# This Week

## Monday

- Modeling and Simulating Uncertainty;
  Lab Exercise: RV Modeling and Simulation

## Wednesday

- NO CLASS

# Topics

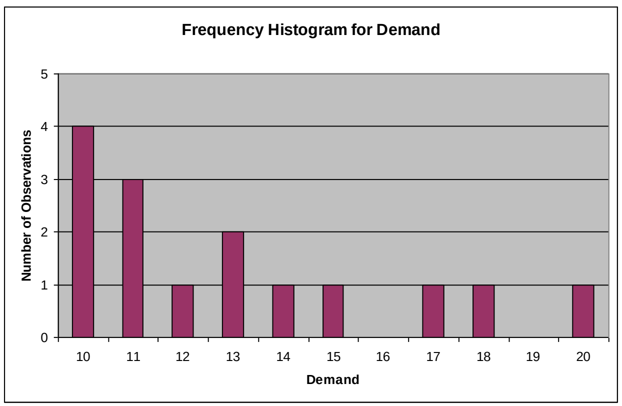
- Characterizing Uncertainty
- Probability Functions in Excel
- Generating Random Variables in Excel

# A Problem Solving Framework

1. *Define the Problem*
2. *Collect and Organize Data*
3. *Characterize Uncertainty and Data Relationships*
4. *Build an Evaluation Model*
5. *Formulate a Solution Approach*
6. *Evaluate Potential Solutions*
7. *Recommend a Course of Action*

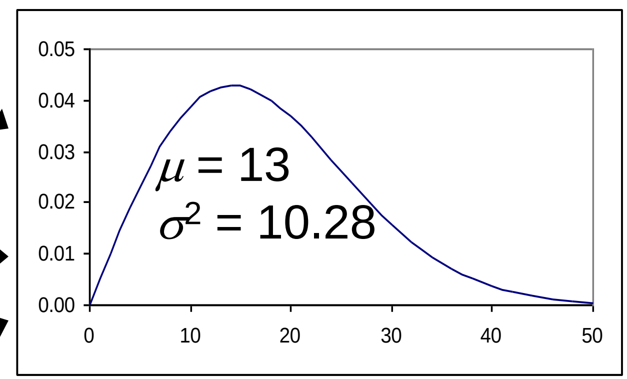
# A Problem Solving Framework

3. ***Characterize uncertainty for modeling purposes*** by *choosing specific probability functions or processes* to represent sources of uncertainty within the context of the modeling framework:



