Assignment 1
Statistics and Samples

1. Identify whether the following variables are numerical or categorical. If numerical, state whether the variable is discrete or continuous. If categorical, state whether the categories have a natural order (ordinal) or not (nominal).

a. Number of sexual partners in a year
b. Petal area of rose flowers
c. Heart beats per minute of a Tour de France cyclist, averaged over the duration of the race
d. Birth weight
e. Stage of fruit ripeness (e.g., underripe, ripe, or overripe)
f. Angle of flower orientation relative to position of the sun
g. Tree species
h. Year of birth
i. Gender

ANSWER

1. **(a)** numerical, discrete. **(b)** numerical, continuous. **(c)** numerical, continuous. **(d)** numerical, continuous. **(e)** categorical, ordinal. **(f)** numerical, continuous. **(g)** categorical, nominal. **(h)** numerical, discrete. **(i)** categorical, nominal.

2. Not all telephone polls carried out to estimate voter or consumer preferences make calls to cell phones. One reason is that in the USA automated calls ("robocalls") to cell phones are not permitted, and interviews conducted by humans are more costly.

a. How might the strategy of leaving out cell phones affect the goal of obtaining a random sample of voters or consumers?
b. Which criterion of random sampling is most likely to be violated by the problems you identified in part (a): equal chance of being selected, or the independence of the selection of individuals?
c. Which attribute of estimated consumer preference is most affected by the problem you identified in (a): accuracy or precision?

ANSWER

2. **(a)** cell phone users may differ in important ways from people who use land lines. **(b)** equal chance of being selected. **(c)** accuracy.

3. In each of the following examples, indicate which variable is the explanatory variable and which is the response variable.

a. The anticoagulant warfarin is often used as a pesticide against house mice, *Mus musculus*. Some populations of the house mouse have acquired a mutation in the *vkorc1* gene from hybridizing with the Algerian mouse, *M. spretus* (Song et al. 2011). In the Algerian mice, this gene confers resistance to warfarin. In a hypothetical follow-up study, researchers collected a sample of house mice to determine whether presence of the *vkorc1* mutation is associated with warfarin resistance in house mice as well. They fed warfarin to all the mice in a sample and compared survival between the individuals possessing the mutation and those not possessing the mutation.

b. Cooley et al. (2009) randomly assigned either of two treatments, naturopathic care (diet counseling, breathing techniques, vitamins and a herbal medicine) or standardized psychotherapy (psychotherapy with breathing techniques and a placebo added), to 81 individuals having moderate to severe anxiety. Anxiety scores decreased an average of 57% in the naturopathic group and 31% in the psychotherapy group.

c. Individuals highly sensitive to rewards tend to experience more food cravings and are more likely to be overweight or develop eating disorders than other people. Beaver et al. (2006) recruited 14 healthy volunteers and scored their reward sensitivity using a questionnaire (they were asked to answer "yes" or "no" to questions like: "I'm always willing to try something new if I think it will be fun"). The subjects were then presented with images of appetizing foods (e.g., chocolate cake, pizza) while activity of their fronto–striatal–amygdala–midbrain was measured using functional MRI. Reward sensitivity of subjects was found to correlate with brain activity in response to the images.

d. Endostatin is a naturally occurring protein in humans and mice that inhibits the growth of blood vessels. O'Reilly et al. (1997) investigated its effects on growth of cancer tumors, whose growth and spread requires blood vessel proliferation. Mice having lung carcinoma tumors were randomly divided into groups that were treated with doses of either 0, 2.5, 10, and 20 mg/kg of endostatin injected once daily. They found that higher doses of endostatin led to inhibition of tumor growth.

ANSWER

3. **(a)** explanatory: mutation vs no mutation, response: survival.
(b) explanatory: treatment type, response: anxiety score.
(c) explanatory: reward sensitivity, response: brain activity.
(d) explanatory: endostatin dose, response: tumor growth.

4. For each of the studies presented in problem 3, indicate whether the study is an experimental or observational study.

ANSWER

4. **(a)** observational, because the researchers had no control over which subjects (mice) had the mutation and which did not, **(b)** experimental, **(c)** observational, **(d)** experimental

Displaying data

5. The Cambridge Study in Delinquent Development was undertaken in north London (UK) to investigate the links between criminal behavior in young men and the socioeconomic factors of their upbringing (Farrington 1994). A cohort of 395 boys was followed for about 20 years, starting at the age of eight or nine. All of the boys attended six schools located near the research office. The following table shows the total number of criminal convictions by the boys between the start and end of the study.

