# Class Exercise 3 — Solutions

1. **Chebyshev's Inequality vs. The Empirical Rule**

A manufacturer of avionics equipment tests the operational lifetime of an autonomous drone navigation sensor before recalibration is required. The sensor lifetime has a mean of $\mu = 600$ hours and a standard deviation of $\sigma = 40$ hours.

(a) Suppose nothing is known about the shape of the distribution of sensor lifetimes.

(i) What interval is guaranteed by Chebyshev's inequality to contain at least 75% of the lifetimes?

(ii) What minimum percentage of sensors will have lifetimes between 480 hours and 720 hours?

(b) Now suppose historical quality records confirm that lifetimes are **bell-shaped and symmetric**.

(i) According to the Empirical Rule, approximately what percentage fall between 520 and 680 hours?

(ii) Approximately what percentage fall between 480 and 720 hours?

(c) A technician remarks: *"Chebyshev guarantees at least 75% between 520 and 680 hours, whereas the Empirical Rule states about 95%. Since they disagree, one rule must be flawed."* Explain why this is incorrect.

## Solution:

**(a) Any Distribution (Chebyshev's Inequality, Theorem 3.1):**

(i) Set $1 - \frac{1}{k^2} = 0.75 \implies \frac{1}{k^2} = 0.25 \implies k = 2$. The guaranteed interval is:

$$\mu \pm 2\sigma = 600 \pm 2(40) = 600 \pm 80 = [\mathbf{520} \text{ hours}, \mathbf{680} \text{ hours}].$$

(ii) Distance from mean: $|480 - 600| = |720 - 600| = 120$ hours $\implies k = \frac{120}{40} = 3$. By Chebyshev's inequality:

$$\text{Minimum percentage} = 1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9} \approx \mathbf{88.89}\% \quad (\text{or } 88.9\%).$$

**(b) Bell-Shaped and Symmetric Distribution (Empirical Rule, Formula 3.2):**

(i) $[520, 680]$ represents $\mu \pm 2\sigma$. By the Empirical Rule, this interval contains approximately **95**% of the data.

(ii) $[480, 720]$ represents $\mu \pm 3\sigma$. By the Empirical Rule, this interval contains approximately **99.7**% of the data.

**(c) Flaw in the Technician's Reasoning:** Chebyshev's inequality guarantees a universal **floor** ("at least 75%") that holds for *all* distributions, regardless of shape. The Empirical Rule provides a sharper estimate ($\approx$ 95%) because it assumes a bell-shaped, symmetric distribution. Since 95% $\geq$ 75%, there is no contradiction: the Empirical Rule sharpens the universal floor when the bell shape is verified.

2. **Standard Deviation from a Frequency Table**

A telecommunications engineer monitored the number of transmission packet errors observed during 10-minute testing windows across a satellite data downlink. A random sample of $n = 25$ testing windows yielded the following frequency distribution:

| Transmission Errors ($x$) | Frequency / Number of Windows ($f$) |
|---|---|
| 0 | 4 |
| 1 | 8 |
| 2 | 1 |
| 3 | 8 |
| 4 | 4 |

(a) Calculate the sample mean $\bar{x}$ number of transmission errors per window.

(b) Construct a computational table showing columns for $x$, $f$, $fx$, and $fx^2$, and use it to compute the sample variance $s^2$.

(c) Determine the sample standard deviation $s$ (round to four decimal places).

(d) A student attempts to calculate the variance by taking the five values $\{0, 1, 2, 3, 4\}$ and calculating their sample variance as $s^2 = 2.5$. Identify the conceptual error in this approach.

## Solution:

**(a) Sample Mean $\bar{x}$:** Total sample size $n = \sum f = 4 + 8 + 1 + 8 + 4 = 25$.

$$\sum fx = 0(4) + 1(8) + 2(1) + 3(8) + 4(4) = 50 \implies \bar{x} = \frac{50}{25} = \mathbf{2.0} \text{ errors per window.}$$

**(b) Computational Table and Sample Variance $s^2$:**

Using Formula 5.1 from the lecture notes with $n - 1 = 25 - 1 = 24$:

| Errors ($x$) | Frequency ($f$) | $fx$ | $fx^2$ |
|---|---|---|---|
| 0 | 4 | 0 | 0 |
| 1 | 8 | 8 | 8 |
| 2 | 1 | 2 | 4 |
| 3 | 8 | 24 | 72 |
| 4 | 4 | 16 | 64 |
| **Total** | $n = \sum f = 25$ | $\sum fx = 50$ | $\sum fx^2 = 148$ |

$$s^2 = \frac{1}{n-1}\left[\sum fx^2 - \frac{(\sum fx)^2}{n}\right] = \frac{1}{24}\left[148 - \frac{50^2}{25}\right] = \frac{1}{24}[148 - 100] = \frac{48}{24} = \mathbf{2.0} \text{ errors}^{\mathbf{2}}.$$

*Definitional check:* $\sum f(x - \bar{x})^2 = 4(-2)^2 + 8(-1)^2 + 1(0)^2 + 8(1)^2 + 4(2)^2 = 48 \implies s^2 = \frac{48}{24} = 2.0$.