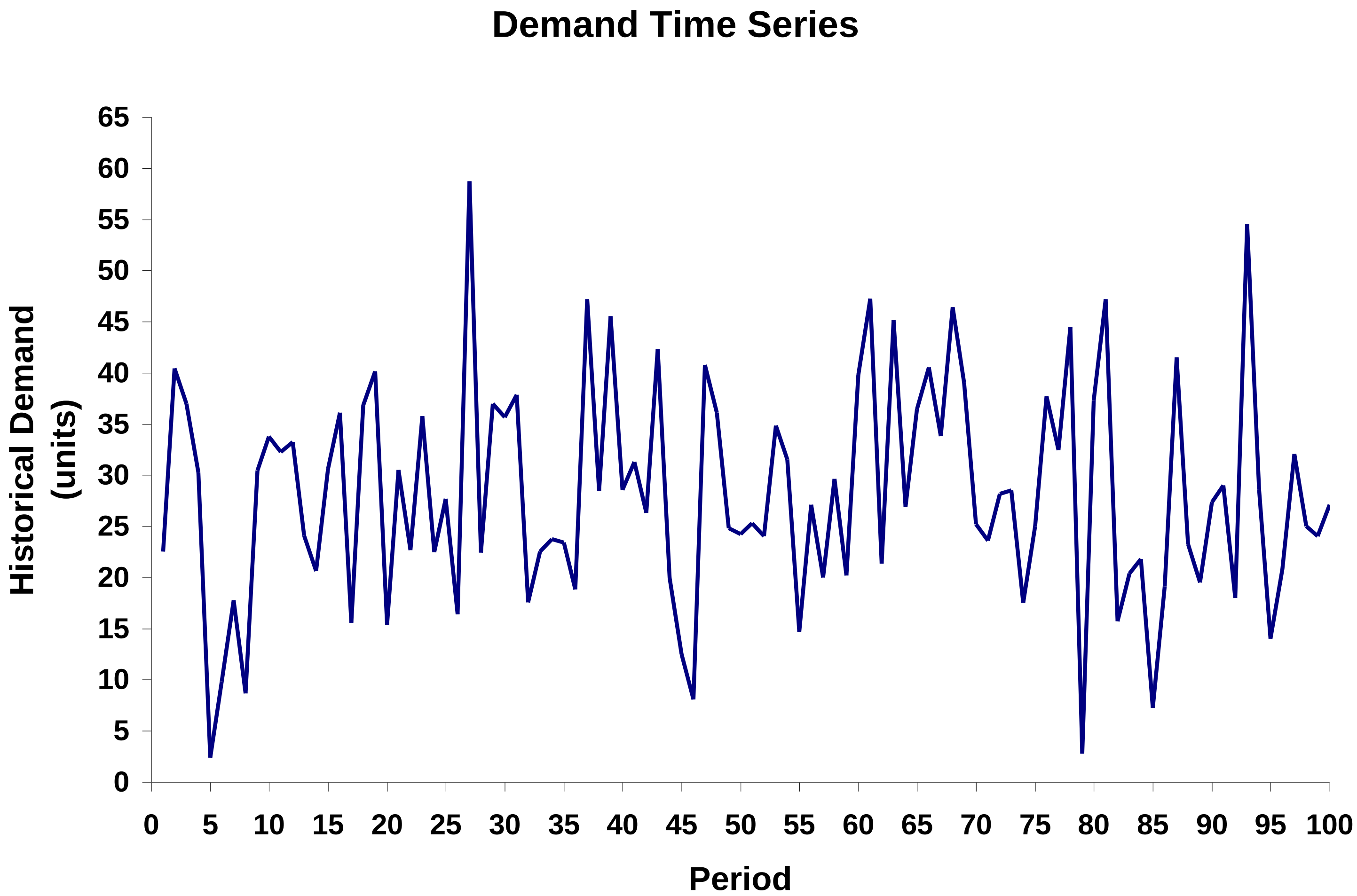

$$\overline{X} = 13, \; s^2 = 10.28$$

**Problem Definition and Parameters**

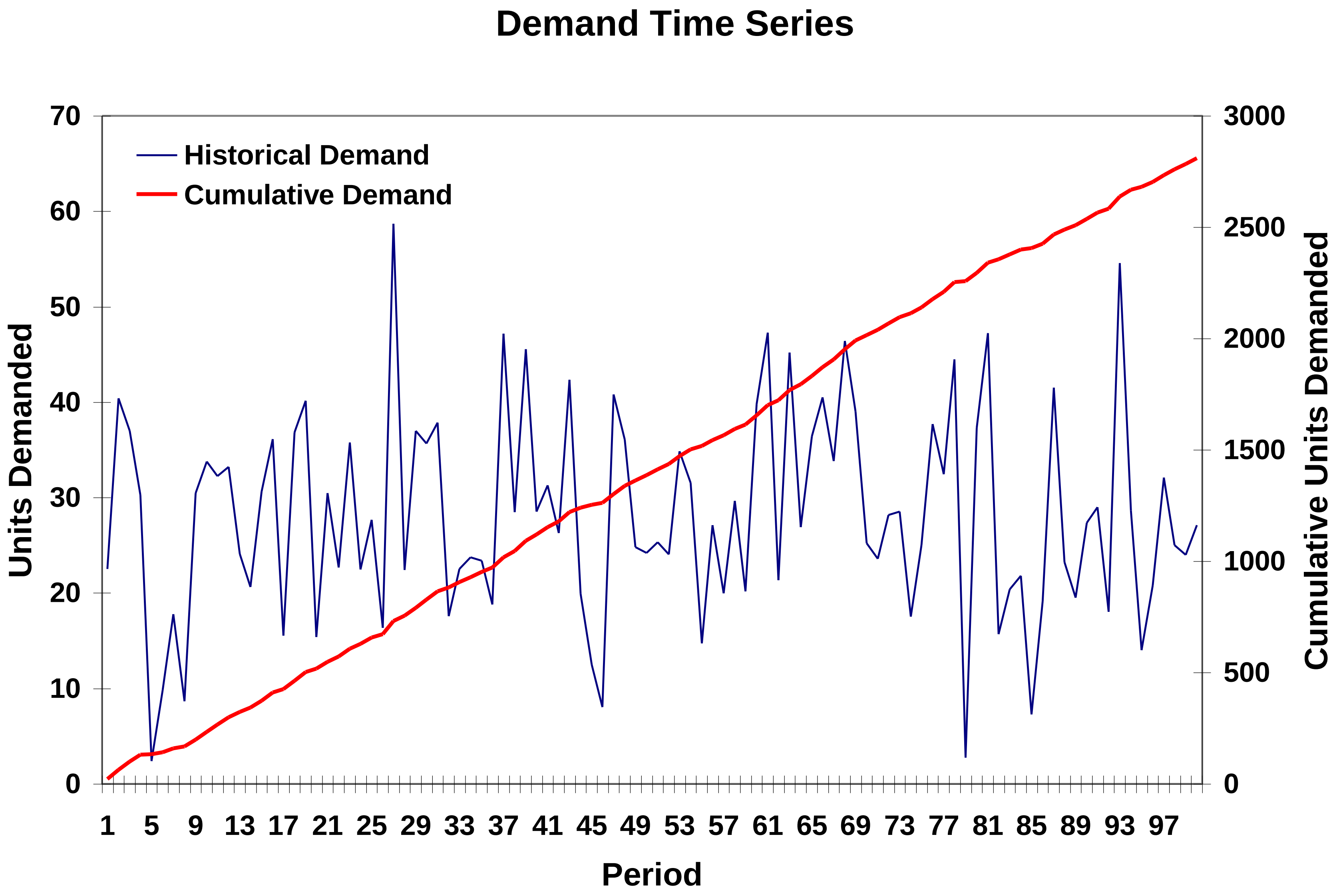


# Example: Demand Planning

***How can we characterize the uncertainty associated with future demand?***

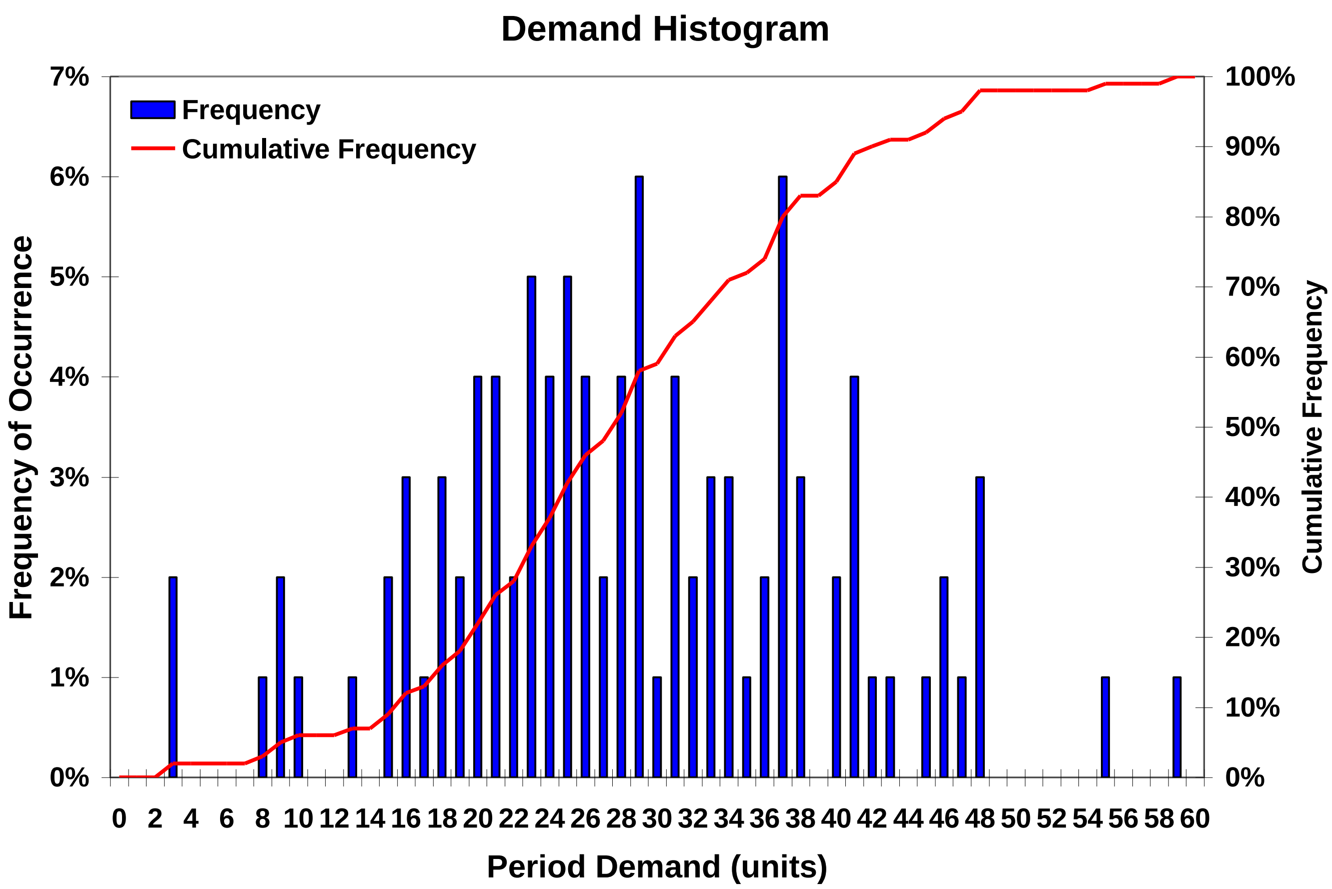


