Expected values

Statistical Inference

Brian Caffo, Jeff Leek, Roger Peng
Johns Hopkins Bloomberg School of Public Health

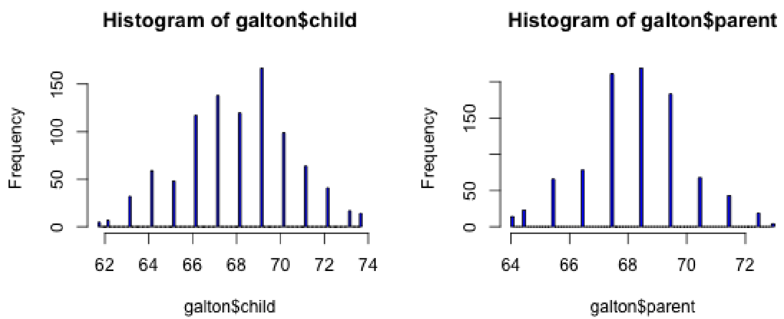
Expected values

- The **expected value** or **mean** of a random variable is the center of its distribution
- For discrete random variable X with PMF p(x), it is defined as follows E[X] = \sum_x xp(x). where the sum is taken over the possible values of x
- E[X] represents the center of mass of a collection of locations and weights, \{x, p(x)\}

Example

Find the center of mass of the bars

```
## Loading required package: MASS
```

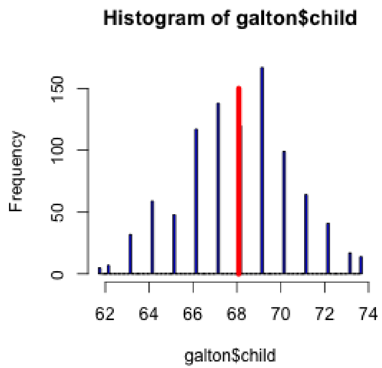


Using manipulate

```
library(manipulate)
myHist <- function(mu){
  hist(galton$child,col="blue",breaks=100)
  lines(c(mu, mu), c(0, 150),col="red",lwd=5)
  mse <- mean((galton$child - mu)^2)
  text(63, 150, paste("mu = ", mu))
  text(63, 140, paste("Imbalance = ", round(mse, 2)))
}
manipulate(myHist(mu), mu = slider(62, 74, step = 0.5))
```

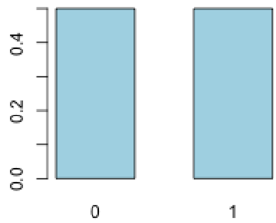
The center of mass is the empirical mean

```
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 a coin is flipped and X is declared 0 or 1 corresponding to a head or a tail, respectively
- What is the expected value of X? E[X] = .5 \times 0 + .5 \times 1 = .5
- Note, if thought about geometrically, this answer is obvious; if two equal weights are spaced at 0 and 1, the center of mass will be .5



Example

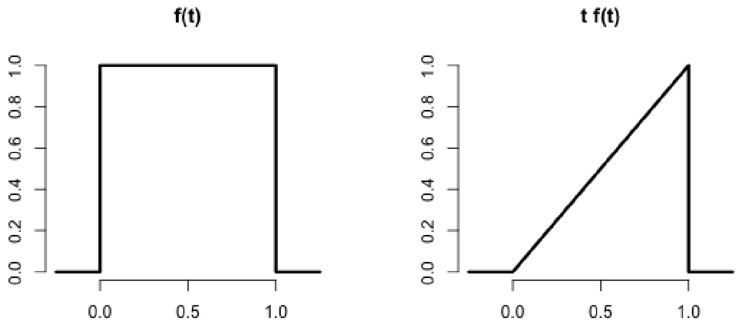
- Suppose that a die is rolled and X is the number face up
- What is the expected value of X? $E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$
- Again, the geometric argument makes this answer obvious without calculation.

Continuous random variables

- For a continuous random variable, X, with density, f, the expected value is defined as follows E[X] = \mbox{the area under the function}~~~ t f(t)
- This definition borrows from the definition of center of mass for a continuous body

Example

- Consider a density where $f(x) = 1$ for x between zero and one
- (Is this a valid density?)
- Suppose that X follows this density; what is its expected value?



Rules about expected values

- The expected value is a linear operator
- If a and b are not random and X and Y are two random variables then
  - $E[aX + b] = a\,E[X] + b$
  - $E[X + Y] = E[X] + E[Y]$

Example

- You flip a coin, X and simulate a uniform random number Y, what is the expected value of their sum? $E[X + Y] = E[X] + E[Y] = .5 + .5 = 1$
- Another example, you roll a die twice. What is the expected value of the average?
- Let X_1 and X_2 be the results of the two rolls $E[(X_1 + X_2) / 2] = \frac{1}{2}(E[X_1] + E[X_2]) = \frac{1}{2}(3.5 + 3.5) = 3.5$

Example

1. Let X_i for $i=1,\ldots,n$ be a collection of random variables, each from a distribution with mean μ
2. Calculate the expected value of the sample average of the X_i
$$\begin{eqnarray*} E\left[\frac{1}{n}\sum_{i=1}^n X_i\right] & = & \frac{1}{n} E\left[\sum_{i=1}^n X_i\right] \\ & = & \frac{1}{n} \sum_{i=1}^n E\left[X_i\right] \\ & = & \frac{1}{n} \sum_{i=1}^n \mu = \mu. \end{eqnarray*}$$

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 X is a random variable with mean μ, the variance of X is defined as

$$Var(X) = E[(X - \mu)^2]$$

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 $Var(X) = E[X^2] - E[X]^2$
- If a is constant then $Var(aX) = a^2 Var(X)$
- The square root of the variance is called the **standard deviation**
- The standard deviation has the same units as X

Example

- What's the sample variance from the result of a toss of a die?
  - $E[X] = 3.5$
  - $E[X^2] = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6} = 15.17$
- $Var(X) = E[X^2] - E[X]^2 \approx 2.92$

Example

- What's the sample variance from the result of the toss of a coin with probability of heads (1) of p?
  - $E[X] = 0 \times (1 - p) + 1 \times p = p$
  - $E[X^2] = E[X] = p$
- $Var(X) = E[X^2] - E[X]^2 = p - p^2 = p(1 - p)$

Interpreting variances

- Chebyshev's inequality is useful for interpreting variances
- This inequality states that P(|X - \mul \geq k\sigma) \leq \frac{1}{k^2}
- For example, the probability that a random variable lies beyond k standard deviations from its mean is less than 1/k^2 \begin{eqnarray*} 2\sigma & \rightarrow & 25\% \\ 3\sigma & \rightarrow & 11\% \\ 4\sigma & \rightarrow & 6\% \end{eqnarray*}
- 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 100 and a sd of 15
- What is the probability of a randomly drawn person having an IQ higher than 160 or below 40?
- Thus we want to know the probability of a person being more than 4 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 4 standard deviations from the mean is on the order of 10^{-5} (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 $1/6^2 \approx 3\%$
- If a bell curve is assumed, the probability of a "six sigma" event is on the order of 10^{-9} (one ten millionth of a percent)