

In this tutorial, we're going to learn about describing scatterplots. Now, you may remember this war describing from before. When we were talking about univariate data, that is one variable data when we were making histograms and dot plots and all of those other ones, we would talk about the shape, center, and spread of a distribution.

Now, here, there are two variables. And so on a scatterplot, it's a little hard to talk about the shape. I mean, in the center and spread, it's all very confusing to talk about if maybe the QB salary is very spread out, and the total salary is maybe not so spread out. It would be hard to talk about spread.

So we're actually not going to talk about those. We're actually going to describe in a scatterplot, form, direction, and strength. And we'll go through each of these individually.

So first, form. In the form, we look for a pattern. Is the pattern linear or do the data show a curve? Do they start low and then peak and then end low or do they start low and end high? And how do they curve when they do that or do they rise quickly and then tail off? There's a lot to look at.

In this particular situation, with quarterback salary and total team salary for the National League football teams, it appears to be fairly linear. It starts kind of low, ends pretty high. Are there clusters or gaps we should be aware of?

There's kind of a gap right here. This value out here may be an outlier, so all of that would be something we would take into account when we talk about form. In this case, it's fairly linear with a possible outlier way out here.

Direction. So in this case, the direction is how does the y-axis variable, that is total salary, how does that respond to as you move to the right on the x-axis variable, which is quarterback salary? So here, does it increase as you move to the right or does total payroll decrease as the quarterback salary increases?

In this case, total team payroll tends to increase as the quarterback salary increases. Now, that's not a hard and fast rule. For instance, this particular team had a very high team payroll, a much higher team payroll than this team, even though their quarterback salary was a lot lower.

So it's not a hard and fast rule. All we're saying is that what tends to be the case. And in this particular example, it tends to be positive. Notice, I'm using this sort of green oval over the data to see an overall tendency of what the data will do.

And then, finally strength. How well do the points follow that indicated form? How well do these points stack up on

a line?

Well, they seem to go up as they go to the right, and they don't seem to like sweep up, or go up quickly and then tail off. They do seem to go up fairly steadily, but there's quite a bit of variation in the y-coordinates. It's not like they're all sitting exactly on a line. So I would say this is a moderate linear association.

Again, this green oval makes its appearance. If this were a stronger association, this oval would be thinner. And if it was a weaker association, this oval would be wider. So let's take a look. So this is the 1970 and 1980 price of different seafood in cents per pound. So let's talk about the form, direction, and strength.

It looks like the form is fairly linear. This one is a little bit low for the line that we would look at for the rest of the data points. Also, it appears to have an outlier on the high side.

This is sea scallops that were expensive in 1970, and expensive in 1980 as well. So this one appears to be an outlier. And we'll describe in another tutorial exactly what to look for with outliers.

But the direction is positive, which means that as the 1970 price increases, so does the 1980 price. That's not surprising, because you would expect that the ones that are less expensive before would be less expensive in 1980, and the ones that were more expensive in '70 would be more expensive in '80. And then the strength, it's very strong. It's pretty predictable what is going to happen with these prices, based on they're very close to a line.

We can also look at these different forms. So this is a strongly non-linear form. This looks kind of like exponential growth. This is a sort of a weak linear, but it is in fact linear form. And you could have something like this, sort of like a cloud of points that has no association.

And again, different strengths. Imagine the oval that you could put over these. So this strong association, you could put a very long, thin oval over.

This moderate association, you could put a kind of a wider oval over it, but it would still be longer than it is wide. And over this weak association, the oval is almost more like a circle. And so the idea is, if you can encase the points in an oval, the stronger associations will have a longer, thinner oval.

And so to recap, rather than looking at shape, center, and spread, when we were looking at one variable data, in two variable data, we analyze form, direction, and strength. So linear or not linear? Are there unusual features, gaps, clusters, outliers?

And all that goes into form. How well do they follow that form? That's strength. And what happens as the x-axis variable increases?

Does the y-axis variable go up, down, or does it stay the same? Or is there really no association at all? And that's the direction of the association. And so we talked about form, strength, and direction of a scatterplot. Good luck, and we'll see you next time.