# Cumulative View of Historical Demand



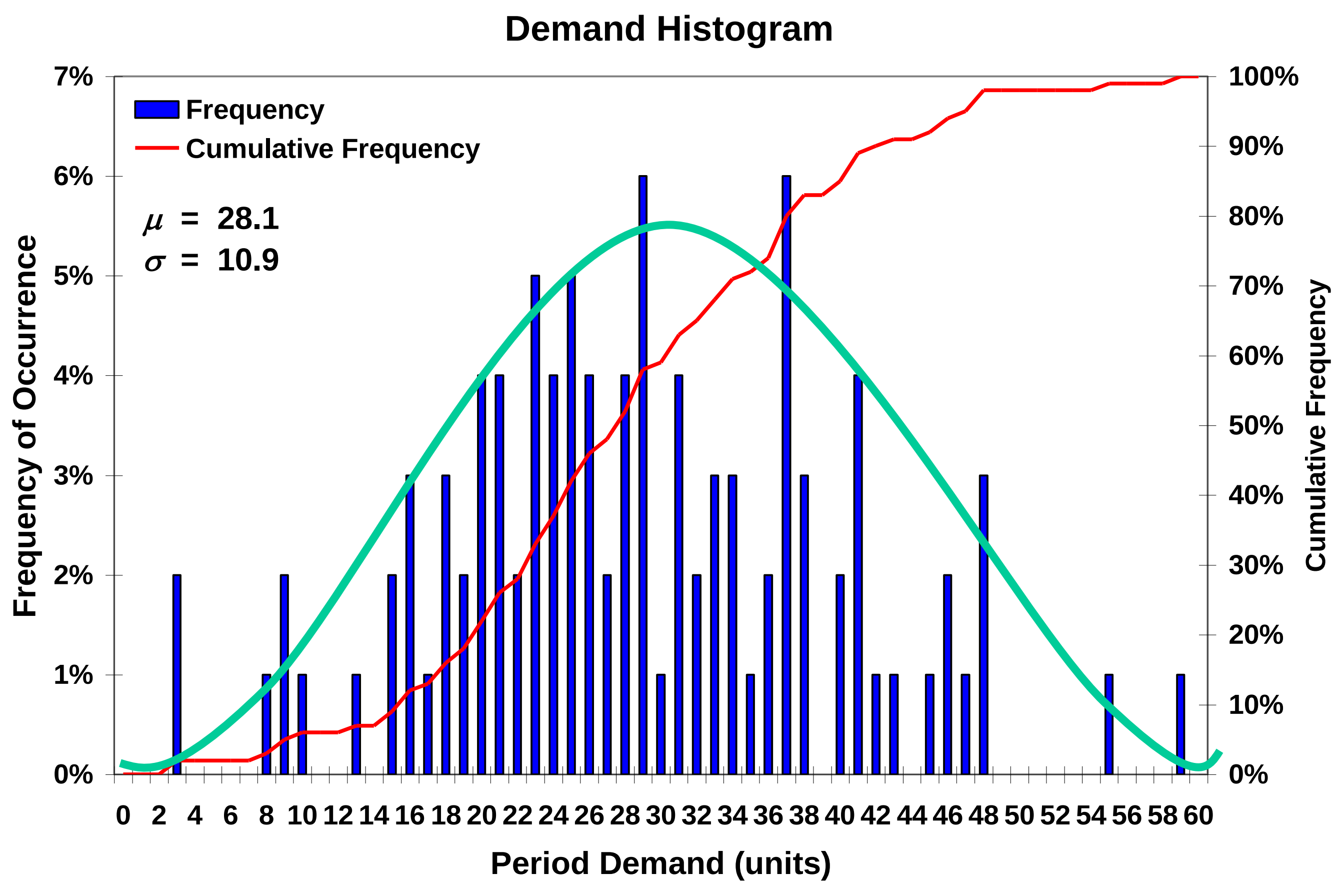
# Characterizing Demand Uncertainty

**<u>One option</u>**: Using some number of past periods, construct a ***frequency histogram*** and use these frequencies to represent the probability function for future demand.



