# Statistical distributions

#### Wendy Rosettini

November 2023

## 1 Statistical distributions, Continuos, Discrete

A discrete distribution is one in which the data can only take on certain values, for example integers. A continuous distribution is one in which data can take on any value within a specified range (which may be infinite). For a discrete distribution, probabilities can be assigned to the values in the distribution – for example, "the probability that the web page will have 12 clicks in an hour is 0.15." In contrast, a continuous distribution has an infinite number of possible values, and the probability associated with any particular value of a continuous distribution is null. Therefore, continuous distributions are normally described in terms of probability density, which can be converted into the probability that a value will fall within a certain range.<sup>1</sup>

## Properties of Discrete and Continuous Distributions

### **Discrete Distribution:**

- 1. **Countable Outcomes:** Discrete distributions deal with countable outcomes, meaning that the set of possible values is finite or countably infinite.
- 2. Probability Mass Function (PMF): Describes the probability of each possible outcome. P(X = x) gives the probability of the random variable X taking the value x.
- 3. **Probability Space:** The sum of all probabilities for all possible outcomes equals 1.  $\sum P(X = x) = 1$ .
- 4. Cumulative Distribution Function (CDF):  $F(x) = P(X \le x)$  gives the probability that the random variable X is less than or equal to x.
- 5. Discrete Random Variables: The random variable X takes on distinct, separate values.
- 6. Examples: Coin tosses, dice rolls, the number of students in a class.

 $<sup>^1</sup>$ urly.it/3yj2n

#### **Continuous Distribution:**

- 1. Uncountable Outcomes: Continuous distributions deal with uncountably infinite outcomes, typically over a range of real numbers.
- 2. Probability Density Function (PDF): Describes the likelihood of a random variable taking on a particular value in a continuous range.  $P(a \le X \le b) = \int_a^b f(x) dx$ , where f(x) is the PDF.
- 3. **Probability Space:** The area under the entire PDF curve is equal to 1.  $\int_{-\infty}^{\infty} f(x) dx = 1.$
- 4. Cumulative Distribution Function (CDF):  $F(x) = P(X \le x)$  gives the probability that the random variable X is less than or equal to x.
- 5. Continuous Random Variables: The random variable X can take any value within a range.
- 6. Examples: Height of individuals, time taken for a process, temperature.

#### **Commonality:**

- 1. Expectation (Mean): Both types have an expected value, which represents the average value.
- 2. Variance: Both types have a variance, which measures the spread of the distribution.

## 2 Simulations

As a simulation, we can take the homework  $2^2$ , where we took a survey and we had to choose 3 variables. I chose:

- Hard Worker(0-5) as the quantitative discrete variable
- Age as the quantitative continuous variable
- Background (degree) ad the qualitative variabl

