# Using @Risk for simulation modeling

Ashley Hill

# Stochastic simulation model in Excel

- Each iteration:
  - Cells show random values drawn from probability distributions
  - Spreadsheet calculates outputs and results stored
- After many iterations:
  - Stored values for each output graphed, analyzed
  - Result represents *approximate* probability distribution of future outcomes

# Monte Carlo simulation in spreadsheet models

- @RISK gives Excel the ability to allow certain quantities to be poorly known or variable
- 3 capabilities of @RISK:
  - Creating input distributions for uncertain quantities
  - Running a simulation
  - Analyzing results

# @RISK functions

- @RISK functions take the format: =RiskXxx()
- 3 categories of @RISK functions:
  - Distributions:
    - RiskNormal(), RiskBinomial(), RiskTriang()
  - Statistics (allow reporting into Excel sheet):
    - RiskMean(), RiskCurrentIter(), RiskResultsGraph()
  - Inputs (which effect actions on input distributions)
    - RiskIndepc(), RiskCollect(), RiskTruncate()
- We will focus primarily on distributions

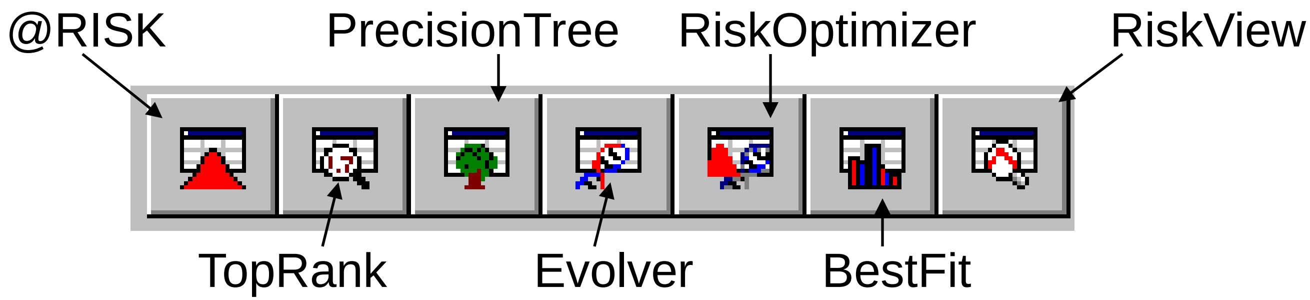
# Introduction to @Risk

- Accessing @Risk in Excel
- Toolbars
- @Risk menus
- Model window
- Results window