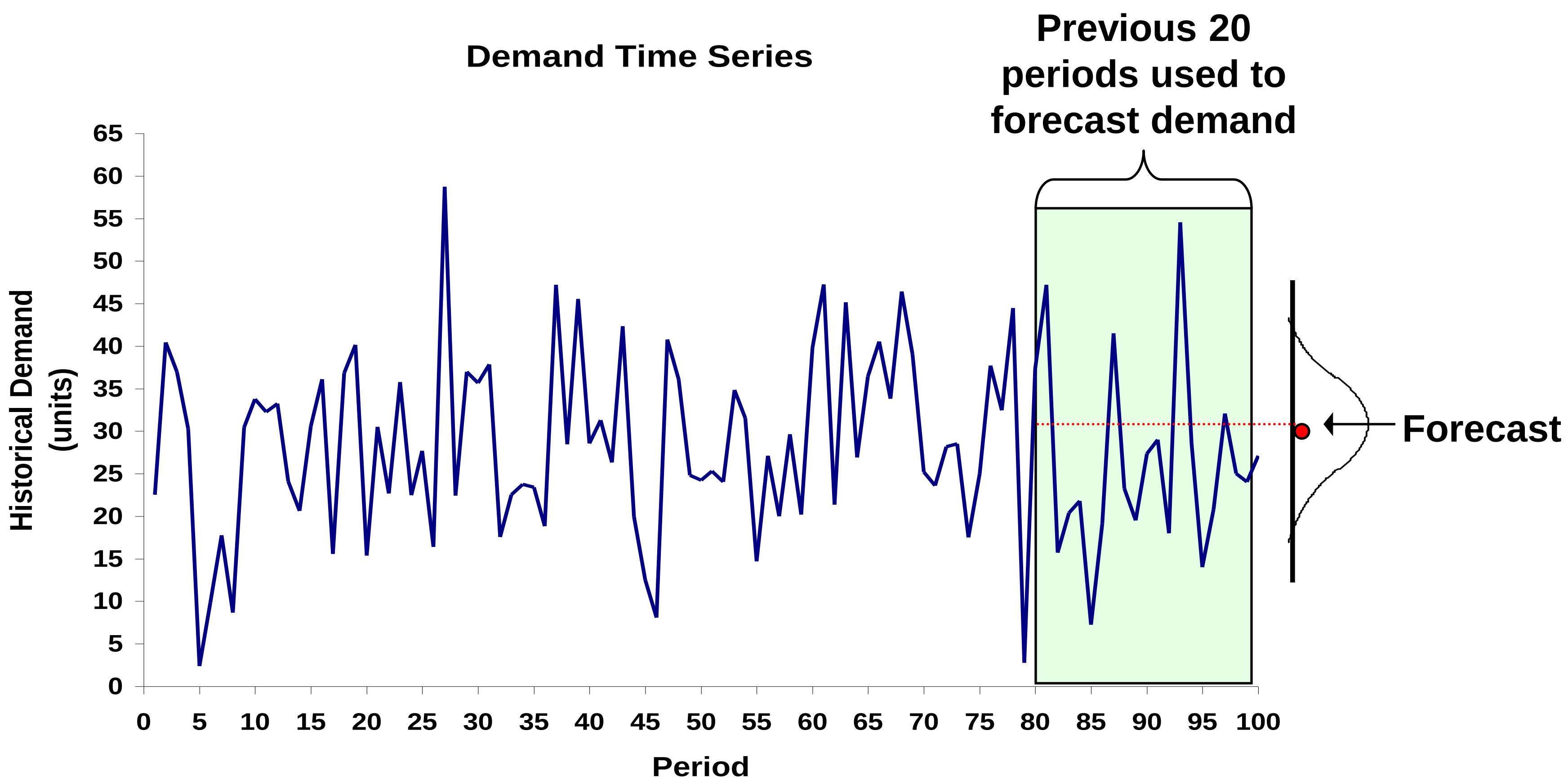
# Characterizing Demand Uncertainty

**<u>Another option</u>**: Use a ***well-defined probability function*** whose shape resembles the frequency histogram (e.g. Normal) to represent future demand uncertainty.



# Characterizing Demand Uncertainty

<u>**A third option**</u>: Use a ***well-defined function*** of past periods' demand to predict future demand and a ***well-defined probability function*** to represent the ***forecast errors***.



# Bottom Line on Modeling Uncertainty

- If the chosen model/function ***closely resembles*** the true likelihoods, then we can:
  - ***Accurately assess the risk*** of different events.
  - State that our conclusions and decisions are valid with a ***high level of confidence***.

- If the chosen model/function ***does not resemble*** the true likelihoods, then:
  - Our analyses and assessment of risk may be ***inaccurate***.
  - Our conclusions/decisions may be ***wrong***.

# Commonly Used Families of Probability Functions

| Family | Type | Range | Parameters |
|---|---|---|---|
| Normal | Continuous | $(-\infty,+\infty)$ | 2 ($\mu$ and $\sigma$) |
| Binomial | Discrete | [0,1,…,n] | 2 ($n$ and $p$) |
| Poisson | Discrete | $[0,1,\ldots+\infty)$ | 1 ($\lambda$) |
| Exponential | Continuous | $(0,+\infty)$ | 1 ($\lambda$) |

**Notation: $X \sim N(\mu=10, \sigma=2)$**

This means that the random variable $X$ has a normal distribution with the specified parameters.

Other Examples: $X \sim \text{Bin}(n = 20, p = .05)$, $X \sim \text{Po}(\lambda = 10)$, $X \sim \text{Exp}(\lambda = 0.2)$

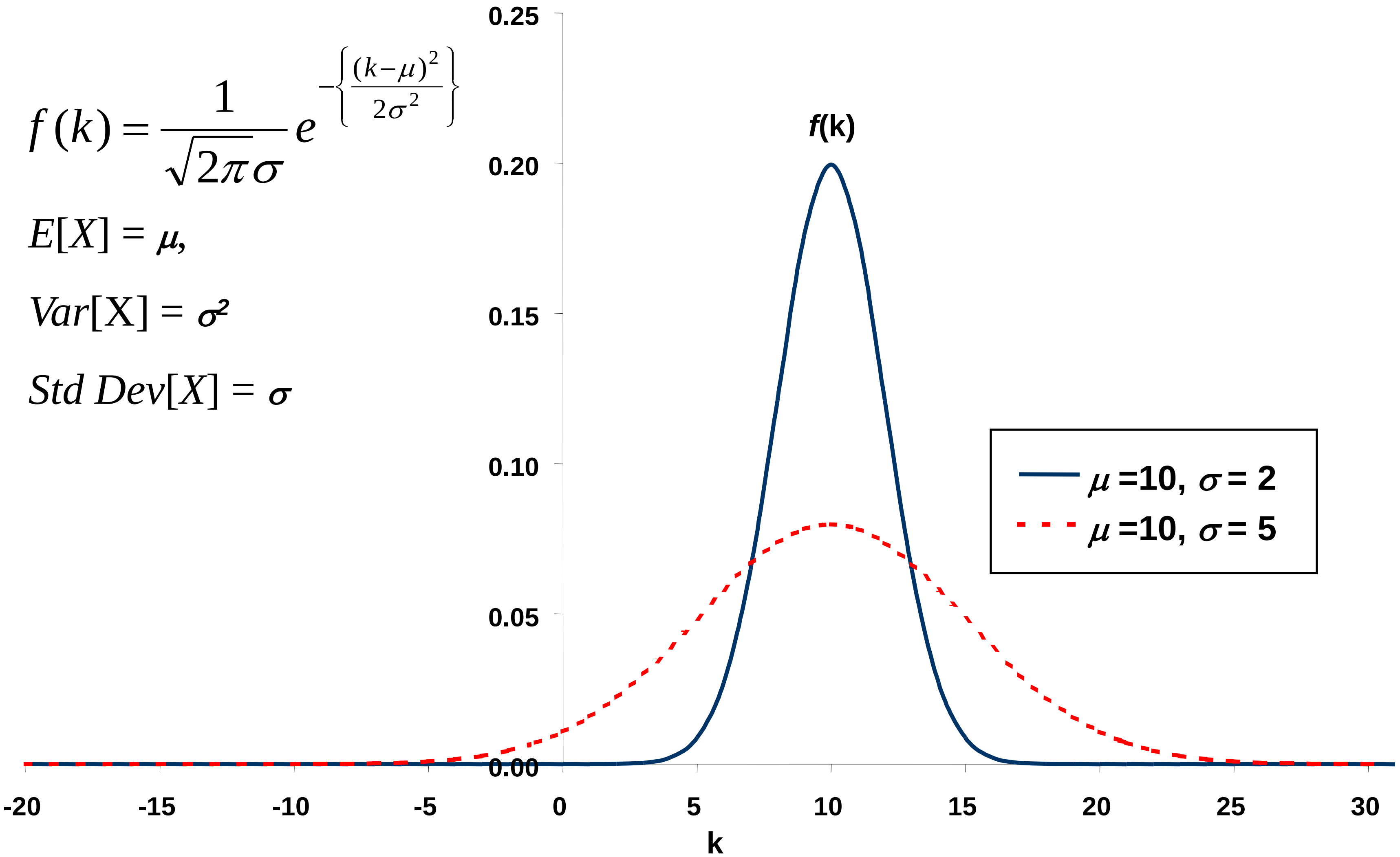
# The Normal Distribution

$$f(k) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left\{\frac{(k-\mu)^2}{2\sigma^2}\right\}}$$

