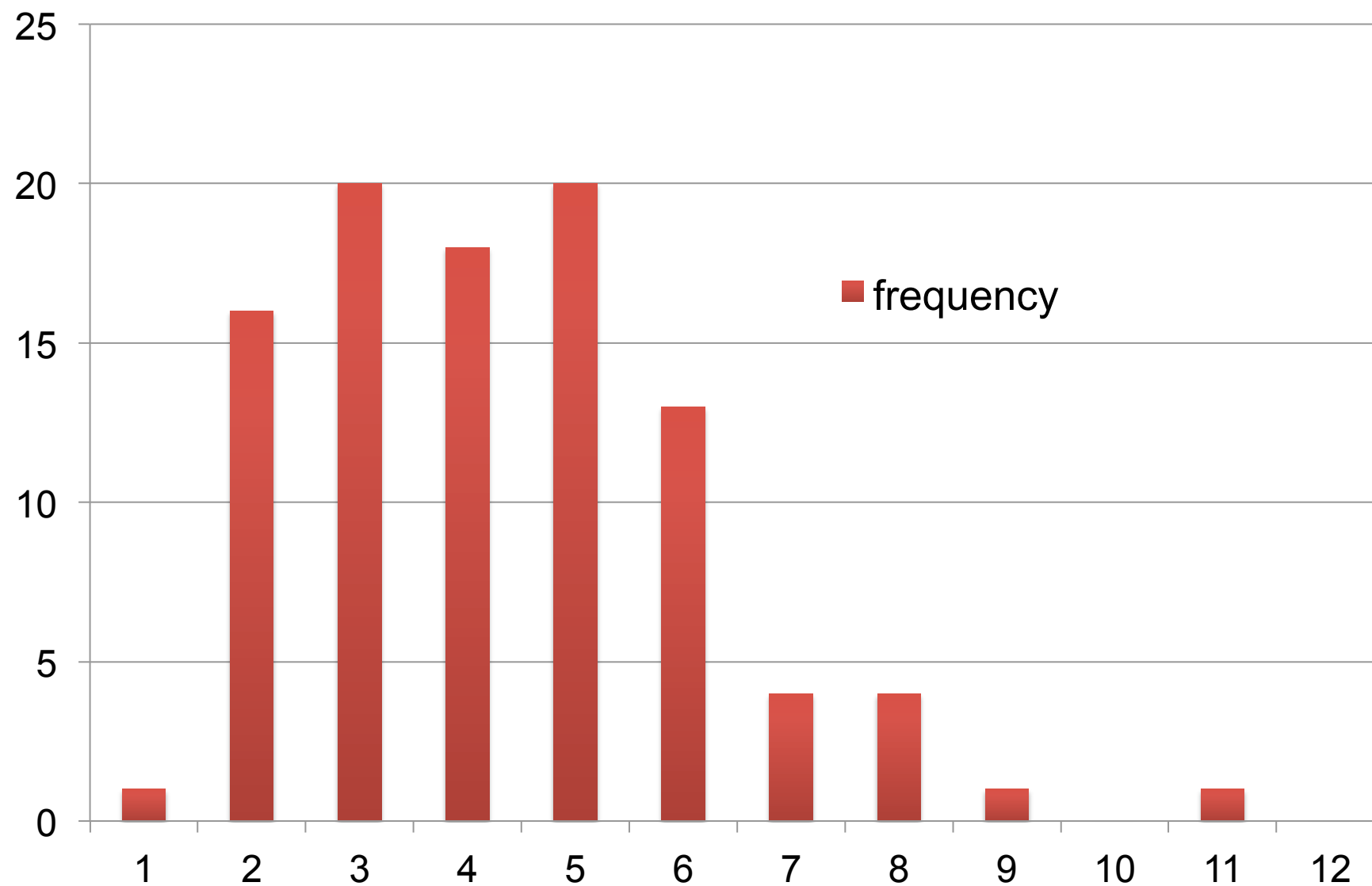
$$p(0) = \frac{e^{-1.5} 1.5^0}{0!} = 0.223$$

