6.10 Descriptive statistics for spread

MATHEMATICAL TECHNIQUES

♠ For the data presented in exercises 6.7.1–6.7.4, find the range, MAD, the variance both the direct way and with the computational formula, the standard deviation, and the coefficient of variation.

- EXERCISE **6.10.1**

  Experiment a.
- EXERCISE **6.10.2**

  Experiment b.
- EXERCISE **6.10.3**

  Experiment c.
- EXERCISE **6.10.4**

  Experiment d.

♠ Consider the following random variables that take only two values with the given probabilities. For each, find MAD, the variance, the standard deviation, and the coefficient of variation.

- EXERCISE **6.10.5**

  A Bernoulli random variable with $p = 1/3$.
- EXERCISE **6.10.6**

  A Bernoulli random variable with $p = 0.9$.
- EXERCISE **6.10.7**

  A random variable that takes the value 10 with probability 1/3 and the value of 0 with probability 2/3. Compare your answers with the answer to exercise 6.10.5.
- EXERCISE **6.10.8**

  A random variable that takes the value 20 with probability 0.9 and the value of 10 with probability 0.1. Compare your answers with the answer to exercise 6.10.6.

♠ Find the quartiles of a random variable with the given probability density function. Illustrate the areas on a graph of the p.d.f.

- EXERCISE **6.10.9**

  The probability density function of a random variable X is $f(x) = 2x$ for $0 \leq x \leq 1$ (as in exercises 6.7.17 and 6.8.5).
- EXERCISE **6.10.10**

  The probability density function of a random variable X is $f(x) = 1 - \frac{x}{2}$ for $0 \leq x \leq 2$ (as in exercises 6.7.18 and 6.8.6).
- EXERCISE **6.10.11**

  The probability density function of a random variable T is $h(t) = \frac{1}{t}$ for $1 \leq t \leq e$ (as in exercises 6.7.19 and 6.8.7).
- EXERCISE **6.10.12**

  The probability density function of a random variable T is $g(t) = 6t(1-t)$ for $0 \leq t \leq 1$ (as in exercises 6.7.20 and 6.8.8). This requires a computer (or Newton's method) to solve the equations.

♠ Find the variance and standard deviation of a continuous random variable with the given probability density function.

- EXERCISE **6.10.13**

  The probability density function of a random variable X is $f(x) = 2x$ for $0 \leq x \leq 1$ (as in exercises 6.7.17 and 6.8.5).
- EXERCISE **6.10.14**

  The probability density function of a random variable X is $f(x) = 1 - \frac{x}{2}$ for $0 \leq x \leq 2$ (as in exercises 6.7.18 and 6.8.6).
- EXERCISE **6.10.15**

  The probability density function of a random variable T is $h(t) = \frac{1}{t}$ for $1 \leq t \leq e$ (as in exercises 6.7.19 and 6.8.7).
- EXERCISE **6.10.16**

  The probability density function of a random variable T is $g(t) = 6t(1-t)$ for $0 \leq t \leq 1$ (as in exercises 6.7.20 and 6.8.8).

♠ Find MAD for a continuous random variable with the given probability density function. How does it compare with the standard deviation found in the earlier problem?

- EXERCISE **6.10.17**

  The probability density function of a random variable X is $f(x) = 2x$ for $0 \le x \le 1$ (as in exercise 6.10.13).

- EXERCISE **6.10.18**

  The probability density function of a random variable X is $f(x) = 1 - \frac{x}{2}$ for $0 \le x \le 2$ (as in exercise 6.10.14).

♠ Find the probability that the random variable has a value less than the one standard deviation below the expectation and less than two standard deviations below the expectation. How do the results compare with the rules of thumb for a bell-shaped distribution?

- EXERCISE **6.10.19**

  The probability density function of a random variable X is $f(x) = 2x$ for $0 \le x \le 1$ (as in exercise 6.10.13).

