Дисциплины: «Теория вероятностей», «Математическая статистика», «Теория вероятностей и математическая статистика»

Тема: Оценки параметров генеральной совокупности

Московский
Технологический
Институт

Домашнее задание (проверка)

16. Для вариационного ряда

Xi	1	3	7
ni	2	5	3

Найдем математическое ожидание, дисперсию, вариацию:

$$\bar{x} = \frac{1\cdot 2 + 3\cdot 5 + 7\cdot 3}{2+5+3} = \frac{2+15+21}{10} = \frac{39}{10} = 3{,}9$$

$$\overline{x^2} = \frac{1^2\cdot 2 + 3^2\cdot 5 + 7^2\cdot 3}{2+5+3} = \frac{2+45+147}{10} = \frac{194}{10} = 19{,}4$$

$$S_x^2 = \overline{x^2} - \bar{x}^2 = 19{,}4 - 3{,}9^2 = 19{,}4 - 15{,}21 = 4{,}19$$

$$S_x = \sqrt{S_x^2} = \sqrt{4{,}19} \cong 2{,}047$$

$$v = \frac{S_x}{\bar{x}}\cdot 100\% = \frac{2{,}047}{3{,}9}\cdot 100\% = 0{,}5249\cdot 100\% = 52{,}49\%$$

Московский Технологический Институт

Точечные оценки параметров

Пусть случайная величина X имеет закон распределения, зависящий от параметра θ (тэта): F(x,θ). О величине параметра можно судить по конечной выборке $x_1, x_2, \ldots, x_n$ из генеральной совокупности.

Оценкой $\hat{\theta}$ параметра θ называется любая функция от значений выборки $\hat{\theta} = \varphi_n(x_1, \ldots, x_n)$, т.е. **статистика**.

Статистику $\hat{\theta}$ можно рассматривать как случайную величину. Ее нужно выбирать таким образом, чтобы ее значения точнее оценивали значение неизвестного параметра θ.

Оценка $\hat{\theta}$ называется **несмещенной**, если ее математическое ожидание $M(\hat{\theta}) = \theta$. Для несмещенных оценок устраняется возможность появления систематической ошибки при оценивании параметра θ.

Оценка $\hat{\theta}$ называется **состоятельной,** если она удовлетворяет закону больших чисел, т.е. предел по вероятности $\underset{n \to \infty}{p\lim}(\hat{\theta}) = \theta$.

Несмещенная оценка $\hat{\theta}$ называется **эффективной**, если она имеет наименьшую дисперсию среди всех $V(\hat{\theta}) = M\left[(\hat{\theta} - \theta)^2\right]$

Московский Технологический Институт
МТИ

Точечные оценки параметров

Оценки $\hat{\theta}$ называются **точечными**, так как они оценивают одно численное значение параметра (точку).

Пусть генеральные параметры распределения для случайной величины X будут $a = M(X)$ (математическое ожидание) и $\sigma^2 = V(X)$ (дисперсия). Тогда для <u>повторной</u> выборки:

1. выборочное среднее является несмещенной, состоятельной и эффективной оценкой параметра ***a:*** $\hat{a} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

2. выборочная дисперсия является смещенной, состоятельной оценкой параметра σ^2: $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, причем $M(S^2) = \frac{n-1}{n}\sigma^2$,

3. исправленная выборочная дисперсия является несмещенной, состоятельной оценкой параметра σ^2: $\hat{\sigma}^2 = \bar{S}^2 = \frac{n \cdot S^2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Московский Технологический Институт

Точечные оценки параметров

4. выборочная доля является несмещенной, состоятельной и эффективной оценкой генеральной доли **p**: $\hat{p} = w = \frac{m}{n}$

Для указанных оценок справедливы формулы:

1) Для повторной выборки дисперсии

$$V(\bar{x}) = \frac{\sigma^2}{n} \qquad V(w) = \frac{pq}{n}$$

2) Для бесповторной выборки дисперсия

$$V(\bar{x}) = \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right) \qquad V(w) = \frac{pq}{n}\left(1 - \frac{n}{N}\right)$$

Московский
Технологический
Институт

Пример 1:

Из 1500 деталей отобрано 250, распределение которых по размеру X задано в таблице:

xi	7,8-8,0	8,0-8,2	8,2-8,4	8,4-8,6	8,6-8,8	8,8-9,0
ni	5	20	80	95	40	10

Найти точечные оценки для среднего и дисперсии, а также дисперсию оценки среднего при повторном и бесповторном отборах.

Решение. Вычислим по формулам (используем середины интервалов ci, число интервалов r=6, объем выборки n=250):

$$\hat{a} = \bar{x} = \frac{1}{n}\sum_{i=1}^{r} c_i = \frac{7{,}9\cdot 5 + 8{,}1\cdot 20 + 8{,}3\cdot 80 + 8{,}5\cdot 95 + 8{,}7\cdot 40 + 8{,}9\cdot 10}{250} = \frac{2110}{250} = 8{,}44,$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{r} c_i^2 = \frac{7{,}9^2\cdot 5 + 8{,}1^2\cdot 20 + 8{,}3^2\cdot 80 + 8{,}5^2\cdot 95 + 8{,}7^2\cdot 40 + 8{,}9^2\cdot 10}{250} = \frac{17818{,}9}{250} = 71{,}2756,$$

$$S^2 = \overline{x^2} - \bar{x}^2 = 71{,}2756 - (8{,}44)^2 = 71{,}2756 - 71{,}2336 = 0{,}042$$

$$\overline{S^2} = \frac{nS^2}{n-1} = \frac{250\cdot 0{,}042}{249} \approx 0{,}042$$

$$\hat{\sigma} = \sqrt{\overline{S^2}} = \sqrt{0{,}042} = 0{,}205$$

Московский
Технологический
Институт

Пример 1 (продолжение):

Вычислим дисперсию оценки среднего:

1)для повторной выборки:

$$V(\bar{x}) = \frac{\sigma^2}{n} \approx \frac{\hat{\sigma}^2}{n} = \frac{0{,}042}{250} = 0{,}000168$$

2) для бесповторной выборки

$$V(\bar{x}) \approx \frac{\hat{\sigma}^2}{n}\left(1 - \frac{n}{N}\right) = \frac{0{,}042}{250}\left(1 - \frac{250}{1500}\right) = \frac{0{,}042 \cdot 1250}{250 \cdot 1500} = 0{,}00014$$

Пример 2:

Выборочно обследовали партию кирпича. Из 100 проб в 12 случаях кирпич оказался бракованным. Найти оценку доли бракованного кирпича и дисперсию этой оценки.

Решение. По условию задачи число бракованных изделий m=12, объем выборки n=100, тогда оценкой доли бракованных является выборочная доля

$$\hat{p} = w = \frac{m}{n} = \frac{12}{100} = 0{,}12$$

Дисперсия этой оценки для повторной выборки равна

$$V(w) = \frac{pq}{n} = \frac{w(1-w)}{n} = \frac{0{,}12 \cdot 0{,}88}{100} = 0{,}001056$$

А среднее квадратическое отклонение этой оценки равно

$$\sqrt{V(w)} = \sqrt{0{,}001056} = 0{,}0325$$

Метод наименьших квадратов для нахождения точечных оценок:

Исследуется зависимость двух случайных величин Y и X по их выборкам $y_1, y_2, ..., y_n$ и $x_1, x_2, ..., x_n$. Пусть выбранный вид функции φ, устанавливающей эту зависимость, содержит параметры θ^i, i=1,2,…,k, тогда их оценки выбираются так, чтобы функция

$$S(\theta^1, \theta^2, ..., \theta^k) = \sum \left(y_i - \varphi(x_i, \theta^1, \theta^2, ..., \theta^k) \right)^2$$

принимала минимальное значение.

