Like peas in a pod

LS2201

Assignment 3

Sharon K

19MS015

14/02/21

Introduction

Variation may be defined as any difference between similar biological entities. These entities may be cells in the same organism, organisms in a species, groups of organisms, or, in this case, peas in a pod. These variations may be caused by genetic differences; that is, difference in the genetic compositions of the genomes of the entities we are comparing, or can be due to the effect of environmental factors.

"Like two peas in a pod" is a popular idiom used to express the very close resemblance of some things. We can check if this corresponds to reality by performing an experiment in which collect a number of pea pods and check if the peas inside them are homogeneous. Since pea is commonly available, and has easily measurable properties, it is a good model organism to study variation.

Theory

We propose the *null hypothesis* that

> All peas in a pod are identical.

The *alternate hypothesis* is that

> All peas in a pod are not identical.

Based on the data collected we may check if we are able to find significant variation within peas in a pod, in which case we can reject the null hypothesis.

Method

A number of pea pods were collected. Care was taken to ensure that the pods were not too distinct from each other, and that there were not any significant deformities. This mitigates the possibility of outliers in the data. Some properties that are straightforward to measure were chosen:

- Color
- Shape
- Texture
- Circumference

Once we have this data, for the discrete variables color, shape and texture, we can compute the *majority fraction* of each of the properties for each pod. This majority fraction is a measure of uniformity. If the majority fraction is 1 then all the peas in that pod are identical with respect to that character. As the majority fraction gets closer to 0, the more heterogeneous the peas in that pod are.

For the continuous variable, that is circumference, we can get a measure of the variability by calculating the standard deviation for each pod. The higher the standard deviation for a particular pod, the higher the variability in the circumference property.

Observations

The data for 15 peas are tabulated below.

Pod No.	No. of pods	Shape		Texture		Circumference (cm)	Color
1	6	5 1 0 0	□ ⬭ ⬠ ◯	6 0	S W	3 3.3 3.2 2.9 3.1 3	
2	6	2 2 2 0	□ ⬭ ⬠ ◯	4 2	S W	1.6 1.7 1.3 1.5 2 2	
3	4	1 1 0 2	□ ⬭ ⬠ ◯	2 2	S W	1.5 3 1.9 1.9	
4	6	3 3 0 0	□ ⬭ ⬠ ◯	6 0	S W	3 2 2.9 3.1 2.7 2.9	
5	6	1 2 1 2	□ ⬭ ⬠ ◯	5 1	S W	2 3 2.7 1 3 2.9	

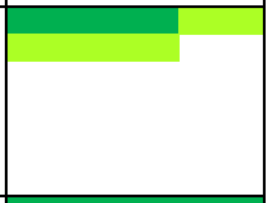
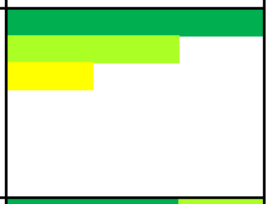
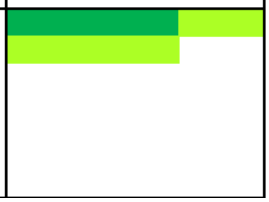
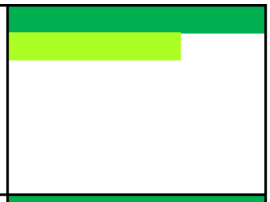
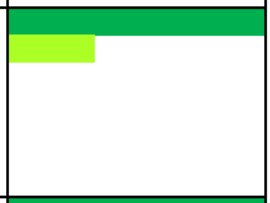
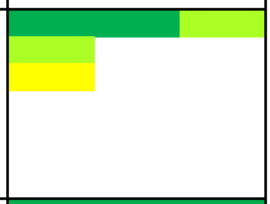
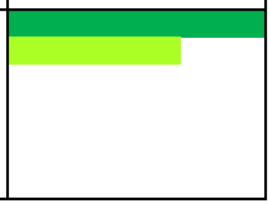
6	7	3 2 0 2	□ ⬭ ⬠ ◯	6 1	S W	3.2 3.5 3.1 3.9 3.2 3 3	
7	5	3 1 0 1	□ ⬭ ⬠ ◯	5 0	S W	3 3.5 3.1 3.5 3.6	
8	5	2 0 0 3	□ ⬭ ⬠ ◯	2 3	S W	3.2 2.6 2.5 3 3.8	
9	6	3 2 0 1	□ ⬭ ⬠ ◯	6 0	S W	3 4 4 3.5 3.8 3.5	
10	5	2 1 1 1	□ ⬭ ⬠ ◯	3 2	S W	3.5 2.5 3.3 3.6 3.5	
11	5	2 2 0 1	□ ⬭ ⬠ ◯	5 0	S W	4 4 3 3.9 2.9	
12	4	1 2 1 0	□ ⬭ ⬠ ◯	3 1	S W	4 3.6 4 3.6	
13	6	2 1 1 2	□ ⬭ ⬠ ◯	5 1	S W	2.1 3.1 2.6 1.2 2.9 2.9	
14	5	3 0 0 2	□ ⬭ ⬠ ◯	5 0	S W	3.1 3.4 3.1 3.6 3.5	
15	5	3 0 0 2	□ ⬭ ⬠ ◯	2 3	S W	3.1 2.7 2.6 3.1 3.6	

Table (1): Properties for each individual pea

The abbreviations used here is summarised in the following table.

