



Contents lists available at ScienceDirect

Early Human Development

journal homepage: www.elsevier.com/locate/earlhumdev

WASP (write a scientific paper) using Excel – 4: Histograms

Victor Grech

Department of Paediatrics, Mater Dei Hospital, Malta

ARTICLE INFO

Keywords:

Software
Computers
Statistics
Biostatistics

ABSTRACT

Plotting data into graphs is a crucial step in data analysis as part of an initial descriptive statistics exercise since it gives the researcher an overview of the shape and nature of the data. Outlier values may also be identified, and these may be incorrect data, or true and important outliers. This paper explains how to access Microsoft Excel's Analysis Toolpak and provides some pointers for the utilisation of the histogram tool within the Toolpak.

1. Introduction

Karl Pearson FRS (1857–1936) was an English mathematician and biostatistician. He established mathematical statistics and founded the world's first university statistics department in London in 1911 [1]. His work is still used today in that it is of primary importance to understand the properties of data that has been collected and be able to objectively describe it in a universally comprehensible manner that adheres to accepted conventions. This allows the presentation of data and its submission as an abstract for an oral presentation or a poster at a meeting. If research is not published, it is almost as if it has not been done, so this work should eventually lead to a publication in a peer-reviewed journal.

2. Data conversion

Data can be shown pictorially as a graph, such as a histogram or a boxplot and whisker. If data is made available as a set of values separated by delimiters, such as commas, Excel can convert this data to individual values in cells (a form of tabulation). This is done by copying and pasting these values into a cell, and from the data tab, convert text to columns (choose Delimited – Commas options), so that:

Urinary concentration of lead in 16 rural children ($\mu\text{mol}/24\text{ h}$) [2].
0.2, 0.3, 0.6, 0.7, 0.8, 1.5, 1.7, 1.8, 1.9, 1.9, 2.0, 2.0, 2.1, 2.8, 3.1,
3.4

Becomes Table 1:

This row can then be easily converted into a column by selecting the values, copying and using Paste-Paste Special-Paste values and transpose options (Figs. 1 and 2), to obtain Table 2:

3. Excel analysis ToolPak

Excel statistical functions are available using the Analysis ToolPak. This may not be available in Excel by default in which case it may be easily activated, a one-off procedure. Click Data Analysis in the Analysis group on the Data tab. If the Data Analysis command is not available (at the far right), click File tab, Options, Add-Ins category. In the Manage box, select Excel Add-ins, click Go, check the Analysis ToolPak check box and click OK. The data analysis ToolPak is now available in the Data tab.

4. Histogram function

If a series of cut-off values are required, Excel has a histogram function that allows the apportioning of data using preset values as well as the creation of an actual histogram. The cut-off values are input into a column and this is referred to by Excel as a “bin”. In this case, data and bin are shown in Table 3 [2]:

The creation of a histogram is via Data Analysis, Histogram and choose options (Fig. 3). The Labels check box is ticked so as to inform Excel that the first row in each column is a header. Tick chart output. Output range specifies where the analysis and graph will appear. The output is a frequency distribution that is plotted as a histogram (Fig. 4). With some manipulation via the Chart Elements Button (which appears as a plus sign to the top right of a graph whenever a graph is selected), a more presentable histogram can be obtained (Fig. 5).

If more data is collected (urinary concentration of lead in $\mu\text{mol}/24\text{ h}$ [2]) as per Table 4 and plotted, a graph such as Fig. 6 will be obtained.

E-mail address: victor.e.grech@gov.mt.

<https://doi.org/10.1016/j.earlhumdev.2018.01.005>

0378-3782/ © 2018 Published by Elsevier B.V.

Table 1
Urinary lead concentration in rows ($\mu\text{mol}/24\text{ h}$).

0.2	0.3	0.6	0.7	0.8	1.5	1.7	1.8	1.9	1.9	2	2	2.1	2.8	3.1	3.4
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	---	---	-----	-----	-----	-----

Table 2
Urinary lead concentration in column ($\mu\text{mol}/24\text{ h}$).

0.2
0.3
0.6
0.7
0.8
1.5
1.7
1.8
1.9
1.9
2
2
2.1
2.8
3.1
3.4

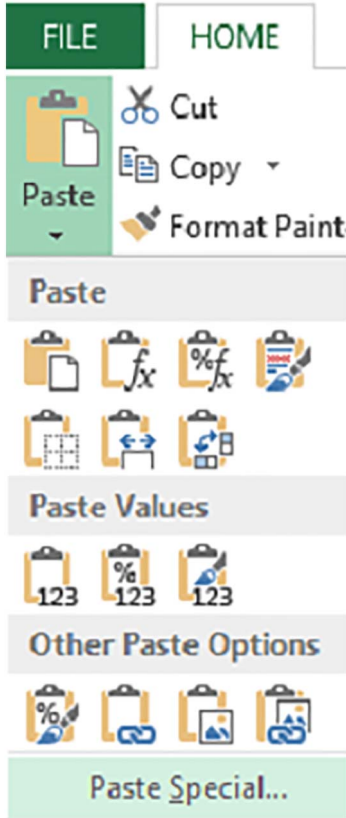


Fig. 1. Paste special options.

Table 3
Urinary lead concentration ($\mu\text{mol}/24\text{ h}$) and bin.

Data	Bin
0.2	0
0.3	0.4
0.6	0.8
0.7	1.2
0.8	1.6
1.5	2
1.7	2.4
1.8	2.8
1.9	3.2
1.9	3.6
2	2.4
2	2.8
2.1	3.2
2.8	3.6
3.1	4
3.4	4.4

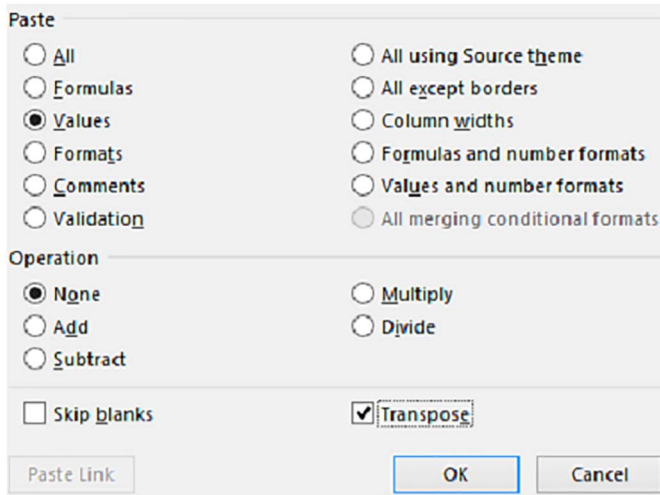


Fig. 2. Paste special options.

5. Values in categories – weighted means

Data may occasionally be available in categories already. In this case, a weighted mean must be calculated by taking the midpoint of the range for each range, multiplying by n for each range, adding the total and dividing that by the total number of observations, as shown in Table 5 (formulas used in calculations are shown at the bottom).

6. Other basic data descriptors

Once data collection is complete, the descriptive statistics exercise includes the acquisition of basic descriptors which may include standard deviation and quartiles. The next paper provides pointers as to how to do this in Microsoft Excel and describes the relationship between the two.

Download English Version:

<https://daneshyari.com/en/article/8777719>

Download Persian Version:

<https://daneshyari.com/article/8777719>

[Daneshyari.com](https://daneshyari.com)