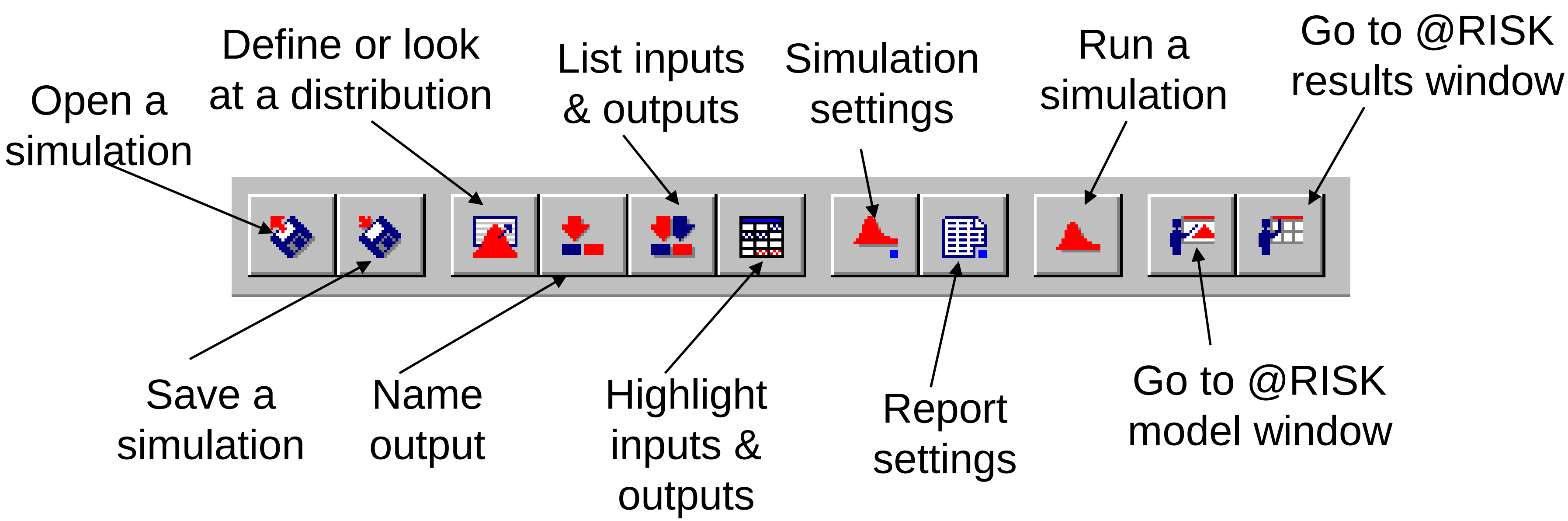
# Working example

- We'll use Aaron's Reed-Frost example from this morning to explore @Risk
  - Population = 101
  - Initial number of susceptibles = 100
  - Initial number of cases = 1
  - $k$ ~ Binomial(4, 0.5 )

# Palisade DecisionTools toolbar




# @RISK toolbar



# Menus

- File
- Edit
- View
- Insert
