

Интернет среди пожилых

1. Анализ мощности.

Анализ мощности позволяет понять, скольких респондентов необходимо опросить, чтобы достичь заданной мощности теста (вероятности того, что тест отвергнет нулевую гипотезу, когда она в реальности тоже не верна).

Зададим мощность теста 90%, возьмем в качестве нулевой гипотезы равенство доли 0.5.

Для того, чтобы задать альтернативную, обратимся к Интернету и найдем статьи на эту тему. В Гане 52,9% пожилых людей пользовались интернетом за три месяца до переписи по данным переписи населения и жилищного фонда Ганы 2021 года согласно статье (Kyei-Arthur, 2024).

Используем это значение как ожидаемую альтернативную долю $\pi_1 = 0.529$ при проверке $H_0: \pi = 0.5$.

```
# one-sample prop-test, двусторонняя альтернатива ( $\pi \neq 0.5$ )

power.prop.test(p1 = 0.5, p2 = 0.529, sig.level = 0.05, power = 0.90,
alternative = "two.sided")

##
##      Two-sample comparison of proportions power calculation
##
##              n = 6239.651
##              p1 = 0.5
##              p2 = 0.529
##      sig.level = 0.05
##              power = 0.9
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```

Результат: минимальный размер выборки $n \approx 6240$ респондентов.

6240 наблюдений слишком много, у нас есть ограничения ресурсов исследования (например, ограничения по времени).

2. Сбор собственных данных:

Было опрошено N = 46 человек 60 лет и старше (30 женщин, 16 мужчин, средний возраст 71.54 лет).

Вопрос: «Пользуетесь ли Вы интернетом?» (ответы: да / нет)

3. Перенос данных в R и их анализ

Вводим случайные величины: жен=1, муж=0; да=1, нет=0

[illegible]

