**Simple Statistical Analysis Using Excel Assignment**
**GPH 719**

While Microsoft Excel is generally not the preferred method for completing statistical analyses in Public Health research, it is essential for data management and is useful for preliminary statistical analyses. For any public health job you will be expected to be proficient in the use of this program. For this assignment you will be using Excel to analyze the data shown in the table below. These data represent weight (in kilograms) of third grade students from a public elementary school in a rural town and an elementary school in an urban town.

| Weight of Children in Third Grade (kg) | | | |
|---|---|---|---|
| Rural School | | Urban School | |
| Girls | Boys | Girls | Boys |
| 26 | 19 | 39 | 25 |
| 28 | 21 | 35 | 29 |
| 19 | 26 | 41 | 33 |
| 25 | 27 | 30 | 48 |
| 32 | 28 | 26 | 40 |
| 22 | 23 | 24 | 41 |
| 29 | 25 | 26 | 32 |
| 30 | 22 | 36 | 28 |
| 28 | 24 | 30 | 29 |
| 35 | 19 | 27 | 23 |
| 21 | 26 | 26 | 30 |
| 36 | 28 | 30 | 28 |
| 18 | 20 | 38 | 25 |
| 25 | 25 | 33 | 38 |
| 31 | 18 | 29 | 36 |

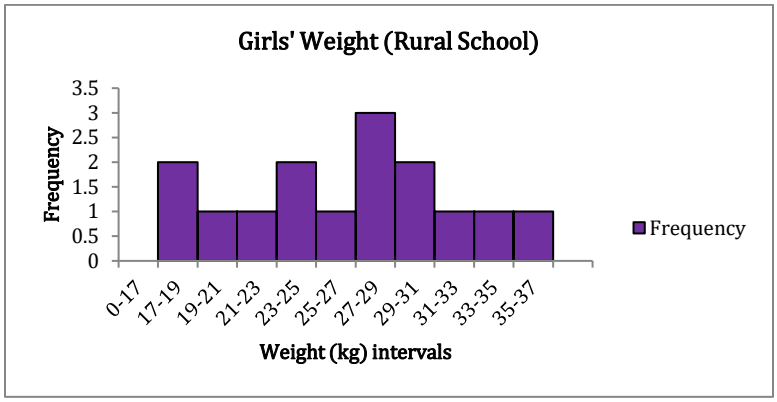
## Descriptive Statistics

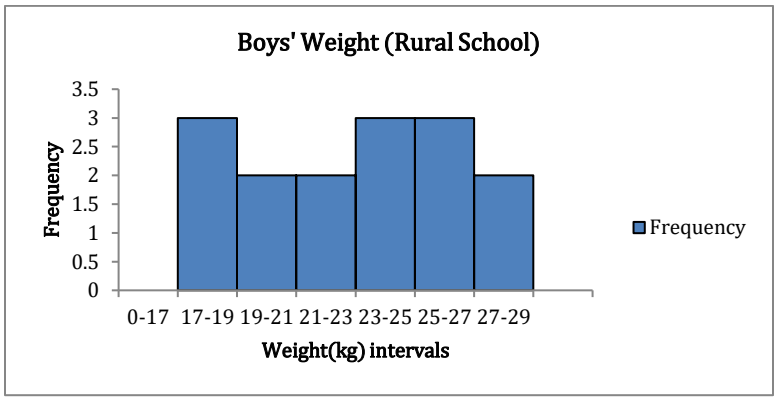
Create a table that includes Mean, Median, Mode, Range and Standard Deviation for each of the four student populations (you can use this Statistical Data Analysis in Excel video as a reference).

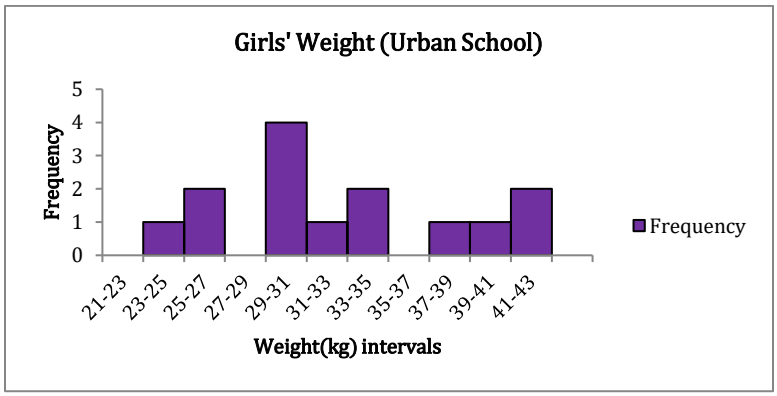
| | Weight of Children in Third Grade (kg) | | | |
|---|---|---|---|---|
| | Rural School | | Urban School | |
| | Girls | Boys | Girls | Boys |
| Mean | 27.0 | 23.4 | 31.3 | 32.3 |
| Median | 28.0 | 24.0 | 30.0 | 30.0 |
| Mode | 28.0 | 19.0 | 30.0 | 25.0 |

