ФГБОУ ВО «УлГПУ им. И.Н. Ульянова»
Теория вероятностей и математическая статистика
лектор Макеева О.В.

Лекция 9

Выборочный метод и статистическое оценивание

1. Выборочный метод
2. Вариационные ряды
3. Меры усреднения выборочных данных
4. Меры разброса выборочных данных
5. Статистическое оценивание
6. Интервальные оценки

Пролог

Математическая статистика опирается на теорию вероятностей, но в отличие от неё изучает не закономерности случайных явлений на основе абстрактного описания действительности, а **оперирует непосредственно к результатам наблюдений** над случайными явлениями.

Используя результаты теории вероятностей математическая статистика позволяет **оценить значения** искомых характеристик и указать **степень точности** выводов, получаемых при обработке данных.

§1. Выборочный метод

В практике статистических наблюдений различают сплошное и выборочное наблюдение. Вся подлежащая изучению совокупность объектов (наблюдений) называется **генеральной совокупностью**. Та часть объектов, которая отобрана из генеральной совокупности непосредственно для изучения, называется **выборочной совокупностью** (выборкой).

Пример сплошного наблюдения.

Пример выборочного наблюдения.

§1. Выборочный метод

Понятие генеральной совокупности в некотором смысле аналогично понятию случайной величины, а выборку можно рассматривать, как некий эмпирический аналог генеральной совокупности.

Число объектов (наблюдений) совокупности называют её **объёмом**. Генеральная совокупность может иметь как конечный, так и бесконечный объём.

Выборка называется **репрезентативной** (представительной), если она хорошо воспроизводит генеральную совокупность.

Репрезентативность выборки обеспечивается **случайным** характером отбора при котором все элементы генеральной совокупности имеют равные возможности быть отобранными в выборку.

§1. Выборочный метод

Сущность **выборочного метода** состоит в том, чтобы по некоторой части генеральной совокупности (по выборке) выносить суждение о свойствах совокупности в целом.

Задача выборочного метода – **оценить** параметры (характеристики) генеральной совокупности по данным выборочной совокупности.

Теоретическим обоснованием выборочного метода является *закон больших чисел* , согласно которому при неограниченном увеличении объёма выборки практически достоверно, что случайные выборочные характеристики как угодно близко приближаются (сходятся по вероятности) к параметрам генеральной совокупности.

§2. Вариационные ряды

Пусть рассматривается некоторый количественный признак (случайная величина) X. Различные наблюдаемые значения признака x называют **вариантами**.

После того как данные наблюдений (экспериментов) собраны их систематизируют. Процесс упорядочения вариант по возрастанию (убыванию) называется **ранжированием**.

Вариационным рядом называется ранжированный в порядке возрастания (убывания) ряд **вариант** с соответствующими им весами (**частотами** или **частостями**).

§2. Вариационные ряды

Вариационный ряд называется **дискретным**, если любые его варианты отличаются на постоянную величину.

Номер варианты	**i**	***1***	***2***	***3***
Варианта	**xi**	3	4	5
Частота варианты	**ni**	10	15	7
Относительная частота варианты	**wi**	0,31	0,47	0,22

§2. Вариационные ряды

Если различные варианты выборочной совокупности различаются сколь угодно мало, их группируют в интервалы. Количество интервалов $\boldsymbol{m}$ определяют по **формуле Стерджеса:**

$$m \approx 1 + 3,322 \cdot \lg n, \qquad (1)$$

где $\boldsymbol{n}$ – объём выборочной совокупности. Тогда длина каждого частичного интервала $\boldsymbol{h}$ будет равна:

$$h = x_{i+1} - x_i \approx \frac{x_{\max} - x_{\min}}{1 + 3,322 \cdot \lg n}, \qquad (2)$$

где $x_{\max} - x_{\min}$ обозначает разность между наибольшим и наименьшим значениями признака.

§2. Вариационные ряды

Вариационный ряд называется **интервальным** (непрерывным) , если варианты могут отличаться одна от другой на сколь угодно малую величину.

