ВШБ Бизнес-информатика: ТВиМС 2025.
Лист задач для самостоятельного решения #10.
Мат. статистика. Выборки. Выборочные распределения и характеристики.

Медиана (середина ранжированного ряда):

$$\text{Median}(x) = \begin{cases} x_{m+1}, & \text{если } n = 2m+1 \\ \frac{x_m + x_{m+1}}{2}, & \text{если } n = 2m \end{cases}$$

Мода: $\text{Mode}(x)$ – значение, встречающееся в выборке чаще всего.

Выборочное среднее $\bar{X}$:
Определение: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
Характеристики: $E[\bar{X}] = \mu$, $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$
Распределение:

- Если $X_i \sim \mathcal{N}(\mu, \sigma^2)$, то $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$
- По центральной предельной теореме (ЦПТ): при $n \geq 30$ выполняется приближенно $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

Выборочная дисперсия S^2:
Определение: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
Характеристики: $E[S^2] = \sigma^2$, $\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$
Распределение: Если $X_i \sim \mathcal{N}(\mu, \sigma^2)$, то $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ (хи-квадрат с $n-1$ степенями свободы)

Выборочная доля $\hat{p}$:
Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$.
Определение: $\hat{p} = \frac{Y}{n}$, где $Y = \sum_{i=1}^{n} X_i$ — количество успешных исходов (имеет биномиальное распределение)
Характеристики: $E[\hat{p}] = p$, $\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$
Распределение:

- По интегральной теореме Муавра-Лапласа (ИТМЛ): при $n > 30$ выполняется приближенно $\hat{p} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$

1. В таблице приведены данные по 10 салонам сотовой связи: первая строка - расстояние от метро, вторая - дневные продажи смартфонов. Для продаж найдите Median (Y), Mode (Y), выборочное среднее $\bar{y}$, выборочное стандартное отклонение s.

x_i	30	50	40	40	60	50	70	100	20	50
y_i	5	7	8	9	4	5	3	5	6	8

2. Предположим, что случайная выборка размера $n = 64$ будет выбрана из генеральной совокупности со средним 40 и стандартным отклонением 5.

   (a) Каковы математическое ожидание и стандартное отклонение выборочного распределения $\bar{X}$? Опишите форму выборочного распределения.

   Решение:

   Пусть $\mathcal{X} = (X_1, \ldots, X_n)$ — случайная выборка с $\mu = E[X_i] = 40$, $\sigma = 5$, $n = 64$.

   Выборочное среднее: $\bar{X} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$.

   Характеристики: $E[\bar{X}] = \mu = 40$, $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n} = \frac{25}{64} = 0.3906$, откуда $\mathrm{SD}(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{5}{8} = 0.625$.

   Так как $n = 64 \geq 30$, по ЦПТ: $\bar{X} \sim \mathcal{N}\left(40, \frac{25}{64}\right)$.

   Ответ: $E[\bar{X}] = 40$, $\mathrm{SD}(\bar{X}) = 0.625$, распределение приближенно нормальное.

   (b) Какова приближённая вероятность того, что $\bar{X}$ будет находиться в пределах 0.5 от среднего генеральной совокупности μ?

   Решение:

   $P(|\bar{X} - \mu| \leq 0.5) = P(39.5 \leq \bar{X} \leq 40.5)$.

   Стандартизируем: $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - 40}{0.625} \sim \mathcal{N}(0, 1)$.

   $P(39.5 \leq \bar{X} \leq 40.5) = P\left(\frac{39.5-40}{0.625} \leq Z \leq \frac{40.5-40}{0.625}\right) = P(-0.8 \leq Z \leq 0.8) = 2\Phi(0.8) - 1 \approx 0.576$.

   Ответ: ≈ 0.576

   (c) Какова приближённая вероятность того, что $\bar{X}$ будет отличаться от μ более чем на 0.7?

   Решение:

   $P(|\bar{X} - \mu| > 0.7) = 1 - P(|\bar{X} - \mu| \leq 0.7) = 1 - P(39.3 \leq \bar{X} \leq 40.7)$.

   $P(39.3 \leq \bar{X} \leq 40.7) = P\left(\frac{39.3-40}{0.625} \leq Z \leq \frac{40.7-40}{0.625}\right) = P(-1.12 \leq Z \leq 1.12) = 2\Phi(1.12) - 1 \approx 0.737$.

   $P(|\bar{X} - \mu| > 0.7) = 1 - 0.737 = 0.263$.

   Ответ: ≈ 0.263

3. Производственный процесс предназначен для изготовления деталей диаметром 0.5 дюйма. Каждый день выбирается случайная выборка из 36 деталей и записываются их диаметры. Если полученное выборочное среднее меньше 0.49 дюйма или больше 0.51 дюйма, процесс останавливается для настройки. Стандартное отклонение диаметра составляет 0.02 дюйма. Какова вероятность того, что производственная линия будет остановлена без необходимости?

   Решение:

   Пусть $\mathcal{X} = (X_1, \ldots, X_n)$ — случайная выборка с $\mu = 0.5$, $\sigma = 0.02$, $n = 36$.