Из необходимого условия экстремума следует решение системы уравнений:

$$\frac{\partial S}{\partial \theta^i} = 0, \quad i = \overline{1, k}$$

Московский Технологический Институт

Пример 3:

Найти оценки параметров *a* и *b* по результатам выборочного наблюдения, если связь между случайными величинами Y и X линейна: $y = a + bx$.

Объем выборки равен n.

Решение. Используем метод наименьших квадратов. Построим функцию

$$S(a,b) = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

и найдем ее минимум. Вычислим частные производные и положим их равными нулю:

$$\begin{cases} \dfrac{\partial S(a,b)}{\partial a} = 2\sum_{i=1}^{n} (y_i - a - bx_i)(-1) = 0 \\ \dfrac{\partial S(a,b)}{\partial b} = 2\sum_{i=1}^{n} (y_i - a - bx_i)(-x_i) = 0 \end{cases}$$

Решим эту систему относительно *a* и *b*:

$$\begin{cases} \sum_{i=1}^{n} (y_i - a - bx_i) = 0 \\ \sum_{i=1}^{n} (y_i x_i - a x_i - b x_i^2) = 0 \end{cases} \Rightarrow \begin{cases} \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} a + b\sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i x_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \end{cases}$$

Пример 3 (продолжение):

Разделим оба равенства на n и обозначим выборочные средние:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \overline{yx} = \frac{1}{n}\sum_{i=1}^{n} y_i x_i \qquad \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Тогда получим систему линейных алгебраических выражений:

$$\begin{cases} a + b\bar{x} = \bar{y} \\ a\bar{x} + b\overline{x^2} = \overline{yx} \end{cases}$$

Эту систему можно решить любым известным методом (Гаусса, Кремера, матричным):

$$\begin{cases} a = \bar{y} - b\bar{x} \\ (\bar{y} - b\bar{x})\bar{x} + b\overline{x^2} = \overline{yx} \end{cases}$$

Окончательно получим оценки:

$$\begin{cases} a = \bar{y} - b\bar{x} \\ b = \dfrac{\overline{yx} - \bar{y}\cdot\bar{x}}{\overline{x^2} - (\bar{x})^2} \end{cases}$$

Московский Технологический Институт

Интервальные оценки параметров

- Интервальная оценка параметра дает возможность определить точность и надежность его оценки.
- **Интервальной оценкой** параметра θ называется интервал (α,β), который с заданной вероятностью γ (гамма) накрывает неизвестное значение этого параметра.
- Интервал (α,β) называется **доверительным интервалом**, вероятность γ - **доверительной вероятностью** или **уровнем**

Московский Технологический Институт

Интервальные оценки параметров

- Обычно доверительный интервал симметричен относительно точечной оценки $\hat{\theta}$, т.е. имеет вид $(\hat{\theta}-\Delta, \hat{\theta}+\Delta)$, где Δ - предельная ошибка выборки. Причем вероятность $P(|\hat{\theta}-\theta|<\Delta)=\gamma$.
- Рассмотрим генеральную совокупность объема *N* и выборку из нее $x_1, x_2, ..., x_n$. Для нее имеем:

выборочное среднее – $\bar{x}$

выборочную дисперсию – S_x^2

выборочную долю признака – $w=\frac{m}{n}$,

которым в выборке обладают m элементов.

- Рассмотрим следующие интервальные оценки:

Московский Технологический Институт

1. Доверительный интервал для генеральной средней *a*

$$\bar{x} - \Delta < a < \bar{x} + \Delta$$

а) для повторной выборки $\Delta = t\frac{S_x}{\sqrt{n}}$

б) для бесповторной выборки $\Delta = t\frac{S_x}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$

Величина **t** определяется:

- при n>30 из функции Лапласа Ф(t)=γ,
- при n≤30 из вероятности $P(|\xi| < t) = \gamma$,

где ξ имеет распределение Стьюдента для (n-1) степени свободы.

