Basic Probability Concepts

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Multilevel Regression Modeling, 2009

An Introduction to R

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Random Variables

Random Variables

- The term *random variable* has a technical definition that we discussed in Psychology 310
- For our purposes, it will suffice to consider a random variable to be a random process with numerical outcomes that occur according to a distribution law

Example (Uniform (0,1) Random Variable)

A random process that generates numbers so that all values between 0 and 1, inclusive, are equally likely to occur is said to have a U(0,1) distribution.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Random Variables

Random Variables

- The term *random variable* has a technical definition that we discussed in Psychology 310
- For our purposes, it will suffice to consider a random variable to be a random process with numerical outcomes that occur according to a distribution law

Example (Uniform (0,1) Random Variable)

A random process that generates numbers so that all values between 0 and 1, inclusive, are equally likely to occur is said to have a U(0,1) distribution.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Random Variables

Random Variables

- The term *random variable* has a technical definition that we discussed in Psychology 310
- For our purposes, it will suffice to consider a random variable to be a random process with numerical outcomes that occur according to a distribution law

Example (Uniform (0,1) Random Variable)

A random process that generates numbers so that all values between 0 and 1, inclusive, are equally likely to occur is said to have a U(0,1) distribution.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Random Variables

Random Variables

- The term *random variable* has a technical definition that we discussed in Psychology 310
- For our purposes, it will suffice to consider a random variable to be a random process with numerical outcomes that occur according to a distribution law

Example (Uniform (0,1) Random Variable)

A random process that generates numbers so that all values between 0 and 1, inclusive, are equally likely to occur is said to have a U(0,1) distribution.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Manifest and Latent Variables

Manifest and Latent Variables

- In advanced applications, we will refer to *manifest* and *latent* random variables
- A variable is manifest if it can be measured directly
- A variable is latent if it is an assumed quantity that cannot be measured directly
- The dividing line between manifest and latent variables is often rather imprecise

Example (Manifest Variable)

Your grade on an exam is a manifest random variable.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Manifest and Latent Variables

Manifest and Latent Variables

- In advanced applications, we will refer to *manifest* and *latent* random variables
  - A variable is manifest if it can be measured directly
  - A variable is latent if it is an assumed quantity that cannot be measured directly
- The dividing line between manifest and latent variables is often rather imprecise

Example (Manifest Variable)

Your grade on an exam is a manifest random variable.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Manifest and Latent Variables

Manifest and Latent Variables

- In advanced applications, we will refer to *manifest* and *latent* random variables
- A variable is manifest if it can be measured directly
- A variable is latent if it is an assumed quantity that cannot be measured directly
- The dividing line between manifest and latent variables is often rather imprecise

Example (Manifest Variable)

Your grade on an exam is a manifest random variable.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Manifest and Latent Variables

Manifest and Latent Variables

- In advanced applications, we will refer to *manifest* and *latent* random variables
- A variable is manifest if it can be measured directly
- A variable is latent if it is an assumed quantity that cannot be measured directly
- The dividing line between manifest and latent variables is often rather imprecise

Example (Manifest Variable)

Your grade on an exam is a manifest random variable.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Manifest and Latent Variables

Manifest and Latent Variables

- In advanced applications, we will refer to *manifest* and *latent* random variables
- A variable is manifest if it can be measured directly
- A variable is latent if it is an assumed quantity that cannot be measured directly
- The dividing line between manifest and latent variables is often rather imprecise

Example (Manifest Variable)

Your grade on an exam is a manifest random variable.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Manifest and Latent Variables

Manifest and Latent Variables

- In advanced applications, we will refer to *manifest* and *latent* random variables
- A variable is manifest if it can be measured directly
- A variable is latent if it is an assumed quantity that cannot be measured directly
- The dividing line between manifest and latent variables is often rather imprecise

Example (Manifest Variable)

Your grade on an exam is a manifest random variable.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Continuous and Discrete Random Variables

- A continuous random variable has an uncountably infinite number of possible outcomes because it can take on all values over some range of the number line
- A discrete random variable takes on only a countable number of discrete outcomes
- As we saw in Psychology 310, discrete random variables can assign a probability to a particular numerical outcome, while continuous random variables cannot

Example (Discrete Random Variable)

Suppose you assign the number 1 to all people born male, and 2 to all people born female. This random variable is discrete, because it takes on only the values 1 and 2.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Continuous and Discrete Random Variables

- A continuous random variable has an uncountably infinite number of possible outcomes because it can take on all values over some range of the number line
- A discrete random variable takes on only a countable number of discrete outcomes
- As we saw in Psychology 310, discrete random variables can assign a probability to a particular numerical outcome, while continuous random variables cannot

Example (Discrete Random Variable)

Suppose you assign the number 1 to all people born male, and 2 to all people born female. This random variable is discrete, because it takes on only the values 1 and 2.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Continuous and Discrete Random Variables

- A continuous random variable has an uncountably infinite number of possible outcomes because it can take on all values over some range of the number line
- A discrete random variable takes on only a countable number of discrete outcomes
- As we saw in Psychology 310, discrete random variables can assign a probability to a particular numerical outcome, while continuous random variables cannot

Example (Discrete Random Variable)

Suppose you assign the number 1 to all people born male, and 2 to all people born female. This random variable is discrete, because it takes on only the values 1 and 2.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Continuous and Discrete Random Variables

- A continuous random variable has an uncountably infinite number of possible outcomes because it can take on all values over some range of the number line
- A discrete random variable takes on only a countable number of discrete outcomes
- As we saw in Psychology 310, discrete random variables can assign a probability to a particular numerical outcome, while continuous random variables cannot

Example (Discrete Random Variable)

Suppose you assign the number 1 to all people born male, and 2 to all people born female. This random variable is discrete, because it takes on only the values 1 and 2.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Informal Definition
Manifest and Latent Random Variables
Continuous and Discrete Random Variables