```
summary(data) #делаем анализ всех трех собранных переменных
```

##	age	net	sex
##	Min. :65.00	Min. :0.0000	Min. :0.0000
##	1st Qu.:68.00	1st Qu.:1.0000	1st Qu.:0.0000
##	Median :71.00	Median :1.0000	Median :1.0000
##	Mean :71.54	Mean :0.8478	Mean :0.6522
##	3rd Qu.:74.75	3rd Qu.:1.0000	3rd Qu.:1.0000
##	Max. :83.00	Max. :1.0000	Max. :1.0000

4. Построение доверительного интервала для доли.

- Рассчитаем долю пользователей интернета в выборке (p).

p=0.8478

- Построим доверительный интервал для этой доли при различных уровнях значимости (90%, 95%, 99%).

$$\left(p - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{N}}; p + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{N}}\right)$$

Для 90%:

$$100(1 - \alpha)$$

$$100(1 - \alpha) = 90$$

$$1 - \alpha = 0.9$$

$\alpha = 0.1$

$$\left(p - z_{\frac{0,1}{2}} \cdot \sqrt{\frac{p(1-p)}{N}}; p + z_{\frac{0,1}{2}} \cdot \sqrt{\frac{p(1-p)}{N}} \right)$$

$$\left(p - z_{0,05} \cdot \sqrt{\frac{p(1-p)}{N}}; p + z_{0,05} \cdot \sqrt{\frac{p(1-p)}{N}} \right)$$

$$\left(0.8478 - 1,64 \cdot \sqrt{\frac{0.8478(1 - 0.8478)}{46}}; 0.8478 + 1,64 \cdot \sqrt{\frac{0.8478(1 - 0.8478)}{46}} \right)$$

Доверительный интервал: (0.76, 0.93)

Это означает, что с вероятностью 90% доверительный интервал от 76% до 93% “поймал” истинную доля пожилых людей, пользующихся интернетом.

Для 95%:

$$100(1 - \alpha)$$

$$100(1 - \alpha) = 95$$

$$1 - \alpha = 0.95$$

$$\alpha = 0.05$$

$$\left(p - z_{\frac{0,05}{2}} \cdot \sqrt{\frac{p(1-p)}{N}}; p + z_{\frac{0,05}{2}} \cdot \sqrt{\frac{p(1-p)}{N}} \right)$$

$$\left(p - z_{0,025} \cdot \sqrt{\frac{p(1-p)}{N}}; p + z_{0,025} \cdot \sqrt{\frac{p(1-p)}{N}} \right)$$

$$\left(0.8478 - 1,96 \cdot \sqrt{\frac{0.8478(1 - 0.8478)}{46}}; 0.8478 + 1,96 \cdot \sqrt{\frac{0.8478(1 - 0.8478)}{46}} \right)$$

Доверительный интервал: (0.74, 0.95)

Это означает, что с вероятностью 95% доверительный интервал от 74% до 95% “поймал” истинную доля пожилых людей, пользующихся интернетом.

Для 99%:

$$100(1 - \alpha)$$

$$100(1 - \alpha) = 99$$

$$1 - \alpha = 0.99$$

$$\alpha = 0.01$$

$$\left(p - \frac{z_{0,01}}{2} \cdot \sqrt{\frac{p(1-p)}{N}}; p + \frac{z_{0,01}}{2} \cdot \sqrt{\frac{p(1-p)}{N}} \right)$$

$$\left(p - z_{0,005} \cdot \sqrt{\frac{p(1-p)}{N}}; p + z_{0,005} \cdot \sqrt{\frac{p(1-p)}{N}} \right)$$

$$\left(0.8478 - 2,57 \cdot \sqrt{\frac{0.8478(1 - 0.8478)}{46}}; 0.8478 + 2,57 \cdot \sqrt{\frac{0.8478(1 - 0.8478)}{46}} \right)$$

Доверительный интервал: (0.71, 0.98)

Это означает, что с вероятностью 99% доверительный интервал от 71% до 98% “поймал” истинную долю пожилых людей, пользующихся интернетом.

- *Интерпретация*

0.5 не попадает ни в один доверительный интервал, уверенно можно утверждать, что большинство пожилых людей пользуются интернетом, выводы для цифровой политики состоят в том, что доля пожилых людей, пользующихся интернетом, уже достаточно велика и дальнейшее ее увеличение может стоить слишком много, если специально не таргетировать остаточную долю.

5. Проверка гипотезы.

Проверьте гипотезу о том, что менее половины пожилых людей пользуются интернетом:

1. Формулировка гипотезы:

Нулевая гипотеза: доля пожилых людей, пользующихся интернетом, равна 0,5.

$$H_0 : \pi = 0.5$$

Альтернативная гипотеза: реальная доля ниже 0,5 (левосторонняя).

$$H_1 : \pi < 0.5 \text{ (односторонний тест)}$$

2. Выбор и расчет тестовой статистики.

Так как элементов в выборке больше 30, то можем использовать z-статистику для одной доли.

$$P = 0.8478, n = 46$$
$$\delta = \frac{(P - \pi_0)}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}} \sim N(0; 1)$$
$$\delta = \frac{(0.8478 - 0.5)}{\sqrt{\frac{(0.5(1 - 0.5))}{46}}} \approx 4,72$$

3. Критическая область или p-value:

$$z_{кр} = -\Phi_0^{-1}\left(\frac{1}{2} - \alpha\right), \alpha = 0.01$$
$$z_{кр} = -\Phi_0^{-1}(0,5 - 0,01) = -\Phi_0^{-1}(0,49) \approx -2,32$$

4. Статистическое решение

$\sigma > z_{кр}$, то есть статистика не попала в критическую область. Значит, статистическое решение: нельзя отвергнуть гипотезу о том, что половина пожилых пользуется интернетом.

5. Интерпретация

Можно сделать социальный вывод, что интернет уже затрагивает значимую долю группы 60+. Основные барьеры, такие как навыки, стоимость устройств, мотивация, часто могут быть устранены при помощи родственников. Предпринимать меры для повышения цифровой грамотности стоит, но уже следует вводить обучение навыкам более продвинутых интернет-пользователей, например, способности взаимодействия с генеративным искусственным интеллектом, или адресно масштабировать на оставшиеся офлайн-группы существующие государственные и корпоративные программы (курсы Сбербанка, «Бабушка-онлайн»).

Источники

Kyei-Arthur, F. (2024). Prevalence and predictors of internet use among Ghanaian older adults: evidence from the Ghana 2021 Population and Housing Census. *Humanities and Social Sciences Communications*, 11, 1309. <https://doi.org/10.1057/s41599-024-03788-7>