

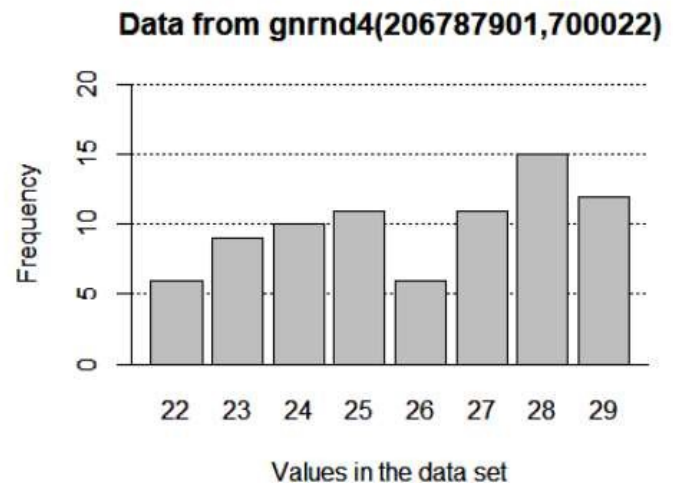
Topic 7j: Discussing Frequency Tables

Suppose we had the following data.

`gnrnd4(206787901, 700022)`

23	25	29	24	28	22	27	25	23	28	27	24	22	23	24	27	22	25	22	25	25	27	26	25
28	23	28	24	29	26	29	23	27	29	28	26	25	24	29	28	23	23	29	28	23	26	28	27
28	27	27	29	25	24	28	24	27	29	28	22	28	22	26	28	27	25	28	25	29	29	29	25
26	27	24	28	24	23	29	24																

Assuming that we generate the data using `gnrnd4(206787901, 700022)` to get those values into `L1`, then we know how to get a picture of the frequency of each value in the data by using the `barplot(table(L1))` function. Here is an example of such a picture, with a few enhancements.



Remember that we used `table(L1)` to get the frequency of each distinct value in the data. In fact, if we just use the command `table(L1)` we get the console output shown at the right.

```
> table(L1)
 22 23 24 25 26 27 28 29
  6  9 10 11  6 11 15 12
```

Conceptually, let us start by rewriting the output that we just saw.

Data value	22	23	24	25	26	27	28	29
Frequency	6	9	10	11	6	11	15	12

Then, we can compute the **relative frequency** for each value by taking the frequency and dividing it by the total number of data points, in this case 80. That gives us a new table.

Data value	22	23	24	25	26	27	28	29
Frequency	6	9	10	11	6	11	15	12
Relative Frequency	0.0750	0.1125	0.1250	0.1375	0.0750	0.1375	0.1875	0.1500

Another way to organize our view of the frequencies is to compute the **cumulative frequency**. That means that we find the "running" total of the frequencies as we read across the table.

Data value	22	23	24	25	26	27	28	29
Frequency	6	9	10	11	6	11	15	12
Relative Frequency	0.0750	0.1125	0.1250	0.1375	0.0750	0.1375	0.1875	0.1500
Cumulative frequency	6	15	25	36	42	53	68	80

At first this may seem to be just an exercise in arithmetic but it is actually quite useful. For example, it is now easy to see that there are indeed 80 items. Or, we can see that there are 53 items that have the value 27 or less. In fact, since we see that there are 36 items that are 25 or less, and knowing that there are 80 items, we see that there are $80 - 36$ or 44 items that are more than 25.

We expand the table again by adding a row for the **cumulative relative frequency** which is the same as the **relative cumulative frequency**. That is, the new row can be computed by getting the "running" total of the relative frequencies, or by dividing the cumulative frequencies by 80. We get the same value either way.

Data value	22	23	24	25	26	27	28	29
Frequency	6	9	10	11	6	11	15	12
Relative Frequency	0.0750	0.1125	0.1250	0.1375	0.0750	0.1375	0.1875	0.1500
Cumulative frequency	6	15	25	36	42	53	68	80
Cumulative relative frequency	0.0750	0.1875	0.3125	0.4500	0.5250	0.6625	0.8500	1.0000

As long as we are doing these computations we might as well add another easy one, namely, we can compute the number of degrees to allocate in a pie chart to each of the different data values. To do this we merely multiply the relative frequency by 360.

Data value	22	23	24	25	26	27	28	29
Frequency	6	9	10	11	6	11	15	12
Relative Frequency	0.0750	0.1125	0.1250	0.1375	0.0750	0.1375	0.1875	0.1500
Cumulative frequency	6	15	25	36	42	53	68	80
Cumulative relative frequency	0.0750	0.1875	0.3125	0.4500	0.5250	0.6625	0.8500	1.0000
Degrees in a pie chart	27.0	40.5	45.0	49.5	27.0	49.5	67.5	54.0

The table that we just developed is organized as rows for the different types of computations and columns for the different values in the data. In our current case we have 8 columns for the 8 different values in the data. You can see that this would be hard to display on a printed page if we had many more different values in the data. However, there was no requirement to organize the information in this way. Instead, we could arrange the table so that the columns held the different kinds of computations and the rows held the different values in the data.

Data value	Frequency	Relative Frequency	Cumulative frequency	Cumulative relative frequency	Degrees in a pie chart
22	6	0.0750	6	0.0750	27.0
23	9	0.1125	15	0.1875	40.5
24	10	0.1250	25	0.3125	45.0
25	11	0.1375	36	0.4500	49.5
26	6	0.0750	42	0.5250	27.0
27	11	0.1375	53	0.6625	49.5
28	15	0.1875	68	0.8500	67.5
29	12	0.1500	80	1.0000	54.0

This arrangement would have a fixed width and we could have many more different values in the data by just adding more rows.