Continuous and Discrete Random Variables

- A continuous random variable has an uncountably infinite number of possible outcomes because it can take on all values over some range of the number line
- A discrete random variable takes on only a countable number of discrete outcomes
- As we saw in Psychology 310, discrete random variables can assign a probability to a particular numerical outcome, while continuous random variables cannot

Example (Discrete Random Variable)

Suppose you assign the number 1 to all people born male, and 2 to all people born female. This random variable is discrete, because it takes on only the values 1 and 2.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Using Probability Distributions

Using Probability Distributions

- Probability distributions are frequently used to provide succinct models for quantities of scientific interest
- We observe distributions of data, and assess how well the distributions conform to the specified model
- While observing the distribution of the data, we may hypothesize the general family of the distribution, but leave open the question of the values of the parameters
- In that case, we talk of *free parameters* to be estimated

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Using Probability Distributions

Using Probability Distributions

- Probability distributions are frequently used to provide succinct models for quantities of scientific interest
- We observe distributions of data, and assess how well the distributions conform to the specified model
- While observing the distribution of the data, we may hypothesize the general family of the distribution, but leave open the question of the values of the parameters
- In that case, we talk of *free parameters* to be estimated

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Using Probability Distributions

Using Probability Distributions

- Probability distributions are frequently used to provide succinct models for quantities of scientific interest
- We observe distributions of data, and assess how well the distributions conform to the specified model
- While observing the distribution of the data, we may hypothesize the general family of the distribution, but leave open the question of the values of the parameters
- In that case, we talk of *free parameters* to be estimated

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Using Probability Distributions

Using Probability Distributions

- Probability distributions are frequently used to provide succinct models for quantities of scientific interest
- We observe distributions of data, and assess how well the distributions conform to the specified model
- While observing the distribution of the data, we may hypothesize the general family of the distribution, but leave open the question of the values of the parameters
- In that case, we talk of *free parameters* to be estimated

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Using Probability Distributions

Using Probability Distributions

- Probability distributions are frequently used to provide succinct models for quantities of scientific interest
- We observe distributions of data, and assess how well the distributions conform to the specified model
- While observing the distribution of the data, we may hypothesize the general family of the distribution, but leave open the question of the values of the parameters
- In that case, we talk of *free parameters* to be estimated

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Using Probability Distributions
More Complex Applications

Using Probability Distributions

- In more complex applications, such as multilevel modeling, we may model data emanating from a particular distribution family at one level (say kids within a school)
- At another level, we might model the parameters for the schools as having a distribution across schools
- For example, we might hypothesize that the parameters across schools have a normal distribution
- In that case, the size of the variance of that distribution would indicate how much the schools show variation on a particular characteristic
- In the slides that follow, we shall examine some of the more useful distributions we will encounter early in the course

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Using Probability Distributions
More Complex Applications

Using Probability Distributions

- In more complex applications, such as multilevel modeling, we may model data emanating from a particular distribution family at one level (say kids within a school)
- At another level, we might model the parameters for the schools as having a distribution across schools
- For example, we might hypothesize that the parameters across schools have a normal distribution
- In that case, the size of the variance of that distribution would indicate how much the schools show variation on a particular characteristic
- In the slides that follow, we shall examine some of the more useful distributions we will encounter early in the course

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Using Probability Distributions
More Complex Applications

Using Probability Distributions

- In more complex applications, such as multilevel modeling, we may model data emanating from a particular distribution family at one level (say kids within a school)
- At another level, we might model the parameters for the schools as having a distribution across schools
- For example, we might hypothesize that the parameters across schools have a normal distribution
- In that case, the size of the variance of that distribution would indicate how much the schools show variation on a particular characteristic
- In the slides that follow, we shall examine some of the more useful distributions we will encounter early in the course

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Using Probability Distributions
More Complex Applications

Using Probability Distributions

- In more complex applications, such as multilevel modeling, we may model data emanating from a particular distribution family at one level (say kids within a school)
- At another level, we might model the parameters for the schools as having a distribution across schools
- For example, we might hypothesize that the parameters across schools have a normal distribution
- In that case, the size of the variance of that distribution would indicate how much the schools show variation on a particular characteristic
- In the slides that follow, we shall examine some of the more useful distributions we will encounter early in the course

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Using Probability Distributions
More Complex Applications

Using Probability Distributions

- In more complex applications, such as multilevel modeling, we may model data emanating from a particular distribution family at one level (say kids within a school)
- At another level, we might model the parameters for the schools as having a distribution across schools
- For example, we might hypothesize that the parameters across schools have a normal distribution
- In that case, the size of the variance of that distribution would indicate how much the schools show variation on a particular characteristic
- In the slides that follow, we shall examine some of the more useful distributions we will encounter early in the course

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Using Probability Distributions
More Complex Applications

Using Probability Distributions

- In more complex applications, such as multilevel modeling, we may model data emanating from a particular distribution family at one level (say kids within a school)
- At another level, we might model the parameters for the schools as having a distribution across schools
- For example, we might hypothesize that the parameters across schools have a normal distribution
- In that case, the size of the variance of that distribution would indicate how much the schools show variation on a particular characteristic
- In the slides that follow, we shall examine some of the more useful distributions we will encounter early in the course

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Normal Distribution

The Normal Distribution

- The *normal distribution* is a widely used continuous distribution
- The normal distribution family is a two-parameter family
- Each normal distribution is characterized by two parameters, the mean μ and the standard deviation σ.
- Shaped like a bell, the normal pdf is sometimes referred to as the *bell curve*
- The *central limit theorem*, discussed on pages 13–14 of Gelman & Hill, explains why many quantities have a distribution that is approximately normal
- The normal distribution family is *closed under linear transformations*, i.e., any normal distribution may be transformed into any other normal distribution by a linear transformation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Normal Distribution

The Normal Distribution

- The *normal distribution* is a widely used continuous distribution
- The normal distribution family is a two-parameter family
- Each normal distribution is characterized by two parameters, the mean μ and the standard deviation σ.
- Shaped like a bell, the normal pdf is sometimes referred to as the *bell curve*
- The *central limit theorem,* discussed on pages 13–14 of Gelman & Hill, explains why many quantities have a distribution that is approximately normal
- The normal distribution family is *closed under linear transformations,* i.e., any normal distribution may be transformed into any other normal distribution by a linear transformation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Normal Distribution

