The Analysis of Biological Data - Whitlock and Schluter

Solutions to assignment problems - PLEASE DO NOT POST

Chapter 1

10. **(a)** Discrete **(b)** Continuous **(c)** Continuous **(d)** Discrete **(e)** Continuous

11. Observational study. The researcher has no control over which women have miscarriages and which lose their fetus from other causes.

12. **(a)** numerical, discrete (the variable, if not the partners)
    (b) numerical , continuous
    (c) categorical, ordinal
    (d) numerical, continuous
    (e) categorical, ordinal
    (f) numerical, continuous
    (g) categorical, nominal
    (h) numerical, discrete
    (i) categorical, nominal
    (j). numerical, continuous

13. **(a)** Observational study: the individual fish were not assigned to subspecies by the researcher.
    (b) Subspecies of fish and wavelength of maximum sensitivity.
    (c) The explanatory variable is the subspecies, the response variable is the wavelength of maximum retina sensitivity.

14. **(a)** No. The 500 households selected to receive the survey might be a random sample, but the low completion rate (< 20%) makes the sample a volunteer sample.

    (b) Volunteer bias Those who volunteer to respond to a survey on recycling might have different opinions of the program than those who did not respond.

15. **(a)** Omitting cell phones could bias a sample. If younger individuals are more likely to use a cell phone, omitting cell phones would bias the sample towards older individuals. **(b)** Equal chance of being selected.

16. **(a)** The population of interest is coastal Californian population of piñon pine trees.

    (b) A single plot was randomly sampled, but trees were not randomly sampled. The multiple trees within the same plot might not be independent, if they are related, of similar age, or share the same environment.

17. The 60 samples are not a random sample. The 6 dives measured on each bird are not independent. The six dive results measured from each bird are likely to be more similar than dive results obtained from six different birds sampled randomly from the population.

Chapter 2

14. **(a)** Between 12 and 13 mm.
    (b) Approximately 50% of the finches are at the modal beak width.
    (c) Changing the widths of intervals or "bins" of the histogram can alter its shape. Draw several histograms with the data, using wider and narrower intervals, is needed to determine whether a second peak is present. **(d)** Bimodal.

15. **(a)** Touching the first segment of the hind leg led to the greatest response. Touching the thorax or distal portions of any of the legs resulted in the lowest response.
    (b) Map.

16. **(a)** Frequency table.
    (b) A single variable (number of convictions).
    (c) 21.
    (d) 265 of 395 (the fraction 0.67) had no convictions.
    (e)
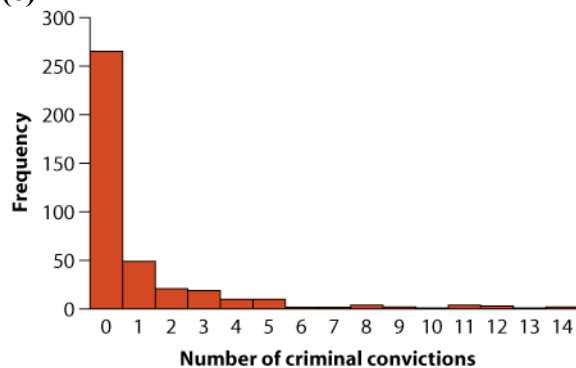
Histogram – it is the easiest way to visualize the frequency distribution for a numerical variable. (The cumulative frequency distribution is also an appropriate graph).
   **(f)** Skewed (right) and unimodal (mode is 0 convictions). There are no outliers.
    (g) The sample was six schools near the research office — not a random sample of British boys or any other population.

17. **(a)** This is a contingency table.
    (b)
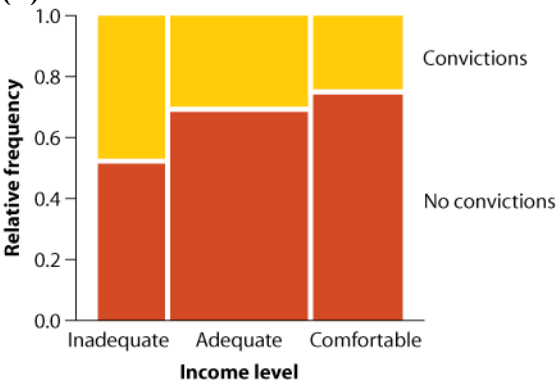
    (c) Categorical, ordered. Groups should be arranged by increasing income.
    (d) The relative frequency of conviction decrease as available income increases.

(e) The mosaic plot made it easier to see the pattern. Whereas the table gives the frequencies, the graph visualizes the association between the variables.