   Процесс останавливается, если $\bar{X} < 0.49$ или $\bar{X} > 0.51$. При правильной работе процесса ($\mu = 0.5$) это ложная тревога.

   По ЦПТ (так как $n = 36 \geq 30$): $\bar{X} \sim \mathcal{N}\left(0.5, \frac{0.02^2}{36}\right) = \mathcal{N}(0.5, 0.0000111)$.

   $P(\bar{X} < 0.49$ или $\bar{X} > 0.51) = P(\bar{X} < 0.49) + P(\bar{X} > 0.51)$.

   Стандартизируем: $Z = \frac{\bar{X} - 0.5}{0.02/\sqrt{36}} = \frac{\bar{X} - 0.5}{0.00333} \sim \mathcal{N}(0, 1)$.

   $P(\bar{X} < 0.49) = P\left(Z < \frac{0.49-0.5}{0.00333}\right) = P(Z < -3) \approx 0.00135$.

   $P(\bar{X} > 0.51) = P\left(Z > \frac{0.51-0.5}{0.00333}\right) = P(Z > 3) \approx 0.00135$.

   Итого: $P(\text{остановка}) \approx 0.0027$.

   Ответ: ≈ 0.0027

4. Сообщается, что в крупном исследовании, проведённом в штате Нью-Йорк, примерно 30% участников исследования жили в пределах 1 мили от опасного места захоронения отходов. Пусть p обозначает долю всех жителей Нью-Йорка, которые живут в пределах 1 мили от такого места, и предположим, что $p = 0.3$.

   (a) Каковы математическое ожидание и стандартное отклонение выборочной доли $\hat{p}$ на основе случайной выборки размера $n = 400$?

   Решение:

   Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с $P(X_i = 1) = p = 0.3$, $n = 400$.

   Выборочная доля: $\hat{p} = \frac{Y}{n}$, где $Y = \sum\limits_{i=1}^{n} X_i$.

   Характеристики: $E[\hat{p}] = p = 0.3$, $\text{Var}(\hat{p}) = \frac{p(1-p)}{n} = \frac{0.3 \cdot 0.7}{400} = 0.000525$.

   $\text{SD}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{0.000525} \approx 0.0229$.

   Ответ: $E[\hat{p}] = 0.3$, $\text{SD}(\hat{p}) \approx 0.0229$

   (b) При $n = 400$ чему равно $P(0.25 \leq \hat{p} \leq 0.35)$?

   Решение:

   По ИТМЛ (так как $n = 400 > 30$): $\hat{p} \sim \mathcal{N}\left(0.3, \frac{0.3 \cdot 0.7}{400}\right)$.

   Стандартизируем: $Z = \frac{\hat{p} - 0.3}{0.0229} \sim \mathcal{N}(0, 1)$.

   $P(0.25 \leq \hat{p} \leq 0.35) = P\left(\frac{0.25-0.3}{0.0229} \leq Z \leq \frac{0.35-0.3}{0.0229}\right) = P(-2.18 \leq Z \leq 2.18) = 2\Phi(2.18) - 1 \approx 0.971$.

   Ответ: ≈ 0.971

   (c) Будет ли вероятность, вычисленная в предыдущем пункте, больше или меньше, чем в случае, если $n = 500$?

   Решение:

   При $n = 500$: $\text{SD}(\hat{p}) = \sqrt{\frac{0.3 \cdot 0.7}{500}} \approx 0.0205$.

   $P(0.25 \leq \hat{p} \leq 0.35) = P\left(\frac{0.25-0.3}{0.0205} \leq Z \leq \frac{0.35-0.3}{0.0205}\right) = P(-2.44 \leq Z \leq 2.44) = 2\Phi(2.44) - 1 \approx 0.985$.

   При увеличении n дисперсия $\hat{p}$ уменьшается, поэтому вероятность попадания в тот же интервал увеличивается.

   Ответ: Больше, ≈ 0.985

5. Производитель компьютерных принтеров закупает пластиковые картриджи с чернилами у поставщика. Когда поступает крупная партия, выбирается случайная выборка из 200 картриджей, и каждый картридж проверяется. Если выборочная доля бракованных картриджей больше 0.02, вся партия возвращается поставщику.

   (a) Какова приближённая вероятность того, что партия будет возвращена, если истинная вероятность брака картриджа равна 0.05?

   Решение:

   Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с $P(X_i = 1) = p = 0.05$, $n = 200$.

   По ИТМЛ: $\hat{p} \sim \mathcal{N}\left(0.05, \frac{0.05 \cdot 0.95}{200}\right) = \mathcal{N}(0.05, 0.0002375)$.

   Партия возвращается, если $\hat{p} > 0.02$.

   $P(\hat{p} > 0.02) = P\left(\frac{\hat{p}-0.05}{\sqrt{0.0002375}} > \frac{0.02-0.05}{\sqrt{0.0002375}}\right) = P(Z > -1.95) = 1 - \Phi(-1.95) = \Phi(1.95) \approx 0.974$.

   Ответ: ≈ 0.974

   (b) Какова приближённая вероятность того, что партия не будет возвращена, если истинная вероятность брака картриджа равна 0.10?