The Normal Distribution

- The *normal distribution* is a widely used continuous distribution
- The normal distribution family is a two-parameter family
- Each normal distribution is characterized by two parameters, the mean μ and the standard deviation σ.
- Shaped like a bell, the normal pdf is sometimes referred to as the *bell curve*
- The *central limit theorem*, discussed on pages 13–14 of Gelman & Hill, explains why many quantities have a distribution that is approximately normal
- The normal distribution family is *closed under linear transformations*, i.e., any normal distribution may be transformed into any other normal distribution by a linear transformation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Normal Distribution

The Normal Distribution

- The *normal distribution* is a widely used continuous distribution
- The normal distribution family is a two-parameter family
- Each normal distribution is characterized by two parameters, the mean μ and the standard deviation σ.
- Shaped like a bell, the normal pdf is sometimes referred to as the *bell curve*
- The *central limit theorem*, discussed on pages 13–14 of Gelman & Hill, explains why many quantities have a distribution that is approximately normal
- The normal distribution family is *closed under linear transformations*, i.e., any normal distribution may be transformed into any other normal distribution by a linear transformation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Normal Distribution

The Normal Distribution

- The *normal distribution* is a widely used continuous distribution
- The normal distribution family is a two-parameter family
- Each normal distribution is characterized by two parameters, the mean μ and the standard deviation σ.
- Shaped like a bell, the normal pdf is sometimes referred to as the *bell curve*
- The *central limit theorem,* discussed on pages 13–14 of Gelman & Hill, explains why many quantities have a distribution that is approximately normal
- The normal distribution family is *closed under linear transformations*, i.e., any normal distribution may be transformed into any other normal distribution by a linear transformation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Normal Distribution

The Normal Distribution

- The *normal distribution* is a widely used continuous distribution
- The normal distribution family is a two-parameter family
- Each normal distribution is characterized by two parameters, the mean μ and the standard deviation σ.
- Shaped like a bell, the normal pdf is sometimes referred to as the *bell curve*
- The *central limit theorem*, discussed on pages 13–14 of Gelman & Hill, explains why many quantities have a distribution that is approximately normal
- The normal distribution family is *closed under linear transformations*, i.e., any normal distribution may be transformed into any other normal distribution by a linear transformation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Normal Distribution

The Normal Distribution

- The *normal distribution* is a widely used continuous distribution
- The normal distribution family is a two-parameter family
- Each normal distribution is characterized by two parameters, the mean μ and the standard deviation σ.
- Shaped like a bell, the normal pdf is sometimes referred to as the *bell curve*
- The *central limit theorem*, discussed on pages 13–14 of Gelman & Hill, explains why many quantities have a distribution that is approximately normal
- The normal distribution family is *closed under linear transformations*, i.e., any normal distribution may be transformed into any other normal distribution by a linear transformation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Multivariate Normal Distribution

The Multivariate Normal Distribution

- The *multivariate normal distribution* is a continuous multivariate distribution having two matrix parameters, the vector of means $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$

- Any linear combination of multi-normal variables has a normal distribution

- As we saw in Psychology 310, the mean and variance of the linear combination is determined by $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and the linear weights

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Multivariate Normal Distribution

The Multivariate Normal Distribution

- The *multivariate normal distribution* is a continuous multivariate distribution having two matrix parameters, the vector of means $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$

- Any linear combination of multi-normal variables has a normal distribution

- As we saw in Psychology 310, the mean and variance of the linear combination is determined by $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and the linear weights

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Multivariate Normal Distribution

The Multivariate Normal Distribution

- The *multivariate normal distribution* is a continuous multivariate distribution having two matrix parameters, the vector of means $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$

- Any linear combination of multi-normal variables has a normal distribution

- As we saw in Psychology 310, the mean and variance of the linear combination is determined by $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and the linear weights

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Multivariate Normal Distribution

The Multivariate Normal Distribution

- The *multivariate normal distribution* is a continuous multivariate distribution having two matrix parameters, the vector of means $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$

- Any linear combination of multi-normal variables has a normal distribution

- As we saw in Psychology 310, the mean and variance of the linear combination is determined by $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and the linear weights

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Lognormal Distribution

The Lognormal Distribution

- If X is normally distributed, then $y = e^x$ is said to have a *lognormal* distribution. If Y is lognormally distributed, the logarithm of Y has a normal distribution
- In R, `dlnorm` gives the density, `plnorm` gives the distribution function, `qlnorm` gives the quantile function, and `rlnorm` generates random deviates

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Lognormal Distribution
Some Basic Facts

The Lognormal Distribution

- It is common, when referring to a normal distribution, to use the abbreviations $N(\mu, \sigma)$ or $N(\mu, \sigma^2)$.
- It is important to realize that, when referring to a lognormal distribution for a variable Y, the convention is to refer to the parameters μ and σ *from the corresponding normal variable* $X = \ln(Y)$
- In this case, the actual mean and variance of Y are not μ and σ^2, but rather are

$$E(Y) = e^{\mu + \frac{1}{2}\sigma^2},$$
$$Var(Y) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$$

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Lognormal Distribution
Some Basic Facts

The Lognormal Distribution

- It is common, when referring to a normal distribution, to use the abbreviations $N(\mu, \sigma)$ or $N(\mu, \sigma^2)$.
- It is important to realize that, when referring to a lognormal distribution for a variable Y, the convention is to refer to the parameters μ and σ *from the corresponding normal variable* $X = \ln(Y)$
- In this case, the actual mean and variance of Y are not μ and σ^2, but rather are

$$E(Y) = e^{\mu + \frac{1}{2}\sigma^2},$$
$$Var(Y) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$$

