**Business Statistics – Probability & Probability Distributions**

**Discrete Probability Distributions**

|  |  |
| --- | --- |
|  | Introduction to  Probability Definitions: Frequency, Classical, Subjective  Distributions: Discrete, Continuous, Sampling  Central Limit Theorem, CLT |
| ► | Discrete Distributions, Empirical, Binomial, Poisson |
|  | Continuous Distributions, Empirical, Normal, Exponential |
|  | Bayesian Probability Analysis |

|  |  |  |
| --- | --- | --- |
|  | **Empirical Distribution** |  |

Suppose a questionnaire was given to 5 people containing the following question.

|  |
| --- |
| 1. How many miles did you travel to get here today? \_\_\_\_\_ |

Let an **Event** be one subject selected at random and observe the survey results of the subject. Consider the likelihood or **probability** of an event. If we select one of the subjects at random, consider probability statements about the miles the subject traveled? Let the random variable, X=Miles

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Subject | Miles |  | X | Frequency | Probability |  | An Empirical Distribution is defined using the frequency of observations. Probabilities from an empirical distribution is called empirical probabilities. Also called “Frequentist” definition of probabilities. |
| 1 | 1 |  | 1 | 1 | 0.2 |  |
| 2 | 3 |  | 3 | 2 | 0.4 |  |
| 3 | 3 |  | 5 | 1 | 0.2 |  |
| 4 | 5 |  | 8 | 1 | 0.2 |  |
| 5 | 8 |  | Sum | 5 | 1.0 |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | Frequency | Probability |  | Frequency | Probability | Empirical Distribution | | | | | | | | | |
| 1 | 1 | 0.2 |  |  | 0.5 |  |  |  |  |  |  |  |  |  |  |
| 3 | 2 | 0.4 |  | 2 | 0.4 |  |  |  |  |  |  |  |  |  |  |
| 5 | 1 | 0.2 |  |  | 0.3 |  |  |  |  |  |  |  |  |  |  |
| 8 | 1 | 0.2 |  | 1 | 0.2 |  |  |  |  |  |  |  |  |  |  |
| Sum | 5 | 1.0 |  |  | 0.1 |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | X |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |  |

|  |
| --- |
| Probability Statements |
| P[X=3]=0.4 |
| P[X<=3]=0.6 |
| P[X>5]=0.2 |
| P[1<X<4]=0.4 |
| P[1<=X<=4]=0.6 |

|  |  |  |
| --- | --- | --- |
|  | **Binomial Distribution** |  |

|  |
| --- |
| Classical definition of probability. |

Binomial. X~B(n,p).

Definition.

Given a trial with two outcomes, success or failure.

Given a probability of success of an outcome of the trial is constant, p.

Given the outcome of the trials are independent.

Given a large number of such trials, n.

Given the number of successes in n trials, X.

Then, the probability of X is Binomial, X~B(n,p)

Examples. Calculations using Excel function. “=binom.dist(X,n,p,cumulative(0=no,1=yes))”

1. If five people are selected at random from a population of half seniors, what is the probability that less than 2 people are seniors? X=number of seniors

Answer: n=5, p=0.5, X~B(n=5,p=0.5), P[X<2]=P[X≤1]=binom.dist(1,5,0.5,1)≈0.1875

2. If only 10 employees out of 50 own compact cars, then what is the probability that at least one employee owns a compact car out of a random sample of three employees? X=number of seniors

Answer: n=3, p=10/50=0.2, X~B(n=3,p=0.2), P[X>=1]=1–P[X=0]=1–binom.dist(0,3,0.2,0) ≈0.5120

3. Suppose a major at a university contains 25 sophomores, 15 juniors, and 10 seniors. What is the probability that a focus group of 13 students selected at random from the major contains no seniors? X=number of seniors

Answer: n=13, p=10/(25+15+10)=0.2, X~B(n=12,p=0.2), P[X=0]=binom.dist(0,13,0.2,0) ≈0.0550



|  |  |  |
| --- | --- | --- |
|  | **Binomial Examples** |  |

1. If half of all shoppers at a mall are seniors, then from a random sample of three people, determine the probabilities. Let X=number of seniors in the sample. Then, X~B(n=3, p=0.5)

Excel function: “=binom.dist(X, n=3, p=0.5, cumulative(0=no,1=yes))”

|  |  |  |  |
| --- | --- | --- | --- |
| In the random sample, what is the probability that . . . | | | |
| a. exactly one person is senior? | P[X=1] ≈ | 0.375 | =binom.dist( 1,3,0.5,0 ) |
| b. at least one person is senior? | P[X≥1]=1–P[X=0] ≈ | 0.875 | =1–binom.dist( 0,3,0.5,0 ) |
| c. at most two people are seniors? | P[X≤2] ≈ | 0.875 | =binom.dist( 2,3,0.5,1 ) |
| d. less than two people are seniors? | P[X<2]=P[X≤1] ≈ | 0.5 | =binom.dist( 1,3,0.5,1 ) |
| e. greater than one person is not a senior? | P[X≤1] ≈ | 0.5 | =binom.dist( 1,3,0.5,1 ) |
| f. no more than two people are not seniors? | P[X=3] ≈ | 0.125 | =binom.dist( 3,3,0.5,0 ) |
| g. no less than one person is not a senior? | P[X=0] ≈ | 0.125 | =binom.dist( 0,3,0.5,0 ) |
| h. exactly two people are not seniors? | P[X=1] ≈ | 0.375 | =binom.dist( 1,3,0.5,0 ) |