- EXERCISE **6.10.20**

  The probability density function of a random variable X is $f(x) = 1 - \frac{x}{2}$ for $0 \le x \le 2$ (as in exercise 6.10.14).

- EXERCISE **6.10.21**

  The probability density function of a random variable T is $h(t) = \frac{1}{t}$ for $1 \le t \le e$ (as in exercise 6.10.15).

- EXERCISE **6.10.22**

  The probability density function of a random variable T is $g(t) = 6t(1-t)$ for $0 \le t \le 1$ (as in exercise 6.10.16).

♠ The following step outline the proof of the computational formula for the variance.

- EXERCISE **6.10.23**

  Multiply out the squared term into three terms and break the sum into three sums.

- EXERCISE **6.10.24**

  Factor constants out of the sums. Remember that $\bar{X}$ is a constant. Recognize certain sums to be equal to the mean. Write in terms of the mean and group together like terms.

♠ There is a very general inequality about any random variable X, called **Chebyshev's inequality**. Suppose X has mean μ and standard deviation σ. Then

$$\Pr(|X - \mu| \ge k\sigma) \le \frac{1}{k^2}$$

for any value of k.

- EXERCISE **6.10.25**

  What does this mean for $k = 1$? Does this tell us anything?

- EXERCISE **6.10.26**

  What does this mean for $k = 2$? How much of the probability must lie within 2 standard deviations of the mean?

- EXERCISE **6.10.27**

  How much of the probability must lie within 3 standard deviations of the mean?

- EXERCISE **6.10.28**

  Compare the result of **b** with the second rule of thumb for bell-shaped distributions. Which gives more precise information?

APPLICATIONS

♠ Consider again the salaries presented in exercises 6.9.27 and 6.9.28. For each, find MAD, the variance and coefficient of variation. Which statistics are most sensitive to large values? What are the units of each statistic?

- EXERCISE **6.10.29**

  With the top salary of $450,000.

- EXERCISE **6.10.30**

  With the top salary of $4,500,000.

♠ For the data presented in exercises 6.7.35–6.7.38, find the variance and standard deviation. How many of the values lie within two standard deviations of the mean?

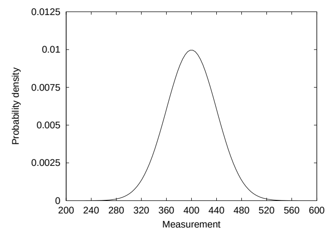
- EXERCISE **6.10.31**

  Experiment a (exercise 6.8.21 computes the expectation).

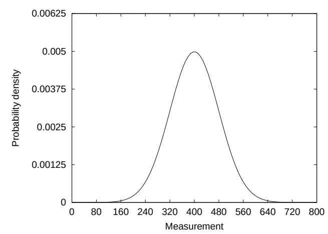
- EXERCISE **6.10.32**
  Experiment b (exercise 6.8.22 computes the expectation).
- EXERCISE **6.10.33**
  Experiment c (exercise 6.8.23 computes the expectation).
- EXERCISE **6.10.34**
  Experiment d (exercise 6.8.24 computes the expectation).

♠ Estimate the standard deviation, coefficient of variation, 2.5th percentile and 97.5th percentile from the following figures.

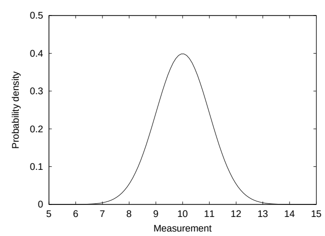
- EXERCISE **6.10.35**



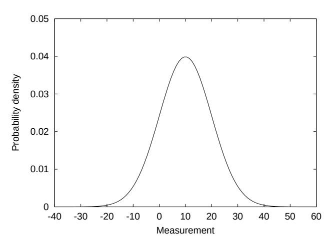
- EXERCISE **6.10.36**



- EXERCISE **6.10.37**



- EXERCISE **6.10.38**



♠ Draw bell-shaped p.d.f.'s with the following properties.

