## Correlation and regression analysis

## Correlation\& association Analysis

- Correlations are relationships between two or more variables or sets of variables. In statistics the correlation is referred to Correlation Coefficient and denoted by $r$.
- Association: two variables are associated if one of them provides information about the likely value of another.


## Correlation

- A number between -1 and +1 is used to 'quantify' the correlation of the variables.
- +1 is the highest positive correlation between two variables: $r=+1,100 \%$ positive correlation.
- -1 is the highest negative correlation between two variables: $1=-1,100 \%$ negative correlation.


## Warning

If the $r$ value is 0.5 that does not mean the correlation between the two variable is $50 \%$. SO, $r$ does not indicated to the percent of correlation.

## Pearson's Product Moment Correlation

 CoefficientCorrelation coefficient is a standardised covariance. Covariance is a measure of joint variances of two variables.

- $\frac{1}{N-1} \times \frac{\sum(x-\text { mean })(y-\text { mean })}{\left(s_{\chi}\right)\left(s_{\gamma}\right)}$

How to calculate the Pearson's correlation using SPSS

1. First we must make sure that the data following the normal distribution or linearity.
2. If the data following the normal distribution, we will follow the next steps to calculate the $r$.

Analyze $\Rightarrow$ Correlate $\Rightarrow$ Bivariate $\Rightarrow$ move the variables to the variable box $\Rightarrow$ Pearson $\rightarrow$ OK

## Example

- Example
- Find the correlation between the two following variables:

| $X$ | $Y$ |
| :--- | :---: |
| 1 | 93 |
| 2 | 87 |
| 3 | 76 |
| 4 | 70 |
| 5 | 62 |
| 6 | 45 |
| 7 | 40 |
| 8 | 32 |
| 9 | 25 |
| 10 | 10 |

Find the Correlation between HB and PCV for the femals blood samples? As it available in Correlation 2 excell folder?

## How to report the results?

- There is a significant negative relationship between x and y variable $r=0 \ldots . ., p<0.00$


## Spearman's correlation

- Spearman's correlation coefficient is a non-Parametric statistic based on ranked data. It is useful to reduced the effect of outliers values.
- Manually, the first step to calculate the Spearman's correlation is the ranking the data (ascending or descending).
- Use the same steps as in the Pearson's coefficient, except you have to choose the Spearman instead of Pearson.



Types of data distribution


Population trend of Dunlin at Swansea Bay for 39 years

## Example

- the depth of a river does not progressively increase the further from the river bank.
- Width cm Depth cm

$50 \quad 10$
$150 \quad 28$
$200 \quad 42$
$250 \quad 59$
$300 \quad 51$
$350 \quad 73$
$400 \quad 85$
450104
50096


## Results Reporting

- Similar to the Pearson's correlation report:
- There is a significant (+ or -) relationship between x and y variable $r=0 \ldots ., p<0.00$


## Regression

- Regression or association, we mentioned the definition of association in the beginning of this lecture. And we said that:
- "the two variables are associated if one of them provides information about the likely value of another".
- We understand from this short statement that in association or regression relationship there are two type of variables: dependent and independent.


## The conditions of Using the Linear

## regression

1. Determine which variable is dependent and which one is independent.
2. The dependent variable must follow the linearity distributed.
3. The data scattered about the regression line.

## Equation

- The regression equation is:
- $y=a+b x$,
- where:
$y=$ Dependent variable, $x=$ Independent variable, $a=$ Intercept and $b=$ slope.

The intercept represents the estimated average value of $y$ when $x$ equals zero and the slope represents the estimated average change in y when x increases/decreases by one unit. Slope and intercept are derived using the least-square method.

## Steps

$>$ Testing normality
$>$ If the dependent variable follow the normal distribution then the linear regression model will be applied as follow:

Analyze $\rightarrow$ regression $\rightarrow$ linear $\rightarrow$ Move the dependent variable to its box $\longrightarrow$ Move the independent variable to its box $\rightarrow$ OK

## Example

- Let us work out an example for calculating b and a . Body weight of babies measured in different months is given in Table 10.1. Month is the independent variable ( x ) and the body weight is the dependent variable (y).

| Age(month) | Weight(kg) |
| :---: | :---: |
| x | y |
| 1 | 3.8 |
| 2 | 4.2 |
| 3 | 4.8 |
| 4 | 5.7 |
| 5 | 6.4 |
| 6 | 6.9 |
| 7 | 7.1 |
| 8 | 7.8 |
| 9 | 8.6 |
| 10 | 10.4 |