Номер интервала	**Интервал вариант**	**Частота**	**Относительная частота**
i	**[xi; xi+1)**	**ni**	**wi**
1	**[26; 37)**	6	0,19
2	**[37; 48)**	4	0,12
3	**[48; 59)**	12	0,38
4	**[59; 70)**	5	0,16
5	**[70; 81)**	3	0,09
6	**[81; 92)**	2	0,06

§2. Вариационные ряды

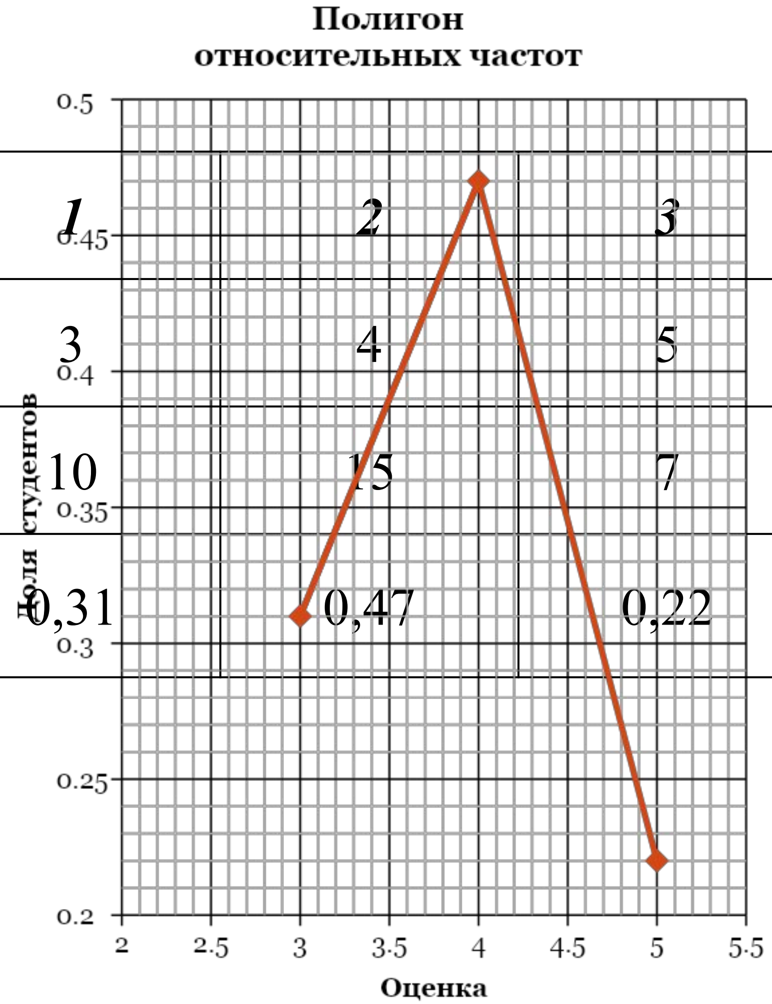
Вариационные ряды можно представить графически.

Для визуализации дискретного вариационного ряда используют **полигон** частот – ломаную, концы которой имеют координаты (x_i, n_i). Можно также построить полигон относительных частот – ломаную с концами в точках (x_i, w_i).

Здесь x_i – значение варианты, n_i – частота варианты, $w_i = n_i / n$ – относительная частота варианты.

§2. Вариационные ряды

Номер варианты	**i**	*1*	*2*	*3*
Варианта	**x_i**	3	4	5
Частота варианты	**n_i**	10	15	7
Относительная частота варианты	**w_i**	0,31	0,47	0,22



