Solution Set to Statistics Problem Set

Naureen Ghani

May 5, 2018

1. $\frac{1+9+14}{3} = 8$
2. $\frac{5+8+10+15+27}{5} = 13$
3. $\frac{2+4+4+6+8+12}{6} = 6$
4. 3.7
5. 14
6. The correlation of two random variables can be at most **1** if they are perfectly correlated (i.e. one is just some constant times the other plus another constant).
7. **They have the same correlation.** Because both data sets have points that lie on the same line and are upward sloping, this means that both have a perfect correlation of 1.
8. 0.75
9. Correlation is indepedent of the units used to measure something. As a quick argument why, imagine that it weren't. Then, we would be able to choose a unit either so small or so large that the correlation would be forced to be greater than 1 (or less than -1) which we know is impossible. Thus, the correct answer is **0.2**.
10. **Less than 1 but greater than 0**: First, we note that since the relationship between area and side length is quadratic, the relationship cannot be linear and thus correlation cannot be 1 or -1. However, as side length increases, area necessarily increases as well, and so the correlation must be positive, putting it somewhere between 0 and 1.
11. $E(X) = 1 \times 0.1 + 2 \times 0.2 + 3 \times 0.3 + 4 \times 0.4 = 3$ $\;V(X) = 1^2 \times 0.1 + 2^2 \times 0.2 + 3^2 \times 0.3 + 4^2 \times 0.4 - 3.0^2 = 1$

    $V(Y) = V(9X + 14) = 9^2 \cdot V(X) = 81$
12. Since $E(X) = 3$ and $E(Y) = 9$, from the properties of expected value we have $E(X) = E(aX+b) = a \cdot E(X)+b = a \cdot 3 + b$ $V(Y) = V(aX + b) = a^2 \cdot V(X) = a^2 \cdot 9$ Since we are given $E(Y) = 6, V(Y) = 9, a > 0$, we have

    $E(Y) = a \cdot 3 + b = 6, V(Y) = a^2 \cdot 9 = 9, a = 1, b = 3$
13. $V(2X + 7) = 2^2 \cdot V(X) = 2^2 \cdot 4 = 16$
14. $V(X) = E(X^2) - (E(X))^2$ so $(E(X))^2 = E(X^2) - V(X) = 10 - 3 = 7$
15. $V(2X + 4) = 2^2 \cdot V(X) = 2^2 \cdot 5 = 20$
16. 79
17. 50: The overall increases in the observations greater than the median do not change the ranking of the median.
18. 2: If we ignore x, then the range of the set would be $42.4 - (-3.3) < 50$. Hence, x must be either the smallest element, or the largest element. If $x = 42.4 - 50$, we get a range of 50. If $x = -3.3 + 50$, we get a range of 50. Thus, there are 2 values of x which would make the range equal to 50.
19. 6
20. 80

21. 9
22. 20
23. 18
24. 10
25. 24
26. When we multiply them all by three, we multiply the mean by 3 and the variance by 9, for a new mean of 15 and a new variance of 18. Then, squaring these values results in a mean equal to the sum of the old mean squared and the variance which is, $15^2 + 18 = 225 + 18 = 243$
27. First of all, we know that the standard deviation cannot be negative, so if we succeed in getting it down to 0, we have gotten it as small as possible. To do this, simply consider the data set consisting of two points: -10 and 10. It has the standard deviation 10, but if we take their absolute values, the data set becomes 10 and 10, which has a standard deviation of 0.
28. **125:** We recall that the standard deviation is defined as the square root of the variance, which means that the variance is thus $5^2 = 25$. Moreover, if we call the data points X, then: $25 = \mu_{X^2} - (\mu_X)^2 \implies \mu_{X^2} - 10^2 \implies$ $\mu_{X^2} = 100 + 25 = 125$
29. **at least 25, but possibly more:** First of all, it is entirely possible for the median to be 25. However, we can also consider the data set $-10, 5, 10$ which when squared becomes $100, 25, 100$ with median 100. If we generalize this to $-x, 5, x$, we can make the median be whatever x^2 we please so long as $x^2 \geq 25$. However, we cannot ever get less than 25. This is because in order for 5 to be the median of the original data set, more than half of the values need to be at least 5. And all the values that were originally greater than 5 must be, when squared, greater than or equal to 25. Thus, 25 is a lower bound for the median.
30. $E[X^2] = E[X]^2 + Var(X) = 5^2 + 4 = 29$
31. A helpful inequality to tackle this problem is that $|x| \geq x$ for all x. When x is positive, we have equality, and when x is negative, a positive number is always greater than a negative number. Thus, the numerical average of the absolute values must be at least 5, i.e.

$$5 = \frac{x_1 + ... + x_n}{n} \leq \frac{|x_1| + ... + |x_n|}{n}$$

Thus, we know not to look for a possible average smaller than 5. However, 5 is possible (i.e. a data set with just one element: 5), and thus, it is the smallest possible average of the transformed data.