12.1: Discrete random variables
the mean

The *mean* of a sequence of numbers $a_1, a_2, \ldots, a_n$ is the average:

$$\mu = \frac{1}{n}(a_1 + \ldots + a_n) = \frac{1}{n}\sum_{i=1}^{n} a_i$$

Example

The quiz scores from one of our sections on a recent quiz were as follows:

8,10,10,10,10,10,10,5,10,0,10,10,10,10,10, 10,10,10,10,10.

Compute the mean quiz score for this section.

Solution

There were $n = 20$ scores in total. The sum is
$8 + 10 + 10 + 10 + 10 + 10 + 10 + 5 + 10 + 0 + 10 + 10 + 10 + 10 + 10 + 10 + 10 + 10 + 10 + 10 = 183$. Thus, the mean is $\mu = \frac{183}{20} = 9.15$.

Variance

The *variance* measures the extent to which individual data points differ from the mean. As with the sum of squares of errors we used to measure the fit of a recursion line to a data set, the variance is defined as the average of the sum of squares of "errors" where we treat the difference between a data point and the mean as an "error"

Formally, the variance, σ^2, of the sequence of numbers $a_1, \ldots, a_n$ having mean μ is

$$\sigma^2 := \frac{1}{n} \sum_{i=1}^{n} (a_i - \mu)^2$$

The number $\sigma := \sqrt{\sigma^2}$ is called the *standard deviation*.

Example

Compute the standard deviation for the quiz scores described above.

Solution

We compute the variance as

$\sigma^2 = \frac{1}{20}(8 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 + (5 - 9.15)^2 + (10 - 9.15)^2 + (0 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 + (10 - 9.15)^2 = 5.2725.$

And the standard deviation is $\sqrt{5.2725} \approx 2.393$.

Frequency table

We may present our data as a frequency table rather than as a list.

Given a list of numbers $a_1, \ldots, a_n$ taking possible values $v_1, \ldots, v_m$ we define the relative frequency of the value v to be the number of data points a_i for which $a_i = v$ divided by n.

Conventionally, this is written as

$$p_j := \frac{\text{the number of indices } i \text{ for which } a_i = v}{n}$$

Note: $\sum_{j=1}^{m} p_j = 1$ and $0 \le p_j \le 1$ for every j.

Example

Score	Number of instances with this score	Relative frequency
0	1	0.05
1	0	0
2	0	0
3	0	0
4	0	0
5	1	0.05
6	0	0
7	0	0
8	1	0.05
9	0	0
10	17	0.85

Random Variables

We may organize the information from a relative frequency table into a function, called a *random variable*.

Given a set of possible values V and a sequence of numbers $a_1, \ldots, a_n$ from V, the random variable X corresponding to this sequence is the function defined by $X(v) :=$ the relative frequency of the value v.

More generally, a random variable X (on V) is a function with domain V having the properties:

- $0 \le X(v) \le 1$ for all v in V

- $\sum_{v \text{ in } V} X(v) = 1$

Example

Find the random variable expressing the relative frequency of the values for the sum of the numbers shown on two dice.

That is, list all the possible pairs of dice throws, $\langle 1, 1 \rangle, \langle 1, 2 \rangle, \ldots, \langle 6, 5 \rangle, \langle 6, 6 \rangle$ and then take the data points to the sums $2, 3, \ldots, 11, 12$ and find a random variable expressing the relative frequencies for these data.

Solution

Value	Sums giving this value
2	$1 + 1$
3	$1 + 2, \, 2 + 1$
4	$1 + 3, \, 2 + 2, \, 3 + 1$
5	$1 + 4, \, 2 + 3, \, 3 + 2, \, 4 + 1$
6	$1 + 5, \, 2 + 4, \, 3 + 3, \, 4 + 2, \, 5 + 1$
7	$1 + 6, \, 2 + 5, \, 3 + 4, \, 4 + 3, \, 5 + 2, \, 6 + 1$
8	$2 + 6, \, 3 + 5, \, 4 + 4, \, 5 + 3, \, 6 + 2$
9	$3 + 6, \, 4 + 5, \, 5 + 4, \, 6 + 3$
10	$4 + 6, \, 5 + 5, \, 6 + 4$
11	$5 + 6, \, 6 + 5$
12	$6 + 6$

Solution, continued

So $X(2) = \frac{1}{36}$, $X(3) = \frac{1}{18}$, $X(4) = \frac{1}{12}$, $X(5) = \frac{1}{9}$, $X(6) = \frac{5}{36}$, $X(7) = \frac{1}{6}$, $X(8) = \frac{5}{36}$, $X(9) = \frac{1}{9}$, $X(10) = \frac{1}{12}$, $X(11) = \frac{1}{18}$, and $X(12) = \frac{1}{36}$.

Expected value

One may compute the mean of a data set from its corresponding random variable. (Called in this case the *expected value* of X, or $E(X)$).

Let $a_1, \ldots, a_n$ be a sequence of numbers with corresponding random variable X and possible values $v_1, \ldots, v_m$.

$$
\begin{aligned}
\mu \quad &= \quad \frac{1}{n} \sum_{i=1}^{n} a_i \\
&= \quad \frac{1}{n} \sum_{j=1}^{m} v_j \times \text{the number of } i \text{ with } a_i = v_j \\
&= \quad \sum_{j=1}^{m} v_j X(v_j) \\
&=: \quad E(X)
\end{aligned}
$$

Variance of a random variable

Likewise, we may compute the variance of a data set from its random variable. Keeping the notation of the previous example,

$$
\begin{aligned}
\sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}(a_i - \mu)^2 \\
&= \sum_{j=1}^{m} X(v_j)(v_j - E(X))^2 \\
&= \sum_{j=1}^{m} (X(v_j)v_j^2 - 2v_j X(v_j)E(X) + X(v_j)E(X)^2 \\
&= (\sum_{j=1}^{m}(X(v_j)v_j^2)) - E(X)^2 \\
&=: E(X^2) - E(X)^2 \\
&=: \mathrm{Var}(X)
\end{aligned}
$$

Example

Compute the variance for the dice example.

Solution

$$E(X^2) = 2^2\frac{1}{36} + 3^2\frac{2}{36} + 4^2\frac{3}{36} + 5^2\frac{4}{36}$$

$$6^2\frac{5}{36} + 7^2\frac{6}{36} + 8^2\frac{5}{36} + 9^2\frac{4}{36}$$

$$10^2\frac{3}{36} + 11^2\frac{2}{36} + 12^2\frac{1}{36}$$

$$= \frac{1}{36}(1 + 18 + 48 + 100 + 180 + 294$$

$$+320 + 324 + 300 + 242 + 144)$$

$$= \frac{1974}{36}$$

$$= 54\frac{5}{6}$$

while $E(X) = 7$. Thus, $\text{Var}(X) = 54\frac{5}{6} - 49 = 5\frac{5}{6}$ and the standard

deviation is $\sqrt{\frac{35}{6}} \approx 2.4$.