J. Jedediah Smith

BIFX 551

Assignment 1

2/19/2022

1. Unify in a single Table the information from the tabs “Axis 1 measures” and “Axis 2 measures” using name as the ID (use the columns that contain same name formats).

Preview of unified table. See code document for full results.

A screenshot of a computer

Description automatically generated with low confidence

1. Find the statistical description and the variable correlation of the unified Table.

Preview of statistical description using summary(). See code document for full results. Average area is about 888,079 pixels. Average max Axis 1 length is about 1459 pixels, compared to the average max Axis 2 length which was 832.9 pixels. This seems to indicate that the average max Axis 1 length is approximately 40% larger than the max Axis 2 length.

A picture containing text, receipt

Description automatically generated

Preview of variable correlation using cor(). See code document for full results. Area seems to be more correlated with Axis 1 measurements than Axis 2, which makes sense given that Axis 1 max length is larger on average. The axis moments seem to be strongly correlated with their own ellipse and max length, while the std are only slightly negatively correlated with them.

Application, table, Excel

Description automatically generated

1. Plot the pairwise relationship from “Axis 1 measures” and “Axis 2 measures”. What you can conclude from this graph.

Initially I created the pairwise relationship from all 20 numeric variables. But this was understandably unreadable. The only thing I could conclude was that the Axis 1 lengths have a different distribution than the Axis 2. Which would make sense, since fly wings are longer than they are wide. This might also tie into the Axis 1 max being larger than Axis 1.

To get a more clear picture of what was going on, I cut out the moment and std measurements, then used just the area and length measurements. It appears clear that the ellipse and max lengths have similar linear distributions within their respective axis, though Axis 1 has a gap in the middle and Axis 2 are parallel. But when comparing across axis, they are more significantly different. This makes sense, given the previously established difference in length between each axis. The Axis 1 lengths appear to match a similarly linear distribution as the area, though there is a gap in the middle. This seems to match with our correlation plot results. The pairwise relationship plot using pairs() can be found below.

A picture containing diagram

Description automatically generated

1. Use bars to display all the variable distributions of “Experimental brood” tab.

I made bar charts for the experimental brood sex, genotype, and phenotype distributions. There are slightly more male than female. For genotypes, there are similar amounts of e-s, s-s, and a-e, with a-e have slightly more than the other two. The a-s has clearly the most representation. For phenotypes, elegans has the most, followed closely by aurora, and then silvana last.

Chart, bar chart

Description automatically generated

1. Use bars to display the sex, genotype, and area distributions of “Axis 1 measures” tab. Describe your process.

I set up bar charts to show the sex and genotype distributions. They seemed to largely match the brood distributions. There were slightly more male than female. For genotype, a-s took a clear lead, while e-s and s-s were equal, with a-e is just slightly above the two equal ones.

I set up a histogram for area since that will automatically bin the data. I split it up into just a handful of columns. I was aiming for 3, but when I was changing the code it would only ever display as 2 or 5. Since 2 didn’t really capture the distribution I wanted, so I went with 5.

Chart, bar chart

Description automatically generated