- EXERCISE **6.10.39**
  A p.d.f. with the same standard deviation of 10, but with a mean of 500. Calculate the coefficient of variation.
- EXERCISE **6.10.40**
  A p.d.f. with the same standard deviation of 10, but with a mean of 5. Calculate the coefficient of variation.
- EXERCISE **6.10.41**
  A p.d.f. with mean of 50 and coefficient of variation of 0.4. Calculate the standard deviation.

• EXERCISE **6.10.42**

A p.d.f. with mean of 50 and coefficient of variation of 0.1. Calculate the standard deviation.

♠ Suppose a population follows the rule

$$N_{t+1} = R_t N_t$$

where R_t is a random variable that takes on the value 1.5 with probability 0.6 and 0.5 with probability 0.4. Suppose $N_0 = 1$.

• EXERCISE **6.10.43**

Find the variance of the random variable N_1.

• EXERCISE **6.10.44**

Find the variance of the random variable $\ln(N_1)$.

Chapter 7

Answers

6.10.1. The range is 0 to 4. We found that the expectation is 2.1. Therefore,

$$\text{MAD} = |0 - 2.1|0.1 + |1 - 2.1|0.2 + |2 - 2.1|0.3 + |3 - 2.1|0.3 + |4 - 2.1|0.1 = 0.92.$$

With the direct method, the variance is

$$\text{Var}(M) = (0 - 2.1)^2 0.1 + (1 - 2.1)^2 0.2 + (2 - 2.1)^2 0.3 + (3 - 2.1)^2 0.3 + (4 - 2.1)^2 0.1 = 1.29.$$

With the computational formula,

$$\text{Var}(M) = 0^2 0.1 + 1^2 0.2 + 2^2 0.3 + 3^2 0.3 + 4^2 0.1 - 2.1^2 = 1.29.$$

The standard deviation is $\sqrt{(}1.29) = 1.136$. The coefficient of variation is $1.136/2.1 = 0.541$.
6.10.3. The range is 0 to 4. We found that the expectation is 1.6. Therefore,

$$\text{MAD} = |0 - 1.6|0.3 + |1 - 1.6|0.2 + |2 - 1.6|0.2 + |3 - 1.6|0.2 + |4 - 1.6|0.1 = 1.2.$$

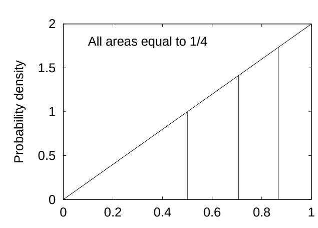
With the direct method, the variance is

$$\text{Var}(M) = (0 - 1.6)^2 0.3 + (1 - 1.6)^2 0.2 + (2 - 1.6)^2 0.2 + (3 - 1.6)^2 0.2 + (4 - 1.6)^2 0.1 = 1.84.$$

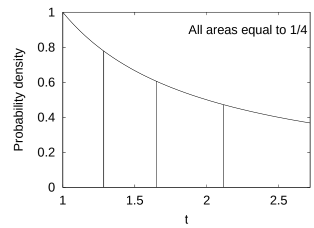
With the computational formula,

$$\text{Var}(M) = 0^2 0.3 + 1^2 0.2 + 2^2 0.2 + 3^2 0.2 + 4^2 0.1 - 1.6^2 = 1.84.$$

