Bernoulli: Outcome of a coin flip, Sex of a newborn, Disease occurrence

Written as if X is a Bernoulli random variable with probability of success p.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Binomial: Relief of allergies, Coin flips, Responses to a survey question

Written as if X is a Binomial random variable with n trials and probability of success p.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Poisson: Car accidents per day in a city, The number of typographical errors found in a book, The number of telephone calls per minute in a small business

Written as if X is a Poisson random variable with rate parameter .

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Geometric: Likelihood that a batter earns a hit before three strikes, Number of customers entering a store before one makes a purchase

Written as if X is a geometric random variable with probability of success p.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Uniform: Outcome of a single dice roll, Wait time for a bus given that it’s last arrival time is unknown, Resulting angular position of a spinner

Written as or if X is a uniform random variable distributed over the interval [a,b].

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Normal: Distribution of height in a population, Measurement errors, Stock volatility

Written as or if X is a normal random variable with mean and variance (or standard deviation )

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Exponential: The amount of time until an earthquake occurs, Time between arrive times of Poisson events such as time until a phone call, Time until a radioactive particle decays

Written as if X is an exponential random variable with rate parameter , cam ne .

Note that the exponential distribution is just the gamma distribution with n = 1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Binomial Example:

Let be number of passengers who arrive

Bayes’ Theorem Example:

and

If and , what is ?

PDF Constraints:

Terminology:

Mode: most frequent value

Median: middle most whole number

Range: difference between smallest and largest

Multiplication Rule: | Unless Dependent:

Disjoint: | | |

Subset: | | | |

Identities: | | |

Addition Rule: | Unless Disjoint:

De Morgan’s Laws: | | |

Conditional: | If independent: | | |

Total Probability: | | |

Bayes’ Theorem:

Combinatorics: | Ordered: | Unordered:

Probability Mass Function:

Probability Density Function:

Cumulative Distribution Function: Probility that random variable is less than or equal to a particular realization . is a valid CDF if:

1. is monotonically increasing
2. is continuous

Standard Deviation:

Population Variance: (where mean, population size)  
Population Standard Deviation:

Sample Variance: (where mean, sample size)

Sample Standard Deviation:

Population Skew:

Sample Skew:

If skew is positive, the data is right skewed where the ‘tail’ of the histogram is longer on the right. If the skew is negative, the data is left skewed where the ‘tail’ of the histogram is longer on the left.

Covariance (Numerical summary for two columns of data):

* If is positive, that means when is large, the will tend to be large as well and vice versa.
* If is negative, that means when is large, will tend to be small and vice versa.
* If is close to 0, that means there is no relationship between and .

Correlation:

Z-score: or

Perform test to see if tai chi is more effective across the 33 patients in each group.

Assume Type I error rate of .

The hypotheses are:

That is that the proportion of patients who did Tai Chi and experienced reduced pain () is less than or equal to the proportion of patients who took the wellness education classes (). This is a one-sided test so the rejection region is:

Because we reject the null hypothesis.

Method of Moments:

Method of Maximum Likelihood:

* Discrete:
* Continuous:

(If given means & size use this, otherwise t-test)

Confidence Coefficients:

Confidence Intervals:

CI for multiple samples:

( = Standard Error of , in case of : )

Single Population Proportion:

(First check that )

| |

Difference Between two Population Means:

(Small Sample) | (Large)

Estimating Difference between two population proportions:

(First check that )

CI for difference between two population proportions:

Null Hypothesis:

Note that these are equivalent to:

Test Statistic for Sample Mean/Proportion:



(Where is sample estimate of parameter , is the hypothesized value of and is standard error of )

* (Multiple)

(Where )

Null Hypothesis about difference between population proportions:

Note that these are equivalent to:

The null hypotheses assume that , the estimator for difference in proportions is:

|

:

Otherwise:

Poisson Distribution :

MLE of the Poisson Distribution:

Find the maximum by taking the derivative of the log-likelihood and solve for 0

words and authors. If is rate of occurrence for the word in the column, if both were the same the null hypothesis is:

vs

The best estimator for is

The expected number in cell is

Recall that for a Gaussian random variable with observed value is

The test statistic is where and where .

and . Since we reject the null hypothesis that the imitation was not the same style as the original

Inference about population mean for small samples:

Small Sample Sampling Statistic Distribution:

If , small sample confidence interval can be computed as:  **or**

Inference about the difference in population means for small samples: (is X greater than Y)

||

-test degrees of freedom (2) are

Unequal Variances and Correction to degrees of freedom:

If true, the pooled variance is inappropriate and each population variance is estimated by its sample variance.

As long as the sample sizes are small, the degree of freedom adjusted to account for unequal variances is

Approximate Tests:

The test of Homogeneity and Independence:

Under the null hypothesis that they are independent, then the best estimate (MLE) of the cell probabilities is

| |

Is there a difference between the two? Assume Type I Error Rate of .

The hypotheses are

First check the assumption of equal variances:

Given unequal variance,

We reject the null hypothesis.

Six samples from the new process are created weighing: carats.

Perform a hypothesis test of the hypotheses:

Assume a Type I error rate of .

And the hypotheses are:

Thus we would reject if our test statistic if

Is greater than where . From the table we see that this critical value is

We fail to reject the null hypothesis.