- Simulation
- Fitting/Results
- Graph
- Window
- Help

# Model window

- List of inputs & outputs in current model
- Insert menu
- Distribution fitting (we will use in later example)

# Results window

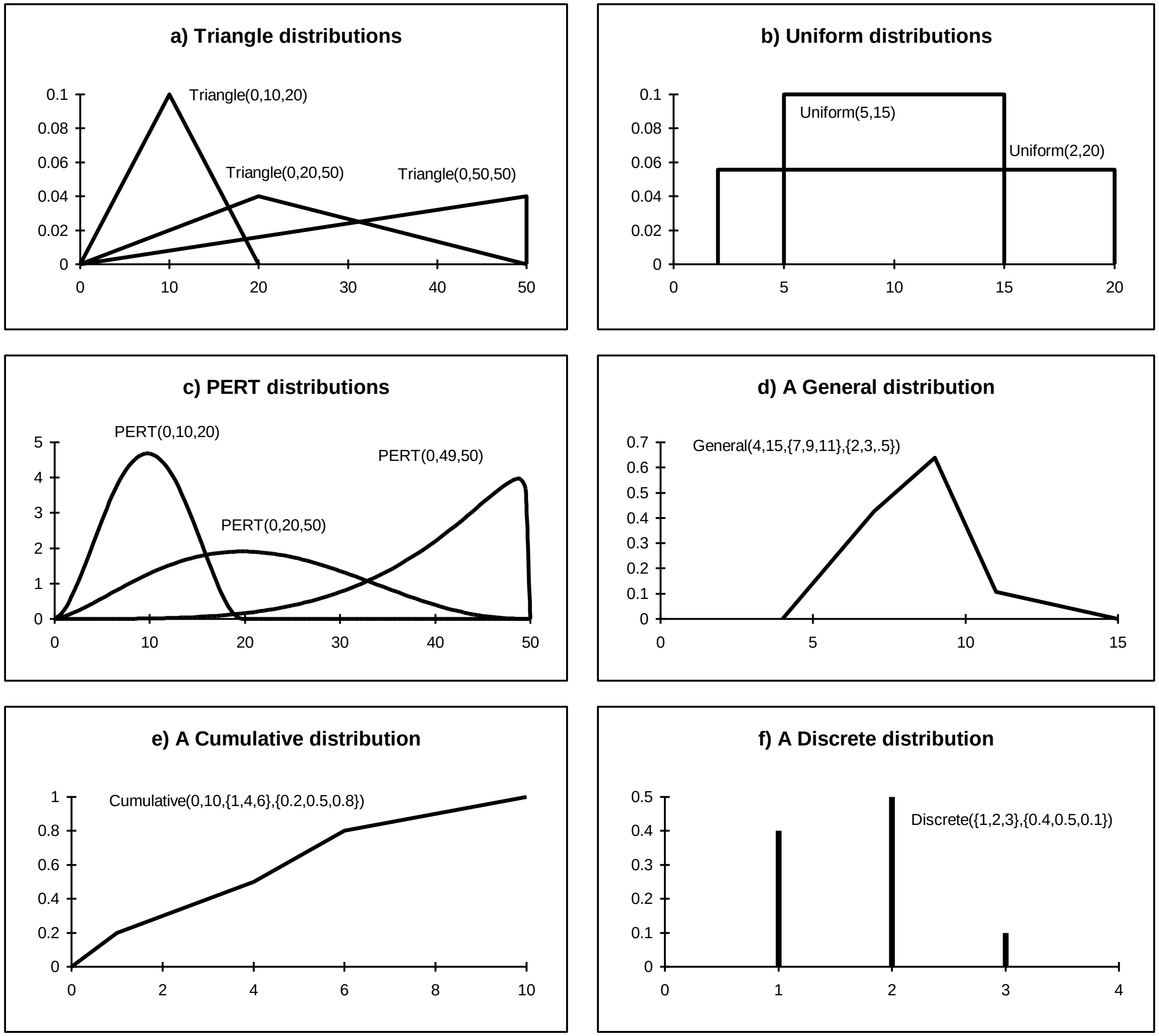
- List of inputs/outputs
- Summary statistics
- Insert menu
  - Detailed statistics
  - Data
  - Graph