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Lognormal Distribution
Some Basic Facts
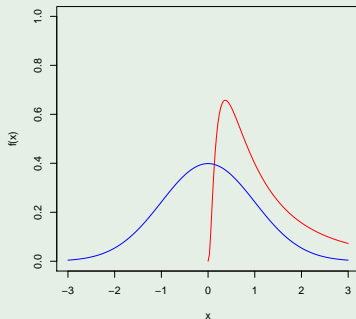
The Lognormal Distribution

- It is common, when referring to a normal distribution, to use the abbreviations $N(\mu, \sigma)$ or $N(\mu, \sigma^2)$.
- It is important to realize that, when referring to a lognormal distribution for a variable Y, the convention is to refer to the parameters μ and σ *from the corresponding normal variable* $X = \ln(Y)$
- In this case, the actual mean and variance of Y are not μ and σ^2, but rather are

$$E(Y) = e^{\mu + \frac{1}{2}\sigma^2},$$
$$Var(Y) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$$

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Lognormal Distribution
Some Basic Facts

The Lognormal Distribution

- It is common, when referring to a normal distribution, to use the abbreviations $N(\mu, \sigma)$ or $N(\mu, \sigma^2)$.
- It is important to realize that, when referring to a lognormal distribution for a variable Y, the convention is to refer to the parameters μ and σ *from the corresponding normal variable $X = \ln(Y)$*
- In this case, the actual mean and variance of Y are not μ and σ^2, but rather are

$$E(Y) = e^{\mu + \frac{1}{2}\sigma^2},$$
$$Var(Y) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$$

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Lognormal Distribution

Example (The Lognormal Distribution)

Here is a picture comparing the lognormal and corresponding normal distribution.

The Lognormal Distribution
Applications

Applications of the Lognormal

- When independent processes combine multiplicatively, the result can be lognormally distributed
- For a detailed and entertaining discussion of the lognormal distribution, see the article by Limpert, Stahel, and Abbt (2001) in the reading list

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Lognormal Distribution
Applications

Applications of the Lognormal

- When independent processes combine multiplicatively, the result can be lognormally distributed
- For a detailed and entertaining discussion of the lognormal distribution, see the article by Limpert, Stahel, and Abbt (2001) in the reading list

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Lognormal Distribution
Applications

Applications of the Lognormal

- When independent processes combine multiplicatively, the result can be lognormally distributed
- For a detailed and entertaining discussion of the lognormal distribution, see the article by Limpert, Stahel, and Abbt (2001) in the reading list

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Binomial Distribution

The Binomial Distribution

- This discrete distribution is one of the foundations of modern categorical data analysis
- The binomial random variable X represents the number of "successes" in N outcomes of a *binomial process*
- A binomial process is characterized by

  - N independent trials
  - Only two outcomes, arbitrarily designated "success" and "failure"
  - Probabilities of success and failure remain constant over trials

- Many interesting real world processes only approximately meet the above specifications
- Nevertheless, the binomial is often an excellent approximation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Binomial Distribution

The Binomial Distribution

- This discrete distribution is one of the foundations of modern categorical data analysis
- The binomial random variable X represents the number of "successes" in N outcomes of a *binomial process*
- A binomial process is characterized by
  - N independent trials
  - Only two outcomes, arbitrarily designated "success" and "failure"
  - Probabilities of success and failure remain constant over trials
- Many interesting real world processes only approximately meet the above specifications
- Nevertheless, the binomial is often an excellent approximation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Binomial Distribution

The Binomial Distribution

- This discrete distribution is one of the foundations of modern categorical data analysis
- The binomial random variable X represents the number of "successes" in N outcomes of a *binomial process*
- A binomial process is characterized by
  - N independent trials
  - Only two outcomes, arbitrarily designated "success" and "failure"
  - Probabilities of success and failure remain constant over trials
- Many interesting real world processes only approximately meet the above specifications
- Nevertheless, the binomial is often an excellent approximation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Binomial Distribution

The Binomial Distribution

- This discrete distribution is one of the foundations of modern categorical data analysis
- The binomial random variable X represents the number of "successes" in N outcomes of a *binomial process*
- A binomial process is characterized by
  - N independent trials
  - Only two outcomes, arbitrarily designated "success" and "failure"
  - Probabilities of success and failure remain constant over trials
- Many interesting real world processes only approximately meet the above specifications
- Nevertheless, the binomial is often an excellent approximation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Binomial Distribution

The Binomial Distribution

- This discrete distribution is one of the foundations of modern categorical data analysis
- The binomial random variable X represents the number of "successes" in N outcomes of a *binomial process*
- A binomial process is characterized by
  - N independent trials
  - Only two outcomes, arbitrarily designated "success" and "failure"
  - Probabilities of success and failure remain constant over trials
- Many interesting real world processes only approximately meet the above specifications
- Nevertheless, the binomial is often an excellent approximation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Binomial Distribution

The Binomial Distribution

- This discrete distribution is one of the foundations of modern categorical data analysis
- The binomial random variable X represents the number of "successes" in N outcomes of a *binomial process*
- A binomial process is characterized by
  - N independent trials
  - Only two outcomes, arbitrarily designated "success" and "failure"
  - Probabilities of success and failure remain constant over trials
- Many interesting real world processes only approximately meet the above specifications
- Nevertheless, the binomial is often an excellent approximation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Binomial Distribution

The Binomial Distribution

- This discrete distribution is one of the foundations of modern categorical data analysis
- The binomial random variable X represents the number of "successes" in N outcomes of a *binomial process*
- A binomial process is characterized by
  - N independent trials
  - Only two outcomes, arbitrarily designated "success" and "failure"
  - Probabilities of success and failure remain constant over trials
- Many interesting real world processes only approximately meet the above specifications
- Nevertheless, the binomial is often an excellent approximation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Binomial Distribution

The Binomial Distribution

- This discrete distribution is one of the foundations of modern categorical data analysis
- The binomial random variable X represents the number of "successes" in N outcomes of a *binomial process*
- A binomial process is characterized by
    - N independent trials
    - Only two outcomes, arbitrarily designated "success" and "failure"
    - Probabilities of success and failure remain constant over trials
- Many interesting real world processes only approximately meet the above specifications
- Nevertheless, the binomial is often an excellent approximation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Binomial Distribution

