Math Symbols, Equations, and Examples

|  |  |  |
| --- | --- | --- |
| Name | Symbol | Names |
| Sample | Population |  |
| Size | **N** | **N** |  |
| Proportion | $$\hat{p}$$ | $$π$$ | p-hat, pi |
| Mean | $$\overbar{x}$$ | **µ** | x-bar, myu |
| Standard Deviation | **s** | **σ** | s, sigma |
| Median | **M** | **M** | M |
| Correlation Coefficient  | **r** | Ignore | are |
| Spearman rho | **ρ** | Ignore | rho |
| Quartiles | **Q1, Q2, Q3** | **Q1, Q2, Q3** |  |
| z-score | **zi** | **zi** | zee |

Other Terms:

 FNS Five Number Summary Min, Q1, Median, Q3, Max

 IQR InterQuartileRange Q3 – Q1

 Max Maximum value Max(xi)

 Min Minimum value Min(xi)

Note that all *equations for Sample and Population are the same* (symbols sometimes differ) *except for standard deviation*, and that only differs in the denominator.

**Size**

N is the number of Observational Units (OUs) in the population.

n is the number of OUs in the sample.

**Proportion**

**Sample Proportion**

 For n = n1+ n2, where Ni’s represent the number in a subgroup of the sample. (e.g. n1 might be the number of males, and n2­ the number of females in the sample). A proportion is ni/n. The proportion of n­1 is $\hat{p}=$ n1/n.

**Sample Proportion**

For N = N1+ N2, where Ni’s represent the number in a subgroup. The proportion of N1 is π=N1/N.

Also *see handout* on proportions if this is not sufficient.

**Mean** (Measure of Center)

For a data set {x1, x2,…, xn}

**Sample Mean**



**Population Mean**

$$µ=\frac{x\_{1}+x\_{2}+…+x\_{N}}{N}$$

**Standard Deviation** (Measure of Spread)

**Of a Sample**

 $s= \sqrt{\frac{\sum\_{i=1}^{n}\left(x\_{i}-\overbar{x}\right)^{2}}{n-1}}$

**of a Population**

 $σ= \sqrt{\frac{\sum\_{i=1}^{n}\left(x\_{i}-µ\right)^{2}}{N}} $

Without the sigma notation, equivalently:

 $s= \sqrt{\frac{\left(x\_{1}-\overbar{x}\right)^{2}+\left(x\_{2}-\overbar{x}\right)^{2}+…+\left(x\_{n}-\overbar{x}\right)^{2}}{n-1}} $

 $σ= \sqrt{\frac{\left(x\_{1}-µ\right)^{2}+\left(x\_{2}-µ\right)^{2}+…+\left(x\_{N}-µ\right)^{2}}{N}} $

Note that in this notation, it looks similar to the equations for the mean. Note that the n-1 exists to make sure that *s* is an *unbiased* estimator of σ. You don’t have to understand why, just the fact that *s* is unbiased is important.

**Z-Score**

Z-score is the *standardized* value of a data point.

**Sample** $z\_{i}=\frac{x\_{i}-\overbar{x}}{s}$

**Population** $z\_{i}=\frac{x\_{i}-µ}{σ}$

**Correlation Coefficient** (Measure of Association)

$$r=\frac{1}{n-1}\sum\_{i=1}^{n}\left(\frac{x\_{i}-\overbar{x}}{s\_{x}}\right)\left(\frac{y\_{i}-\overbar{y}}{s\_{y}}\right)$$

In terms of z-scores

$$r=\frac{1}{n-1}\sum\_{i=1}^{n}z\_{i\_{x}}z\_{i\_{y}}$$

So if you calculate the z-scores for a data set (or, equivalently, standardize the data) then the z-scores can be used to calculate the correlation coefficient, r. It’s fine if you just recall the latter equation.

Spearman rho, **ρ**, is a non-parametric correlation coefficient. Use it for *ordered, ranked,* or *non-linear* data.[[1]](#footnote-1)

**Median, Q1, Q3, Min, Max**

How to do these are in the notes. Min and Max are easy, as they are the maximum and minimum values in the dataset.

**EXAMPLES**

Two values are measured in an experiment with ten subjects. A dosage value, x, and a response value, y.