# Creating input distributions

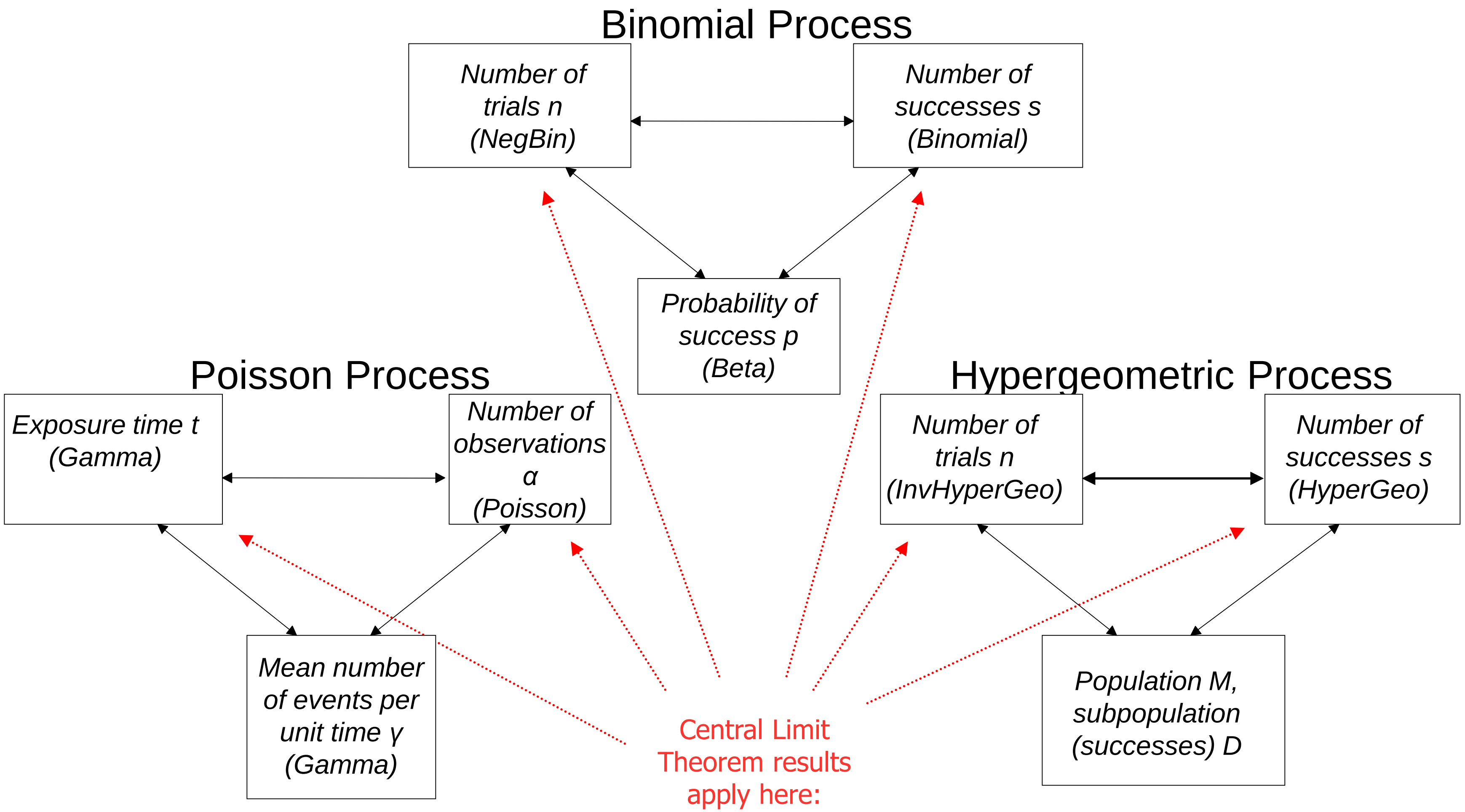
# Review of probability distributions

- Parametric distributions:
  - Binomial Poisson Hypergeometric
  - Gamma NegBin Normal
- Empirical distributions (useful for data, expert opinion)
  - PERT Triangle Beta
  - General Discrete Uniform
  - Bootstrapping (sampling from existing data)
- Knowing which distribution to use requires working knowledge of probability theory

# Review: Empirical probability distributions



# Review: parametric distributions in 3 stochastic processes



# Quick summary of some parametric distributions

| **Binomial process distributions** | |
|---|---|
| **Quantity** | Formula |
| **Number of successes** | s=Binomial(n,p) |
| **Probability of success** | p=Beta(s+1,n-s+1) |
| **Number of trials** | n=s+NegBin(s,p) (last trial is a success) |
| | n=s+NegBin(s+1,p) (last trial is not known) |
| **Poisson process distributions** | |
| **Number of events** | α=Poisson(λt) |
| **Mean number of events per unit exposure** | λ=Gamma(α,1/t) |
| **Time until first event** | t1=Exponential(1/λ) |
| **Time until first α events** | tα=Gamma(α,1/ λ) |
| **Hypergeometric process distributions** | |
| **Number of successes (subpopulation) in sample** | s=HyperGeo(n,D,M) |
| **Number of samples to observe s successes** | n=s+InvHyp(s,D,M) |

# Creating input distributions

- Select appropriate distribution
  - Fit distributions to data (if data available!)
  - Use distribution that is mathematically appropriate
  - Handy tools:
    - BetaBuster
      - Generates Beta parameters for expert opinion on Se, Sp, prevalence or other proportions
      - http://www.epi.ucdavis.edu/diagnostictests/betabuster.html

# Example 1:

- In an adequate contact, the probability of transmitting Disease A from an infected to susceptible animal is 0.4.
  - A single infected animal has 40 adequate contacts with susceptible animals during its infectious period
  - On average, how many animals will an infected animal infect?
  - If 100 infected animals are released, what's the fewest number of new infections from one infected animal? What's the highest number of new infections per animal?

# Example 1

- Working through the problem
  - What probability distribution to use?
  - What are the parameters for the distribution?
  - How many iterations should we run?
  - Insert distribution using @Risk
  - Run iterations
  - Evaluate output

# Example 2

- We think that the probability of infection given adequate contact is 0.4, but when we asked experts, they said it was most likely 0.4, but was definitely less than 0.8
  - What parameter does this affect?
  - Is this uncertainty or variability?
  - How can we incorporate expert opinion into our model?

# Example 3

- The average number of adequate contacts is 40, but anecdotally it ranges from 10 to 60 adequate contacts.
  - How does this change our results?
    - What distribution(s) could we use?
    - How do we add this to our simulation model?