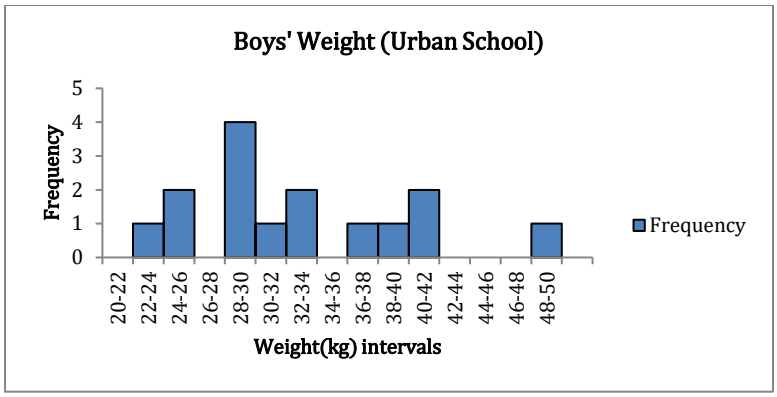
| | | | | |
|---|---|---|---|---|
| Min | 18.0 | 18.0 | 24.0 | 23.0 |
| Max | 36.0 | 28.0 | 41.0 | 48.0 |
| Range | 18.0 | 10.0 | 17.0 | 25.0 |
| Std | 5.3 | 3.3 | 5.2 | 6.8 |

Descriptive data is often represented visually, most commonly with a histogram. For this assignment you should create a histogram that represents each of the four data sets (you can use this Making a Histogram on Excel 2013 video as a reference.






**Inferential Statistics – comparing groups**

**t- tests**: Since you have already calculated the mean and standard deviation for each of your four populations, the simplest way to complete a t-test is by using an online t-test calculator.

For this particular t-test calculator you will:

1. Choose your data entry format – i.e. Enter mean, SD, and N
2. Choose your test (unpaired for paired t-test)
3. Enter your data
4. Click calculate

For this assignment you should create a table of P values for each of the following t-test analyses.

- Compare weights of boys from the urban school to the weights of boys from the rural school.
- Compare weights of girls from the urban school to the weights of girls from the rural school.
- Compare weights of boys from the urban school to the weights of girls from the urban school.
- Compare weights of boys from the rural school to the weights of girls from the rural school.

| Comparison (weight, kgs) | p-values |
|---|---|
| Boys (urban vs. rural) | < 0.0001 |
| Girls (urban vs. rural) | 0.033 |
| Urban (boys vs. girls) | 0.6544 |
| Rural (boys vs. girls) | 0.0337 |

How would you interpret these statistical analyses? What conclusions would you draw?

- At 5% level of significance, boys in urban school weigh significantly more than those in rural school.
- At 5% level of significance, girls in urban school weigh significantly more than those in rural school.
- At 5% level of significance, there is no meaningful weight difference between boys in urban school as compared to girls in urban school
- At 5% level of significance, girls in rural school weigh significantly more than boys in rural school.

**ANOVA:** If you want to compare the means of all four of your groups simultaneously you can complete an ANOVA analysis. Again, since you have already used Excel to calculate the mean and standard deviation you can use an online ANOVA calculator.

When you enter your data (Mean, SD, N) for each group and click calculate, the program will calculate the ANOVA test statistics, including the P value. This P value tells you if there are any statistical differences between the means of your groups. It is important to recognize that this analysis does not tell you which means are statistically different. When the results of an ANOVA analysis indicate statistical difference, you may want to do additional statistical analyses to find out which population means are different. That is beyond the scope of this assignment. For this assignment you should complete an ANOVA of the weights of the four populations from the data set of weights among third grade students and report the P value.

P = 0.000

How would you interpret these statistical analyses? What conclusions would you draw? There is a highly statistically significant difference in mean weights among at least one of the four groups.

**Inferential Statistics – associations between variables**

For this analysis you will look at the association between the number of hours students sleep each night and their grade point average.

**Pearson Correlation:**

Using Excel, calculate the correlation coefficient (r value) for the data in the table below. You can use this Using Excel to calculate a correlation coefficient || interpret relationship between variables video as a reference.

| Hours of Sleep | GPA |
|---|---|
| 9 | 3.8 |
| 7 | 2.9 |
| 8 | 3.5 |
| 9 | 3 |
| 6 | 2.6 |
| 8 | 3.7 |
| 8 | 3.9 |
| 7 | 3 |
| 9 | 3.8 |
| 8 | 2.9 |
| 6 | 2.6 |
| 7 | 2.4 |
| 8 | 2.9 |
| 10 | 3.2 |
| 6 | 3 |
| 8 | 3.7 |
| 9 | 3.5 |
| 7 | 3 |
| 9 | 3.6 |
| 7 | 2.2 |

Once you have calculated your correlation coefficient, you can draw some conclusions about your data – for example, you can determine if there is a positive or negative relationship between the variables, and if the relationship is weak, moderate or strong. However, you do not know if the relationship is statistically significant until you calculate significance (P value).

r = 0.62

What conclusions can you draw from your calculated correlation coefficient? There is a moderately strong positive relationship between hours of sleep and GPA. As hours of sleep increase, the GPA tends to increase.

Again, the simplest way to do that when you are working in Excel is to use an online calculator like the one at this site: https://www.socscistatistics.com/pvalues/pearsondistribution.aspx

In this case you will type in your r value, your sample size, and you're a priori established significance level. When you click calculate the statistical analyses will be completed and the P value will be provided.

Correlation Coefficient  P value:0.003545

How would you interpret these statistical analyses?  What conclusions would you draw? At 5% level of significance, the relationship between hours of sleep and GPA is statistically significant. It is highly unlikely that this is due to chance.