Домашнее задание 5

Чудова Маргарита

Задание 1

Проверим гипотезу о том, что средний вес детали равен 1500 грамм на уровне значимости $\alpha = 0.1$. При условии, что она верна, должно выполняться

$$z = \sqrt{n}\frac{\overline{x} - a}{\sigma} \underset{H_0}{\sim} N(0,1)$$

Распределение симметричное, уровень значимости 0,1, значит $z = z_{0.95} \approx 1.65$

$$z = \sqrt{250}\frac{1505 - 1500}{50} \approx 1.58$$

Таким образом, наша гипотеза верна.

Задание 2

Проверим гипотезу о том, что средний счет гостя не изменился. Поскольку известна только выборочная дисперсия и не гарантирована нормальность, по ЦПТ можем воспользоваться асимптотическим критерием

$$Z = \sqrt{n}\frac{\overline{x} - a}{\sqrt{\overline{s}^2}} \underset{H_0}{\to} N(0,1)$$

Уровени значимости $\alpha = 0.04$, значит $z = z_{0.98} \approx 2.6$

$$z = \sqrt{150}\frac{148 - 150}{\sqrt{64}} \approx -3.06$$

Таким образом, отвергаем нашу гипотезу, поскольку не попали в доверительный интервал.

Задание 3

Сформулируем гипотезы:
H_0: $p = 1/2$ — человек без телепатических способностей $(\xi Bin(p = 1/2))$
H_1: $p > 1/2$ — человек с телепатическими способностями (распределение другое)
По сути, вероятность признать человека без телепатических способностей телепатом — это вероятность ошибки 1 рода. По ЦПТ Муавра-Лапласа знаем, что

$$\frac{\xi - np}{\sqrt{np(1-p)}} \sim N(0,1)$$

Тогда искомый уровень значимости:

$$z_t = \frac{\xi - np}{\sqrt{np(1-p)}} = \frac{70 - 50}{5} = 4$$

По нему находим вероятность ошибки

$$P_1 = \Phi(4) \approx 0.00003$$

Задание 4

Сформулируем гипотезы:
H_0: $a = 120$
H_1: $a = a_1 = 117$

- Точность 98% означает, что когда мы делаем вывод, что условия не соблюдаются, то в 98% это и правда так, а в 2% может быть не так. То есть вероятность ошибки 2 рода равна 0.02.

  $$P_2 = 0.02$$

  Тогда
  $$\sqrt{n}\frac{\overline{x}-a}{\sigma} = \sqrt{n}\frac{\overline{x}-a_1}{\sigma} + \sqrt{n}\frac{a_1-a}{\sigma} = \sqrt{n}\frac{\overline{x}-a_1}{\sigma} + C \sim N(C,1)$$
  $C = \sqrt{n}\frac{a_1-a}{\sigma}$ — константа. То есть распределение будет нормальным смещенным, чтобы привести к нормальному, надо вычесть C. Тогда доверительный интервал изменится: был (-1.95, 1.95), станет $(-1.96-C, 1.96+C)$.
  Что такое вероятность ошибки второго рода? Это (если порисовать) вероятность того, что распределение H_1 лежит в доверительной области H_0. Значит

  $$P_2 = \Phi(1.96-C) - \Phi(-1.96-C) = 0.02$$

  Подобрать C тут можно численно, средствами stats. Тогда $C \approx -4.01$.

  $$\sqrt{n} = C\sigma\frac{1}{a_1-a} = \frac{4.01\cdot 5}{3} \approx 6.68$$

  $$n \approx 45$$

- Разберем пример для 118:

  $$C = \sqrt{n}\frac{a_1-a}{\sigma} = 6.68\frac{118-120}{5} \approx 2.67$$

  Теперь надо подставить получившиеся C и проверить, что они не больше 0.02 Посчитаем все средствами stats, простым кодом

  ```
  a = 118 #наше a_1
  c = 6.68 *(a-120)/5
  st.norm.cdf(1.96-c) - -st.norm.cdf(-1.96-c)
  ```

  Итого:

  $a = 118$; $P = 0.2382305772598008$, не подходит
  $a = 119$; $P = 0.7331958473896388$, не подходит
  $a = 121$; $P = 0.7331958473896388$, не подходит
  $a = 122$; $P = 0.23823057725980085$, не подходит
  $a = 123$; $P = 0.02027999849766412$, в принципе подходит

Задание 5
H_0: выборка имеет распределение $N(a, \sigma^2)$
H_1: выборка имеет распределение $P(\lambda)$

Распределение Пуассона определено только для $k = 0, 1, 2\ldots$, в то время как нормальное распределение может принимать любое значение из $\mathbf{R}$. То есть для распределения Пуассона вероятность попасть в какое-то целое число равна единице, а в нецелое —нулю, а для нормального распределения в точности наоборот (мера целых чисел на $\mathbf{R}$ ноль).

Исходя из этого, построим такой критерий: H_0 не верна, если в выборке есть хотя бы одно целое неотрицательное число. Действительно, рассмотрим ошибку первого рода: пусть H_0 оказалась верна, а мы ее отвергли. Это означает, что при нормальном распределении встретилось целое число, а вероятность этого 0. Для ошибки второго рода аналогично: приняли H_0, что значит, что ни одного целого числа не было, вероятность этого события для распределения Пуассона 0.