Notes on Descriptive Statistics

George Baker

A descriptive statistic is a summary statistic that quantifiably describes or summarises features from a collection of information.

Descriptive statistics is the process of using and analysing those statistics.

Expectation

The **expected value** of a random variable X, denoted $E(X)$ or $E[X]$ is a weighted average of the possible values that X can take, with each value being weighted according to the probability of that event occurring.

I.e. intuitively, **expectation is the arithmetic mean of a random variable** X.

Definition

Discrete

If $F_X(x)$ is the pmf of the discrete random variable X with $R_X = \{x_1, x_2, ...\}$, $E[X]$ is defined as:

$$E[X] = \sum_{x \epsilon R_X} x F_X(x)$$

$$= \sum_{x \epsilon R_X} x p X(x)$$

$$= \sum_{x \epsilon R_X} x P(X = x)$$

Continuous

If $F_X(x)$ is the pdf of the continuous random variable X, $E[X]$ is defined as:

$$E[X] = \int_{-\infty}^{\infty} x F_X(x)\, dx$$

Example

Here's a very **simple** example: What is the expected value when we roll a fair die?

$$\omega = \{1, 2, 3, 4, 5, 6\}$$

$$P(X = x \mid x \,\epsilon\, \omega) = \frac{1}{6}$$

$$E[X] = \sum_{x \epsilon R_X} xP(X = x)$$

$$E[X] = 1\times P(X = 1)+2\times P(X = 2)+3\times P(X = 3)+4\times P(X = 4)+5\times P(X = 5)+6\times P(X = 6)$$

$$E[X] = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = \frac{7}{2}$$

So the expectation is 3.5.

If you think about it, 3.5 is halfway between the possible values the die can take and so this is what you should have expected.

Variance

The **variance** of a random variable X, denoted $Var(X)$ and sometimes σ^2, measures how far a set of numbers is spread out from their average value.

Definition

The variance of a random variable X is the expected value of the squared deviation (difference) from the mean of X. The mean of X is $\mu = E[X]$.

Discrete

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

$$= \sum_{x \epsilon R_X} p_X(x)(x - \mu)^2$$

$$= E[X^2] - E[(X)^2]$$

$$= \sum_{x \epsilon R_X} p_X(x)x^2 - \mu^2$$

Continuous

$$
\begin{aligned}
Var(X) &= E[(X-\mu)^2] \\
&= \int_{-\infty}^{\infty} p_X(x)(x-\mu)^2 \\
&= E[X^2] - E[(X)^2] \\
&= \int_{-\infty}^{\infty} p_X(x)x^2 - \mu^2
\end{aligned}
$$

Where $p_X(x)$ is shorthand notation for pmf.