Московский Технологический Институт

Пример 4:

Для определения среднего процентного содержания белка в зернах пшеницы было отобрано 625 зерен, обследование которых показало, что выборочное среднее равно 16,8, а выборочная дисперсия равна 4. Чему равна с вероятностью 0,988 предельная ошибка выборки?

Решение. По условию задачи $\bar{x} = 16{,}8,\ S^2 = 4,\ \gamma = 0{,}988$. Так как генеральная совокупность бесконечна, то используем формулу для повторной выборки при определении предельной ошибки: $\Delta = t\dfrac{S}{\sqrt{n}}$

Значение t найдем из условия Φ(t)=γ, т.е. Φ(t)=0,988. По таблице значений функции Лапласа найдем: t=2,51. Найдем предельную ошибку

$$\Delta = 2{,}51\frac{\sqrt{4}}{\sqrt{625}} = \frac{2{,}51 \cdot 2}{25} = \frac{5{,}02}{25} = 0{,}2008$$

Целые и десятые доли x	**Сотые доли x**									
	0	1	2	3	4	5	6	7	8	9
2,4	0,9836	0,9841	0,9845	0,9849	0,9853	0,9857	0,9861	0,9865	0,9869	0,9872
2,5	0,9876	0,9879	0,9883	0,9886	0,9889	0,9892	0,9895	0,9898	0,9901	0,9904
2,6	0,9907	0,9910	0,9912	0,9915	0,9917	0,9920	0,9922	0,9924	0,9926	0,9928

Пример 5:

Выборочное среднее квадратическое отклонение десяти измерений некоторой величины равно 10 см. Найти с надежностью γ=0,6 предельную ошибку выборки.

Решение. Здесь n=10<30 и выборка повторная, *S*=10. По таблицам распределения Стьюдента для γ=0,6 и степени свободы n-1=9 находим t=0,88. Тогда получим предельную ошибку выборки

$$\Delta = t\frac{S}{\sqrt{n}} = 0{,}88\frac{10}{\sqrt{10}} = 0{,}88\sqrt{10} = 0{,}88\cdot 3{,}162 = 2{,}78$$

Число степеней свободы	Вероятность γ											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,90	0,95	0,98	0,99
9	0,13	0,26	0,40	0,54	0,70	0,88	1,10	1,38	1,83	2,26	2,82	3,25

Московский Технологический Институт

Пример 6:

Из партии в 5000 электрических ламп было отобрано 300 по схеме бесповторной выборки. Средняя продолжительность горения ламп в выборке оказалась равной 1450 часам, а дисперсия – 4000. Найти доверительный интервал для среднего срока горения лампы с надежностью 0,9996.

Решение. По условию задачи γ=0,9996 и объем выборки n=300>30, тогда по таблице значений функции Лапласа находим t из условия Ф(t)=0,9996: t=3,57. Применим формулу $\Delta = t\dfrac{S}{\sqrt{n}}$, где $S=\sqrt{4000}$ и вычислим предельную ошибку

$$\Delta = t\frac{S}{\sqrt{n}} = 3{,}57\sqrt{\frac{4000}{300}} = 3{,}57\sqrt{\frac{40}{3}} = 3{,}57\cdot 3{,}653 \cong 13$$

Искомый доверительный интервал будет равен:

$$1450-13 < a < 1450+13$$

$$1437 < a < 1463$$

Целые и десятые доли x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
3,4	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995	0,9995
3,5	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997	0,9997
3,6	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998	0,9998

2. Доверительный интервал для генеральной доли признака p:

$$w - \Delta < p < w + \Delta$$

а) для повторной выборки

$$\Delta = t\sqrt{\frac{w(1-w)}{n}}$$

б) для бесповторной выборки

$$\Delta = t\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$$

Величина **t** определяется из функции Лапласа Φ(t)=γ.

