ANALYSIS OF SIMULATION DATA:
INPUT MODELING

Lecture # 28

Lecture by Engr. Sidra

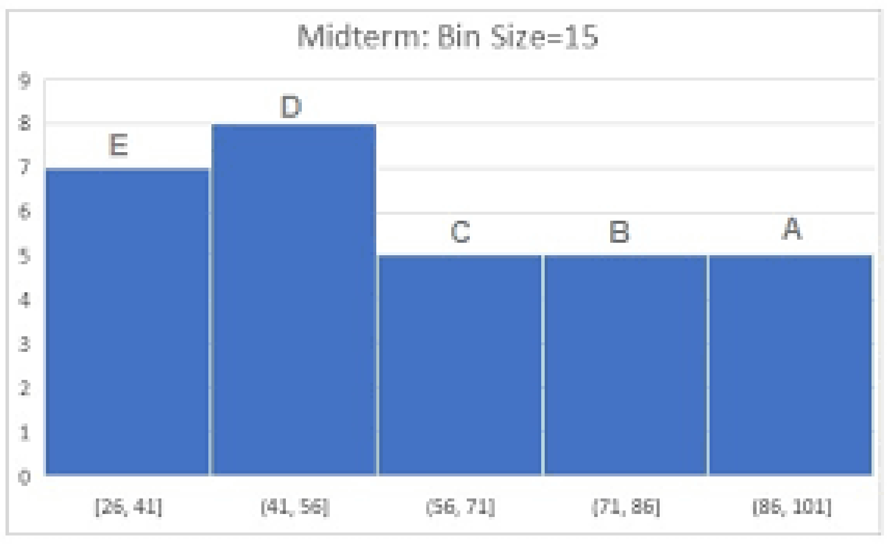
2. Identifying the distribution with Data

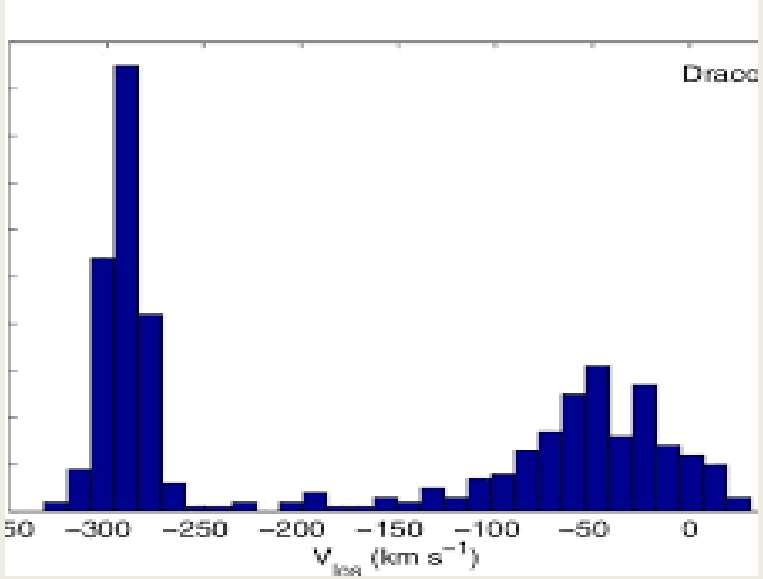
1. Histogram:

- A frequency distribution or histogram is useful in identifying the shape of a distribution.
- A histogram is constructed as follows:
  1. *Divide the range of the data into intervals.*
  2. *Label the horizontal axis to conform to the intervals selected.*
  3. *Find the frequency of occurrences within each interval.*
  4. *Label the vertical axis so that the total occurrences can be plotted for each interval.*
  5. *Plot the frequencies on the vertical axis.*

1. Histogram:

- The number of class intervals depends on the number of observations and on the amount of scatter or dispersion in the data.
  - *If the intervals are too wide, the histogram will be coarse, or blocky, and its shape and other details will not show well.*
  - *If the intervals are too narrow, the histogram will be ragged and will not smooth the data.*