The standard deviation is $\sqrt{(}1.84) = 1.357$. The coefficient of variation is $1.357/1.6 = 0.848$.
6.10.5. Call the random variable B. Then $\Pr(B = 1) = 1/3$ and $\Pr(B = 0) = 2/3$. Then $\text{E}(B) = 1/3$, $\text{E}(B^2) = 1/3$, so $\text{Var}(B) = \frac{1}{3} - (\frac{1}{3})^2 = \frac{2}{9}$. The standard deviation is $\sqrt{29} = 0.471$, and the coefficient of variation is $0.471/0.333 = 0.157$. MAD$=2/3 \cdot 1/3 + 1/3 \cdot 2/3 = 4/9$.
6.10.7. Call the random variable B. Then $\Pr(B = 10) = 1/3$ and $\Pr(B = 0) = 2/3$. Then $\text{E}(B) = 10/3$, $\text{E}(B^2) = 100/3$, so $\text{Var}(B) = \frac{100}{3} - (\frac{10}{3})^2 = \frac{200}{9}$. The standard deviation is $\sqrt{2009} = 4.71$, and the coefficient of variation is $4.71/3.33 = 0.157$. MAD$=1 \cdot 0.9 + 9 \cdot 0.1 = 1.8$. Compared with exercise 6.10.6, the standard deviation and MAD increased by a factor of 10, the variance increased by a factor of 100, and the expectation and coefficient of variation changed in less predictable ways.
6.10.9. The c.d.f. is $F(x) = x^2$. The lower quartile solves $F(x) = 0.25$, and is equal to $\sqrt{0.25} = 0.5$. The median solves $F(x) = 0.5$, and is equal to $\sqrt{0.5} = 0.707$. The upper quartile solves $F(x) = 0.75$, and is equal to $\sqrt{0.75} = 0.866$.



6.10.11. The c.d.f. is $H(t) = \ln(t)$. The lower quartile solves $H(t) = 0.25$, and is equal to $e^{0.25} = 1.284$. The median solves $H(t) = 0.5$, and is equal to $e^{0.5} = 1.649$. The upper quartile solves $H(t) = 0.75$, and is equal to $e^{0.75} = 2.117$.



6.10.13. We found that $\mathrm{E}(X) = 2/3$. Using the computational formula,

$$\mathrm{Var}(X) = \int_0^1 x^2 f(x)dx - \mathrm{E}(X)^2 = \int_0^1 2x^3 dx - \mathrm{E}(X)^2 = \frac{x^4}{2}\Big|_0^1 - (\frac{2}{3})^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}.$$

The standard deviation is $\sqrt{1/18} = 0.236$.
6.10.15. We found that $\mathrm{E}(T) = e - 1$. Using the computational formula,

$$\mathrm{Var}(T) = \int_1^e t^2 h(t)dt - \mathrm{E}(T)^2 = \int_1^e tdt - \mathrm{E}(T)^2 = \frac{t^2}{2}|_1^e - (e-1)^2 = \frac{e^2}{2} - \frac{1}{2} - (e-1)^2 = 0.242.$$

The standard deviation is $\sqrt{0.242} = 0.492$.
6.10.17. We found that $\mathrm{E}(X) = 2/3$. Then

$$\begin{aligned}\mathrm{MAD}(X) &= \int_0^1 |x - 2/3| f(x)dx \\ &= \int_0^{\frac{2}{3}} 2x(2/3 - x)dx + \int_{\frac{2}{3}}^1 2x(x - 2/3)dx \\ &= \frac{2}{3}x^2 - \frac{2x^3}{3}|_0^{\frac{2}{3}} + \frac{2x^3}{3} - \frac{2}{3}x^2|_{\frac{2}{3}}^1 = 16/81.\end{aligned}$$

