

ОП «Политология», 2023-24**Введение в ТВиМС****Доверительные интервалы. Проверка гипотез о доле и среднем (24 апреля)***А. А. Макаров, А. А. Тамбовцева*

Задача 1. В ходе исследования изучалось, сколько часов в неделю жители региона N в возрасте от 35 до 45 лет посвящают просмотру телевизора. Были получены данные (случайная выборка):

4 8 9 12 0 16 10 8 7 6

- (a) Найдите выборочное среднее и выборочное стандартное отклонение.
- (b) Найдите стандартную ошибку среднего. Как изменится значение стандартной ошибки, если при прочих равных условиях объём выборки увеличится в 2 раза?
- (c) Постройте 95%-ный доверительный интервал для среднего времени просмотра телевизора. Проинтерпретируйте полученный интервал.

Решение.

- (a) среднее: $\bar{x} = 8$
дисперсия: $s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1} = \frac{(4-8)^2 + (8-8)^2 + \dots + (6-8)^2}{9} = \frac{170}{9} \approx 18.89$
стандартное отклонение: $s = \sqrt{18.89} \approx 4.35$
- (b) стандартная ошибка среднего: $se = \frac{s}{\sqrt{n}} = \frac{4.35}{\sqrt{10}} \approx 1.38$
- (c) уровень доверия $\beta = 0.95$, значит, $\alpha = 1 - \beta = 0.05$
 $t^* = t(1 - \frac{\alpha}{2}, df = n - 1) = t(0.975, df = 9) = 2.262$

границы доверительного интервала:

$$[\bar{x} - t^* \cdot se; \bar{x} + t^* \cdot se] \rightarrow [8 - 2.262 \cdot 1.38; 8 + 2.262 \cdot 1.38] \rightarrow [4.88; 11.12]$$

С 95%-ной уверенностью можно утверждать, что среднее время, которое жители региона N в возрасте от 35 до 45 лет посвящают просмотру телевизора в недели, лежит в интервале примерно от 5 до 11 часов.

Задача 2. Предположим, на основе данных предыдущей задачи мы хотим проверить, можно ли считать среднее время, которое люди данной возрастной категории тратят на просмотр телевизора, равным 10 часам в неделю.

- (a) Сформулируйте нулевую гипотезу и одностороннюю альтернативную гипотезу (направление выберите, исходя из данных).
- (b) Вычислите наблюдаемое значение статистики критерия.
- (c) Постройте критическую область, приняв уровень значимости равным 5%. Сделайте статистический и содержательный вывод относительно нулевой гипотезы.

Решение.

- (a) $H_0 : \mu = 10$ и $H_1 : \mu < 10$ (так как $\bar{x} = 8$ менее 10)

$$(b) t_{\text{набл}} = \frac{\bar{x} - a}{\text{se}} = \frac{\bar{x} - a}{\frac{s}{\sqrt{n}}} = \frac{8 - 10}{1.38} \approx -1.45$$

- (с) Уровень значимости $\alpha = 0.05$, альтернативная гипотеза левосторонняя, поэтому критическое значение статистики равно $-t(1 - \alpha, n - 1) = -t(0.95, 9) = -1.833$. Критическая область – область отвержения нулевой гипотезы – слева от значения $t_{\text{крит}} = -1.833$, то есть от $-\infty$ до -1.833 . В нашем случае наблюдаемое значение статистики критерия равно -1.45 , в критическую область не попадает, значит, является типичным при условии, что нулевая гипотеза верна, а значит, эту гипотезу отвергать не стоит.

Статистический вывод: на уровне значимости 5% на имеющихся данных нет оснований отвергнуть нулевую гипотезу.

Содержательный вывод: среднее время просмотра телевизора можно считать равным 10 часам.

Задача 3. Студенты в рамках проекта по социологии опросили 100 человек и выяснили, что 48% опрошенных хотя бы раз в жизни мечтали заняться экстремальным видом спорта. Можно ли считать, что доля людей, хотя бы раз в жизни мечтавших заняться экстремальным видом спорта, равна 0.5?

- (а) Сформулируйте подходящую нулевую гипотезу, которую необходимо проверить для ответа на этот вопрос. Сформулируйте одностороннюю альтернативную гипотезу (направление выберите, исходя из данных).
- (б) Какое распределение имеет статистика критерия? Найдите наблюдаемое значение статистики критерия.
- (с) Постройте критическую область, приняв уровень значимости равным 5%. Сделайте статистический и содержательный вывод относительно нулевой гипотезы.

Решение.

- (а) $H_0 : p = 0.5$ и $H_1 : p < 0.5$
- (б) Стандартное нормальное распределение $Z \sim N(0, 1)$.

$$z_{\text{набл}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.48 - 0.5}{\sqrt{\frac{0.5 \cdot 0.5}{100}}} \approx -0.4$$

- (с) Уровень значимости $\alpha = 0.05$, альтернативная гипотеза левосторонняя, поэтому критическое значение статистики равно $-z_{1-\alpha} = -z_{0.95} = -1.65$. Критическая область – область отвержения нулевой гипотезы – слева от значения $z_{\text{крит}} = -1.65$, то есть от $-\infty$ до -1.65 . В нашем случае наблюдаемое значение статистики критерия равно -0.4 , в критическую область не попадает, значит, является типичным при условии, что нулевая гипотеза верна, а значит, эту гипотезу отвергать не стоит.

Статистический вывод: на уровне значимости 5% на имеющихся данных нет оснований отвергнуть нулевую гипотезу.

Содержательный вывод: долю людей, хотя бы раз в жизни мечтавших заняться экстремальным видом спорта, можно считать равной 0.5.

Задача 4. Решите предыдущую задачу, считая, что альтернативная гипотеза двусторонняя и уровень значимости равен 1%.

Решение. Нулевая и альтернативная гипотезы: $H_0 : p = 0.5$ и $H_1 : p \neq 0.5$. Так как данные те же, наблюдаемое значение статистики $z_{\text{набл}} = -0.4$ то же, а вот критическая область меняется.

Уровень значимости $\alpha = 0.01$, альтернативная гипотеза двусторонняя, поэтому критических значений статистики здесь два, положительное и отрицательное.

Положительное: $z_{1-\frac{\alpha}{2}} = z_{0.995} = 2.58$

Отрицательное: $-z_{1-\frac{\alpha}{2}} = -z_{0.995} = -2.58$

Итого: критическая область – от $-\infty$ до -2.58 и от 2.58 до $+\infty$, а зона типичных значений с точки зрения H_0 , это отрезок от -2.58 до 2.58 . Наблюдаемое значение не попадает в критическую область, нулевая гипотеза по-прежнему не отвергается.