**(c) Sample Standard Deviation $s$:** $s = \sqrt{s^2} = \sqrt{2.0} \approx \mathbf{1.4142}$ transmission errors.

**(d) Conceptual Error in the Student's Approach:** This is **Common Error 3** ("ignoring $f$ in a frequency table") from Section 7 of the lecture notes. Treating $\{0, 1, 2, 3, 4\}$ as unweighted values assumes each occurred once, ignoring the frequencies (e.g., $x = 1$ and $x = 3$ occurred 8 times, while $x = 2$ occurred once). It also incorrectly divides by $5 - 1 = 4$ rather than $n - 1 = 25 - 1 = 24$.

3. **Effects of Shifting and Scaling on Spread**

A civil engineering laboratory measures the daily thermal expansion of a set of steel bridge joints during summer temperature swings. The measured expansions $x$ (in millimetres) have a sample mean of $\bar{x} = 15$ mm, a sample variance of $s^2 = 9$ mm$^2$, and a sample standard deviation of $s = 3$ mm.

(a) **Shifting (Sensor Calibration):** Due to a calibration error, each sensor had an initial reading of +4 mm. The engineers subtract 4 mm from each reading: $y_i = x_i - 4$. Find the new mean $\bar{y}$, the new variance $s_y^2$, and the new standard deviation $s_y$. Explain conceptually why the standard deviation does not change.

(b) **Scaling (Unit Conversion):** To report measurements in centimetres, each observation is divided by 10 (multiplied by 0.1): $u_i = 0.1\,x_i$. Find the new mean $\bar{u}$, the new variance $s_u^2$, and the new standard deviation $s_u$ (include units).

(c) **Combined Linear Transformation with Negative Sign:** A normalized risk index is computed using $z_i = 100 - 4\,x_i$. Calculate the mean $\bar{z}$, the variance $s_z^2$, and the standard deviation $s_z$. (*Caution: Pay careful attention to the sign of the standard deviation!*)

## Solution:

**(a) Shifting by $a = -4$ mm ($y_i = x_i - 4$):**

- **New Mean:** $\bar{y} = \bar{x} - 4 = 15 - 4 = \mathbf{11}$ mm.
- **New Variance:** $s_y^2 = s_x^2 = \mathbf{9}$ $\mathbf{mm^2}$.
- **New Standard Deviation:** $s_y = s_x = \mathbf{3}$ mm.
- **Conceptual Reason:** Subtracting 4 shifts both every data point and the mean by $-4$ mm. The deviations $(y_i - \bar{y}) = (x_i - 4) - (\bar{x} - 4) = x_i - \bar{x}$ are identical. Since dispersion measures distance from the center, rigid shifts leave spread completely unchanged.

**(b) Scaling by $c = 0.1$ ($u_i = 0.1\,x_i$):**

- **New Mean:** $\bar{u} = c\,\bar{x} = 0.1(15) = \mathbf{1.5}$ cm.
- **New Standard Deviation:** $s_u = |c|\,s_x = 0.1(3) = \mathbf{0.3}$ cm.
- **New Variance:** $s_u^2 = c^2\,s_x^2 = (0.1)^2(9) = 0.01(9) = \mathbf{0.09}$ $\mathbf{cm^2}$ (note that $0.3^2 = 0.09$).

**(c) Combined Linear Transformation ($z_i = 100 - 4\,x_i$, with $a = 100, c = -4$):**

- **New Mean:** $\bar{z} = a + c\,\bar{x} = 100 - 4(15) = 100 - 60 = \mathbf{40}$.
- **New Variance:** $s_z^2 = c^2\,s_x^2 = (-4)^2(9) = 16 \times 9 = \mathbf{144}$.
- **New Standard Deviation:** By Formula 6.1, $s_z = |c|\,s_x = |-4|(3) = 4 \times 3 = \mathbf{12}$.
- **Key Insight on Sign:** Standard deviation is a root-mean-square distance and is **never negative** ($s \geq 0$). While the data is reflected by the negative sign, dispersion stretches by $|c| = |-4| = 4$, yielding +12, not $-12$.