The Binomial Distribution

- This discrete distribution is one of the foundations of modern categorical data analysis
- The binomial random variable X represents the number of "successes" in N outcomes of a *binomial process*
- A binomial process is characterized by
  - N independent trials
  - Only two outcomes, arbitrarily designated "success" and "failure"
  - Probabilities of success and failure remain constant over trials
- Many interesting real world processes only approximately meet the above specifications
- Nevertheless, the binomial is often an excellent approximation

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Characteristics of the Binomial Distribution

Characteristics of the Binomial Distribution

- The binomial distribution is a two-parameter family, N is the number of trials, p the probability of success
- The binomial has pdf

$$Pr(X = r) = \binom{N}{r} p^r (1 - p)^{N-r}$$

- The mean and variance of the binomial are

$$E(X) = Np$$
$$Var(X) = Np(1 - p)$$

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Characteristics of the Binomial Distribution

Characteristics of the Binomial Distribution

- The binomial distribution is a two-parameter family, N is the number of trials, p the probability of success
- The binomial has pdf

$$Pr(X = r) = \binom{N}{r} p^r (1 - p)^{N-r}$$

- The mean and variance of the binomial are

$$E(X) = Np$$

$$Var(X) = Np(1 - p)$$

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Characteristics of the Binomial Distribution

Characteristics of the Binomial Distribution

- The binomial distribution is a two-parameter family, N is the number of trials, p the probability of success
- The binomial has pdf

$$Pr(X = r) = \binom{N}{r} p^r (1-p)^{N-r}$$

- The mean and variance of the binomial are

$$E(X) = Np$$
$$Var(X) = Np(1-p)$$

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Characteristics of the Binomial Distribution

Characteristics of the Binomial Distribution

- The binomial distribution is a two-parameter family, N is the number of trials, p the probability of success
- The binomial has pdf

$$Pr(X = r) = \binom{N}{r} p^r (1 - p)^{N-r}$$

- The mean and variance of the binomial are

$$E(X) = Np$$
$$Var(X) = Np(1 - p)$$

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Characteristics of the Binomial Distribution

Normal Approximation to the Binomial

- The $B(N, p)$ distribution is well approximated by a $N(Np, Np(1 - p))$ distribution as long as p is not too far removed from .5 and N is reasonably large
- A good rule of thumb is that both Np and $N(1 - p)$ must be greater than 5
- The approximation can be further improved by *correcting for continuity*

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Characteristics of the Binomial Distribution

Normal Approximation to the Binomial

- The $B(N, p)$ distribution is well approximated by a $N(Np, Np(1 - p))$ distribution as long as p is not too far removed from .5 and N is reasonably large
- A good rule of thumb is that both Np and $N(1 - p$ must be greater than 5
- The approximation can be further improved by *correcting for continuity*

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Characteristics of the Binomial Distribution

Normal Approximation to the Binomial

- The $B(N, p)$ distribution is well approximated by a $N(Np, Np(1-p))$ distribution as long as p is not too far removed from .5 and N is reasonably large
- A good rule of thumb is that both Np and $N(1-p)$ must be greater than 5
- The approximation can be further improved by *correcting for continuity*

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Characteristics of the Binomial Distribution

Normal Approximation to the Binomial

- The $B(N, p)$ distribution is well approximated by a $N(Np, Np(1-p))$ distribution as long as p is not too far removed from .5 and N is reasonably large
- A good rule of thumb is that both Np and $N(1-p$ must be greater than 5
- The approximation can be further improved by *correcting for continuity*

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Poisson Distribution

The Poisson Distribution

- When events arrive without any systematic "clustering," i.e., they arrive with a known average rate in a fixed time period but each event arrives at a time independent of the time since the last event, the exact integer number of events can be modeled with the Poisson distribution

- The Poisson is a single parameter family, the parameter being λ, the expected number of events in the interval of interest

- For a Poisson random variable X, the probability of exactly r events is

$$Pr(X = r) = \frac{\lambda^r e^{-\lambda}}{r!}$$

The Poisson Distribution

The Poisson Distribution

- When events arrive without any systematic "clustering," i.e., they arrive with a known average rate in a fixed time period but each event arrives at a time independent of the time since the last event, the exact integer number of events can be modeled with the Poisson distribution
- The Poisson is a single parameter family, the parameter being λ, the expected number of events in the interval of interest
- For a Poisson random variable X, the probability of exactly r events is

$$Pr(X = r) = \frac{\lambda^r e^{-\lambda}}{r!}$$

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Poisson Distribution

The Poisson Distribution

- When events arrive without any systematic "clustering," i.e., they arrive with a known average rate in a fixed time period but each event arrives at a time independent of the time since the last event, the exact integer number of events can be modeled with the Poisson distribution
- The Poisson is a single parameter family, the parameter being λ, the expected number of events in the interval of interest
- For a Poisson random variable X, the probability of exactly r events is

$$Pr(X = r) = \frac{\lambda^r e^{-\lambda}}{r!}$$

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

The Poisson Distribution

The Poisson Distribution

- When events arrive without any systematic "clustering," i.e., they arrive with a known average rate in a fixed time period but each event arrives at a time independent of the time since the last event, the exact integer number of events can be modeled with the Poisson distribution
- The Poisson is a single parameter family, the parameter being λ, the expected number of events in the interval of interest
- For a Poisson random variable X, the probability of exactly r events is

$$Pr(X = r) = \frac{\lambda^r e^{-\lambda}}{r!}$$

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

Probability Models
The Normal Distribution
The Multivariate Normal Distribution
The Lognormal Distribution
The Binomial Distribution
The Poisson Distribution

Characteristics of the Poisson Distribution

Characteristics of the Poisson Distribution

- The Poisson is used widely to model occurrences of low probability events

- A random variable X having a Poisson distribution with parameter λ has mean and variance given by

$$E(X) = \lambda$$
$$Var(X) = \lambda$$

Sampling Distributions

Sampling Distributions

- As discussed in your introductory course, we frequently sample from a population and obtain a statistic as an estimate of some key quantity
- Over repeated samples, these estimates show variability
- This variability is like noise, degrading the signal that is the parameter
- The known or hypothetical *sampling distribution* of the statistic allows us to gauge how accurate our parameter estimate is (at least in the long run)

