MATH1001 Worksheet III-1 (solution)

1.3.2 (d)

"To study the size distribution of rock cod (Epinephelus puscus) off the coast of southeastern Australia, the lengths and weights were recorded for all cod captured by a commercial fishing vessel on one day (using standard hook-and-line fishing methods)."

Identify the source(s) of sampling bias and describe

(i) how it might affect the study conclusions and

(ii) how you might alter the sampling method to avoid the bias.

Fish caught by a single vessel on one day are **not a random sample**.

(i) If the vessel is in a region that has not been fished recently and thus contains large fish, for example, then the sample average will be too large.

(ii) To avoid this bias, use randomly chosen fishing vessels on randomly chosen days.

2.4.1

Here are the data from Exercise 2.3.10 on the number of virus-resistant bacteria in each of 10 aliquots:
14 15 13 21 15 14 26 16 20 13

(a) Determine the quartiles.

(b) Determine the interquartile range.

(c) How large would an observation in this data set have to be in order to be an outlier?

(a) Putting the data in order, we have 13 13 14 14 15 15 16 20 21 26

The lower half of the distribution is 13 13 14 14 15. Thus, Q1 = 14.

Likewise, the upper half of the distribution is 15 16 20 21 26. Thus, Q3 = 20.

(b) IQR = Q3 - Q1 = 20 - 14 = 6

(c) To be an outlier at the upper end of the distribution, an observation would have to be larger than

Q3 + 1.5 x IQR = 20 + 1.5 x 6 = 20 + 9 = 29, which is the upper fence.

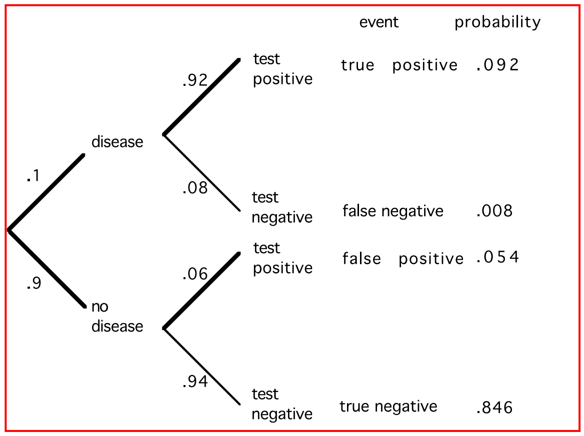
3.2.8

Suppose that a medical test has a 92% chance of detecting a disease if the person has it (i.e., 92% sensitivity) and a 94% chance of correctly indicating that the disease is absent if the person really does not have the disease (i.e., 94% specificity). Suppose 10% of the population has the disease.

(a) What is the probability that a randomly chosen person will test positive?

(b) Suppose that a randomly chosen person does test positive. What is the probability that this person really has the disease?

(a) There are two ways to test positive. A true positive happens with probability (0.1)(0.92) = 0.092. A false positive happens with probability (0.9)(0.06) = 0.054.



Thus,

Pr{test positive}

=0 .092 + 0.054

= 0.146.

(b) Pr{have disease given test positive}

= 0.092/0.146

= 0.63.

3.5.9-10

A group of college students were surveyed to learn how many times they had visited a dentist in the previous year. The probability distribution for Y, the number of visits, is given by the following table:

Y (No. Visits)	Probability
0	0.15
1	0.50
2	0.35
Total	1.00

Calculate the mean and the standard deviation of the number of visits Y.

(0)(0.15) + (1)(0.50) + (2)(0.35) = 1.2

VAR(Y) = $(0 - 1.2)^2(0.15) + (1 - 1.2)^2\ (0.50) + (2 - 1.2)^2\ (0.35) = 0.46$.

Thus, the standard deviation is $\sqrt{0.46} = 0.678$.

3.6.3

In the United States, 44% of the population has type A blood. Consider taking a sample of size 4. Let Y denote the number of persons in the sample with type A blood. Find

(a) $\Pr\{Y = 0\}$.

(b) $\Pr\{Y = 1\}$.

(c) $\Pr\{Y = 2\}$.

(d) $\Pr\{0 \le Y \le 2\}$.

(e) $\Pr\{0 < Y \le 2\}$.

(a) $0.56^4 = 0.0983$

(b) ${}_4C_1 (0.44^1)(0.56^3) = 4 (0.44^1)(0.56^3) = 0.3091$

(c) ${}_4C_2 (0.44^2)(0.56^2) = 6 (0.44^2)(0.56^2) = 0.3643$

(d) $0.0983 + 0.3091 + 0.3643 = 0.7717$

(e) $0.3091 + 0.3643 = 0.6734$