2. If 20% of all shoppers at a mall are seniors, then from a random sample of three people, determine the probabilities. Let X=number of seniors in the sample. Then, X~B(n=3, p=0.2)

Excel function: “=binom.dist(X, n=3, p=0.2, cumulative(0=no,1=yes))”

|  |  |  |
| --- | --- | --- |
| In the random sample, what is the probability that . . . | | |
| a. exactly one person is senior? | P[X=1] ≈0.384 | =binom.dist( 1,3,0.2,0 ) |
| b. at least one person is senior? | P[X≥1]=1–P[X=0] ≈0.488 | =1–binom.dist( 0,3,0.2,0 ) |
| c. at most two people are seniors? | P[X≤2] ≈0.992 | =binom.dist( 2,3,0.2,1 ) |
| d. less than two people are seniors? | P[X<2]=P[X≤1] ≈0.896 | =binom.dist( 1,3,0.2,1 ) |
| e. greater than one person is not a senior? | P[X≤1] ≈0.896 | =binom.dist( 1,3,0.2,1 ) |
| f. no more than two people are not seniors? | P[X=3] ≈0.008 | =binom.dist( 3,3,0.2,0 ) |
| g. no less than one person is not a senior? | P[X=0] ≈0.512 | =binom.dist( 0,3,0.2,0 ) |
| h. exactly two people are not seniors? | P[X=1] ≈0.384 | =binom.dist( 1,3,0.2,0 ) |

3. If 5% of items are defective coming off an assembly line, then from a random sample of three items, determine the probabilities. Let X=number of defectives in the sample. Then, X~B(n=3, p=0.05)

Excel function: “=binom.dist(X, n=3, p=0.05, cumulative(0=no,1=yes))”

|  |  |
| --- | --- |
| In the random sample, what is the probability that . . . | |
| a. exactly one item is defective? P[X=1] ≈0.1354 | =binom.dist( 1,3,0.05,0 ) |
| b. at least one item is defective? P[X≥1]=1–P[X=0] ≈0.1426 | =1–binom.dist( 0,3,0.05,0 ) |
| c. at most two items are defective? P[X≤2] ≈0.9999 | =binom.dist( 2,3,0.05,1 ) |
| d. less than two items are defective? P[X<2]=P[X≤1] ≈0.9928 | =binom.dist( 1,3,0.05,1 ) |
| e. greater than one item is not defective? P[X≤1] ≈0.9928 | =binom.dist( 1,3,0.05,1 ) |
| f. no more than two items are not defective? P[X=3] ≈0.0001 | =binom.dist( 3,3,0.05,0 ) |
| g. no less than one item is not defective? P[X=0] ≈0.8574 | =binom.dist( 0,3,0.05,0 ) |
| h. exactly two items are not defective? P[X=1] ≈0.1354 | =binom.dist( 1,3,0.05,0 ) |

Suppose a random sample of three items was repeated 3 times. What is the probability of no defectives found? [ (1-(1-0.05)3 )3 ≈0.003 ]

|  |  |  |
| --- | --- | --- |
|  | **Poisson Distribution** |  |

|  |
| --- |
| Classical definition of probability. |

Poisson. X~Pssn().

Definition.

Let X represent the number of random, independent occurrences in an interval of consideration.

The probability of X is Poisson where  is the average number of occurrences in the interval.

Examples. Calculations using Excel function. “=poisson.dist(X,mean,cumulative(0=no,1=yes))”

1. If the average number of auto accidents in downtown on a particular holiday is 1.2/hour, then what is the probability that there are less than 2 accidents in any given hour?

Answer: =1.2, X~Pssn(=1.2), P[X<2]=P[X≤1]=poisson.dist(1,1.2,1) ≈0.6626

2. If the average number of clerical errors for a tax accounting firm is 0.7 for every tax return, then what is the probability that there are more than 2 errors out of 5 returns?