$E[X] = \mu,$

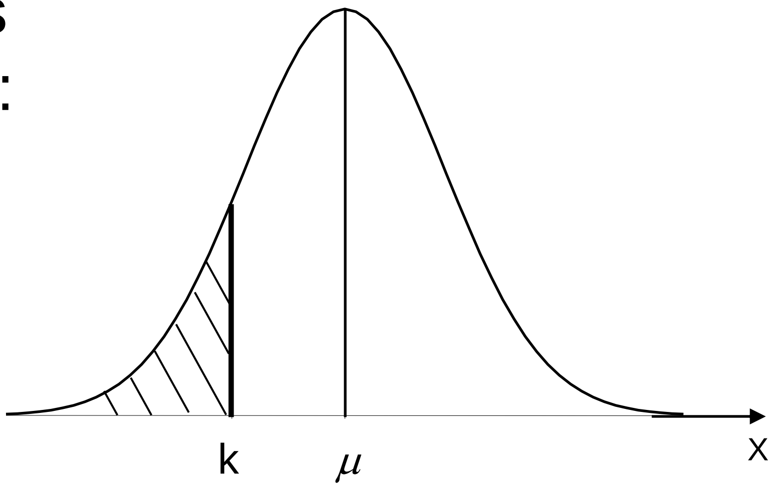
$Var[X] = \sigma^2$

$Std\ Dev[X] = \sigma$



# Normal Distribution in Excel

Excel has the following functions related to the normal distribution: $X \sim N(\mu, \sigma)$




- **=NORM.DIST(k,$\mu$,$\sigma$,TRUE)** returns P(X ≤ k). (If last parameter is **FALSE**, then *f*(k) is returned.)
- **=NORM.INV(*percentile*,$\mu$,$\sigma$)** returns the value k such that P(X ≤ k) = *percentile*.
- **=NORM.S.DIST(k)** and **=NORM.S.INV(*percentile*)** return analogous values for $Z \sim N(0, 1)$.
- **=STANDARDIZE(x,$\mu$,$\sigma$)** returns the standardized value $z = (x - \mu)/\sigma$.

# A Decision Problem

- A management consulting firm has interviewed several candidates for entry-level positions and would like to make job offers to some of them. For logistical reasons, all job offers must be issued ***at the same time***.
- Ideally, the firm would like to hire ***between 5 and 7*** new people.
- Past experience has shown that on average, ***80% of job offers are accepted*** (20% are rejected).

***How many offers should the firm make to have the highest likelihood of meeting their hiring target?***

# The Binomial Distribution

$$f(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

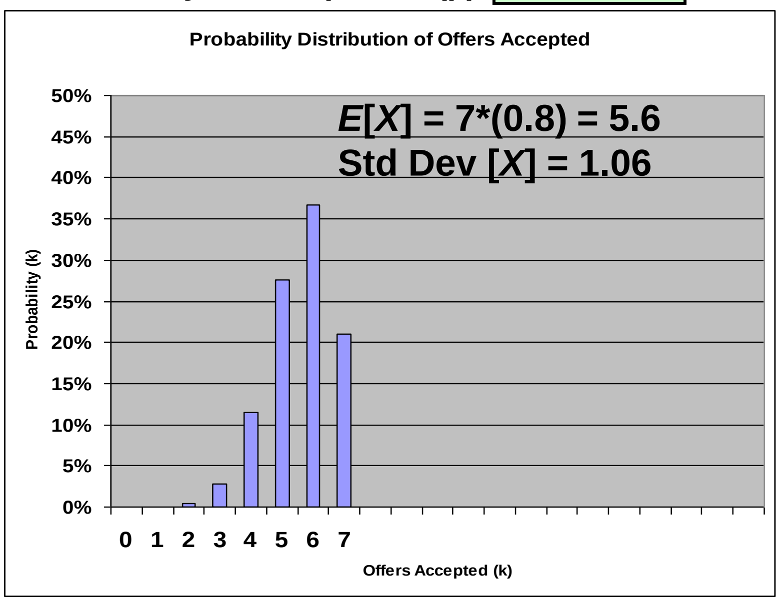
$$E[X] = np$$

$$Var[X] = np(1-p)$$

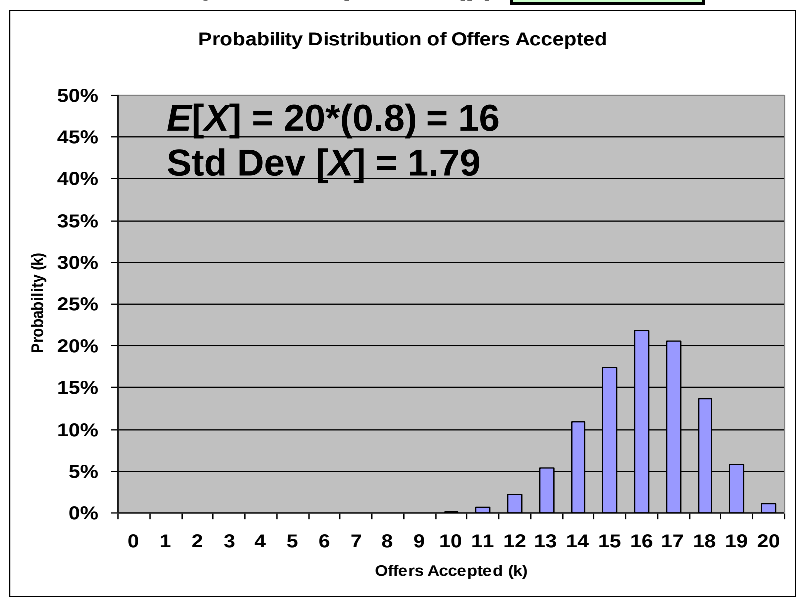
$$StdDev[X] = \sqrt{np(1-p)}$$

Used to model the total ***number of successes*** that occur in ***n*** trials where each trial has the same success probability ***p***.