Shape		Texture		Colour		Size	
Irregular	⬠	Smooth	S	Green		Continuous	
Flat on two sides	□	Wrinkled	W	Yellow Green			
Oval	⬭			Yellow			
Spherical	◯						

Table (2): Abrreviations and notation

The majority fractions calculated is displayed below.

Majority fraction					
Pod No.	Shape		Texture		Color
1	0.83	□	1	S	0.5 G
2	0.33	□	0.66	S	0.66 YG
3	0.5	◯	0.5	S	1 G
4	0.5	□	1	S	0.83 G
5	0.33	◯	0.83	S	0.5 G
6	0.42	□	0.85	S	0.71 G
7	0.6	□	1	S	0.6 YG
8	0.6	◯	0.6	W	0.8 G
9	0.5	□	1	S	0.5 G
10	0.4	□	0.6	S	0.6 YG
11	0.4	□	1	S	0.6 G
12	0.5	⬭	0.75	S	0.75 G
13	0.33	□	0.83	S	0.5 G
14	0.6	□	1	S	0.4 G
15	0.6	□	0.6	W	0.6 G

Table (3): Majority fractions of discrete variables

The higher values for the majority fraction, a measure of homogeneity, are indicated in a stronger shade of blue.

The histograms of majority fractions of the three discrete properties are displayed below, separately.

