UNIVERSITATEA DE STAT DE MEDICINĂ ŞI FARMACIE „NICOLAE TESTEMIŢANU”

CATEDRA MANAGEMENT ŞI PSIHOLOGIE

Adjusted Rate: Direct Method of Standardization. Correlation Link

Nicov Irina, asistent universitar

Globa Nina, asistent universitar

Adjusted Rate

- Crude rates can be used to make comparisons between two different populations only if the populations are homogeneous in all characteristics.
- Therefore if the populations are different by factors such as gender, age, etc. instead of crude rate must be used **adjusted rate** by gender, age, etc. for comparison making; otherwise comparison will not be valid.

Steps of calculation of adjusted rates

- 1. To compute the rates for each comparison group
- 2. To select the standard
- 3. To compute the expected number for each group
- 4. To calculate adjusted rates for each group

Example

1. Total Nr of workers in Factory "A"- 250

Males – 50 (number of diseases -1)

Females – 200 (number of diseases -10)

2. Total Nr of workers in Factory “B"- 200

Males – 170 (number of diseases -4)

Females – 30 (number of desiases -3)

Step1.

To *compute the sex specific viral hepatitis morbidity rates (VHMR) for each comparison group and total:*

- **Factory A (males) 1/50 *x 100 = 2%***
- **Factory A (females)=10/200 *x 100 = 5%***
- **Factory A (VHMR) =11/250 x 100 = 4.4%**

Step1.

To *compute the sex specific viral hepatitis morbidity rates (VHMR) for each comparison group and total:*

- **Factory B (males) 4/170** ***x 100 = 2,35%***
- **Factory B (females)=3/30** ***x 100 = 10%***
- **Factory B (VHMR) =7/200 x 100 = 3.5%**

Step 2

- **To select the standard distribution**

- **the standard for males 220 (50+170)**
- **the standard for female 230 (200+30)**
- **the total standard population 450 (220+230).**

Step3

- **To compute the expected number for each group**

It is necessary to find out how many sick individuals will be expected in the standard males population of 220, having the same sex-specific viral hepatitis morbidity rate as at factory 'A" is (2%):

100 - 2

220 - X;

- X= 220 x 2 / 100=4,4 .
- So, expected number of sick males workers for standard population of 220 is 4.4. and for females – 11.5
- Total expected number factory "A"= 4.4+11.5 = 15.9

Step4

- **To calculate adjusted rate for each group**
- The sex-adjusted viral hepatitis morbidity rate for each factory is calculated by dividing its total expected number of sick individuals by the total number of standard population:
- **Factory "A"=15.9 / 450 X 100% = 3.5%**
- **Factory “B" =28.06 / 450X 100% = 6.2%**

Interpretation

- **Specific rate for Factory A – 4.4 %**
- **Specific rate for Factory B – 3.5%**

- **Adjusted rate for Factory A – 3.5%**
- **Adjusted rate for Factory B – 6.2%**

Conclusions

- **Making comparison of sex-adjusted viral hepatitis morbidity rate: This is the opposite of what we observed when we looked at the specific rates, implying that this specific morbidity rates were indeed influenced by the gender structure of the underlying groups (Factory "A" and "B").**

Correlation

- **A lot of medical researches are related to relationship between two or more characteristics.**
- **For this kind of purpose is appropriate to use <u>correlation</u> that is able to examining the relationship between two variables.**

Definition

1. **Relationship, mutual relation between two or more things or phenomena; a relationship in which one of the terms can not exist without the other.**

2. Mutual dependence, relation of two phenomena or processes between which of there variations is a certain relation.

Types of correlations

- ***<u>Functional or mathematical</u>* correlations- are perfect, expressing the cause –to- effect link between fenomena(ex. Correlation between speed, time and distance)**
- ***<u>Statistical correlations</u>***

In the case of the statistical correlation, each numerical values of the variable X correspond not to one but more values of the variable Y, a statistical sum of this value, which is grouped around the average xY (ex. Correlation between height of persons and their weight - people of the same height may have different weight values that are grouped around their average)

Correlation Link

- By the shape (forma) can be :

1. <u>Linear</u>

In the case of linear correlation to uniform changes in the mean values of a variable, occur equal changes of another variable (eg changes in systolic and diastolic blood pressure).

2. <u>Nonlinear</u> (curvilinear)

For the nonlinear correlation to uniform changes of a value corresponds to the average values of another variable, which has a growth or decrease character.

Correlation Link

- **By the meaning designated can be :**

1. **direct (positive) - are established between phenomena that evolve in the same direction. For example, increasing the height of children increases their weight.**

2. reversed (negative) - are established between phenomena that evolve in the opposite direction. It raises a phenomenon and decreases the one with which it has a connection of dependence; or decreases a phenomenon and increases the one with which it correlates. For example, increasing of the children's age decreases their mortality

Direct statistical correlations

- **<u>Direct statistical correlations</u> - imply the evolution of correlated phenomena in the same sense, but not with the same unit of measure.**

For example: raising the living standards of a community's population twice, increases the body's resistance to disease, but not to the same measure because, besides living standards, the body's receptivity to disease is determined by other factors (biological, climatic, geographies, etc.).

Reverse statistical correlations

- **Reverse statistical correlations** imply the growth of a phenomenon and the decrease of the correlated one, but not to the same extent.