MAD is $16/81 = 0.197$, which is slightly smaller than the standard deviation.
6.10.19. We have that $\mathrm{E}(X) = 2/3 = 0.667$ and $\sigma = 0.236$. Then one standard deviation below the mean is $0.667 - 0.236 = 0.431$ and two standard deviations below the mean is $0.667 - 2 \cdot 0.236 = 0.195$. Using the c.d.f./ of $F(x) = x^2$, we have that $F(0.431) = 0.186$ and $F(0.195) = 0.038$. This p.d.f. has no points of inflection, but the percentiles are pretty close to those expected for a bell-shaped curve, where they would be 0.16 and 0.025.
6.10.21. We found that $\mathrm{E}(T) = e - 1 = 1.718$ and $\sigma = 0.492$. Then one standard deviation below the mean is $1.718 - 0.492 = 1.226$ and two standard deviations below the mean is $1.718 - 2 \cdot 0.492 = 0.734$. Using the c.d.f./ of $H(t) = \ln(t)$, we have that $H(1.226) = 0.204$. Because the value that is two standard deviations below the mean is not in the range, its percentile is not defined. This p.d.f. has no points of inflection, but the percentile for the first point is close to that expected for a bell-shaped curve, while the second is completely different.
6.10.23. $(x_i - \bar{X})^2 = x_i^2 - 2x_i\bar{X} + \bar{X}^2$, and

$$\sum_{i=1}^n (x_i - \bar{X})^2 p_i = \sum_{i=1}^n x_i^2 p_i - \sum_{i=1}^n 2x_i\bar{X}p_i + \sum_{i=1}^n \bar{X}^2 p_i.$$

6.10.25. It says that the probability that the value of the random variable is more than 1 standard deviation σ away from the mean μ is less than $1/1^2 = 1$. This is useless – all probabilities are less than 1.
6.10.27. At least 8/9, or 89%.
6.10.29. With the top salary of \$450,000, MAD=\$45,888, $\sigma^2 = 7,441,338$ square dollars. σ =\$86,000 and CV = 1.51 (no units).

6.10.31. Using the computational formula, and the calculation of the expectation as 7.53,

$$\text{Var}(P) = 4^2 \cdot 0.02 + 5^2 \cdot 0.08 + 6^2 \cdot 0.09 + 7^2 \cdot 0.31 + 8^2 \cdot 0.22 + 9^2 \cdot 0.22 + 10^2 \cdot 0.06 - 7.53^2 = 1.949.$$

The standard deviation is $\sqrt{1.949} = 1.396$. Only the two values of 4 lie more than two standard deviations away from the mean, or 2 out of 100.

6.10.33. Using the computational formula, and the calculation of the expectation as 4.98,

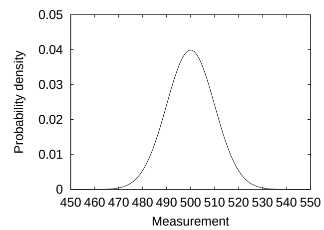
$$\text{Var}(P) = 1^2 \cdot 0.02 + 2^2 \cdot 0.05 + 3^2 \cdot 0.10 + 4^2 \cdot 0.21 + 5^2 \cdot 0.27 + 6^2 \cdot 0.19 + 7^2 \cdot 0.08 + 8^2 \cdot 0.05 + 9^2 \cdot 0.03 - 4.98^2 = 2.820.$$

The standard deviation is 1.679. The values of 1 and 9 lie more than two standard deviations away from the mean, or 5 out of 100.

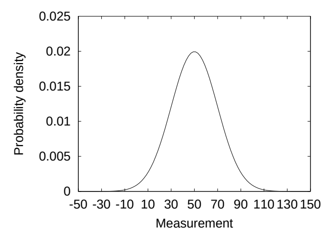
6.10.35. Mean is about 400, $\sigma = 40$, CV=0.1, 2.5th percentile at 320, 97.5th percentile at 480.

6.10.37. Mean is about 10, $\sigma = 1$, CV=0.1, 2.5th percentile at 8, 97.5th percentile at 12.

6.10.39. The coefficient of variation is 0.02.



6.10.41. The standard deviation is $0.4 \cdot 50 = 20$.



6.10.43. N_1 is exactly the same as R_0. The mean is 1.1 and the variance is

$$\text{Var}(N_1) = 1.5^2 \cdot 0.6 + 0.5^2 \cdot 0.4 - 1.1^2 = 0.24.$$