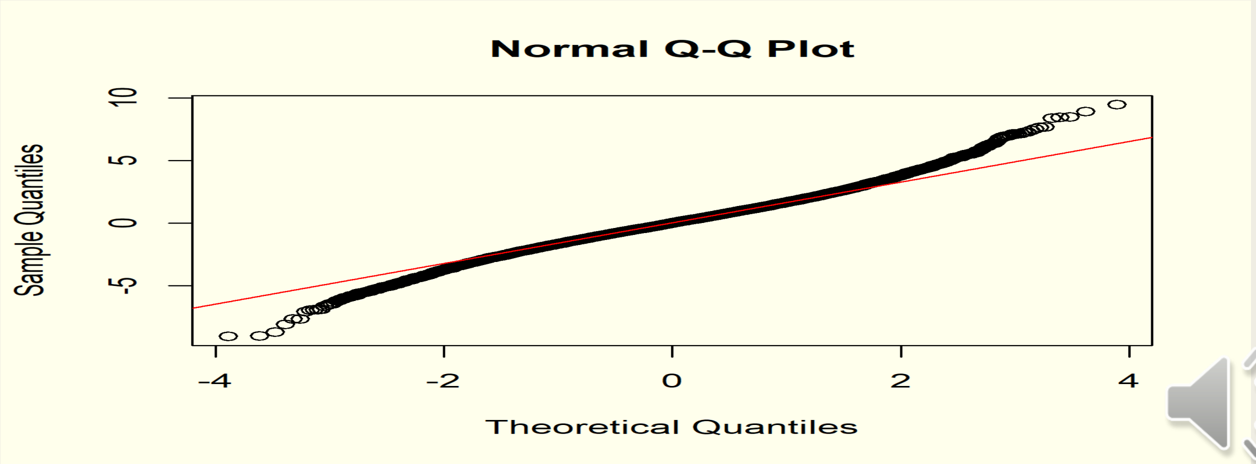


2. Summary Statistics:

- *Selecting distributions is to use the physical basis of the distributions as a guide.*

3. Quantile-Quantile plot (q-q plot):

- *In statistics, a Q–Q plot is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other.*
- *If both sets of quantiles came from the same distribution, we should see the points forming a line that's roughly straight*
- *a useful tool for evaluating distribution fit, one that does not suffer from those problems.*



3. Parametric Estimation

- After a family of distributions has been selected, the next step is to estimate the parameters of the distribution.
- Preliminary Statistics: Sample Mean and Sample Variance
  - *When discrete or continuous raw data are available*

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1}$$

• Preliminary Statistics: Sample Mean and Sample Variance

– *When data are discrete and grouped in frequency distribution*

$$\bar{X} = \frac{\sum_{j=1}^{k} f_j X_j}{n} \qquad S^2 = \frac{\sum_{j=1}^{k} f_j X_j^2 - n\bar{X}^2}{n-1}$$

- Where k is distinct values of X and fj is observed frequency of values Xj of X.

– *When the data are discrete or continuous and have been placed in class intervals.*

$$\bar{X} \doteq \frac{\sum_{j=1}^{c} f_j m_j}{n} \qquad S^2 \doteq \frac{\sum_{j=1}^{c} f_j m_j^2 - n\bar{X}^2}{n-1}$$

- Where fj is the observed frequency in the j^{th} class interval, mj is the midpoint of j^{th} interval and c is number of class intervals.

Suggested Estimators

- Numerical estimates of the distribution parameters are needed to reduce the family of distributions to a specific distribution.

Distribution	*Parameter(s)*	*Suggested Estimator(s)*
Poisson	α	$\hat{\alpha} = \bar{X}$
Exponential	λ	$\hat{\lambda} = \frac{1}{\bar{X}}$
Gamma	β, θ	$\hat{\beta}$ (see Table A.9) $\hat{\theta} = \frac{1}{\bar{X}}$
Normal	μ, σ^2	$\hat{\mu} = \bar{X}$ $\hat{\sigma}^2 = S^2$ (unbiased)
Lognormal	μ, σ^2	$\hat{\mu} = \bar{X}$ (after taking ln of the data) $\hat{\sigma}^2 = S^2$ (after taking ln of the data)

4. Goodness-of-Fit Tests

- Goodness-of-fit tests provide helpful guidance for evaluating the suitability of a potential input model
- It is especially important to understand the effect of sample size.
- If very little data are available, then a goodness-of-fit test is unlikely to reject any candidate distribution; but if a lot of data are available, then a goodness-of-fit test will likely reject all candidate distributions.
- Therefore, failing to reject a candidate distribution should be taken as one piece of evidence in favor of that choice, and rejecting an input model as only one piece of evidence against the choice.