Data Analysis and Representation

Sheet 1

This table provides information on hospital episodes by month, and is typical of the format in which you would find 'raw' data downloaded from an online questionnaire service or similar.

1. Imagine that your end goal in analysing this data is to find the percentage of episodes each month that occur in each different type of unit (APC, A&E, and Outpatient). 'Clean up' the data sheet in a way that allows you to find this information.
2. Imagine that your end goal in analysing this data is to track the rise and fall of each different type of episode that occurs in A&E units month-by-month. 'Clean up' the data sheet in a way that allows you to find this information.

Sheet 2

This table provides the number of complaints brought forward from the previous year, and the number of new complaints brought, against each NHS Organisation in the UK.

1. Sum up the complaints brought forward from the previous year and the new complaints from this year to find the total number of complaints outstanding against each NHS Organisation. (Pink Cells)
2. Sort the data so that it is ordered by the NHS Region Name.
3. Tabulate the total complaints for each NHS Region outstanding against NHS Organisations, as well as the total complaints outstanding against the NHS UK-wide. (Blue Cells)
4. Use the total complaints to calculate the percentage of complaints per region. (Green Cells)
5. Create a pie chart showing the percentage of the total outstanding complaints by region. Why is a pie chart the most meaningful way to show this data? Is there another type of chart you would choose instead?

Sheet 3

This table provides the number of arrests in the United States, broken down by US State.

1. Find the per capita arrest rate for each US State. (Orange Cells)
    a. Find the mean US State per capita arrest rate. (Pink Cell)
    b. Find the median US State per capita arrest rate. (Yellow Cell)
    c. Find the mode US State per capita arrest rate. (Blue Cell)
    d. Which of these numbers do you think is the most meaningful? Which captures or provides the most information about the data? Why?
2. Create a bar chart showing the per capita arrest rate for each US State. Also create a line graph showing the per capita arrest rate for each US State.
    a. Which of these charts do you think provides the most meaningful visual representation of the data and its importance? Why?

Sheet 4

This table provides the occurrence of crime in the United States. Crimes are broken down by category, and by community type.

1. Find the crime rate per 100,000 inhabitants of each category of crime, within each community type. (Pink Cells)
2. Find the per capita crime rate for each category of crime, within each community type. (Blue Cells)
3. What is the difference between these two rates? What is the significance of this difference in terms of representing data?
4. Produce bar charts showing:
    a. The incidence of each category of crime in each community type.
    b. The crime rate per 100,000 inhabitants of each category of crime in each community type.
    c. The crime rate per capita of each category of crime in each community type.
5. Compare your three bar charts. Which do you think provides the most meaningful representation of the data and its importance? Why?

Sheet 5

This table provides the number of cancer cases in the United States from 2009-2013, broken down by the patient's age range, gender, race, and region of residence.

1. Fill in an "observed value" table and an "expected value" table in preparation for running a Chi Squared test comparing the actual and expected value of cases in each region among males and females.
    a. Remember to use formulas to save time when counting your different categories.
2. Run a Chi Squared test to determine whether there is a significant difference between what has been observed and what is expected.

Sheet 6

This table provides the number of "biscuits" eaten by a group of rats, both with and without receiving an appetite stimulant.

1. Is this data set suitable for an independent t-test or a paired t-test? Why?
2. Run your chosen test. Is there a difference in biscuits eaten by rats with or without the appetite stimulant? Is this difference significant? How do we know this?

Sheet 7

This table provides the exam scores for three groups of students. One group was given a memory-enhancing drug prior to studying for the exam. The second group was given a placebo drug. The third group was given nothing at all.

1. What type of ANOVA test should be run on this data set to determine whether there is a statistically significant difference between the groups: one-way, two-way with replication, or two way without replication?

2. Run your chosen test. Is there a difference between the test scores of the groups? Is this difference significant? How do we know this?

<u>Sheet 8</u>

This table provides the unemployment rate, inflation rate, and presidential approval rating for post-WWII United States.

1. Is there a correlation between the unemployment rate and the presidential approval rating? If so, is it a positive or a negative correlation? How strong is the correlation?
2. Can you run a regression analysis, to determine how much of the presidential approval rating is determined by the inflation rate? How confident are you in this relationship?