Statistics Solutions

Measures, Probability, Permutations and Combinations

1. Consider the following dataset: 2, 0, 5, 0, 4, 8, 2, 1, 1, 7.
   (a) Calculate the mean, median and mode of this set.
   The mean is 3, the median is 2. There are three modes in this set: 0, 1 and 2.
   (b) Is the dataset symmatrical? If not, is this set skewed to the left or to the right?
   Because the mean is greater than the median, the dataset is not symmetrical, but skewed to the right.
   (c) Calculate the range of the dataset.
   The range of the dataset is 8-0 = 8.
2. Consider the following dataset: 3, 5, 6, 1, 6, 5, 8, 4, 7.
   (a) Calculate the mean, median and mode of this set.
   The mean as well as the median of this dataset is 5. This set has two modes: 5 and 6.
   (b) Calculate the variance and the standard deviation.
   The variance = 4, the standard deviation is 2.
   (c) Verify that at least 75% of the measurements differ from the mean less than twice the standard deviation (Tchebichevs Rule).
   This means that the measurements should lie between 1 and 9. The only element that does not follow this requirement is 1. The percentage lying within the interval is therefore 8/9 * 100% = 88,9%. The rule therefore holds $(88,9 > 75)$.
   (d) Calculate the range of the dataset.
   The range of this set is 8-1 = 7.
   (e) Calculate the standard score of the 100th percentile.
   The 100th percentile is the last (highest) measurement = 8. The standard score then is $(8 - 5)/2 = 1.5$.
3. Consider the following eight letters: a, c, f, g, i, t, x, w.
   (a) How many permutations are there of the eight letters?
   Since $n = k$, this is a permutation of 8 elements. Therefore, $8! = 40320$.
   (b) Of the permutations in part (a) how many start with the letter t?
   In this case, the first letter is fixed so we can ignore it. Remaining is the permutation of 7 letters. Therefore $7! = 5040$.
   (c) Of the permutations in part (a) how many start with the letter g and end with the letter c?
   In this case, the first and last letters are fixed, so we only need to calculate the number for permutations for the remaining 6 letters: $6! = 720$.
4. A computer science professor has seven different programming books on a bookshelf, three of them dealing with C++ and the other four with Prolog. In how many ways can the books be arranged on the shelf if
   (a) there are no restrictions,
   With no restrictions we have 7 books to arrange in a row so there are 7! permutations of the books = 5040.

(b) if the languages must alternate,
If the languages must alternate then we must have PCPCPCP so the only variable is the ordering of the 3 C++ books in their slots and the ordering of the 4 Prolog books in their slots. Thus the number of orderings of the books is $3! * 4! = 144$.

(c) if all the C++ books must be next to each other, and
If all the C++ books are together then we treat the 3 C++ books as a unit and, overall, the number of arrangements is 5!. But the 3 C++ books can be arranged in 3! ways so the total number of arrangements is $5! * 3! = 720$.

(d) if all the C++ books must be next to each other and all the Prolog books must be next to each other?
If all the C++ books must be next to each other and all the Prolog books next to each other then there are two blocks of books to arrange. That can be done in 2 ways. Then the arrangements of the two sets of book within the blocks can be done in 4! and 3! ways. So the total number of arrangements is $2 * 4! * 3! = 288$.

5. How many ordered words of 3 different letters can be created from the set {m, r, a, f, t}, including meaningless words?
This is a variation where $n = 5$ and $k = 3$, therefore: 5*4*3 = 60 possibilities.

6. In a certain population, 30% of the persons smoke and 8% have a certain type of heart disease. Moreover, 12% of the persons who smoke have the disease. Use Bayes' theorem to compute:

   (a) the percentage of the population that smoke and have the disease
   For a person chosen at random from the population, let S denote the event that the person smokes and D the event that the person has the disease. Then $P(D \wedge S) = P(D|S) * P(S) = 0.12 * 0.3 = 0.036$

   (b) the percentage of the population with the disease that also smoke
   $P(S|D) = P(S \wedge D)/P(D) = 0.036/0.08 = 0.45$

7. Two coins are flipped simultaneously. One has a probability of heads equal to 0.6 and the other has a probability of heads equal to 0.7. What is the probability that the coin flip will be both heads or both tails?
$P((Head, Head) or (Tail, Tail)) = 0.6 \times 0.7 + (1 - 0.6)(1 - 0.7) = 0.42 + 0.12 = 0.54$

8. For a very large group of students, the probability of receiving the mark 10 in the course 'Mathematical Techniques in Computer Science' was 30 percent ($P(MTCS)$ is 0.3), and the probability of a 10 in the course 'Elementary Maths for Game and Media Technology' was 20 percent ($P(EMGMT)$ is 0.2). Also 15 percent of the students had 10 in both classes ($P(MTCS$ and $EMGMT)$ is 0.15). All students followed both courses.

   (a) Build a table with columns $P(MTCS)$ and $P(\neg MTCS)$ and rows $P(EMGMT)$ and $P(\neg EMGMT)$, and fill the table elements with $P(MTCS \wedge EMGMT)$, $P(MTCS \wedge \neg EMGMT)$ etc.

   (b) What is the probability that a randomly selected student got exactly one 10?
   $0.15 + 0.05 = 0.2$

   (c) What is the probability that a student got no 10?
   0.65

   (d) If someone got a 10 in EMGMT, what is the probability that she or he got a 10 in MTCS?
   $P(EMGMT \text{and } MTCS) = P(EMGMT)P(MTCS|EMGMT)$, so $P(MTCS|EMGMT) = 0.15/0.2 = 0.75$.

   (e) Does the data precisely agree with the assumption that the event of getting a 10 in MTCS is independent of getting a 10 in EMGMT?
   No: $P(MTCS)P(EMGMT) = 0.3 \times 0.2 = 0.06 \neq 0.15$.