# Example 4

- A graduate student collected daily data on the number of adequate contacts between 40 animals.
  - How can we incorporate this data into our model?
    - Do we need any other information?
    - Fit a distribution
    - Bootstrap

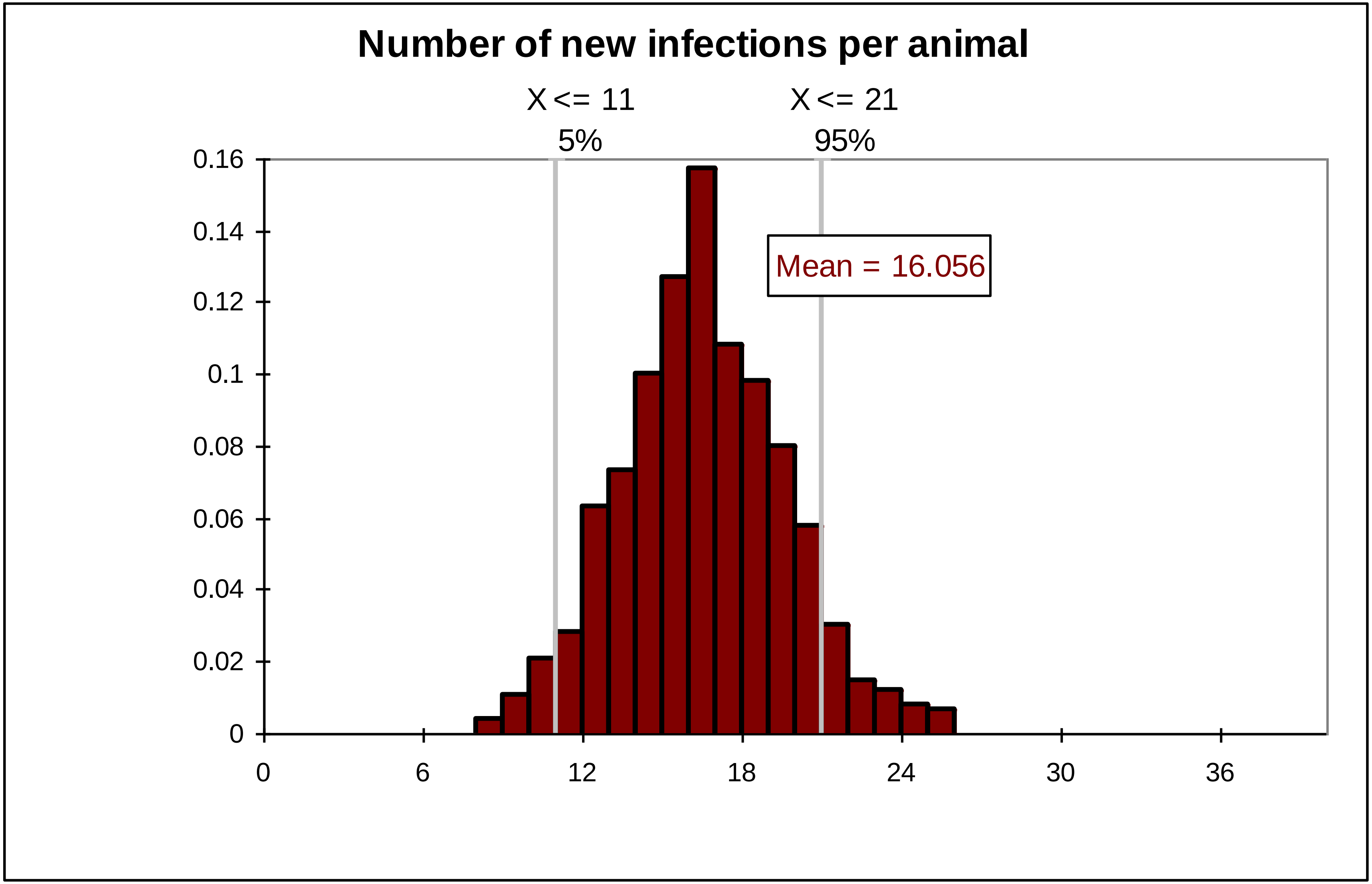
# Distribution fitting

- Distribution fitting
  - Methods of comparison
    - Visual
    - Difference
    - P-P and Q-Q
  - Statistics and Goodness of Fit
    - Statistics: Descriptives on data and distribution
    - GOF: Results of Chi-square, Anderson-Darling, and Kolmogorov-Smirnov testing