Пример 7:

В партии, содержащей 5000 изделий, проверено 400. Среди них оказалось 300 изделий высшего сорта. Найти с надежностью 0,95 доверительный интервал для доли изделий высшего сорта в случаях повторной и бесповторной выборок.

Решение. По условию задачи имеем:

$$N = 5000, \quad n = 400, \quad \gamma = 0{,}95, \quad w = \frac{300}{400} = 0{,}75.$$

По значению функции Лапласа Ф(t)=0,95 определим t=1,96.

Целые и десятые доли x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
1,9	0,9426	0,9439	0,9451	0,9464	0,9476	0,9488	0,9500	0,9512	0,9523	0,9533

1) Для повторной выборки предельная ошибка доли равна

$$\Delta = t\sqrt{\frac{w(1-w)}{n}} = 1{,}96\sqrt{\frac{0{,}75 \cdot 0{,}25}{400}} = 1{,}96 \cdot 0{,}02165 = 0{,}0424$$

Тогда доверительный интервал равен: $0{,}75 - 0{,}0424 < p < 0{,}75 + 0{,}0424$

$$0{,}7076 < p < 0{,}7924$$

2) Для бесповторной выборки предельная ошибка доли равна

$$\Delta = t\sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)} = 1{,}96\sqrt{\frac{0{,}75 \cdot 0{,}25}{400}\left(1-\frac{400}{5000}\right)} = 1{,}96 \cdot 0{,}02165 \cdot 0{,}959 = 0{,}0407.$$

Тогда доверительный интервал равен:

$$0{,}75 - 0{,}0407 < p < 0{,}75 + 0{,}0407 \Rightarrow \qquad 0{,}7093 < p < 0{,}7907$$

Пример 8:

Среди стандартных изделий одной фабрики в среднем 15% относится ко второму сорту. С какой вероятностью можно утверждать, что процент изделий второго сорта среди 1000 стандартных изделий данной фабрики отличается от 15% не более чем на 2%?

Решение. По условию задачи имеем n=1000, w=15%/100%=0,15, Δ=2%/100%=0,02.

Требуется найти вероятность

$$P = \left(\left|p - w\right| < \Delta\right) = P(\left|p - 0{,}15\right| \le 0{,}02) = \Phi\left(t\right)$$

Найдем t из формулы

$$\Delta = t\sqrt{\frac{w(1-w)}{n}}$$

, тогда

$$t = \Delta\sqrt{\frac{n}{w(1-w)}} = 0{,}02\sqrt{\frac{1000}{0{,}15 \cdot 0{,}85}} = 0{,}02\sqrt{\frac{1000}{0{,}1275}} = 0{,}02 \cdot 88{,}56 = 1{,}77$$

Используя значения из таблицы функции Лапласа найдем

$$P(\left|p - 0{,}15\right| \le 0{,}02) = \Phi\left(1{,}77\right) = 0{,}9233$$

Целые и десятые доли x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
1,6	0,8904	0,8926	0,8948	0,8969	0,8990	0,9011	0,9031	0,9051	0,9070	0,9090
1,7	0,9109	0,9127	0,9146	0,9164	0,9181	0,9199	0,9216	0,9233	0,9249	0,9265

3. **Доверительный интервал для генеральной дисперсии** σ^2

$$\frac{nS_x^2}{z_2} < \sigma^2 < \frac{nS_x^2}{z_1}$$

Где z_1 и z_2 определяются из условия $P(z_1 < \chi^2_{n-1} < z_2) = \gamma$

Обычно они определяются так, чтобы

$$P(\chi^2_{n-1} < z_1) = P(\chi^2_{n-1} > z_2) = (1-\gamma)/2$$

Тогда по таблице распределения Хи-квадрат со степенью свободы (n-1) они определяются из условий

$$P(\chi^2_{n-1} > z_1) = (1+\gamma)/2,$$

$$P(\chi^2_{n-1} > z_2) = (1-\gamma)/2.$$

Московский
Технологический
Институт

Пример 9:

Признак *X* генеральной совокупности распределен нормально. Имеется выборка в виде таблицы

xi	0,1	0,2	0,3	0,4	0,5
ni	2	4	7	6	1

Найти доверительный интервал, накрывающий среднее квадратическое отклонение с вероятностью 0,99.