Sampling Distributions

Sampling Distributions

- As discussed in your introductory course, we frequently sample from a population and obtain a statistic as an estimate of some key quantity
- Over repeated samples, these estimates show variability
- This variability is like noise, degrading the signal that is the parameter
- The known or hypothetical *sampling distribution* of the statistic allows us to gauge how accurate our parameter estimate is (at least in the long run)

Sampling Distributions

Sampling Distributions

- As discussed in your introductory course, we frequently sample from a population and obtain a statistic as an estimate of some key quantity
- Over repeated samples, these estimates show variability
- This variability is like noise, degrading the signal that is the parameter
- The known or hypothetical *sampling distribution* of the statistic allows us to gauge how accurate our parameter estimate is (at least in the long run)

Sampling Distributions

Sampling Distributions

- As discussed in your introductory course, we frequently sample from a population and obtain a statistic as an estimate of some key quantity
- Over repeated samples, these estimates show variability
- This variability is like noise, degrading the signal that is the parameter
- The known or hypothetical *sampling distribution* of the statistic allows us to gauge how accurate our parameter estimate is (at least in the long run)

Sampling Distributions

Sampling Distributions

- As discussed in your introductory course, we frequently sample from a population and obtain a statistic as an estimate of some key quantity
- Over repeated samples, these estimates show variability
- This variability is like noise, degrading the signal that is the parameter
- The known or hypothetical *sampling distribution* of the statistic allows us to gauge how accurate our parameter estimate is (at least in the long run)

Sampling Distributions
An Example

Sampling Distributions — An Example

- Suppose we take an opinion poll of $N = 100$ people at random, and 47% of them favor some position

- The question is, what does that tell us about the proportion of people in the population favoring the position?

Sampling Distributions
An Example

Sampling Distributions — An Example

- Suppose we take an opinion poll of $N = 100$ people at random, and 47% of them favor some position

- The question is, what does that tell us about the proportion of people in the population favoring the position?

Sampling Distributions
An Example

Sampling Distributions — An Example

- Suppose we take an opinion poll of $N = 100$ people at random, and 47% of them favor some position

- The question is, what does that tell us about the proportion of people in the population favoring the position?

Sampling Distributions
An Example

Sampling Distributions — An Example

- In your introductory course, you learned as a simple consequence of the binomial distribution that if the population proportion is p, the sample proportion $\hat{p}$ has a sampling distribution that is approximately normal, with mean p and variance $p(1-p)/N$
- For any hypothesized value of p, this tells us, through our knowledge of the normal distribution, how likely we would be to observe a value of .47
- We can use this, in turn, to evaluate which values of p are "reasonable" in some sense

Sampling Distributions
An Example

Sampling Distributions — An Example

- In your introductory course, you learned as a simple consequence of the binomial distribution that if the population proportion is p, the sample proportion $\hat{p}$ has a sampling distribution that is approximately normal, with mean p and variance $p(1-p)/N$

- For any hypothesized value of p, this tells us, through our knowledge of the normal distribution, how likely we would be to observe a value of .47

- We can use this, in turn, to evaluate which values of p are "reasonable" in some sense

Sampling Distributions
An Example

Sampling Distributions — An Example

- In your introductory course, you learned as a simple consequence of the binomial distribution that if the population proportion is p, the sample proportion $\hat{p}$ has a sampling distribution that is approximately normal, with mean p and variance $p(1-p)/N$
- For any hypothesized value of p, this tells us, through our knowledge of the normal distribution, how likely we would be to observe a value of .47
- We can use this, in turn, to evaluate which values of p are "reasonable" in some sense

Sampling Distributions
An Example

Sampling Distributions — An Example

- In your introductory course, you learned as a simple consequence of the binomial distribution that if the population proportion is p, the sample proportion $\hat{p}$ has a sampling distribution that is approximately normal, with mean p and variance $p(1-p)/N$

- For any hypothesized value of p, this tells us, through our knowledge of the normal distribution, how likely we would be to observe a value of .47

- We can use this, in turn, to evaluate which values of p are "reasonable" in some sense

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals

Confidence Intervals

- A *confidence interval* is a numerical interval constructed on the basis of data
- Such an interval is called a 95% (or .95) confidence interval if it is constructed so that it contains the true parameter value at least 95% of the time *in the long run*
- There are a variety of methods available for constructing confidence intervals

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals

Confidence Intervals

- A *confidence interval* is a numerical interval constructed on the basis of data
- Such an interval is called a 95% (or .95) confidence interval if it is constructed so that it contains the true parameter value at least 95% of the time *in the long run*
- There are a variety of methods available for constructing confidence intervals

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals

Confidence Intervals

- A *confidence interval* is a numerical interval constructed on the basis of data
- Such an interval is called a 95% (or .95) confidence interval if it is constructed so that it contains the true parameter value at least 95% of the time *in the long run*
- There are a variety of methods available for constructing confidence intervals

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals

Confidence Intervals

- A *confidence interval* is a numerical interval constructed on the basis of data
- Such an interval is called a 95% (or .95) confidence interval if it is constructed so that it contains the true parameter value at least 95% of the time *in the long run*
- There are a variety of methods available for constructing confidence intervals

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Simple Normal Theory Confidence Intervals

Normal Theory Confidence Intervals

- In Psychology 310 we leared about simple symmetric confidence intervals based on the normal distribution
- If a statistic $\hat{\theta}$ used to estimate a parameter θ has a normal sampling distribution with mean θ and sampling variance $Var(\hat{\theta})$, then we may construct a 95% confidence interval for θ as

$$\hat{\theta} \pm 1.96\sqrt{Var(\hat{\theta})}$$

- In general, a consistent estimator $\widehat{Var}(\hat{\theta})$ may be substituted for $Var(\hat{\theta})$ in the above