The data appears in the following table.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| OU# | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| X | 0 | 3 | 1 | 2 | 4 | 5 | 3 | 6 |
| Y | 1 | 2 | 0 | 2 | 4 | 10 | 4 | 10 |

1. Find the sample means for x and y:

$$\overbar{x}=\frac{x\_{1}+x\_{2}+…+x\_{n}}{n}= \frac{0+3+1+2+…+6}{8}= 3$$

$$\overbar{y}=\frac{y\_{1}+y\_{2}+…+y\_{n}}{n}= \frac{1+2+0+2+…+10}{8}= 4$$

1. Find the sample standard deviations for x and y:

(using the values for $\overbar{x}$ and $\overbar{y}$ previously calculated)

$s\_{x}= \sqrt{\frac{\sum\_{i=1}^{n}\left(x\_{i}-\overbar{x}\right)^{2}}{n-1}}=$ $\sqrt{\frac{\left(x\_{1}-\overbar{x}\right)^{2}+\left(x\_{2}-\overbar{x}\right)^{2}+…+\left(x\_{n}-\overbar{x}\right)^{2}}{n-1}}=\sqrt{\frac{\left(0-3\right)^{2}+\left(3-3\right)^{2}+…+\left(6-3\right)^{2}}{7}}$

$s\_{x}=\sqrt{\frac{28}{7}}$ = $\sqrt{4}$ = 2

Likewise, *s*y=4

1. Calculate z-scores for the x and y values:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| OU# | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| X | 0 | 3 | 1 | 2 | 4 | 5 | 3 | 6 |
| Zix | -1.5 | 0 | -1 | -0.5 | .5 | 0 | 1 | 1.5 |
| Y | 1 | 2 | 0 | 2 | 4 | 10 | 4 | 10 |
| Ziy | -1 | -0.5 | -1 | -0.5 | 0 | 1.5 | 0 | 1.5 |

We’ll do the first few for x, and using $\overbar{x}=3$ and $s\_{x}=2$, with $z\_{xi}=\frac{x\_{i}-\overbar{x}}{s}$

Where $\overbar{x}=3$, $z\_{ix}=0$ (because $x\_{i}-\overbar{x}$ = 3 - 3 = 0), so $z\_{2x}$ and $z\_{7x}$ are done. For

 $z\_{1x}$= (0-3)/2 = -3/2 = -1.5

 $z\_{3x}$= (1-3)/2 = 2/2 = 1

 $z\_{4x}$= (2-3)/2 = -1/2 = -0.5

And so on.

Do the same for y and the ziy’s, with $\overbar{y}=4$ and $s\_{y}=4$, and get the values for the last row.

1. Calculate the regression coefficient, r, for x and y.

Using the values for zix and ziy, and the equation for r:
$$r=\frac{1}{n-1}\sum\_{i=1}^{n}z\_{i\_{x}}z\_{i\_{y}}$$

$$r=\frac{1}{7}\left(z\_{1\_{x}}z\_{1\_{y}}+z\_{2\_{x}}z\_{2\_{y}}+z\_{3\_{x}}z\_{3\_{y}}++z\_{4\_{x}}z\_{4\_{y}}+…+z\_{n\_{x}}z\_{n\_{y}}\right)$$

$$r=\frac{1}{7}\left(\left(-1.5\right)\left(-1\right)+\left(0\right)\left(-0.5\right)+\left(-1\right)\left(-1\right)+\left(-0.5\right)\left(-0.5\right)+…+\left(1.5\right)\left(1.5\right)\right)$$

$$r=\frac{1}{7}\*6.5= 0.928571$$

That’s it.

Recall that the use of r is quite restrictive. This is discussed in the lecture notes.

I the data is non-linear or ranked or ordered, then a non-parametric measure of association such as Spearman’s rho must be used.

Sigma Operator

Symbols in math, like +, -, ÷, etc. are called *operators.* ∑ is the ‘sigma operator’

The sigma operator sums whatever is in the argument based on the indices, i, the lower bound (in this case 1) and n, which marks the upper bound. So, for data set {z1, z2, z3,…zn},

$\sum\_{i=1}^{n}z\_{i}$ means “sum all z’s starting from 1 and continue to n,” or

$$\sum\_{i=1}^{n}z\_{i}= z\_{1}+z\_{2}+z\_{3}+…+z\_{n}$$

Notice that the lower and upper bounds can be anything you choose, and the indice can also be used in the argument in a mathematic sense (not just and index):

$$\sum\_{i=2}^{4}z\_{i}= z\_{2}+z\_{3}+z\_{4}$$

 $$\sum\_{k=1}^{6}kz\_{k}= z\_{1}+2z\_{2}+3z\_{3}+4z\_{4}-5z\_{5}+6z\_{6}$$

 I replaced *i* with *k*, just to show you can index with any letter. *i* is a popular choice.

1. There are better non-parametric measures of association for non-linear data, but for now just Spearman’s rho will do. [↑](#footnote-ref-1)