ILLAWARRA TUTORING

This worksheet focuses on determining the variance and standard deviation of a distribution in order to measure how spread out the outcomes are. Work through the questions carefully and show all your working.

Easy Questions

1. Calculate the variance and standard deviation of the data set $2, 4, 4, 4, 6, 8$.
2. Given the frequency table below, calculate the variance and standard deviation.

x	Frequency
1	2
3	3
5	5

3. Use the formula $\sigma^2 = \dfrac{1}{n}\sum(x-\mu)^2$ to compute the variance of the data set $3, 7, 7, 19$.
4. Explain why the variance and standard deviation of the set $5, 5, 5, 5$ are zero.
5. Describe in your own words what the variance and standard deviation tell us about a data distribution.

Intermediate Questions

6. For the distribution with outcomes and frequencies given below, compute the mean, variance and standard deviation:

x	Frequency
2	4
4	6
7	2

7. Use the shortcut formula $\sigma^2 = \dfrac{1}{n}\sum x^2 - \mu^2$ to compute the variance for the data set $1, 3, 5, 7, 9$.
8. A discrete random variable X has the following probability mass function:

x	$P(X = x)$
0	0.1
2	0.3
4	0.4
6	0.2

Determine the mean, variance and standard deviation of X.

9. Show that rearranging the order of data in the set $8, 3, 5, 7$ does not affect its variance and standard deviation by calculating these measures before and after rearrangement.

10. Given a data set $a_1, a_2, \ldots, a_n$ with variance σ^2, explain and demonstrate mathematically how the variance changes if a constant c is added to every term.

11. Demonstrate how multiplying every value in the data set $2, 4, 6, 8$ by 3 affects the variance and standard deviation. Calculate the original and new values.

12. For the data set $1, 2, 3, 4, 5$, explicitly write out the summation for $\sum(x-\mu)^2$ and use it to compute the variance.

13. Two data sets, $A = 2, 4, 6, 8$ and $B = 3, 5, 7, 9$, have the same mean. Compute the variance of each data set and comment on their spreads.

14. Calculate the variance and standard deviation for the data set $10, 12, 14, 16, 18$ and explain how these measures indicate the dispersion of the data.

15. Consider the two sets $S_1 = 5, 6, 7, 8, 9$ and $S_2 = 5, 6, 7, 8, 20$. Calculate the variance for each and discuss the effect of an outlier on the spread.

16. A discrete random variable Y takes values according to the following table:

y	$P(Y = y)$
-2	0.2
0	0.5
2	0.3

Determine the mean, variance and standard deviation of Y.

17. Verify that for any data set, the formula $\sigma^2 = \frac{1}{n}\sum x^2 - \mu^2$ holds by applying it to the data set $3, 3, 7, 7$.

18. Two classes recorded the following test scores:

$$\text{Class 1: } 70, 75, 80, 85, 90, \quad \text{Class 2: } 60, 80, 80, 80, 100.$$

Calculate the variance and standard deviation for each class and compare the spread in performance.

19. For the data set $4, 8, 6, 10$, compute the deviations $(x - \mu)$ for each x and then use them to verify the calculation of the variance.

20. If X is a discrete random variable with mean μ and variance σ^2, show that the variance of a new variable defined by $Z = 2X+3$ is $4\sigma^2$. (Provide a brief derivation.)

Hard Questions

21. Prove that the variance of any discrete data set is always non-negative. Provide a rigorous explanation.

22. Derive the formula for variance $\sigma^2 = \dfrac{1}{n}\sum(x-\mu)^2$ starting from the definition of mean. Show each step in your derivation.

23. Consider a discrete random variable W with the following probability mass function:

w	$P(W = w)$
-3	0.1
0	0.4
3	0.3
6	0.2

Determine the mean, variance and standard deviation of W.

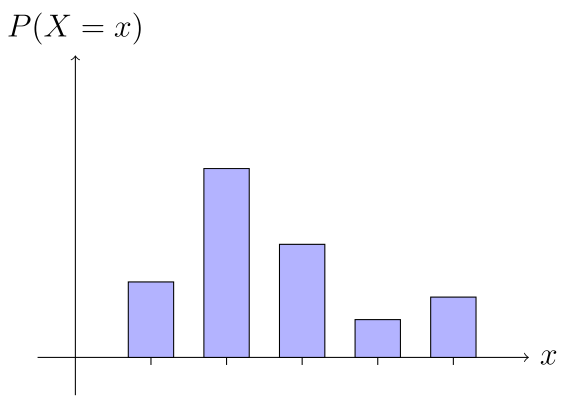
24. Given a data set $k, k+2, k+4, k+6$, where k is a constant, express the variance in terms of k and explain why the variance does not depend on k.

25. A discrete distribution has outcomes a, b, c with probabilities p, q, r such that $a = k-1$, $b = k$ and $c = k+1$. Determine the value of k that minimises the variance, given that p, q, r are fixed probabilities.

26. A company measures the daily number of defective items produced over a week: $3, 5, 2, 6, 4, 8, 3$. Compute the variance and standard deviation and discuss what these values suggest about the consistency of production.

27. Suppose a discrete random variable X has variance σ^2. If we define a new variable $Y = -4X + 7$, show algebraically that the variance of Y is $16\sigma^2$.

28. The following diagram represents a discrete probability distribution.



Assuming these probabilities are correctly scaled, explain how you would estimate the mean and variance from this diagram.

29. Discuss why the variance is measured in squared units and how the standard deviation provides a measure in the same units as the original data. Support your discussion with appropriate examples.

30. Two different discrete distributions have the same mean but different variances. Describe a scenario where this might occur and calculate a simple example to illustrate why variance is a better measure than the mean alone for understanding data spread.