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Simple Normal Theory Confidence Intervals

Normal Theory Confidence Intervals

- In Psychology 310 we leared about simple symmetric confidence intervals based on the normal distribution
- If a statistic $\hat{\theta}$ used to estimate a parameter θ has a normal sampling distribution with mean θ and sampling variance $Var(\hat{\theta})$, then we may construct a 95% confidence interval for θ as

$$\hat{\theta} \pm 1.96\sqrt{Var(\hat{\theta})}$$

- In general, a consistent estimator $\widehat{Var}(\hat{\theta})$ may be substituted for $Var(\hat{\theta})$ in the above

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Simple Normal Theory Confidence Intervals

Normal Theory Confidence Intervals

- In Psychology 310 we leared about simple symmetric confidence intervals based on the normal distribution
- If a statistic $\hat{\theta}$ used to estimate a parameter θ has a normal sampling distribution with mean θ and sampling variance $Var(\hat{\theta})$, then we may construct a 95% confidence interval for θ as

$$\hat{\theta} \pm 1.96\sqrt{Var(\hat{\theta})}$$

- In general, a consistent estimator $\widehat{Var}(\hat{\theta})$ may be substituted for $Var(\hat{\theta})$ in the above

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Simple Normal Theory Confidence Intervals

Normal Theory Confidence Intervals

- In Psychology 310 we leared about simple symmetric confidence intervals based on the normal distribution
- If a statistic $\hat{\theta}$ used to estimate a parameter θ has a normal sampling distribution with mean θ and sampling variance $Var(\hat{\theta})$, then we may construct a 95% confidence interval for θ as

$$\hat{\theta} \pm 1.96\sqrt{Var(\hat{\theta})}$$

- In general, a consistent estimator $\widehat{Var}(\hat{\theta})$ may be substituted for $Var(\hat{\theta})$ in the above

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals on Linear Combinations

Confidence Intervals on Linear Combinations

- As we saw in Psychology 310, frequently linear combinations of parameters are of interest
- In that case, we can construct appropriate point estimates, standard errors, test statistics, and confidence intervals
- Methods are discussed in detail in the Psychology 310 handout, *A Unified Approach to Some Common Statistical Tests*

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals on Linear Combinations

Confidence Intervals on Linear Combinations

- As we saw in Psychology 310, frequently linear combinations of parameters are of interest
- In that case, we can construct appropriate point estimates, standard errors, test statistics, and confidence intervals
- Methods are discussed in detail in the Psychology 310 handout, *A Unified Approach to Some Common Statistical Tests*

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals on Linear Combinations

Confidence Intervals on Linear Combinations

- As we saw in Psychology 310, frequently linear combinations of parameters are of interest
- In that case, we can construct appropriate point estimates, standard errors, test statistics, and confidence intervals
- Methods are discussed in detail in the Psychology 310 handout, *A Unified Approach to Some Common Statistical Tests*

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals on Linear Combinations

Confidence Intervals on Linear Combinations

- As we saw in Psychology 310, frequently linear combinations of parameters are of interest
- In that case, we can construct appropriate point estimates, standard errors, test statistics, and confidence intervals
- Methods are discussed in detail in the Psychology 310 handout, *A Unified Approach to Some Common Statistical Tests*

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation

Confidence Intervals Via Simulation

- In some cases, we are interested in a function of parameters
- We know the distribution of individual parameter estimates, but we don't have a convenient expression for the distribution of the function of the parameter estimates
- In this case, we can simulate the distribution of the function of parameter estimates using random number generation
- To generate the 95% confidence interval, we extract the .025 and .975 quantiles of the resulting simulated data

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation

Confidence Intervals Via Simulation

- In some cases, we are interested in a function of parameters
- We know the distribution of individual parameter estimates, but we don't have a convenient expression for the distribution of the function of the parameter estimates
- In this case, we can simulate the distribution of the function of parameter estimates using random number generation
- To generate the 95% confidence interval, we extract the .025 and .975 quantiles of the resulting simulated data

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation

Confidence Intervals Via Simulation

- In some cases, we are interested in a function of parameters
- We know the distribution of individual parameter estimates, but we don't have a convenient expression for the distribution of the function of the parameter estimates
- In this case, we can simulate the distribution of the function of parameter estimates using random number generation
- To generate the 95% confidence interval, we extract the .025 and .975 quantiles of the resulting simulated data

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation

Confidence Intervals Via Simulation

- In some cases, we are interested in a function of parameters
- We know the distribution of individual parameter estimates, but we don't have a convenient expression for the distribution of the function of the parameter estimates
- In this case, we can simulate the distribution of the function of parameter estimates using random number generation
- To generate the 95% confidence interval, we extract the .025 and .975 quantiles of the resulting simulated data

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation

Confidence Intervals Via Simulation

- In some cases, we are interested in a function of parameters
- We know the distribution of individual parameter estimates, but we don't have a convenient expression for the distribution of the function of the parameter estimates
- In this case, we can simulate the distribution of the function of parameter estimates using random number generation
- To generate the 95% confidence interval, we extract the .025 and .975 quantiles of the resulting simulated data

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation
An Example

Example (Confidence Intervals Via Simulation)

- An example of the simulation approach can be found on page 20 of Gelman & Hill
- They assume that, with $N = 500$ per group, the distribution of the sample proportion can be approximated very accurately with a normal distribution
- In the problem of interest, the experimenter has observed sample proportions $\hat{p}_1$ and $\hat{p}_2$, each based on samples of 500
- However, the experimenter wishes to construct a confidence interval on p_1/p_2.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation
An Example

Example (Confidence Intervals Via Simulation)

- An example of the simulation approach can be found on page 20 of Gelman & Hill
- They assume that, with $N = 500$ per group, the distribution of the sample proportion can be approximated very accurately with a normal distribution
- In the problem of interest, the experimenter has observed sample proportions $\hat{p}_1$ and $\hat{p}_2$, each based on samples of 500
- However, the experimenter wishes to construct a confidence interval on p_1/p_2.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation
An Example