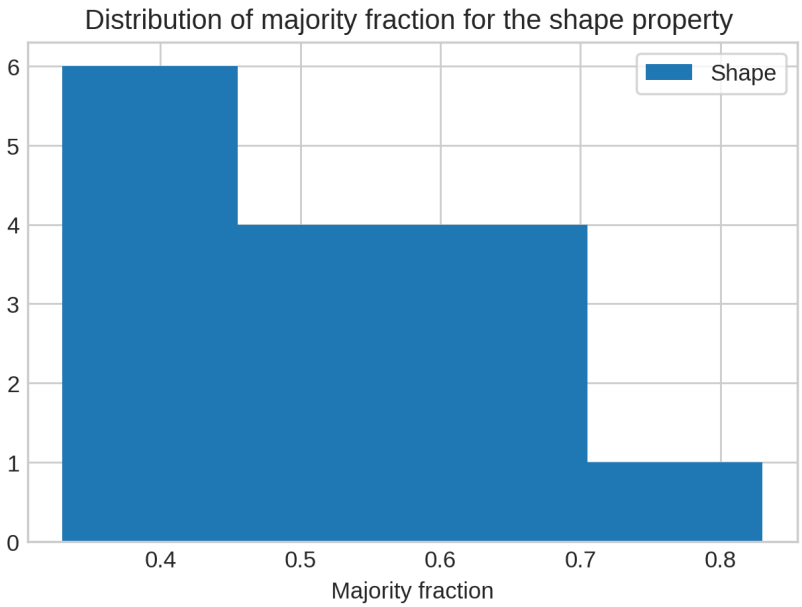


Figure (1): Majority fraction for shape property

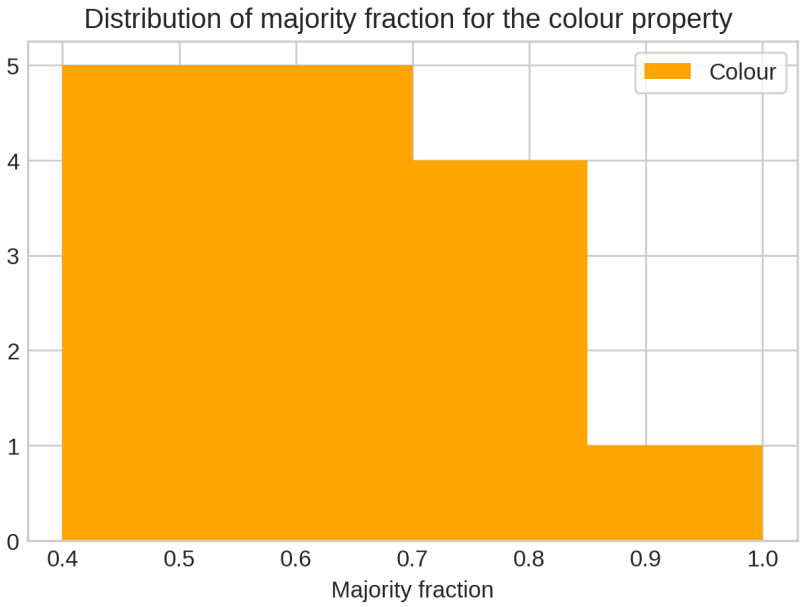


Figure (2): Majority fraction for colour property

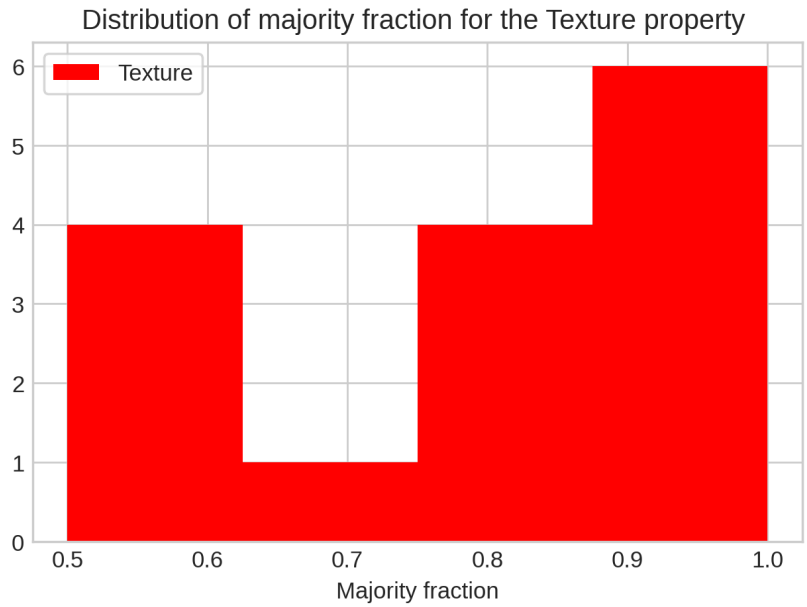


Figure (3): Majority fraction for texture property

The histogram for the standard deviation of each of the pods is given below. Here, the average of the standard deviations is marked by the dotted line. In contrast to the previous plots, in this case the higher the standard deviation, the more the variability within a pod.

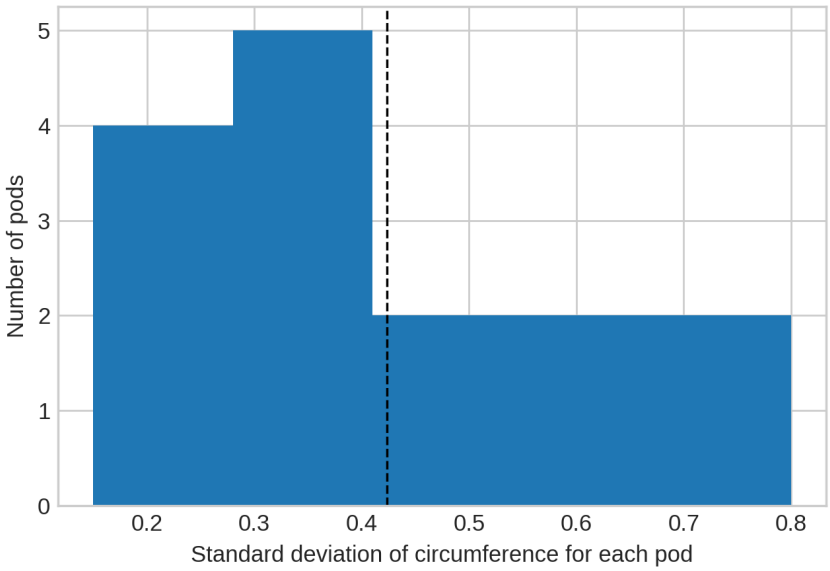


Figure (4): Standard deviation of size for each pod

Results

From table (3) and figure (3), it can be seen that **texture** has the lowest variability, as it has the largest number of majority fractions closest to 1. Texture is followed by **color** and then by **shape**, in homogeneity.

For the circumference property, the average of the standard deviation comes out to be 0.42cm. This can be interpreted as the average deviation between any two pods in the same pea. That is, most of the peas in the same pod are expected to lie between $\pm 0.42\text{cm}$ of the average pea size.

Significant variabiliy is observed for all pods for most properties. *None of the pods are perfectly homogeneous*. Hence, we reject the null hypothesis in favor of the alternate hypothesis, and state that **all peas in a pod are not identical**.

Limitations

Measuring the shapes, texture and color are fairly subjective. Different people may measure different values when in reality they are the same. In addition, none of the three are continuous variables and is hence difficult to do statistics on it. We can eliminate the subjectivity in measuring color using a camera to capture a picture and perform image analysis to extract the exact RGB colors of the pea. As for shape, we can take a picture and calculate its resemblance to a circle to obtain its sphericity, very roughly.

For the circumference property, the sample size is much too small to obtain any significant results. The average standard deviation of 0.42cm is a very rough estimate and is not likely to correspond to real measurements. This discrepancy can be mitigated using a much larger sample size.