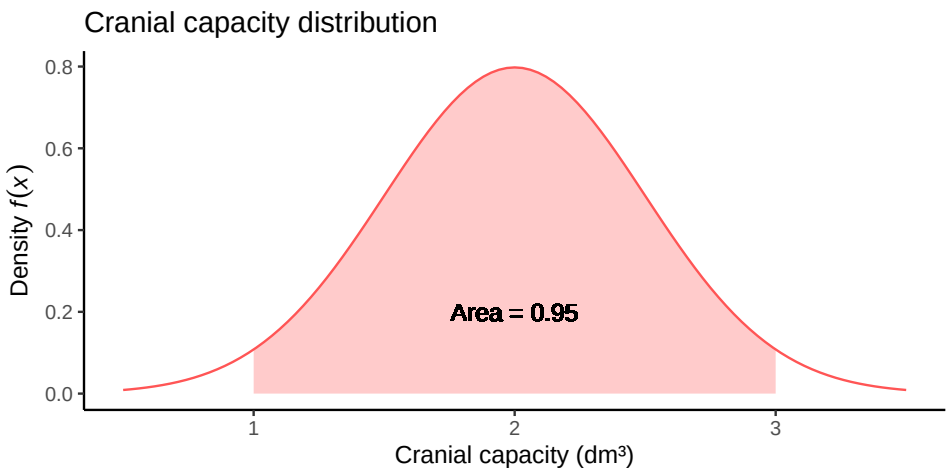
EXAM OF STATISTICS
(PROBABILITY AND RANDOM VARIABLES)

Pharmacy/Biotechnology 1st year — Version A — **November 22, 2021**

Duration: 1 hour.

(2.5 pts.) 1. The cranial capacity (in dm^3) of a primate population follows a normal probability distribution $X \sim N(\mu, \sigma)$. The chart below shows the Gauss bell of X. Observe that the chart shows the area below the bell between 1 and 3.



(a) What is the mean of the cranial capacity distribution?

(b) Is the mean of the cranial capacity representative of the population?

(c) What are the coefficients of skewness and kurtosis?

(d) What is the interquartile range of the cranial capacity?

(e) If a cranial capacity outside of the interval $(Q_1 - 1.5IQR, Q_3 + 1.5IQR)$ is considered an outlier, what is the probability of observing an outlier in the cranial capacity?

Remark: If you are not able to solve parts (a) and (b), use a mean $\mu = 1.5$ dm^3 and a standard deviation $\sigma = 0.25$ dm^3 for the other parts.

Solution

Let X be the cranial capacity of a primate. Then, $X \sim N(\mu, \sigma)$.

(a) $\mu = 2$ dm^3

(b) $\sigma = 0.5$ dm^6 and $cv = 0.25$. As the coef. of variation is small, the mean is representative.

(c) As X follows a normal distribution, $g_1 = 0$ and $g_2 = 0$.

(d) $Q_1 = 1.6628$ dm^3, $Q_3 = 2.3372$ dm^3 and $IQR = 0.6745$ dm^3.

(e) Fences: $f_1 = 0.651$ dm^3 and $f_2 = 3.349$.
$P(X < 0.651) + P(X > 3.349) = 0.007$.

(2.5 pts.) 2. A pharmaceutical company produces the same drug in 5 different laboratories. It has been observed that each laboratory produces, on average, one non-marketable defective batch every three months.

(a) What is the probability that a laboratory produce more than 3 defective batches in one year?

(b) What is the probability that at least 2 laboratories produce no defective batches in one year?

Solution

(a) Let X be the number of defective batches in a year, then $X \sim P(4)$, and $P(X > 3) = 0.5665$.

(b) Let Y be the number of laboratories that produce no defective batches in one year, then $Y \sim B(5, 0.0183)$, and $P(Y \geq 2) = 0.0032$.

(2 pts.) 3. The table below shows the frequencies observed in a random sample from a population for the blood type and SARS-CoV-2 infection:

Blood type	Infection	Persons
O	No	1800
O	Yes	100
A	No	4200
A	Yes	400
B	No	2500
B	Yes	150
AB	No	800
AB	Yes	50

(a) Compute the probability of SARS-CoV-2 infection for a random person.

(b) Compute the probability of having a blood type A and being infected by SARS-CoV-2 for a random person.

(c) Compute the probability of having a blood type A or being infected by SARS-CoV-2 for a random person.

(d) Compute the probability of being infected by SARS-CoV-2 for a person with blood type O.

(e) Compute the probability of having a blood type different from A and B for a person infected by SARS-CoV-2.

(f) Does the SARS-CoV-2 infection depend on the blood type?

Solution

(a) $P(I) = 0.07$.

(b) $P(A \cap I) = 0.04$.

(c) $P(A \cup I) = 0.49$.

(d) $P(I|O) = 0.0526$.

(e) $P(\overline{A} \cap \overline{B}|I) = 0.2143$.

(f) The infection depends on the blood as, for instance, $p(I) \neq P(I|O)$.

(3 pts.) 4. To study the relation between the blood Rh and the SARS-CoV-2 infection a random sample of non-infected people was drawn from a population. The table below shows the number of people infected after one year.

Blood Rh	Infection	Persons
−	Yes	520
−	No	6380
+	Yes	780
+	No	6200

(a) Compute the relative risk and the odds ratio to study the association between the SARS-CoV-2 infection and the blood Rh. Which association measure is more suitable to explain the relation between the SARS-CoV-2 infection and the blood Rh. Interpret it.

(b) A diagnostic test for the SARS-CoV-2 has been developed with a 95% of specificity and a 60% of sensitivity, regardless of blood Rh. In which blood Rh will produce more errors? Which diagnosis will we make if we apply the test to a persons with blood Rh − and we get a positive outcome? Which diagnosis will we make if we apply the test to a persons with blood Rh + and we get a negative outcome?

Solution

Let I be the event of being infected by SARS-CoV-2.

(a) $RR(I) = R_+(I)/R_-(I) = 1.4828$ and $OR(I) = O_+(I)/O_-(I) = 1.5435$.
The relative risk is more suitable as this is a prospective study and the incidence of infection can be estimated. Thus the risk of infection with Rh + is almost one and a half the risk with Rh -.

(b) $P(\text{Error}|\text{Rh-}) = 0.0764$ and $P(\text{Error}|\text{Rh+}) = 0.0891$. Thus, the test will produce more errors in people with Rh+.
Positive predictive value for Rh-: $p(I|+) = 0.4945$. Therefore, we will diagnose no infection.
Negative predictive value for Rh+: $p(\overline{I}|-) = 0.9497$. Therefore, we will predict no infection.
