Математические и статистические методы в психологии
Семинар 9. Решение. (12.03.2019)
А. А. Макаров, А. А. Тамбовцева, Н. А. Василёнок

Задача 1. Случайная величина Z имеет стандартное нормальное распределение. Найдите
(a) $P(-0.85 < Z < 0.23)$;
(b) для $p = 0.2$ квантиль x_p.

Решение. Известно, что случайная величина $Z \sim N(0,1)$. Это значит, что ее математическое ожидание равно 0, а дисперсия равна 1.

(a) Вероятность, что случайная величина попадет в некоторый отрезок $P(a < X < b)$ всегда равна разности между функцией распределения от значения верхней границы отрезка и функцией распределения от значения нижней границы отрезка

$$P(X < b) - P(X < a) = F(b) - F(a)$$

Мы видим, что нижняя граница данного отрезка меньше нуля. Это значит, что мы не сможем найти функцию распределения от этого значения, используя таблицу стандартного нормального распределения. Но мы знаем, что если $a < 0$, то

$$F(a) = P(X < a) = 1 - P(X < |a|)$$

Вернемся к условию задачи:

$$P(-0.85 < Z < 0.23) = F(0.23) - F(-0.85) = F(0.23) - (1 - F(0.85)) =$$

$$= 0.5910 - (1 - 0.8023) = 0.3933$$

(b) Квантиль – это такое *значение* случайной величины, для которого верно следующее равенство (p нам дано по условию, это уровень квантиля, а x_p неизвестно – это и есть квантиль):

$$P(X < x_p) = p$$

Мы знаем, что квантиль уровня $p < 0.5$ всегда будет отрицательным числом. Чтобы найти этот квантиль, нам нужно найти квантиль уровня $1 - p$, а затем умножить найденное значение на -1.

$$x_{0.8} \sim 0.845$$

$$x_{0.2} \sim -0.845$$

Задача 2. Найдите математическое ожидание, дисперсию случайной величины

$$U = 4X - 2Y + 1,$$

если известно, что X и Y – независимые случайные величины, имеющие
(a) стандартное нормальное распределение;
(b) нормальное распределение: $X \sim N(-3, \sigma^2 = 4)$, $Y \sim N(4, \sigma^2 = 1)$.

Решение.

(a) Если сказано, что обе величины имеют стандартное нормальное распределение, это значит, что $X \sim N(0,1)$ и $Y \sim N(0,1)$. Искомая случайная величина U имеет неизвестное произвольное нормальное распределение: $U \sim (\mu, \sigma^2)$.
Найдем математическое ожидание случайной величины U:

$$E(U) = E(4X - 2Y + 1) = 4E(X) - 2E(Y) + 1 = 0 - 0 + 1$$

Найдем дисперсию случайной величины U:

$$D(U) = D(4X - 2Y + 1) = 16D(X) + 4E(Y) = 16 + 4 = 20$$

Следовательно, $U \sim (1, 20)$.

(b) Теперь величины X и Y имеют произвольные нормальные распределения. Найдем математическое ожидание случайной величины U:

$$E(U) = E(4X - 2Y + 1) = 4E(X) - 2E(Y) + 1 = 4 \times (-3) - 2 \times 4 + 1 = -19$$

Найдем дисперсию случайной величины U:

$$D(U) = D(4X - 2Y + 1) = 16D(X) + 4E(Y) = 16 \times 4 + 4 \times 1 = 68$$

Следовательно, $U \sim (-19, 68)$.

Задача 3. Z_1 и Z_2 – независимые случайные величины, имеющие стандартное нормальное распределение. Найти вероятность того, что случайная величина

$$U = 2Z_1 - 3Z_2 - 5$$

примет положительное значение.

Решение. По условию задачи нам необходимо найти $P(U > 0)$.
(*Примечание: будьте внимательны к знакам в неравенствах!*)

Мы знаем, что линейная комбинация *стандартно нормально распределенных случайных величин* имеет *произвольное нормальное распределение* (уже не стандартное). Поэтому данную вероятность нельзя искать по табличке сразу – сначала надо стандартизовать случайную величину U. Для этого нам потребуются ее математическое ожидание и стандартное отклонение. Найдем математическое ожидание случайной величины U:

$$E(U) = E(2Z_1 - 3Z_2 - 5) = -5$$

Найдем дисперсию случайной величины U:

$$D(U) = D(2Z_1 - 3Z_2 - 5) = 4D(Z_1) + 9D(Z_2) = 13$$

Следовательно, $U \sim (-5, 13)$. В таком случае, $\sigma = \sqrt{13} \sim 3.6$. Найдем вероятность:

$$P(U > 0) = P(Z > \frac{0 - (-5)}{3.6}) = P(Z > \frac{5}{3.6}) = P(Z > 1.39) =$$
$$= 1 - P(Z < 1.39) = 1 - 0.9177 = 0.0823$$

Задача 4. Найдите вероятность того, что если бросить монету 200 раз, то орел выпадет от 90 до 110 раз.

Решение. По теореме Муавра-Лапласа известно, что случайная величина количества успехов в серии испытаний Бернулли при увеличении числа n стремится к нормальному распределению с параметрами $S_n \sim N(np, npq)$. Если n достаточно велико (например, больше нескольких десятков), то распределение этой случайной величины считают нормальным.

Нам требуется найти $P(90 < S_n < 110)$. Исходя из условия задачи, мы понимаем, что количество испытаний Бернулли $n = 200$, а вероятность успеха в одном испытании Бернулли $p = 0.5$. Следоваельно, $q = 1 - p = 0.5$.

Чтобы рассчитать вероятность, нам нужно стандартизовать случайную величину. Рассчитаем $E(S_n) = 200 \times 0.5 = 100$. Рассчитаем $D(S_n) = 200 \times 0.5 \times 0.5 = 50$. Рассчитаем $SD(S_n) = \sqrt{50} \sim 7$.

$$P(\frac{90 - np}{\sqrt{npq}} < \frac{S_n - np}{\sqrt{npq}} < \frac{110 - np}{\sqrt{npq}}) = P(\frac{90 - 100}{7} < Z < \frac{110 - 100}{7}) =$$
$$= P(-1.42 < Z < 1.42) = F(1.42) - (1 - F(1.42)) = 0.8444$$

Задача 5. По данным Фонда «Общественное мнение» (2013 г.) 33% москвичей утверждают, что пользуются метро ежедневно. Производится очередной репрезентативный опрос, в ходе которого респондентам задают вопрос о том, ездят ли они в метро каждый день. Используя теорему Муавра-Лапласа, найдите вероятность того, что в выборке объема 1000 человек окажется не более 360 респондентов, которые пользуются метро ежедневно.

Решение. Нам требуется найти $P(S_n < 360)$. Согласно условию задачи, количество испытаний Бернулли $n = 1000$, а вероятность успеха совпадает с долей москвичей, ежедневно пользующихся метро $p = 0.33$. Тогда $q = 0.67$.

Рассчитаем $E(S_n) = 1000 \times 0.33 = 330$. Рассчитаем $D(S_n) = 1000 \times 0.33 \times 0.67 = 221$. Рассчитаем $SD(S_n) = \sqrt{221} \sim 14.87$.

$$P(\frac{S_n - np}{\sqrt{npq}} < \frac{360 - np}{\sqrt{npq}}) = P(Z < \frac{360 - 330}{14.87}) = P(Z < 2.02)$$
$$= F(2.02) = 0.9783$$