Решение. Вычислим выборочные характеристики:

$$\bar{x} = \frac{0{,}1 \cdot 2 + 0{,}2 \cdot 4 + 0{,}3 \cdot 7 + 0{,}4 \cdot 6 + 0{,}5 \cdot 1}{2+4+7+6+1} =$$

$$= \frac{0{,}2 + 0{,}8 + 2{,}1 + 2{,}4 + 0{,}5}{20} = \frac{6}{20} = 0{,}3$$

$$S^2 = \frac{(0{,}1-0{,}3)^2 \cdot 2 + (0{,}2-0{,}3)^2 \cdot 4 + (0{,}3-0{,}3)^2 \cdot 7 + (0{,}4-0{,}3)^2 \cdot 6 + (0{,}5-0{,}3)^2 \cdot 1}{20} =$$

$$= \frac{1}{20}\left(0{,}04 \cdot 2 + 0{,}01 \cdot 4 + 0 + 0{,}01 \cdot 6 + 0{,}04 \cdot 1\right) = \frac{1}{20}\left(0{,}04 \cdot 3 + 0{,}01 \cdot 10\right) =$$

$$= \frac{0{,}12 + 0{,}1}{20} = \frac{0{,}22}{20} = 0{,}011.$$

Московский Технологический Институт

Пример 9 (продолжение):

По условию задачи n=20, γ=0,99.

Доверительный интервал для генеральной дисперсии равен: $\frac{nS^2}{z_2} < \sigma^2 < \frac{nS^2}{z_1}$

Где z_1 и z_2 определяются из условий $P(\chi^2_{n-1} > z_1) = (1+\gamma)/2, \quad P(\chi^2_{n-1} > z_2) = (1-\gamma)/2$

Т.е. $P(\chi^2_{n-1} > z_1) = 0{,}995, \quad P(\chi^2_{n-1} > z_2) = 0{,}005$

Найдем по таблицам критерия Пирсона (Хи-квадрат) величины

Число степеней свободы	Вероятность α												
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	**0,01**
19	7,63	8,57	10,1	11,6	13,7	15,3	18,3	21,7	23,9	27,2	30,1	33,7	36,2

$z_1 = 6{,}8$ (меньше табличного 7,63 для вероятности 0,99),

$z_2 = 38{,}5$ (больше табличного 36,2 для вероятности 0,01),

$$\frac{nS^2}{z_2} = \frac{200{,}011}{38{,}5} = 0{,}006 \qquad \frac{nS^2}{z_1} = \frac{20 \cdot 0{,}011}{6{,}8} = 0{,}032$$

$$0{,}006 < \sigma^2 < 0{,}032$$

$$0{,}077 < \sigma < 0{,}179$$

Московский Технологический Институт

4. Объем выборки **n,** необходимый для достижения требуемой надежности γ

1) При параметре ***a***

повторная выборка – $n = \frac{t^2 S_x^2}{\Delta^2}$

бесповторная выборка – $n = \frac{N t^2 S_x^2}{N\Delta^2 + t^2 S_x^2}$

2) При параметре ***p***

повторная выборка – $n = \frac{t^2 w(1-w)}{\Delta^2}$

бесповторная выборка – $n = \frac{N t^2 w(1-w)}{N\Delta^2 + t^2 w(1-w)}$

Замечание: При N→ в бесконечность, формулы для бесповторной выборки совпадут с формулами для повторной выборки.

