1. Suppose that a dataset X consists of the values $\{17, 20, 18, 15, 18, 17, 15, 14, 16, 19\}$. Compute

(a) mean
(b) variance
(c) standard deviation

of X.

2. Given a dataset of pairs $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ the **(sample) Pearson correlation coefficient** r_{xy} is defined as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where

- n is sample size
- x_i, y_i are the individual sample points indexed with i
- $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean, and similarly for $\bar{y}$.

Item (a): Find the value of the correlation coefficient from the following table:

Person	Age x	Glucose level y
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

- A correlation coefficient of 1 means that for every positive increase in one variable, there is a positive increase of a **fixed proportion** in the other.

- A correlation coefficient of -1 means that for every positive increase in one variable, there is a negative decrease of a **fixed proportion** in the other.

- Zero means that for every increase, there isn't a positive or negative increase. The two just aren't related.

Item (b): What can you tell about the correlation between “Age” and “Glucose level”.?

We can use `pnorm` function of R to compute probabilities from a normal distribution. If X is a normally distributed random variable, with mean $= \mu$ and standard deviation $= \sigma$, then:

- $P(X < U) =$ `pnorm(U, mean =` μ`, sd =` σ`, lower.tail=TRUE)`
- $P(x > D) =$ `pnorm(D, mean =` μ`, sd =` σ`, lower.tail=FALSE)`
- $P(D < X < U) =$ `pnorm(D, mean =` μ`, sd =` σ`, lower.tail=TRUE) - pnorm(U, mean =` μ`, sd =` σ`, lower.tail=TRUE)`

You may use R to solve Exercises 3 and 4.

3. X is a normally normally distributed variable with mean $\mu = 30$ and standard deviation $\sigma = 4$. Find

1. $P(x < 40)$
2. $P(x > 21)$
3. $P(30 < x < 35)$

4. For a certain type of computers, the length of time bewteen charges of the battery is normally distributed with a mean of 50 hours and a standard deviation of 15 hours. John owns one of these computers and wants to know the probability that the length of time will be between 50 and 70 hours.

5. A casino offers a game for a single player in which a fair coin is tossed at each stage. The pot starts at 1 euro and is doubled every time a head appears. The first time a tail appears, the game ends and the player wins whatever is in the pot. Thus the player wins 1 euro if a tail appears on the first toss, 2 euros if a head appears on the first toss and a tail on the second, 4 euros if a head appears on the first two tosses and a tail on the third, 8 euros if a head appears on the first three tosses and a tail on the fourth, and so on. In short, the player wins 2^{k-1} euros if the coin is tossed k times until the first tail appears.

(a) Let X be the amount of money in euros that the player wins. Find $E(X)$.
(b) What is the probability that the player wins at least $2^7 = 128$ euros?
(c) What would be a fair price to pay for entering the game? Why?
(d) Now suppose that the casino only has a finite amount of money. Specifically, suppose that the maximum amount of the money that the casion will pay you is 2^{20} euros (around 1.05 million euros). That is, if you win more than 2^{20} euros, the casino is going to pay you only 2^{20} euros. Let Y be the money that the player wins in this case. Find $E(Y)$.