- For example, if in a community we have carried out a double number of anti-flu vaccinations, the number of people who will get seek the next year will certainly be lower but will not be twice as low as in the previous year

Pearson's Correlation Coefficient

- **Pearson's correlation coefficient is one measure of the relationship between two *numerical characteristics, symbolized by "X" and “Y".***
- ***The correlation coefficient is* denoted by *"r", it is calculated using the formula:***

Formula:

$$r_{xy} = \pm \frac{\Sigma\, d_x d_y}{\sqrt{\Sigma d_x^2 \Sigma d_y^2}}$$

r_{xy} - the correlation coefficient

$\Sigma d_x d_y$ - sum of multiplying between deviations from arithmetic mean of frequency values of the two phenomena (x and y) that correlate

Σdx^2- the sum of squares of deviations from the arithmetic mean of the frequencies of the phenomenon x

Σdy^2 - the sum of squares of deviations from the arithmetic mean of the frequencies of the phenomenon y

Example: Determine the correlation link between time past from access to acute pancreatitis (x) and the number of after surgery complications (y):

Nr	Time x	Number of complications, y	dx (X-9)	dy (Y-14,8)	dx^2	dy^2	dxdy
1	3	6	-6	-8,8	36	77,4	52,8
2	5	8	-4	-6,8	16	46,2	27,2
3	7	12	-2	-2,8	4	7,8	5,6
4	10	19	1	4,2	1	17,6	4,2
5	13	20	4	5,2	16	27	20,8
6	16	84	7	9,2	49	84	64,4
Total	**54**	**89**			122	260	175
Average	54/6= 9	89/6= 14,8					

$$r_{xy} = \pm\frac{\sum dxdy}{\sqrt{\sum dx^2 \sum dy^2}} = \pm\frac{175}{\sqrt{122\times260}} = \pm\frac{175}{178,3} = \pm 0,98$$

Interpretation

- **The maximum values that "r" can achieve is 1, and its minimum values is -1.**
- Therefore for any given data set: $-1 < r < 1$

Coefficient of correlation interpretation is based on the relation noted above:

- ***Positive correlation (+), when coefficient of correlation is 0 < r < 1 : "Y“*** tends to increase in magnitude as "X" increases;
- ***Negative correlation (-), when coefficient of correlation is -1< r < 0 : "Y“*** decreases as "X" increases;

There are following crude rule for *interpreting the size of correlation:*

- • The correlation coefficient equal ± 1 denote a <u>perfect</u> linear *relationship;*
- • Correlations from 0 to 0.25 (or(- 0.25) indicate *<u>little</u> or no relationship;*
- • Correlations from 0.25 to 0.50 (or - 0.25 to - 0.50) indicate *<u>a fair degree</u> of relationship;*
- • Correlations from 0.50 to 0. 75 (or- 0.50 to- 0. 75) indicate *<u>a good</u> relationship;*
- • Correlations greater than 0.75 (or - 0.75 to 0.99) indicate *a <u>strong</u> relationship.*

The coefficient of determination

- Sometimes the correlation is squared (r^2) to form a useful statistic called the *coefficient of determination or r-squared:*
- It is a statistical term that tells us how good one variable is at predicting another.
- <u>$r^2 = 1.0$</u> means given value of one variable can perfectly predict the value for other variable
- <u>$r^2 = 0$</u> means knowing either variable does not predict the other variable

The coefficient of determination

- **The higher r2 value means more correlation there is between two variables. Though coefficient of determination is denoted the common association of the factors that influence the two variables.**
- **In other words, the coefficient of determination indicates the part of total value dispersion of variable can be explained or justified by dispersion of the values the other variable. Sometimes coefficient of determination is presented in percent, being multiply by 100.**

- For the previous example, for "r" = 0.98, "r^2" is 0.96. Usually, the determination coefficient is multiplied by 100 and the expression is converted to percent of the dispersion (96%).
- Therefore, 96% of the variation in the number of complications in pancreatitis is determined by the change in time of access.

Spearman's Rank Correlation Coefficient

- **The Spearman's rank correlation is used when one or both of the relevant variables are ordinal (or one ordinal and one numerical) characteristics and when the numerical observations are skewed with extreme values.**

$$r_s = 1 - \frac{6\,\Sigma d^2}{n \times (n^2 - 1)}$$

1 = absolute value of correlation coefficient;

r_s = correlation coefficient Spearman;

6 = constant value ;

Σd^2 = the sum of differences between ranks;

n = number of pairs.

Example: For Six students in a group know: degrees for biostatistics that students' achieved during the year and degrees (marks) obtained at the biostatistics test:

Nr. student	Year rating, **x**	Mark of final test, **y**	Rank rating, **x**	Rank rating, **y**	Difference between ranks, **d**	Difference, raised to the square, $\mathbf{d^2}$
1	very weak	5	1	2	-1	1
2	weak	3	2	1	1	1
3	satisfactory	6	3	3	0	0
4	good	8	4	4	0	0
5	very good	9	5	5	0	0
6	exceptional	10	6	6	0	0

$$r_s = \pm 1 - \frac{6\sum d^2}{n*(n^2-1)} = \pm 1 - \frac{6*2}{6*(6^2-1)} = \pm 1 - \frac{12}{6(36-1)} = \pm 1 - 0,057 = \pm 0,94$$

Interpretation: Therefore, there is a strong correlation between the annual grade and the exam grade.

Interpretation

- The calculation of the Spearman rank correlation, symbolized as $\mathbf{r_s}$ involves rank ordering the values on each of the characteristics from lower to highest;
- As the Pearson's correlation coefficient, the Spearman rank correlation coefficient ranges in value from -1 to 1. Values of $\mathbf{r_s}$ close to the extremes indicate a high degree of correlation between X and Y; values near 0 imply a lack of linear association between the two variables.

Thank you