Expected values

Statistical Inference

Brian Caffo, Jeff Leek, Roger Peng

## Expected values

* The **expected value** or **mean** of a random variable is the center of its distribution
* For discrete random variable with PMF , it is defined as follows
* where the sum is taken over the possible values of
* represents the center of mass of a collection of locations and weights,

## Example

### Find the center of mass of the bars

## Loading required package: MASS

## Loading required package: HistData

## Loading required package: Hmisc

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

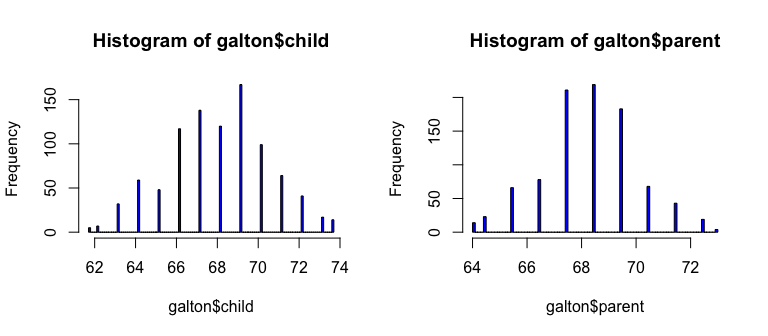
## Loading required package: ggplot2

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':  
##   
## format.pval, round.POSIXt, trunc.POSIXt, units

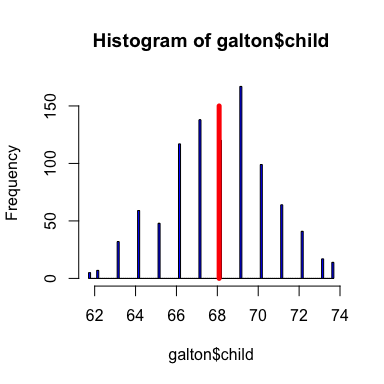
##   
## Attaching package: 'UsingR'

## The following object is masked from 'package:survival':  
##   
## cancer



## The center of mass is the empirical mean

hist(galton$child,col="blue",breaks=100)  
 meanChild <- mean(galton$child)  
 lines(rep(meanChild,100),seq(0,150,length=100),col="red",lwd=5)



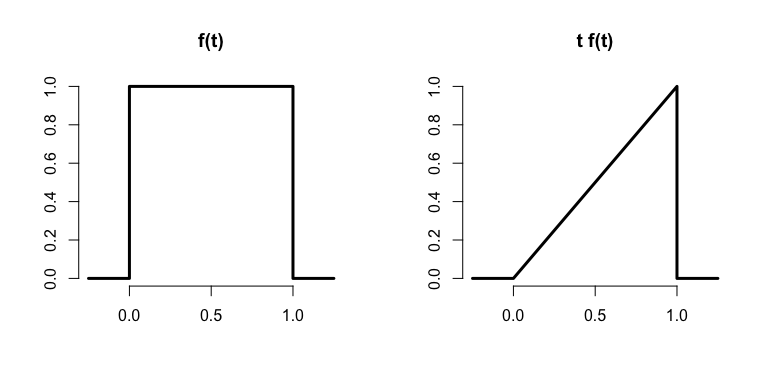
## Example

* Suppose that a die is rolled and is the number face up
* What is the expected value of ?
* Again, the geometric argument makes this answer obvious without calculation.

## Continuous random variables

* For a continuous random variable, , with density, , the expected value is defined as follows
* This definition borrows from the definition of center of mass for a continuous body

## Example

* Consider a density where for between zero and one
* (Is this a valid density?)
* Suppose that follows this density; what is its expected value?  
  

## Rules about expected values

* The expected value is a linear operator
* If and are not random and and are two random variables then

## Example

* You flip a coin, and simulate a uniform random number , what is the expected value of their sum?
* Another example, you roll a die twice. What is the expected value of the average?
* Let and be the results of the two rolls

## Example

1. Let for be a collection of random variables, each from a distribution with mean
2. Calculate the expected value of the sample average of the

## Remark

* Therefore, the expected value of the **sample mean** is the population mean that it's trying to estimate
* When the expected value of an estimator is what its trying to estimate, we say that the estimator is **unbiased**

## The variance

* The variance of a random variable is a measure of *spread*
* If is a random variable with mean , the variance of is defined as

the expected (squared) distance from the mean - Densities with a higher variance are more spread out than densities with a lower variance

* Convenient computational form
* If is constant then
* The square root of the variance is called the **standard deviation**
* The standard deviation has the same units as

## Example

* What's the sample variance from the result of a toss of a die?

## Example

* What's the sample variance from the result of the toss of a coin with probability of heads (1) of ?

## Interpreting variances

* Chebyshev's inequality is useful for interpreting variances
* This inequality states that
* For example, the probability that a random variable lies beyond standard deviations from its mean is less than
* Note this is only a bound; the actual probability might be quite a bit smaller

## Example

* IQs are often said to be distributed with a mean of and a sd of
* What is the probability of a randomly drawn person having an IQ higher than or below ?
* Thus we want to know the probability of a person being more than standard deviations from the mean
* Thus Chebyshev's inequality suggests that this will be no larger than 6%
* IQs distributions are often cited as being bell shaped, in which case this bound is very conservative
* The probability of a random draw from a bell curve being standard deviations from the mean is on the order of (one thousandth of one percent)

## Example

* A former buzz phrase in industrial quality control is Motorola's "Six Sigma" whereby businesses are suggested to control extreme events or rare defective parts
* Chebyshev's inequality states that the probability of a "Six Sigma" event is less than
* If a bell curve is assumed, the probability of a "six sigma" event is on the order of (one ten millionth of a percent)