

DISTRIBUTIONS RANDOM VARIABLES

Example - You have a bag with 8 oranges and 3 of them are rotten. Choose a sample of 3. Let X be the number of rotten oranges in the sample. Write a probability distribution table for the number of rotten oranges in the sample. Graph the probability distribution histogram.

We will need to use counting to find the probabilities. The probability of event E is

$$P(E) = \frac{n(E)}{n(S)} = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } S}$$

Find the probability distribution table:

event	X	$P(X)$
0R,3G	0	$C(3,0)C(5,3)/C(8,3) = 10/56$
1R,2G	1	$C(3,1)C(5,2)/C(8,3) = 30/56$
2R,1G	2	$C(3,2)C(5,1)/C(8,3) = 15/56$
3R,0G	3	$C(3,3)C(5,0)/C(8,3) = 1/56$

Put this into a histogram.

We will have X on the horizontal axis and the probability $P(X)$ on the vertical axis.

We have a rectangle 1 unit wide centered on each X value and its height is the probability:

Example - We are given the following data for the number of a certain magazine sold per week for the past year at a newstand:

number sold	frequency	probability
15	5	
16	4	
17	8	
18	10	
19	8	
20	15	
21	2	
total	52	

Find the probability distribution histogram for this data set:

EXPECTED VALUE

How many rotten oranges would you expect to get in your sample of 3? How many magazines do you expect to sell? We have a formula for the expected value of a probability distribution:

$$E(X) = X_1P(X_1) + X_2P(X_2) + \dots + X_nP(X_n)$$

Look at the oranges:

$$\begin{aligned} E(X) &= 0 \cdot \left(\frac{10}{56}\right) + 1 \cdot \left(\frac{30}{56}\right) + 2 \cdot \left(\frac{15}{56}\right) + 3 \cdot \left(\frac{1}{56}\right) \\ &= \frac{63}{56} = 1.125 \end{aligned}$$

We can show this on the histogram - this is where it would “balance”.

Look at the magazines,

$$\begin{aligned} E(X) &= 15 \cdot \left(\frac{5}{52}\right) + 16 \cdot \left(\frac{4}{52}\right) + 17 \cdot \left(\frac{8}{52}\right) + 18 \cdot \left(\frac{10}{52}\right) \\ &+ 19 \cdot \left(\frac{8}{52}\right) + 20 \cdot \left(\frac{15}{52}\right) + 21 \cdot \left(\frac{2}{52}\right) = 18.25 \end{aligned}$$

What is the “average” number of magazines sold?

One kind of average is the MEAN, \bar{X} ,

$$\begin{aligned} \bar{X} &= \frac{\textit{number of magazines}}{\textit{number of weeks}} \\ &= \frac{15 \cdot 5 + 16 \cdot 4 + \dots + 21 \cdot 2}{52} = 18.25 \end{aligned}$$

Same as the expected value! So we can let the calculator do the work.

There are two other kinds of “averages”.

The MEDIAN is the number in the middle if all the data is lined up in order.

The calculator will list this for you when you do the 1-Var Stats.

The other is the MODE.

This is the data that occurs most often. You must see the frequency distribution to find this.

A set of data may have no mode, one mode or more than one mode.

The mode will work for non-numerical data too.

ODDS

If the probability that event E occurs is $P(E)$, we say that the ODDS IN FAVOR OF E are

$$\frac{P(E)}{1 - P(E)} = \frac{P(E)}{P(E^c)}, P(E) \neq 1$$

The ratio is usually expressed as integers.

example - what are the odds in favor of rolling doubles with two fair dice?

E is the event that we roll doubles,

$$\text{so } P(E) = 6/36$$

E^c is the event that we don't roll doubles,

$$\text{so } P(E^c) = 30/36$$

The odds are then

$$\frac{P(E)}{P(E^c)} = \frac{6/36}{30/36} = \frac{6}{30} = \frac{1}{5}$$

So we would say the odds in favor of rolling a double are 1 to 5 or 1 : 5.

If we are given the odds, we often want to change this to probability.

If the odds in favor of an event are $a : b$, then the probability of the event is

$$P = \frac{a}{a + b}$$

example - we are given the odds in favor of a certain horse winning a race are 6 to 5. What is the probability that the horse will win?

$$P = \frac{6}{6 + 5} = \frac{6}{11} \approx .55$$

Variance and Standard Deviation

The variance and the standard deviation measure the spread in your data.

We can find these values for raw data, grouped data and probability distributions.

The work will be done by the calculator.

Example - You have a class of 10 students. The students take a 20 point quiz and the scores are:

20, 13, 15, 19, 18, 17, 14, 16, 16, 18

Find the mean, median, mode, standard deviation and variance of this data set.

Answer - Put the data values into a list on the calculator

We then find

mean = $\mu = 16.6$ (it is a population)

median = 16.5

standard deviation = $\sigma = 2.1071$

To find the mode, look what score appears the most often.

16 and 18 each appear twice, so this data is bimodal,

modes = 16 and 18

To find the variance, σ^2 , you can use the formula or square the standard deviation,

variance = $\sigma^2 = 2.1071^2 = 4.4400$

Example - Find the mean, median, mode, standard deviation and variance for the number of seeds in a grapefruit from a sample of 50 grapefruits:

number of seeds	frequency
3	5
4	8
5	15
6	12
7	10

the mean is $\bar{x} = 5.28$ (it is a SAMPLE)

the median is 5

the mode is 5

the standard deviation is $S = 1.2461$

Example - find the mean and standard deviation for the number of hearts in a hand of 6 cards. Draw the histogram for this distribution and show the mean and standard deviation on the histogram.

Start with a table for the probability.

We will have $P(E) = n(E)/n(S)$:

number of hearts	probability
0	$C(13, 0)C(39, 6)/C(52, 6) \approx .1602$
1	$C(13, 1)C(39, 5)/C(52, 6) \approx .3677$
2	$C(13, 2)C(39, 4)/C(52, 6) \approx .3151$
3	$C(13, 3)C(39, 3)/C(52, 6) \approx .1284$
4	$C(13, 4)C(39, 2)/C(52, 6) \approx .0260$
5	$C(13, 5)C(39, 1)/C(52, 6) \approx .0025$
6	$C(13, 6)C(39, 0)/C(52, 6) \approx .0001$

the mean = $\mu = 1.5002$ (a probability distributions is ALWAYS a population)

the standard deviation = $\sigma = 1.007$

Draw the histogram:

Chebychev's Theorem

We can ESTIMATE the probability that data lies within $k \cdot \sigma$ of the mean of a population with Chebychev's theorem:

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

Example - You have a probability distribution with a mean of 100 and a standard deviation of 5.

a) Estimate the probability that an outcome of the experiment will be within 5 standard deviations of the mean.

b) Estimate the probability of an outcome of an experiment is between 80 and 120.

c) Find a value c such that guarantees that the probability is at least 75% that an outcome of the experiment lies between $100 - c$ to $100 + c$.