and we did for example the joint distribution of 2 variables

 $<sup>^{2} \</sup>rm http://wendy.altervista.org/frequenze.html$ 

#### Carica un file Excel (.xlsx)

Scegli file Profession	al Life xisx													
(Anonymous) Stat Unit ID	Background (degree)	Expected work sectors	Programming Languages	Hard Worker (0-5)	Ambitious (0-5)	Team leader or Team player ?	Enterpreneurial attitude (0-5)	Preferred Workload	Scalability	Most recent working position	Sex	Age	Main Interests	Main hobbies
1933541	Engineering in computer science	cybersecurity	C, Python, java, JavaScript, PHP, assembly, C++, Ruby on Rails, (other languages: HTML e CSS, mySQL)	5	4	Team Jeader	2	8 hours per day	Yes		F	22	Cybersecurity	
1933529	Computer and System Engineering	cybersecurity	PHP, Java, JavaScript, Python, Ruby on rails, C, C++, assembly,	5	5	Team leader	3	6 hours per day	Yes	- 1	м	22	Cybersecurity	Gaming
31337	Computer Science	cybersecurity	C, Python, Java, PHP, JS, C*,	5	5	Health assurance	Team player	3	8 hours per day	Yes	Penetration Tester	м	22	Vulnerability Research
2111409	Computer Science and Engineering	cybersecurity	c, java, angular, html, cas	s	5	team Jeader	3	8 hours per day	yes	software developer	F	23	Cybersecurity	Crafting
309200	Computer Engineering	Cybersecurity	C, Java, MySQL	s	s	3	6 hours per day	yes	м	23	96	1.76	Cybersecurity	Workout
1234567	Software and Information Engineering	Cybersecurity	C, Java, Python, JavaScritp, PHP, SQL, Haskell, assembly, C++,	5	5	Team player	4	8 hours per day	yes	software developer	м	21	84	1.76
2150048	Computer Engineering	Cybersecurity	Java,Python,PHP,SQL	5	5	Team leader	3	6 hours per day	yes	F	23	Cybersecurity	skiing.roller skating	Piano and Guitar
1600	Computer Engineering	Cybersecurity	C, C++, C#, Javascript, PHP, x86 Assembly, Python, CSS, Typoscript, MySQL	s	s	Team player	3	6 hours per day	Yes	Quality Assurance	м	23	45	1.73
1843916	Computer Science	Cybersecurity	C++, Java, Python, Assembly, Javascript, MySQL	5	4	Team player	3	6 hours per day	Yes	Help desk	м	24	100	1.8

#### Analisi della variabile: Hard Worker (0-5)

Valore	Frequenza Assoluta	Frequenza Relativa	Percentuale	
2	1	0.02	1.85%	
3	2	0.04	3.70%	
4	25	0.46	46.30%	
5	26	0.48	48.15%	
Analisi della variabile: Age				
Valore	Frequenza Assoluta	Frequenza Relativa	Percentuale	
21	7	0.13	12.96%	
22	17	0.31	31.48%	
23	14	0.26	25.93%	
24	4	0.07	7.41%	
25	5	0.09	9.26%	
26	2	0.04	3.70%	
27	3	0.06	5.56%	
39	1	0.02	1.85%	
23+	1	0.02	1.85%	

#### Analisi della variabile: Background (degree)

Valore	Frequenza Assoluta	Frequenza Relativa	Percentuale
Engineering in computer science	1	0.02	1.85%
Computer and System Engineering	1	0.02	1.85%
Computer Science	22	0.41	40.74%
Computer Science and Engineering	1	0.02	1.85%
Computer Engineering	7	0.13	12.96%
Software and Information Engineering	1	0.02	1.85%
Computer Engineering	2	0.04	3.70%
Engineering in Computer Science	1	0.02	1.85%
Information Engineering, Computer Science and Statistics	1	0.02	1.85%
Computer science	5	0.09	9.26%
Computer Sciwence	1	0.02	1.85%
International relations	1	0.02	1.85%
cybersecurity	2	0.04	3.70%
Security of information technologies	1	0.02	1.85%
computer science	2	0.04	3.70%
Computer engineering	2	0.04	3.70%
conputer engineering	1	0.02	1.85%
Cybersecurity	1	0.02	1.85%
computer engineering	1	0.02	1.85%

#### Distribuzione Congiunta tra Hard Worker (0-5) e Age

Valore	Frequenza Congiunta	Frequenza Relativa Congiunta	Percentuale Congiunta
5   22	8	0.15	14.81%
5   23	6	0.11	11.11%
5   21	5	0.09	9.26%
5   24	3	0.06	5.56%
4   25	3	0.06	5.56%
5   23+	1	0.02	1.85%
4   21	2	0.04	3.70%
4   23	7	0.13	12.96%
4   24	1	0.02	1.85%
5   25	2	0.04	3.70%
4   22	8	0.15	14.81%
2   22	1	0.02	1.85%
4   27	2	0.04	3.70%
4   39	1	0.02	1.85%
4   26	1	0.02	1.85%
3   23	1	0.02	1.85%
5   27	1	0.02	1.83%
3   26	1	0.02	1.85%