Московский Технологический Институт

Пример 10:

Найти объемы повторной и бесповторной выборок из 10000 банок консервов для определения доли банок, не соответствующих стандарту. Предполагается, что предельная ошибка выборки не превосходит 0,05 с доверительной вероятностью 0,9995.

Решение. По условию задачи N=10000, Δ=0,05, γ=0,9995.

Целые и десятые доли x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
3,4	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995	0,9995
3,5	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997	0,9997

По таблице значений функции Лапласа Ф(t)=0,9995 найдем t=3,5.

1) Для повторной выборки объем равен

$$n=\frac{w(1-w)}{\Delta^2}t^2=\frac{3{,}5^2}{0{,}05^2}w(1-w)=4900w(1-w)$$

Так выборочная доля w по условию задачи неизвестна, тогда выберем его таким, чтобы выражение w(1-w) было максимальным. Это условие достигается при w=0,5 (вычислим производную функции и положим ее равной нулю: (w(1-w))'=1-2w=0). Тогда завышенное значение n будет равно n=4900*0,5*0,5=1225.

Московский Технологический Институт

Пример 10 (продолжение):

Для бесповторной выборки объем равен

$$n = \frac{Nt^2 w(1-w)}{N\Delta^2 + t^2 w(1-w)} = \frac{Nt^2}{\dfrac{N\Delta^2}{w(1-w)} + t^2}$$

В этом случае наибольшее значение выражения w(1-w) соответствует максимальному n. Положим w=0,5, тогда

$$n = \frac{10000 \cdot 3{,}5^2 \cdot 0{,}5 \cdot 0{,}5}{10000 \cdot 0{,}05^2 + 3{,}5^2 \cdot 0{,}5 \cdot 0{,}5} = \frac{10000 \cdot 3{,}0625}{25 + 3{,}0625} = \frac{30625}{28{,}0625} \cong 1091$$

Вопрос: Для расчета средней арифметической статистической совокупности используется формула (*n* – объем выборки, x_i – выборочные значения):

1) $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$;

2) $\bar{x} = \dfrac{n}{\sum_{i=1}^{n} \dfrac{1}{x_i}}$;

3) $\bar{x} = \sqrt{\dfrac{\sum_{i=1}^{n} x_i^2}{n}}$.

Московский Технологический Институт

Тестовые вопросы

1. Характеристикой оценок числовых характеристик по результатам выборочных значений является:

а) репрезентативность оценки;

б) несмещенность оценки;

в) сходимость любой оценки к математическому ожиданию теоретического распределения;

г) независимость оценки от объема выборки.

2. Определение искомой характеристики генеральной совокупности внутри какого-то интервала с заданной вероятностью, называется

а) интервальной оценкой;

б) точечной оценкой;

в) выборочной оценкой;

г) качественной оценкой.

Тестовые вопросы

3. Точечная оценка математического ожидания нормального распределения равна 12. Тогда его интервальная оценка может иметь вид …

а) (10,6; 13,4)

б) (12; 13,7)

в) (10,8; 12)

г) (11,2; 11,8)

4. Точечная оценка математического ожидания нормального распределения равна 15. Тогда его интервальная оценка может иметь вид...

а) (13,8; 15)

б) (13,8; 16,2)

в) (15; 16,2)

г) (13,8; 14,1)

Тестовые вопросы

5. В результате измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты: 10, 13, 13. Тогда несмещенная оценка дисперсии измерений равна:

а) 6;

б) 2;

в) 12;

г) 3.

6. По городской телефонной сети было произведено 100 наблюдений и установлено, что средняя продолжительность телефонного разговора составляет 4 минут при среднеквадратичном отклонении 2 мин. Предельная ошибка выборки с вероятностью 0,954 составляет

а) 0,2;

б) 0,3;

в) 0,4;

г) 0,5.