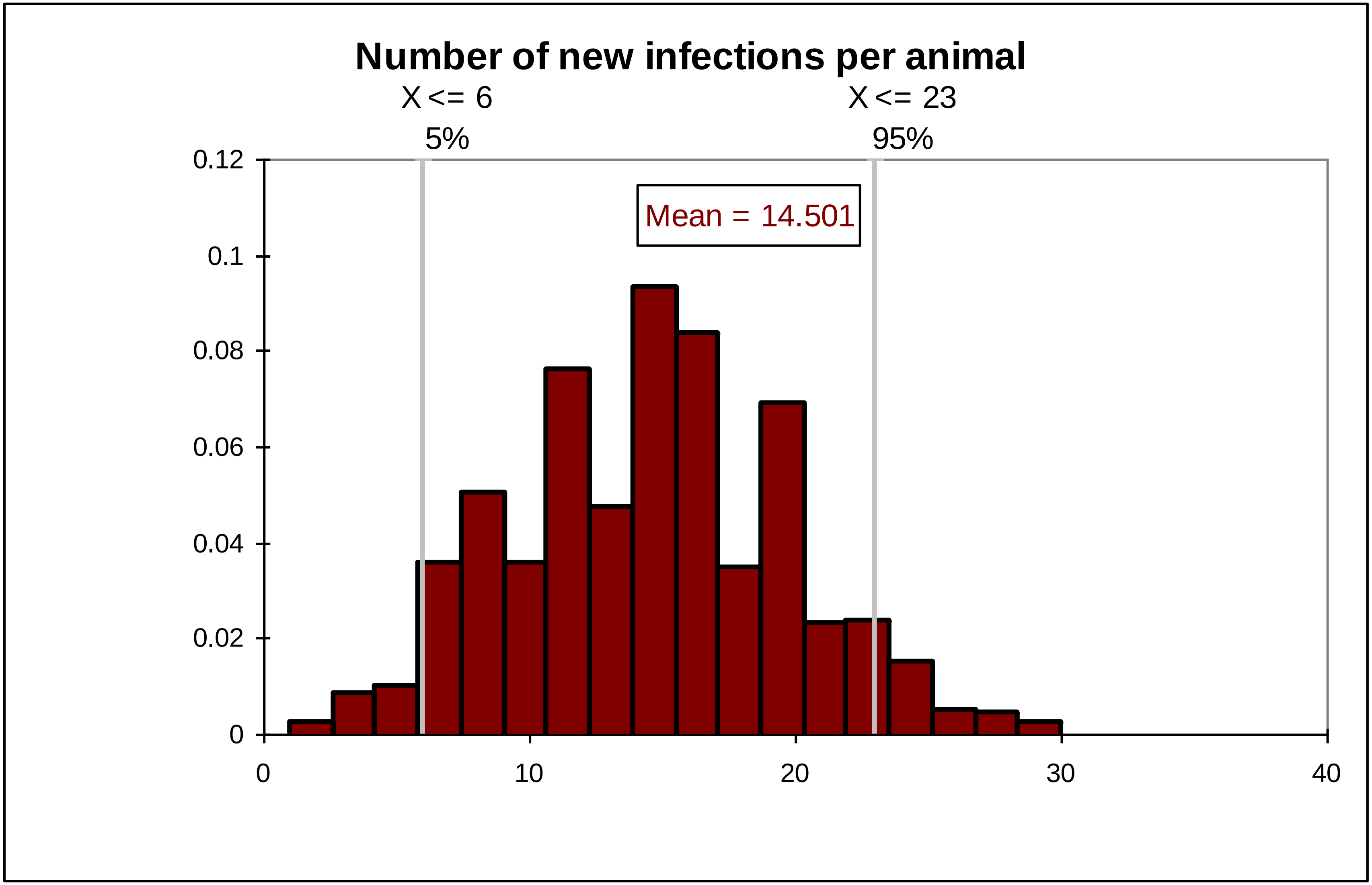
# Comparing results

- Did incorporation of uncertainty about number of adequate contacts change our estimates of the numbers infected?

# Constant number of adequate contacts (40)



# Triangular distribution of adequate contacts (10,40,60)



# Bootstrapped distribution of adequate contacts