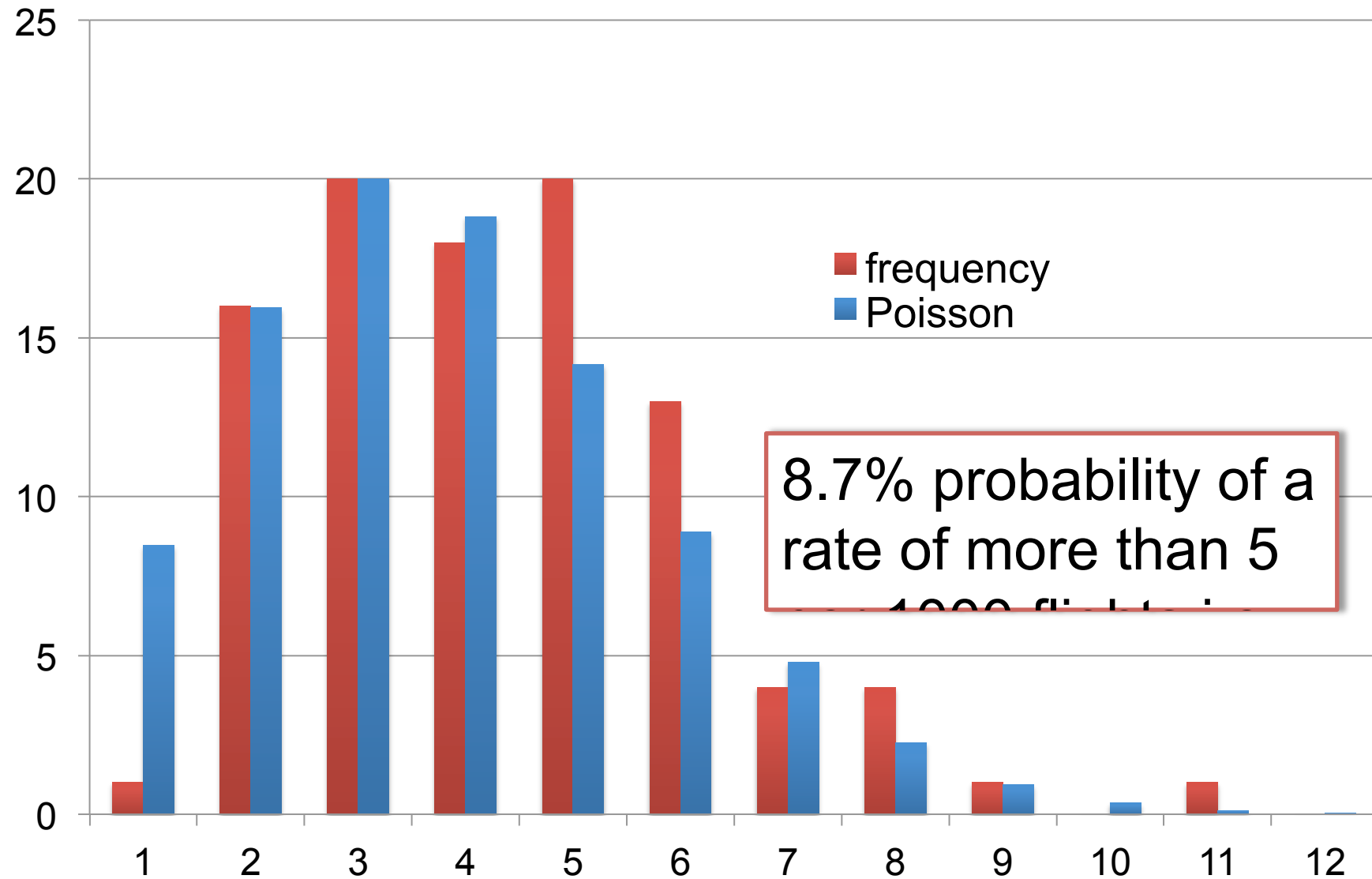


- There are conditions to the use of this approach!

App speed low







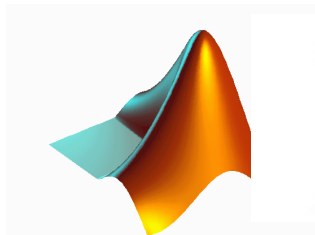
Conclusions

- Be aware of the common traps
- Digging deeper into safety information does not have to be too time consuming
- Don't base safety related decisions on poor analysis!

Where next?



- Hypothesis testing
- Prediction through distribution
- R - Open Source
- Matlab, SPSS, Minitab, Excel, STATISTICA



QUESTIONS?