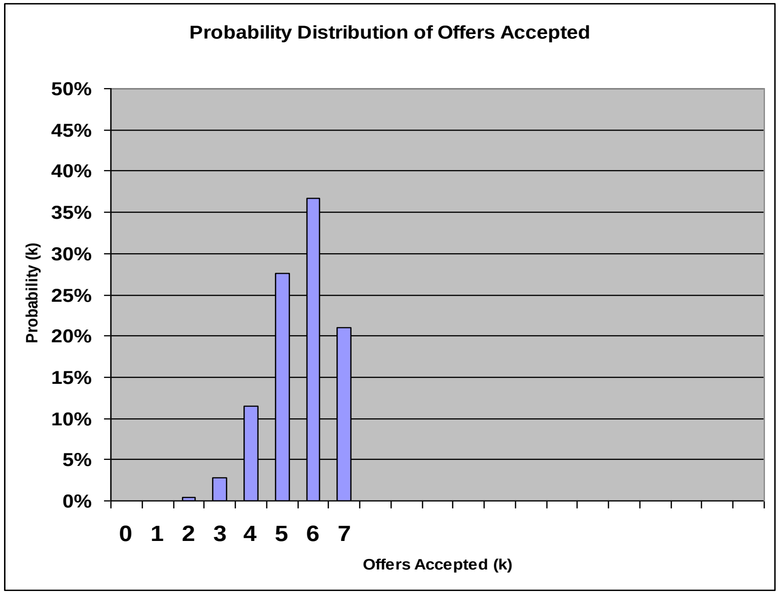
| | |
|---|---|
| **Number of Offers (n):** | **7** |
| **Probability of Acceptance (p):** | **80%** |




| | |
|---|---|
| **Number of Offers (n):** | **20** |
| **Probability of Acceptance (p):** | **80%** |



# Binomial Distribution in Excel




Excel has the following functions related to the binomial distribution:
*X* ~ Bin(*n*,*p*)

- **=BINOM.DIST(k,*n*,*p*,TRUE)** returns P(X ≤ k). (If last parameter is FALSE, then f(k) = P(X = k) is returned.)
- **=BINOM.INV(*n*,*p*,*cutoff_prob*)** returns the smallest value k such that P(X ≤ k) ≥ *cutoff_prob*. This is similar to finding a value that corresponds to a particular percentile, except we are dealing with discrete values.

# The Poisson Distribution

$$f(k) = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

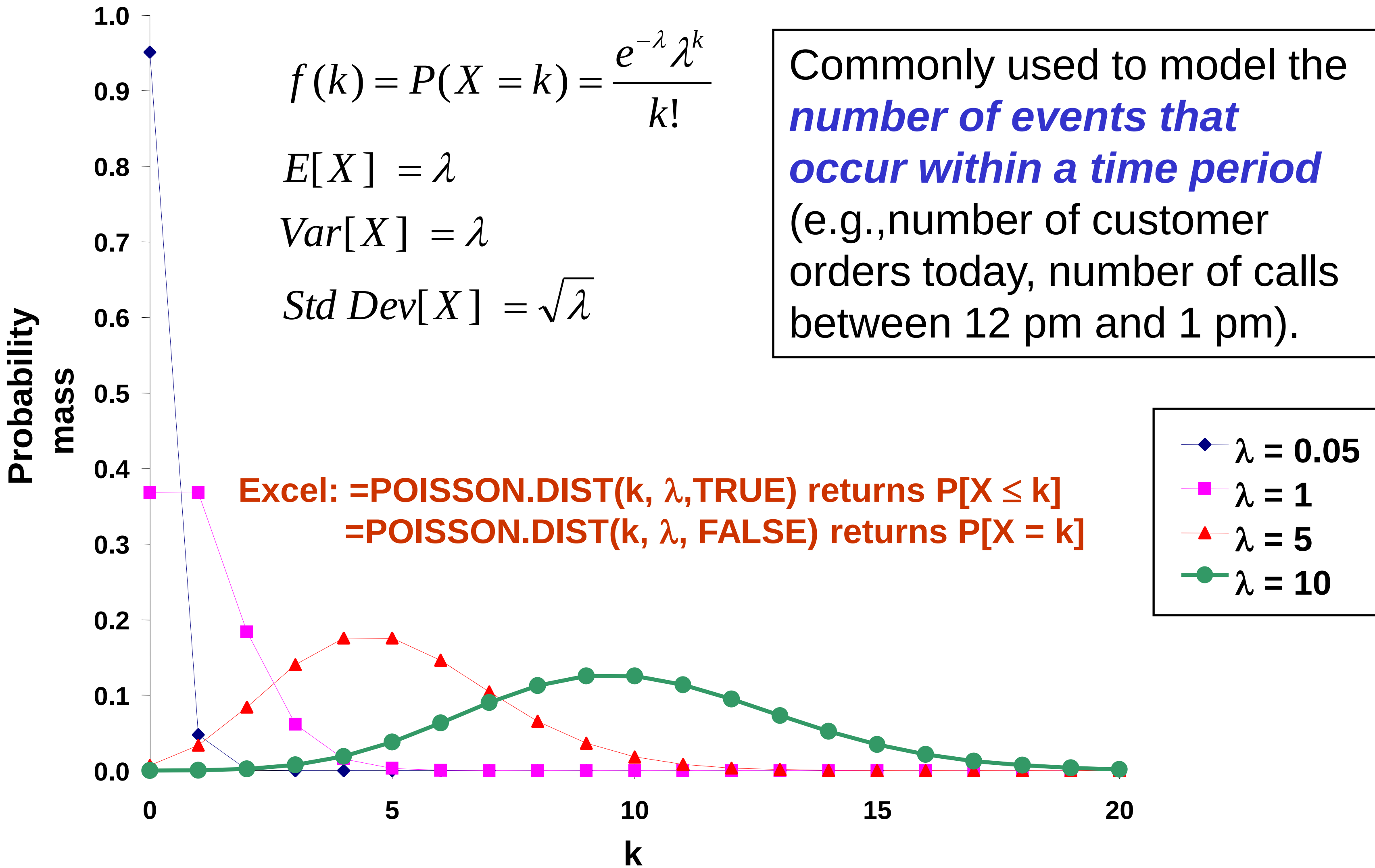
$$E[X] = \lambda$$

$$Var[X] = \lambda$$

$$Std\,Dev[X] = \sqrt{\lambda}$$

Commonly used to model the ***number of events that occur within a time period*** (e.g.,number of customer orders today, number of calls between 12 pm and 1 pm).

