1.6 The distribution of the average

As another example from the doping case, we now consider the mean value of each individual. As it turns out, each individual may have his own mean value and these values can vary considerably between individuals. Say that we introduce a 6 months test period to achieve a higher sensitivity of the antidoping control. In this period, each athlete is tested 5 times, and this average is taken as the estimate of the personal mean and shall be used for setting the limits in the future controls. Say that we know that the values of each individual are normally distributed around his personal mean with variance $\sigma^2 = 5^2$ and that the measurements are mutually independent.

- Let $X_i,\ i = 1, \ldots, n$ be independent, equally distributed random variables. What can we say about the expected value and the variance of their mean? Denote $E(X_i) = \mu$ and $\text{var}(X_i) = \sigma^2$ for each i.

- Calculate the limits around the estimated mean, in which an undoped athlete should stay with probability 0.99 at his next measurement. *Hint: use the result that the sum of independent normal variables is again normal.*

- Is the average of i.i.d. random variables always distributed with the same distribution?

Understanding the ideas in R:

- Assume that the means of the athletes are normally distributed as $N(148, 7.5^2)$ and generate the personal averages for 100 athletes. Further, use the normal distribution $N(0, 5^2)$ and add it to the personal mean of each athlete to generate 6 values for each individual. Estimate the personal means using the first five values and compare their variance to the theoretical result. Check the distribution of the difference between the 6th value and the estimated mean.

1.7 Conditional expected value and variance

Previous research has shown that the hemoglobin of a cyclist outside the competition phase is distributed as $N(150, 7^2)$, whereas the distribution in the competition phase equals $N(140, 11^2)$. Say that the competition phase lasts for 9 months. We are interested in the expected value and the variance

of a randomly taken measurement.
Hint: We are interested in the random variable Y, we know that $\{Y|X = 0\} \sim N(150, 7^2)$ *and* $\{Y|X = 1\} \sim N(140, 11^2)$, $P(X = 1) = 0.75$

- Sketch the distribution of Y and try guessing the expected value and the standard deviation
- Use the example to explain the formula $E(Y) = E[E(Y|X)]$. Is the expected value $E(Y|X)$ a random variable or a constant? Calculate the expected value of the variable Y.
- Calculate the variance of Y
- Find the general expression for the variance ($\{Y|X = 0\} \sim N(\mu_0, \sigma_0^2)$, $\{Y|X = 1\} \sim N(\mu_1, \sigma_1^2)$, $P(X = 1) = p$)
- Calculate the covariance of X and Y. Find the general expression ($\{Y|X = 0\} \sim N(\mu_0, \sigma_0^2)$, $\{Y|X = 1\} \sim N(\mu_1, \sigma_1^2)$, $P(X = 1) = p$). How does the covariance depend on the parameters? What about the correlation?
- What are the values of variance, covariance and correlation if the averages in both phases equal? Are the variables X and Y independent in that case?

Understanding the ideas in R:

- Use R to generate a large number of values and check their distribution
- Try checking each of the above results with R, compare the theoretical values with their estimates.