18. **(a)**
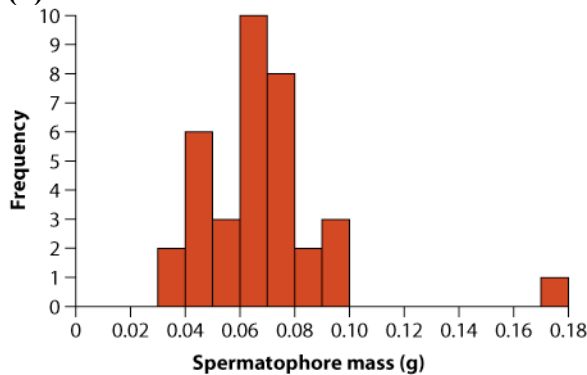
(b) Histogram: it visualizes the frequencies of each spermatophore mass interval very clearly.

(c) The main part of the distribution is fairly symmetric with a mode of 0.06–0.07. There is an extreme measurement at large spermatophore mass.

(d) Outlier.

19. **(a)** Both variables are continuous numeric variables.

(b) Scatter plot.

(c) The relationship is positive but non-linear. As temperature increases the fusion frequency increases.

(d) The 20 measurements are not a random sample because each fish was measured several times and the multiple measurements were all combined.

20. **(a)** Line graph.

(b) The steepness of each segment tells us the net increase in the number of endangered species added in a given year (it is not exactly the total number added, because some species might have been removed from the list in a given year).

(c) The net number of endangered species has been increasing steadily over time, but has tapered off toward the most recent dates.

21. **(a)** Histogram.

(b) The bars of a histogram should not have gaps between them. (A lesser problem is that it is not clear what the ticks on the x-axis refer to.)

(c) The variation in protein similarity is the most interesting feature: some proteins are nearly identical between humans and puffer fish, while others are nearly completely dissimilar.

(d) Skewed left.

(e) The mode is 70% similarity (presumably the interval the number "70" represents is 67.5–72.5).

22. **(a)** Cumulative frequency distribution.

(b) The y-axis indicates the quantile of the variable indicated on the x-axis (annual percent change in human population). The quantile is the fraction of observations less than or equal to the value on the x-axis.

(c) Approximately 10% of the countries had negative change in population size.

(d) The 0.10 quantile is approximately 0 growth, the 0.50 quantile is 1.5% growth, and the 0.90 quantile is 3% growth.

(e) The 60th percentile is approximately 1.75% growth.

23. **(a)** Scatter plot.
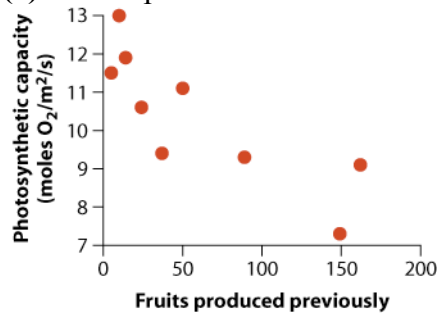
(b) Number of fruits previously produced, because we wish to use it to predict photosynthetic capacity.

(c) Negative association: photosynthetic capacity reduced in trees that produced many fruits previously.

24. **(a)** Grouped histograms. Explanatory variable: genotype at PTC gene. Response variable: Taste sensitivity score. Genotype is categorical variable, taste sensitivity is numerical. **(b)**. Scatter plot. Explanatory variable: migratory activity of parents. Response variable: migratory activity of offspring. Both variables are numerical.

(c) Grouped cumulative frequency distributions. Explanatory variable: year of study. Response variable: density of fine roots. Root density is numerical. While year is a numerical variable, strictly speaking, it is used as a categorical variable in this figure to define the three groups of measurements.

(c) Grouped bar graph. Explanatory variable: HIV status. Response variable: needle sharing. Both variables are categorical.

25. **(a)** Percentage of adults with BMI greater than 25 increased steadily from 1995 to 2002 then it dropped slightly and became steady after 2002.

(b) While cute, the figure does not help the eye visualize the association between year and the percentage of adults with BMI greater than 25.

(c) Line graph.
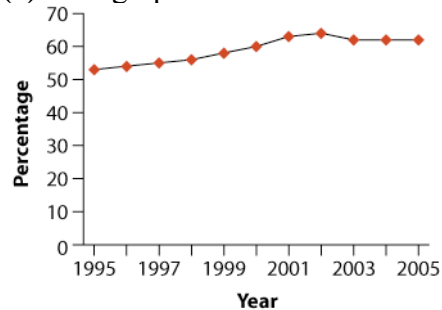
26. **(a)** Contingency table.

	No sneaker	One sneaker	Two or more sneakers	Total
Eggs eaten	61	18	16	95
No eggs eaten	389	17	4	410
Total	450	35	20	505