Excel: =POISSON.DIST(k, λ,TRUE) returns P[X ≤ k]
=POISSON.DIST(k, λ, FALSE) returns P[X = k]



# The Exponential Distribution

$$f(k) = \lambda e^{-k\lambda}$$

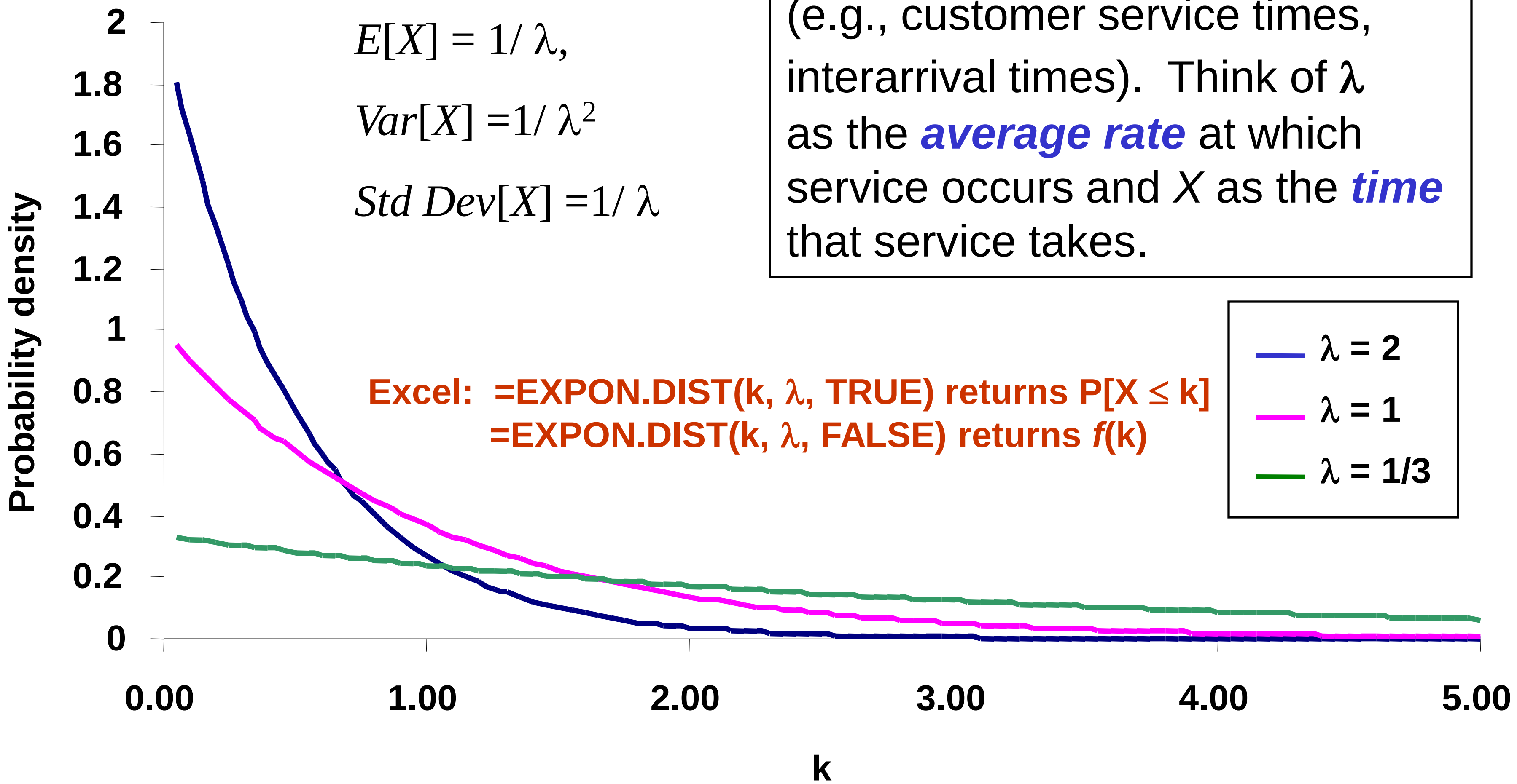
$E[X] = 1/\lambda,$

$Var[X] = 1/\lambda^2$

$Std\ Dev[X] = 1/\lambda$

Commonly used to model ***times*** (e.g., customer service times, interarrival times). Think of λ as the ***average rate*** at which service occurs and *X* as the ***time*** that service takes.

**Excel: =EXPON.DIST(k, λ, TRUE) returns P[X ≤ k]**
**=EXPON.DIST(k, λ, FALSE) returns *f*(k)**



# Other Distributions in Excel

- In addition, Excel has numerous other built-in probability functions. Here are some of the more commonly used:
  - **BETA.DIST** (Beta)
  - **CHISQ.DIST** (Chi-Squared)
  - **F.DIST** (F)
  - **GAMMA.DIST** (Gamma)
  - **HYPGEOM.DIST** (Hypergeometric)
  - **LOGNORM.DIST** (Lognormal)
  - **NEGBINOM.DIST** (Negative Binomial)
  - **T.DIST** (Student's t)
  - **WEIBULL.DIST** (Weibull)

# Generating Random Variables

To generate a random variable ***X*** with cdf ***F***:

- **Inversion Method**
  - Generates a standard uniform random variable ***U*** over [0,1] and returns $X = F^{-1}(U)$. ***We will focus on this method.***
- **Acceptance-Rejection Method**
  - Good for positive, bounded random variables.
  - Uses two uniform random variables $U_1$ and $U_2$ to generate a point ($x$,$y$) in a specially-sized 2D-box, where part of the box is designated "acceptable" and the rest is not.
  - Returns $X = x$ if ($x$,$y$) falls in the "acceptable" region, else tries again.
- **Convolution Method**
  - Good for random variables $X = Y_1 + \ldots + Y_n$ that are the sums of (not too many) other independent random variables.
  - Generates the independent pieces (using one of the other methods), then adds them together.

# Inversion Method in Excel

- **Excel: =RAND()** generates a standard uniform random variable ***U*** over [0,1].
- In Excel, can generate $X = \boldsymbol{F}^{-1}(\boldsymbol{U})$ by computing $\boldsymbol{F}^{-1}(\cdot)$ explicitly or by using one of many built-in functions:
  - **NORM.INV**, **NORM.S.INV** (Inverse of Normal, Standard Normal)
  - **BINOM.INV** (Inverse of Binomial)
  - **BETA.INV** (Inverse of Beta)
  - **CHISQ.INV** (Inverse of Chi-Squared)
  - **F.INV** (Inverse of F)
  - **GAMMA.INV** (Inverse of Gamma)
  - **LOGNORM.INV** (Inverse of Lognormal)
  - **T.INV** (Inverse of Student's t)
- Examples:
  - If $X \sim N(\mu = 10, \sigma = 2)$, then generate $X$ with: **=NORM.INV(RAND(),10,2)**
  - If $X \sim \text{Exp}(\lambda = 0.2)$, then generate $X$ explicitly with: **=-LN(1 - RAND())/0.2**

# Discrete Random Variable Example

- Consider an experiment that is the ***roll of two fair dice***.

$$\Omega = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6),$$
$$(2,1), (2,2), (2,3), (2,4), (2,5), (2,6),$$
$$(3,1), (3,2), (3,3), (3,4), (3,5), (3,6),$$
$$(4,1), (4,2), (4,3), (4,4), (4,5), (4,6),$$
$$(5,1), (5,2), (5,3), (5,4), (5,5), (5,6),$$
$$(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}$$

