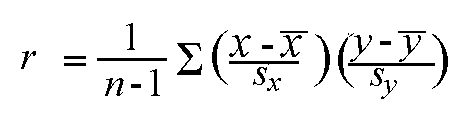
DSC 530 Week 7

**Option 5: What is Pearson’s correlation? What do the results indicate? What is Spearman’s rank correlation? How are these methods related?**

**What is Pearson’s correlation?**

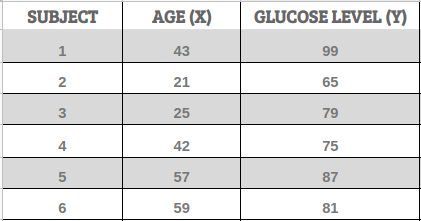
Pearson’s, also known as the Pearson correlation coefficient or Pearson product-moment correlation coefficient, is the standardized covariance r= cov(x,y)/sx sy. It is used to measure the correlation between two variables x and y. It can be expressed as:



It has a range of [-1,1] where 1 means two variables are positively correlated (as one increases the other one increases); -1 means they are negatively correlated (as one increases the other decreases); or 0 means there is no correlation.

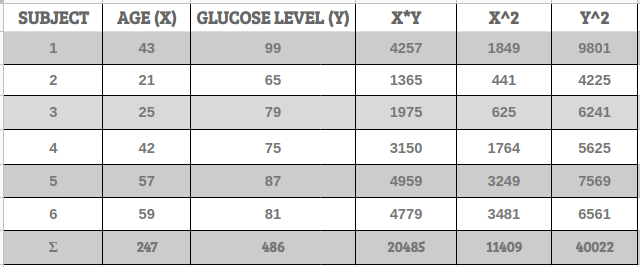
**What would this be used for in real life?**

Below is an example where we can attempt to find a relationship, if any, between glucose level and age in 6 subjects calculating Pearson Coefficient.



We’ll calculate the value of **r** using the formula mentioned above. For using that formula, we need to compute **Σ(X\*Y), Σ(X), Σ(Y), Σ(X²), Σ(Y²)**.

The table below shows the computed values of all the summations mentioned above.



From our table we get:

* Σ(X) = 247
* Σ(Y) = 486
* Σ(X\*Y) = 20,485
* Σ(X²) = 11,409
* Σ(Y²) = 40,022
* n is the sample size, in our case = 6

**r**= 6(20,485) — (247 × 486) / [√[[6(11,409) — (24⁷²)] × [6(40,022) — 48⁶²]]]

**r**= 0.5298.

The range of the correlation coefficient is from **-1** to **+1**. Our result is **0.5298**, which means the variables have a **moderate positive correlation**. This can also be interpreted as the variables age and glucose being related in a way that as age increases, glucose also increases.

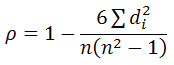
**What is Spearman’s rank correlation? How are these methods related?**

 The Spearman's rank-order correlation is the nonparametric version of the [Pearson product-moment correlation](https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php). Spearman's correlation coefficient, (ρ, also signified by *r*s) measures the strength and direction of association between two ranked variables.

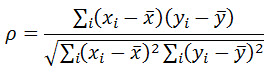
For this test, you need two variables that are either ordinal, interval or ratio. Although you would normally hope to use a Pearson product-moment correlation on interval or ratio data, the Spearman correlation can be used when the assumptions of the Pearson correlation are markedly violated. However, Spearman's correlation determines the strength and direction of the monotonic relationship between your two variables rather than the strength and direction of the linear relationship between your two variables, which is what Pearson's correlation determines. Monotonicity is "less restrictive" than that of a linear relationship. For example, the middle image above shows a relationship that is monotonic, but not linear.

A monotonic relationship is not strictly an assumption of Spearman's correlation. That is, you can run a Spearman's correlation on a non-monotonic relationship to determine if there is a monotonic component to the association. However, you would normally pick a measure of association, such as Spearman's correlation, that fits the pattern of the observed data. That is, if a scatterplot shows that the relationship between your two variables looks monotonic you would run a Spearman's correlation because this will then measure the strength and direction of this monotonic relationship. On the other hand, if, for example, the relationship appears linear (assessed via scatterplot) you would run a Pearson's correlation because this will measure the strength and direction of any linear relationship. You will not always be able to visually check whether you have a monotonic relationship, so in this case, you might run a Spearman's correlation anyway.

There are two methods to calculate Spearman's correlation depending on whether: (1) your data does not have tied ranks or (2) your data has tied ranks. The formula for when there are no tied ranks is:



where di = difference in paired ranks and *n* = number of cases. The formula to use when there are tied ranks is:



where *i* = paired score.

The Spearman correlation coefficient, rs, can take values from +1 to -1. A rs of +1 indicates a perfect association of ranks, a rs of zero indicates no association between ranks and a rs of -1 indicates a perfect negative association of ranks. The closer rs is to zero, the weaker the association between the ranks.

Wicklin R. 2017. *Fisher's transformation of the correlation coefficient.* SAS Blogs. <https://blogs.sas.com/content/iml/2017/09/20/fishers-transformation-correlation.html>

Field, A., Miles., J., Field, Z. (2012) Discovering Statistics Using R. Sage Publications, Ltd., Thousand Oakes, California. [pp. 209-211]

<https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php>