ДЗ 9

Кондраев Дмитрий

1. В чём различие между зависимыми и независимыми выборками? (2 балла)

Между элементами зависимых выборок можно установить взаимооднозначное соответствие, а значит их размеры одинаковы.

Независимые выборки могут быть разной величины.

Пример зависимой пары выборок: вес одной и той же группы испытуемых утром и вечером. Между элементами выборок можно установить соответствие (вес испытуемого до ↔ вес испытуемого после).

Пример независимого набора выборок: время, проведенное в приложении TikTok за сутки среди пользователей одной демографической группы и другой. Даже если в двух таких выборках будет равное количество людей, нельзя установить соответствие между парами величин из этих выборок.

Другими словами, значение в одной из зависимых выборок несет некоторую информацию о значении из другой выборки. Независимые выборки таким свойством не обладают.

2. Когда применяются параметрические статистические критерии, а когда — их непараметрические аналоги? (2 балла)

Параметрические критерии предполагают, что выборка порождена распределением из заданного параметрического семейства (например, нормальное распределение имеет параметры μ и σ (смещение и коэффициент масштаба), а распределение χ^2 имеет параметр k — число степеней свободы). Гипотеза в случае их использования может состоять в указании параметров распределения.

Параметрические критерии применяют, если заранее ясно, какая функция распределения породила выборку. Если это предположение верно, параметрические критерии дают более точные результаты. А если нет, вероятность ошибок (как I, так и II рода) резко увеличивается.

Непараметрические критерии не исходят из предположения о распределении выборки.

Непараметрические критерии используются, если необходимо проверить, что выборка подчиняется той или иной функции распределения, или что две некоторые выборки имеют одну функцию распределения, то есть принадлежат одной и той же генеральной совокупности.

3. Найти доверительный интервал для оценки математического ожидания с надежностью 0.95 (3 балла)

> Известно, что генеральная совокупность распределена нормально с известным квадратическим отклонением $s = 16$, выборочное среднее равно $\overline{x} = 80$, а объем равен $n = 256$.

Надежности 0.95 соответствует уровень доверия $t = 2$.

Доверительный интервал —

$$CI = \overline{x} \pm t\frac{s}{\sqrt{n}} = 80 \pm 2\frac{16}{\sqrt{256}} = 80 \pm 2.$$

Получается интервал $[78; 82]$.

4. Верно ли утверждение продавца, если учитывать, что уровень значимости равен $\alpha = 1\%$? (3 балла)

> Продавец утверждает, что средний вес пачки печенья составляет $\mu = 200$ грамм. Из партии извлечена выборка из $n = 10$ пачек. Вес каждой пачки составляет 202, 203, 199, 197, 195, 201, 200, 204, 194, 190 грамм соответственно. Известно, что их веса распределены нормально.

Применим критерий Стьюдента (т.к. предполагаем, что распределение нормальное).

Гипотеза $H_0 : E(X) = \mu$. Альтернативная: $H_1 : E(X) \neq \mu$.

Доверительная вероятность $1 - \alpha = 0.99$, значит критическое значение для $n = 10$: $t_{\alpha,9} = 3.169$ (таблица из вики, у нас распределение двухстороннее).

Выборочное среднее $\overline{X} = 198.5$ (здесь и ниже код на Python 3).

```
>>> import math
>>> sample = [202, 203, 199, 197, 195, 201, 200, 204, 194, 190]
>>> μ, n = 200, len(sample)
>>> avg = sum(sample) / n
>>> avg
198.5
```

Стандартное отклонение $S = 4.45$.

```
>>> stdev = math.sqrt(sum((x - avg) ** 2 for x in sample) / (n - 1))
>>> stdev
4.453463071962462
```

Найдем t-статистику

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}.$$

```
>>> T = (avg - μ) / (stdev / math.sqrt(n))
>>> T
-1.0651074037450896
```

Получили, что t-значение меньше по модулю, чем критическое, поэтому не можем отвергнуть нулевую гипотезу H_0.