**Sum is 2**
**Sum is 3**
**Sum is 4**
**Sum is 5**
**Sum is 6**
**Sum is 7**

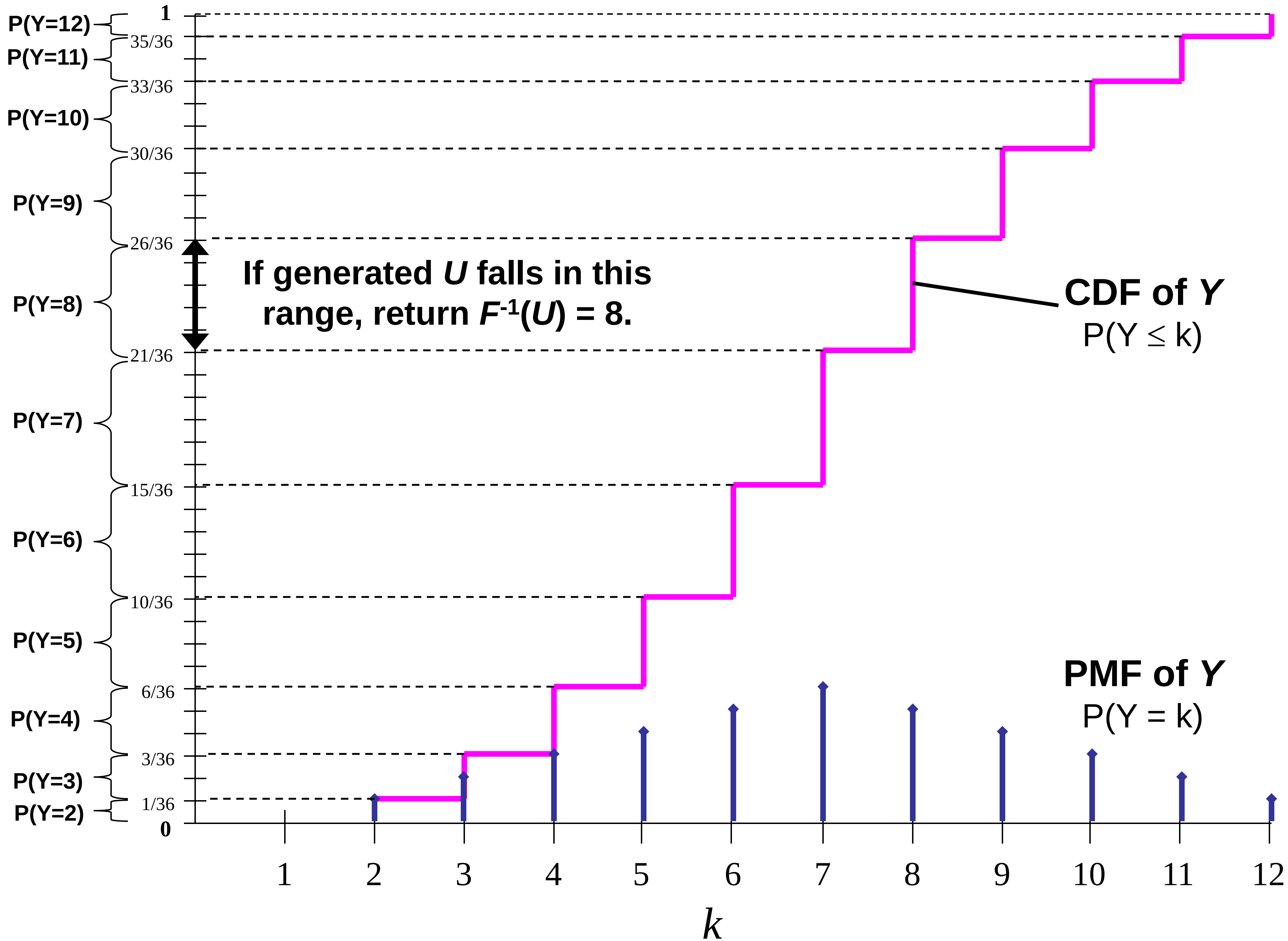
**Sum is 8** **Sum is 9** **Sum is 10** **Sum is 11** **Sum is 12**

- Define the random variable $Y$ to be the ***sum*** of the numbers rolled.

# Example (cont'd)

PMF and CDF of $Y$ (assuming iid fair dice):

| $k$ | $P(Y = k)$ | $P(Y \leq k)$ |
|---|---|---|
| 2 | **1/36** | **1/36** |
| 3 | **2/36** | **3/36** |
| 4 | **3/36** | **6/36** |
| 5 | **4/36** | **10/36** |
| 6 | **5/36** | **15/36** |
| 7 | **6/36** | **21/36** |
| 8 | **5/36** | **26/36** |
| 9 | **4/36** | **30/36** |
| 10 | **3/36** | **33/36** |
| 11 | **2/36** | **35/36** |
| 12 | **1/36** | **1** |



P(Y=12)
P(Y=11)
P(Y=10)
P(Y=9)
P(Y=8)
P(Y=7)
P(Y=6)
P(Y=5)
P(Y=4)
P(Y=3)
P(Y=2)
1
35/36
33/36
30/36
26/36
21/36
15/36
10/36
6/36
3/36
1/36
0
If generated *U* falls in this range, return $F^{-1}(U) = 8$.
CDF of *Y*
P(Y ≤ k)
PMF of *Y*
P(Y = k)
1 2 3 4 5 6 7 8 9 10 11 12
*k*

# Sample Statistics vs. Distribution Parameters

If ***X*** and ***Y*** are ***vectors of sample data***:

$$\bar{X} = \sum_{i=1}^{n} X_i / n$$

$$s_{\boldsymbol{X}}^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$$

$$\mathrm{cov}(\mathbf{X}, \mathbf{Y}) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

$$\mathrm{corr}(\mathbf{X}, \mathbf{Y}) = \frac{\mathrm{cov}(\mathbf{X}, \mathbf{Y})}{s_{\mathbf{X}}\ s_{\mathbf{Y}}}$$

If ***X*** and ***Y*** are ***discrete random variables***:

$$E[X] = \sum_{k=-\infty}^{\infty} k \cdot P(X = k)$$

**Excel: =SUMPRODUCT(values,probs)**

$$\sigma_X^2 = \sum_{k=-\infty}^{\infty} (k - E[X])^2 P(X = k)$$

$$\mathrm{cov}(X,\ Y) = \sum_{\substack{\text{all pairs} \\ (x_i, y_i)}} (x_i - E[X])(y_i - E[Y])\ f_{X,Y}(x_i, y_i)$$

$$\mathrm{corr}(X,\ Y) = \frac{\mathrm{cov}(X,\ Y)}{\sigma_X \sigma_Y}$$