Приложение: Значения Ф(x)

Целые и десятые доли x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
1,3	0,8054	0,8098	0,8132	0,8165	0,8198	0,8230	0,8262	0,8293	0,8324	0,8355
1,4	0,8385	0,8415	0,8444	0,8473	0,8501	0,8529	0,8557	0,8584	0,8611	0,8638
1,5	0,8664	0,8690	0,8715	0,8740	0,8764	0,8789	0,8812	0,8836	0,8859	0,8882
1,6	0,8904	0,8926	0,8948	0,8969	0,8990	0,9011	0,9031	0,9051	0,9070	0,9090
1,7	0,9109	0,9127	0,9146	0,9164	0,9181	0,9199	0,9216	0,9233	0,9249	0,9265
1,8	0,9281	0,9297	0,9312	0,9327	0,9342	0,9357	0,9371	0,9385	0,9392	0,9412
1,9	0,9426	0,9439	0,9451	0,9464	0,9476	0,9488	0,9500	0,9512	0,9523	0,9533
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9916	0,9925	0,9634
2,1	0,9643	0,9651	0,9660	0,9668	0,9676	0,9684	0,9692	0,9700	0,9707	0,9715
2,2	0,9722	0,9729	0,9736	0,9743	0,9749	0,9756	0,9762	0,9768	0,9774	0,9780
2,3	0,9786	0,9791	0,9797	0,9802	0,9807	0,9812	0,9817	0,9822	0,9827	0,9832
2,4	0,9836	0,9841	0,9845	0,9849	0,9853	0,9857	0,9861	0,9865	0,9869	0,9872
2,5	0,9876	0,9879	0,9883	0,9886	0,9889	0,9892	0,9895	0,9898	0,9901	0,9904
2,6	0,9907	0,9910	0,9912	0,9915	0,9917	0,9920	0,9922	0,9924	0,9926	0,9928
2,7	0,9931	0,9933	0,9935	0,9937	0,9939	0,9940	0,9942	0,9944	0,9946	0,9947
2,8	0,9949	0,9951	0,9952	0,9953	0,9955	0,9956	0,9958	0,9959	0,9960	0,9961
2,9	0,9963	0,9964	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972

Задачи для самостоятельного решения

1. С целью определения средней суммы вкладов Q в банке, имеющем 2200 вкладчиков, проведено выборочное обследование (бесповторный отбор), результаты которого имеют вид:

Q, тыс. руб.	**10-30**	**30-50**	**50-70**	**70-90**	**90-110**	**110-130**
Число вкладчиков	**1**	**3**	**10**	**30**	**60**	**7**

Найти с вероятностью 0,96 доверительные границы для Q.

2. При формировании портфеля поставок был произведен случайный повторный отбор 100 поставщиков, осуществлявших поставки ранее. Для процента w несвоевременно отгрузивших сырье поставщиков необходимо определить доверительные границы на уровне 0,997, если в выборке оказалось 25 таких поставщиков.

3. В выборке объемом 500 единиц, произведенной для определения процента всхожести семян, установлена частость доброкачественных семян 0,94. Найти вероятность процента всхожести, если допустимая погрешность в его определении равна 2%.

Задачи для самостоятельного решения

4. Сколько лиц в возрасте от 19 до 24 лет надо опросить, чтобы установить средний процент студентов с точностью до 0,5%?

5. Определить численность выборки при обследовании остатков на расчетных счетах у клиентов банка, чтобы с вероятностью 0,683 предельная ошибка равнялась 5 усл. ед., если усл. ед.

6. Из 2500 ящиков продукции было проверено 10%. Среди них оказалось 80% ящиков с продукцией первого сорта. Найти границы, в которых с вероятностью 0,996 заключена доля ящиков с продукцией первого сорта.

7. По данным 10 измерений некоторой величины найдено ее выборочное среднее значение 20 и выборочная исправленная дисперсия 25. Найти границы, в которых с вероятностью 0,99 заключено истинное значение измеряемой величины. Найти с вероятностью 0,99 доверительный интервал для дисперсии генеральной совокупности этой величины.