Answer: =0.7\*5=3.5, X~Pssn(=3.5), P[X>2]=1–P[X ≤2]=1–poisson.dist(2,3.5,1) ≈0.6792

|  |
| --- |
| . . .    . . . |

|  |  |  |
| --- | --- | --- |
|  | **Poisson Examples** |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X~Poisson(0.1) |  | X~Poisson(0.2) |  | X~Poisson(0.3) |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | pdf | PDF |  | pdf | PDF |  | pdf | PDF |
| X | f(X) | F(X) |  | f(X) | F(X) |  | f(X) | F(X) |
| 0 | .9048 | .9048 |  | .8187 | .8187 |  | .7408 | .7408 |
| 1 | .0905 | .9953 |  | .1637 | .9824 |  | .2223 | .9631 |
| 2 | .0045 | .9998 |  | .0164 | .9988 |  | .0333 | .9964 |
| 3 | .0002 | 1 |  | .0011 | .9999 |  | .0033 | .9997 |
| 4 | .0000 | 1 |  | .0001 | 1 |  | .0003 | 1 |
| Sum | 1 |  |  | 1 |  |  | 1 |  |

1. Suppose call arriving at a switchboard follow a Poisson process with an average of one call every 20 seconds. X=number of calls with l = (1/20) calls/second

|  |  |
| --- | --- |
| What is the probability that . . . | |
| a. no calls in the first 2 seconds? P[X=0] ≈ 0.9048 | =poisson.dist(0,2(1/20),0) |
| b. at least one call in the first 4 seconds? P[X≥1]=1–P[X=0] ≈ 0.1813 | =1–poisson.dist(0,4(1/20),0) |
| c. no more than two calls in the first 6 seconds? P[X≤2] ≈ 0.9964 | =poisson.dist(0,6(1/20),1) |
| d. no calls in the first minute? P[X=0] ≈ 0.0498 | =poisson.dist(0,60(1/20),0) |

2. Suppose potholes in a county road follow a Poisson process where the number of potholes averages one every 10 miles. X=number of potholes with l = (1/10) potholes/mile

|  |  |
| --- | --- |
| a. no potholes in a randomly selected mile of road?  P[X=0] ≈ 0.9048 | =poisson.dist(0,(1/10),0) |
| b. at least one pothole in a random 3 miles of road?  P[X≥1]=1–P[X=0] ≈ 0.2592 | =1–poisson.dist(0,3(1/10),0) |
| c. no more than two potholes in a random 2 miles stretch?  P[X≤2] ≈ 0.9989 | =poisson.dist(0,2(1/10),1) |
| d. at least one pothole within 500 feet either side of a randomly selected junction? P[X≥1]= 1–P[X=0] ≈ 0.0188 | =poisson.dist(0,(1000/5280)(1/10),0) |

3. Suppose the imperfections in the weave of a certain textile follow a Poisson process where the average number of imperfections is one per five square yards.

X=number of imperfections with l = (1/5) imperfections/square-yard

|  |  |
| --- | --- |
| a. no imperfections in 1 yd2? P[X=0] ≈ 0.8187 | =poisson.dist(0,(1/5),0) |
| b. at least one imperfections in ½ yd2? P[X≥1]=1–P[X=0] ≈ 0.0952 | =1–poisson.dist(0,0.5(1/5),0) |
| c. no more than 2 imperfections in 1.5 yd2? P[X≤2] ≈ 0.9964 | =poisson.dist(0,1.5(1/5),1) |
| d. at least one imperfection in 5 square feet? P[X≥1]= 1–P[X=0] ≈ 0.1052 | =poisson.dist(0,(5/9)(1/5),0) |

4. Suppose the occurrence of bacteria colonies of a certain type in a sample of polluted water follows a Poisson process with the mean number of colonies is one in every ten cubic centimeters of water.

X=number of bacteria colonies with l = (1/10) colonies/1cc of water

|  |  |
| --- | --- |
| a. no bacteria colonies in 3 cc of water? P[X=0] ≈ 0.7408 | =poisson.dist(0,3(1/10),0) |
| b. at least one bacteria colony in 1 cc of water?  P[X≥1]=1–P[X=0] ≈ 0.0952 | =1–poisson.dist(0,(1/10),0) |
| c. no more than two bacteria colonies in 2 cc of water? P[X≤2] ≈ 0.9988 | =poisson.dist(0,2(1/10),1) |
| d. at least one out of three 1-cc samples contain  at least one bacteria colony? P[X≥1]= 1–P[X=0] ≈ 0.0952  Y=number of samples.Y~B(n=3,p=P[X≥1]). P[Y≥1]= 1–P[Y=0]≈0.2592 | =1–poisson.dist(0,(1/10),0)  =1–binom.dist(0,3,P[X ≥1],0) |
| Alternatively, [ =0.1, P1=1-e-0.1=0.0952, P2=1-(1-P1)n=1-(1-(1- e-0.1))3=1- e-0.3 ≈0.2592 ] | |

|  |  |  |
| --- | --- | --- |
|  | **Binomial Distribution Summary** |  |
| . . .   |  | | --- | | X~B(n,p)  n=total number of trials  p=probability of success  f(X)= px (1–p)n–x  n!/(x!(n–x)!)  E[x] = np  V[x] = np(1–p)  For p<0.5, skewed right  For p=0.5, symmetric  For p>0.5, skewed left |   . . . | | |

|  |  |  |
| --- | --- | --- |
|  | **Poisson Distribution Summary** |  |
| . . .   |  | | --- | | X~Pssn()  =mean  f(X)= x e / x!  E[x] =   V[x] =   Always skewed right |   . . . | | |