   Решение:

   При $p = 0.10$, $n = 200$: $\hat{p} \sim \mathcal{N}\left(0.10, \frac{0.10 \cdot 0.90}{200}\right) = \mathcal{N}(0.10, 0.00045)$.

   Партия не возвращается, если $\hat{p} \leq 0.02$.

   $P(\hat{p} \leq 0.02) = P\left(\frac{\hat{p}-0.10}{\sqrt{0.00045}} \leq \frac{0.02-0.10}{\sqrt{0.00045}}\right) = P(Z \leq -3.77) = \Phi(-3.77) \approx 0.00008$.

   Ответ: ≈ 0.00008

6. Кабельная компания решает, стоит ли заменить канал "магазин на диване"новой местной телепрограммой. Будет проведён опрос 100 абонентов. Кабельная компания решила оставить канал "магазин на диване если выборочная доля окажется больше 0.25. Какова приближённая вероятность того, что кабельная компания все-таки его оставит,если истинная доля тех, кто его смотрит, составляет только 0.20?

   Решение:

   Пусть $X_1, \dots, X_n$ — выборка из распределения Бернулли с $P(X_i = 1) = p = 0.20$, $n = 100$.

   По ИТМЛ: $\hat{p} \sim \mathcal{N}\left(0.20, \frac{0.20 \cdot 0.80}{100}\right) = \mathcal{N}(0.20, 0.0016)$.

   Канал оставляют, если $\hat{p} > 0.25$.

   $P(\hat{p} > 0.25) = P\left(\frac{\hat{p} - 0.20}{\sqrt{0.0016}} > \frac{0.25 - 0.20}{\sqrt{0.0016}}\right) = P(Z > 1.25) = 1 - \Phi(1.25) \approx 0.106$.

   Ответ: ≈ 0.106

7. Предположим, что случайная выборка будет взята из нормального распределения с неизвестным средним μ и стандартным отклонением $\sigma = 2$. Какого размера должна быть случайная выборка, чтобы $P(|\bar{X} - \mu| \le 0.1) \ge 0.95$ для любого возможного значения μ?

   Решение:

   Пусть $\mathcal{X} = (X_1, \dots, X_n)$ — случайная выборка из $\mathcal{N}(\mu, \sigma^2)$ с $\sigma = 2$.

   Если $X_i \sim \mathcal{N}(\mu, \sigma^2)$, то $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

   Стандартизируем: $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu}{2/\sqrt{n}} \sim \mathcal{N}(0, 1)$.

   $P(|\bar{X} - \mu| \le 0.1) = P\left(|Z| \le \frac{0.1}{2/\sqrt{n}}\right) = P\left(|Z| \le \frac{0.1\sqrt{n}}{2}\right) = 2\Phi\left(\frac{0.1\sqrt{n}}{2}\right) - 1 \ge 0.95$.

   Откуда: $2\Phi\left(\frac{0.1\sqrt{n}}{2}\right) - 1 \ge 0.95 \Rightarrow \Phi\left(\frac{0.1\sqrt{n}}{2}\right) \ge 0.975$.

   $\frac{0.1\sqrt{n}}{2} \ge z_{0.025} = 1.96$, откуда $\sqrt{n} \ge \frac{1.96 \cdot 2}{0.1} = 39.2$, и $n \ge 1536.64$.

   Ответ: $n \ge 1537$

8. Предположим, что случайная выборка будет взята из распределения Бернулли с неизвестным параметром p. Предположим также, что считается, что значение p находится в окрестности 0.2. Найдите приближённо размер случайной выборки, который должен быть взят, чтобы $P(|\bar{X} - p| \leq 0.1) \geq 0.95$ при $p = 0.2$.

   Решение:

   Пусть $X_1, \ldots, X_n$ — выборка из распределения Бернулли с $P(X_i = 1) = p = 0.2$.

   Выборочное среднее: $\bar{X} = \hat{p} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$.

   По ИТМЛ (при больших n): $\bar{X} \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right) = \mathcal{N}\left(0.2, \frac{0.2 \cdot 0.8}{n}\right)$.

   Стандартизируем: $Z = \frac{\bar{X} - p}{\sqrt{p(1-p)/n}} = \frac{\bar{X} - 0.2}{\sqrt{0.16/n}} \sim \mathcal{N}(0, 1)$.

   $P(|\bar{X} - p| \leq 0.1) = P\left(|Z| \leq \frac{0.1}{\sqrt{0.16/n}}\right) = P\left(|Z| \leq \frac{0.1\sqrt{n}}{\sqrt{0.16}}\right) = 2\Phi\left(\frac{0.1\sqrt{n}}{0.4}\right) - 1 \geq 0.95.$

   Откуда: $\Phi\left(\frac{0.1\sqrt{n}}{0.4}\right) \geq 0.975$.

   $\frac{0.1\sqrt{n}}{0.4} = \frac{\sqrt{n}}{4} \geq 1.96$, откуда $\sqrt{n} \geq 7.84$, и $n \geq 61.47$.
   Ответ: $n \geq 62$