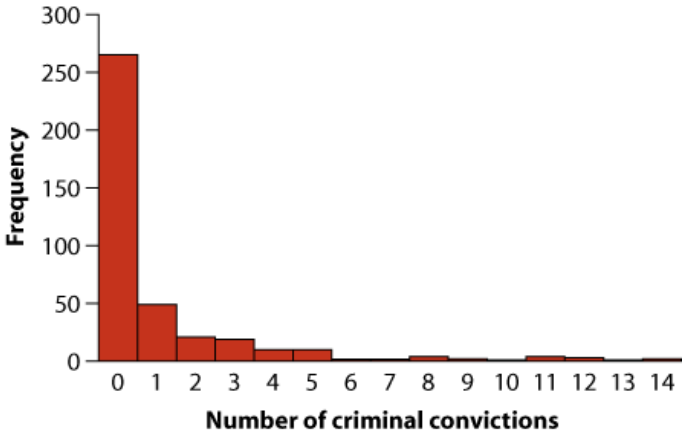
Number of convictions	Frequency
0	265
1	49
2	21
3	19
4	10
5	10
6	2
7	2
8	4
9	2
10	1
11	4
12	3
13	1
14	2
Total	395

a. What type of table is this?
b. How many variables are presented in this table?
c. How many boys had exactly two convictions by the end of the study?
d. What fraction of boys had no convictions?
e. Display the frequency distribution in a graph. Which type of graph is most appropriate? Why?
f. Describe the shape of the frequency distribution. Is it skewed or is it symmetric? Is it unimodal or bimodal? Where is the mode in number of criminal convictions? Are there outliers in the number of convictions?
g. Does the sample of boys used in this study represent a random sample of British boys? Why

or why not?

ANSWER

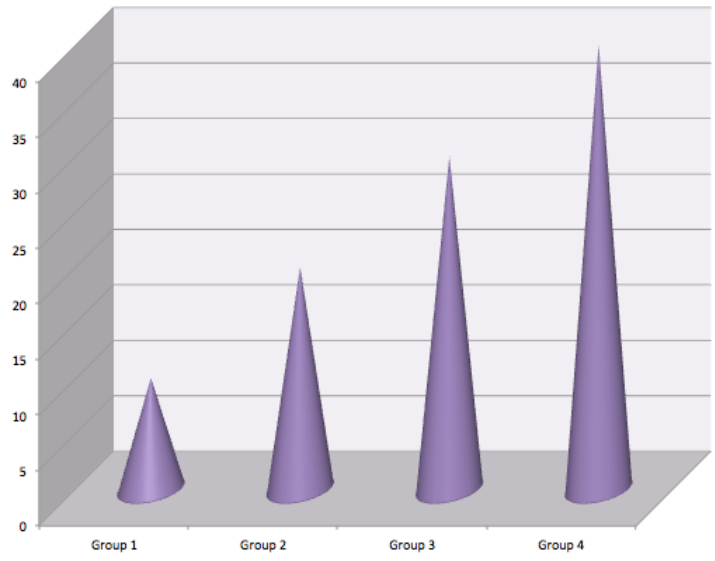
5. **(a)** Frequency table. **(b)** A single variable (number of convictions). **(c)** 21. **(d)** 265 of 395 (the fraction 0.67) had no convictions. **(e)** Histogram – it is the easiest way to visualize the frequency distribution for a numerical variable.



(f) Skewed (right) and unimodal (mode is 0 convictions). There are no outliers.

(g) The sample was six schools near the research office — not a random sample of British boys or any other population.

6. The following graph was drawn using a very popular spreadsheet program in an attempt to show the frequencies of observations in four hypothetical groups. Before reading further, estimate by eye the frequencies in each of the four groups.



a. Identify two features of this graph that cause it to violate the principle, “Make patterns in the data easy to see.”

b. Identify at least two other features of the graph that make it difficult to interpret.

c. The actual frequencies are 10, 20, 30, and 40. Draw a graph that overcomes the problems identified above.

ANSWER

6. **(a)** The 3D effects and the pointed cones make it difficult to judge bar height; cone-shaped “bars” distort magnitudes (cone area not proportional to magnitude).

(b) The graph fails to “draw graphical elements clearly”. Explanatory (horizontal axis) and response (vertical axis) variables are not labeled. The font is too small to read.

(c) A proper bar graph: