5.1 Discrete Random Variables

Dr. Lauren Perry

Goals

1. Discuss discrete random variables using key terminology.
2. Calculate the expected value and standard deviation of a discrete random variable.

Discrete Random Variables

A **random variable** is a quantitative variable whose values are based on chance. By "chance", we mean that you can't *know* the outcome before it occurs.

A **discrete random variable** is a random variable whose possible values can be listed.

Notation

- Variables: x,y,z
- Random variables: X, Y, Z
- The event that the random variable X equals x: $\{X = x\}$
- The probability that the random variable X equals x: $P(X = x)$

Probability Histograms

A **probability histogram** is a histogram where the heights of the bars correspond to the probability of each value.

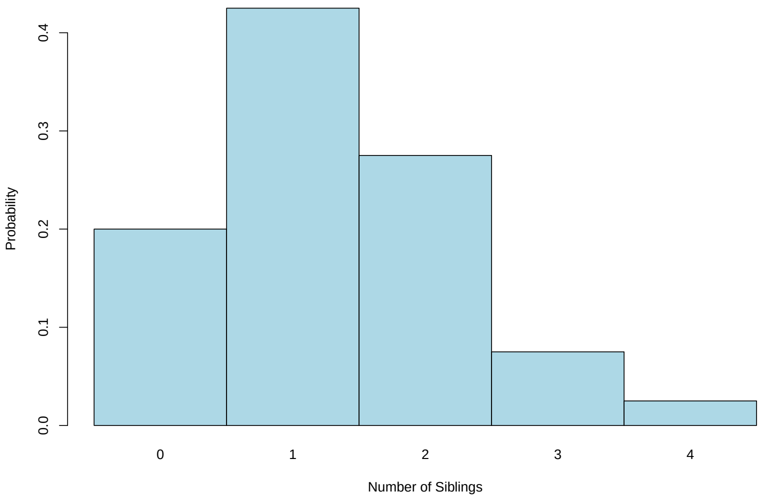
- For discrete random variables, each “bin” is one of the listed values.

Example: Probability Histograms

Number of Siblings, x	0	1	2	3	4
Probability, $P(X = x)$	0.200	0.425	0.275	0.075	0.025

(Assume for the sake of the example that no one has more than 4 siblings.)

Number of Siblings, x	0	1	2	3	4
Probability, $P(X = x)$	0.200	0.425	0.275	0.075	0.025



The Mean of a Discrete Random Variable

The mean of a discrete random variable X is denoted μ_X.

$$\mu_X = x_1 P(X = x_1) + x_2 P(X = x_2) + \cdots + x_n P(X = x_n).$$

If it's clear which random variable we're talking about, we can just write μ.

Example

Number of Siblings, x	0	1	2	3	4
Probability, $P(X = x)$	0.200	0.425	0.275	0.075	0.025

$$\mu = 0(0.200) + 1(0.425) + 2(0.275) + 3(0.075) + 4(0.025) = 1.3$$

Interpretation: in a large number of independent observations of a random variable X, the mean of those observations will approximately equal μ.

The Mean of a Discrete Random Variable

The mean of a random variable is also called the **expected value** or **expectation**.

Since measures of center are meant to identify the most common or most likely, you can think of this as the value we *expect* to see (most often).

Law of Large Numbers

The larger the number of observations, the closer their average tends to be to μ. This is known as the **law of large numbers**.

Example: Law of Large Numbers

Suppose I took a random sample of 10 people and asked how many siblings they have.

$$2, 2, 2, 2, 1, 0, 3, 1, 2, 0$$

In my random sample of 10, $\bar{x} = 2$, which is a reasonable estimate but not that close to the true mean $\mu = 1.3$.

- A random sample of 30 gave me a mean of $\bar{x} = 1.53$.
- A random sample of 100 gave me a mean of $\bar{x} = 1.47$.
- A random sample of 1000 gave me a mean of $\bar{x} = 1.307$.

Standard Deviation of a Discrete Random Variable

$$\sigma_X^2 = [x_1^2 P(X = x_1) + x_2^2 P(X = x_2) + \cdots + x_n^2 P(X = x_n)] - \mu_X^2$$

As before, the standard deviation is the square root of the variance:

$$\sigma = \sqrt{\sigma^2}$$

Example

Calculate the standard deviation of the Siblings variable.

In general, a table is the best way to keep track of a variance calculation:

x	$P(X=x)$	$xP(X=x)$	x^2	$x^2P(X=x)$
0	0.200	0	0	0
1	0.425	0.425	1	0.425
2	0.275	0.550	4	1.100
3	0.075	0.225	9	0.675
4	0.025	0.100	16	0.400
		$\mu = 1.3$		Total $= 2.6$

x	$P(X=x)$	$xP(X=x)$	x^2	$x^2P(X=x)$
0	0.200	0	0	0
1	0.425	0.425	1	0.425
2	0.275	0.550	4	1.100
3	0.075	0.225	9	0.675
4	0.025	0.100	16	0.400
		$\mu = 1.3$		Total $= 2.6$

- The variance is $\sigma^2 = 2.6 - 1.3^2 = 0.9$
- The standard deviation is $\sigma = \sqrt{0.9} = 0.9539$