Example (Confidence Intervals Via Simulation)

- An example of the simulation approach can be found on page 20 of Gelman & Hill
- They assume that, with $N = 500$ per group, the distribution of the sample proportion can be approximated very accurately with a normal distribution
- In the problem of interest, the experimenter has observed sample proportions $\hat{p}_1$ and $\hat{p}_2$, each based on samples of 500
- However, the experimenter wishes to construct a confidence interval on p_1/p_2.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation
An Example

Example (Confidence Intervals Via Simulation)

- An example of the simulation approach can be found on page 20 of Gelman & Hill
- They assume that, with $N = 500$ per group, the distribution of the sample proportion can be approximated very accurately with a normal distribution
- In the problem of interest, the experimenter has observed sample proportions $\hat{p}_1$ and $\hat{p}_2$, each based on samples of 500
- However, the experimenter wishes to construct a confidence interval on p_1/p_2.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation
An Example

Example (Confidence Intervals Via Simulation)

- An example of the simulation approach can be found on page 20 of Gelman & Hill
- They assume that, with $N = 500$ per group, the distribution of the sample proportion can be approximated very accurately with a normal distribution
- In the problem of interest, the experimenter has observed sample proportions $\hat{p}_1$ and $\hat{p}_2$, each based on samples of 500
- However, the experimenter wishes to construct a confidence interval on p_1/p_2.

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation
An Example

Example (Confidence Intervals Via Simulation)

- The experimenter proceeds by constructing 10000 independent replications of $\hat{p}_1$ and 10000 replications of $\hat{p}_2$
- For each pair, the ratio $\hat{p}_1/\hat{p}_2$ is computed
- This creates a set of 10000 replications of the ratio of proportions
- The 95% confidence interval is then constructed from the .025 and .975 quantiles of this set of 10000 ratios

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation
An Example

Example (Confidence Intervals Via Simulation)

- The experimenter proceeds by constructing 10000 independent replications of $\hat{p}_1$ and 10000 replications of $\hat{p}_2$
- For each pair, the ratio $\hat{p}_1/\hat{p}_2$ is computed
- This creates a set of 10000 replications of the ratio of proportions
- The 95% confidence interval is then constructed from the .025 and .975 quantiles of this set of 10000 ratios

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation
An Example

Example (Confidence Intervals Via Simulation)

- The experimenter proceeds by constructing 10000 independent replications of $\hat{p}_1$ and 10000 replications of $\hat{p}_2$
- For each pair, the ratio $\hat{p}_1/\hat{p}_2$ is computed
- This creates a set of 10000 replications of the ratio of proportions
- The 95% confidence interval is then constructed from the .025 and .975 quantiles of this set of 10000 ratios

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation
An Example

Example (Confidence Intervals Via Simulation)

- The experimenter proceeds by constructing 10000 independent replications of $\hat{p}_1$ and 10000 replications of $\hat{p}_2$
- For each pair, the ratio $\hat{p}_1/\hat{p}_2$ is computed
- This creates a set of 10000 replications of the ratio of proportions
- The 95% confidence interval is then constructed from the .025 and .975 quantiles of this set of 10000 ratios

Random Variables
Probability Distributions
Sampling Distributions
Confidence Intervals
Hypothesis Testing

The Classic Normal Theory Approach
Confidence Intervals on Linear Transformations
Confidence Intervals Via Simulation

Confidence Intervals Via Simulation
An Example

Example (Confidence Intervals Via Simulation)

- The experimenter proceeds by constructing 10000 independent replications of $\hat{p}_1$ and 10000 replications of $\hat{p}_2$
- For each pair, the ratio $\hat{p}_1/\hat{p}_2$ is computed
- This creates a set of 10000 replications of the ratio of proportions
- The 95% confidence interval is then constructed from the .025 and .975 quantiles of this set of 10000 ratios

Hypothesis Testing

Hypothesis Testing

- Gelman and Hill make a number of interesting points in their brief discussion
- They suggest viewing a hypothesis as a model about the data
- Testing the hypothesis involves comparing the behavior of the data with the data predicted by the model
- For example, if proportions are showing their standard random variation, this implies something about the size of that variation
- They examine this notion in an extensive example

Hypothesis Testing

Hypothesis Testing

- Gelman and Hill make a number of interesting points in their brief discussion
- They suggest viewing a hypothesis as a model about the data
- Testing the hypothesis involves comparing the behavior of the data with the data predicted by the model
- For example, if proportions are showing their standard random variation, this implies something about the size of that variation
- They examine this notion in an extensive example

Hypothesis Testing

Hypothesis Testing

- Gelman and Hill make a number of interesting points in their brief discussion
- They suggest viewing a hypothesis as a model about the data
- Testing the hypothesis involves comparing the behavior of the data with the data predicted by the model
- For example, if proportions are showing their standard random variation, this implies something about the size of that variation
- They examine this notion in an extensive example

Hypothesis Testing

Hypothesis Testing

- Gelman and Hill make a number of interesting points in their brief discussion
- They suggest viewing a hypothesis as a model about the data
- Testing the hypothesis involves comparing the behavior of the data with the data predicted by the model
- For example, if proportions are showing their standard random variation, this implies something about the size of that variation
- They examine this notion in an extensive example

Hypothesis Testing

Hypothesis Testing

- Gelman and Hill make a number of interesting points in their brief discussion
- They suggest viewing a hypothesis as a model about the data
- Testing the hypothesis involves comparing the behavior of the data with the data predicted by the model
- For example, if proportions are showing their standard random variation, this implies something about the size of that variation
- They examine this notion in an extensive example

Hypothesis Testing

Hypothesis Testing

- Gelman and Hill make a number of interesting points in their brief discussion
- They suggest viewing a hypothesis as a model about the data
- Testing the hypothesis involves comparing the behavior of the data with the data predicted by the model
- For example, if proportions are showing their standard random variation, this implies something about the size of that variation
- They examine this notion in an extensive example