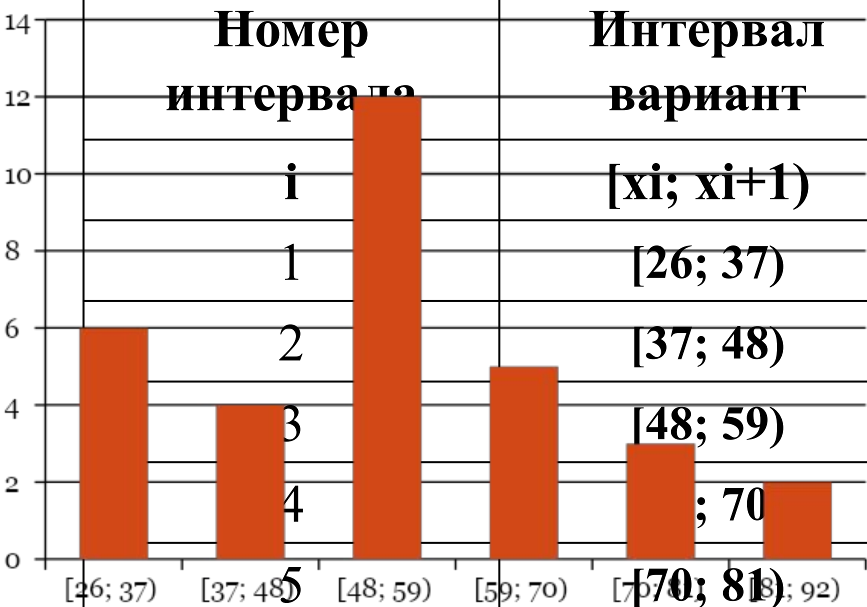
§2. Вариационные ряды

Для визуализации интервального вариационного ряда используют **гистограмму** частот – ступенчатую фигуру, составленную из прямоугольников; основание каждого прямоугольника совпадает с интервалом значений признака $[x_i\ ;\ x_{i+1})$, а высота прямоугольника равна n_i – сумме частот вариант, попавших в интервал. Можно также построить гистограмму относительных частот, в которой высоты прямоугольников равны относительным частотам интервалов вариант w_i .

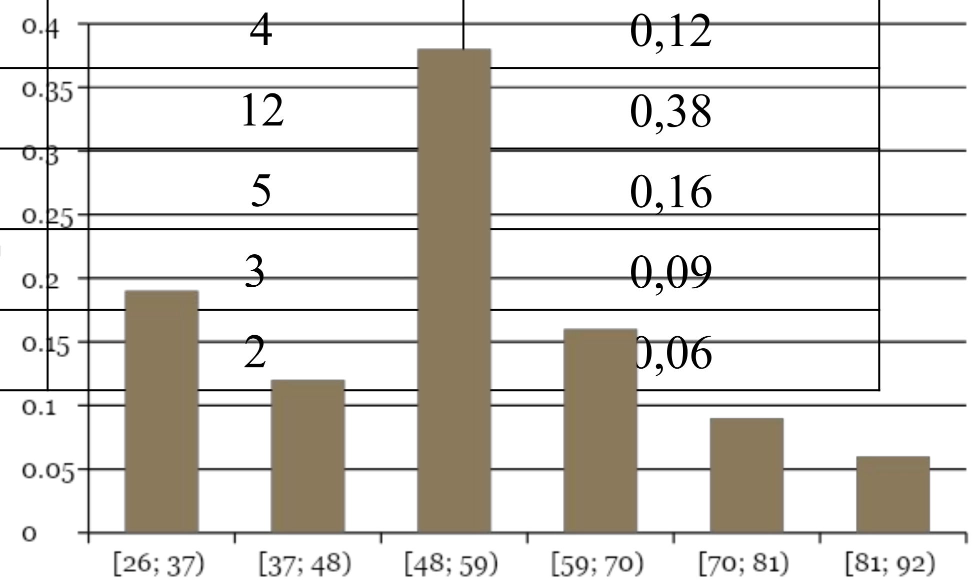
§2. Вариационные ряды

Гистограмма частот

Номер интервала	Интервал вариант	Частота	Частость
i	[xi; xi+1)	ni	wi
1	[26; 37)	6	0,19
2	[37; 48)	4	0,12
3	[48; 59)	12	0,38
4	[illegible]; 70	5	0,16
5	[70; 81)	3	0,09
6	[81; 92)	2	0,06



Гистограмма относительных частот



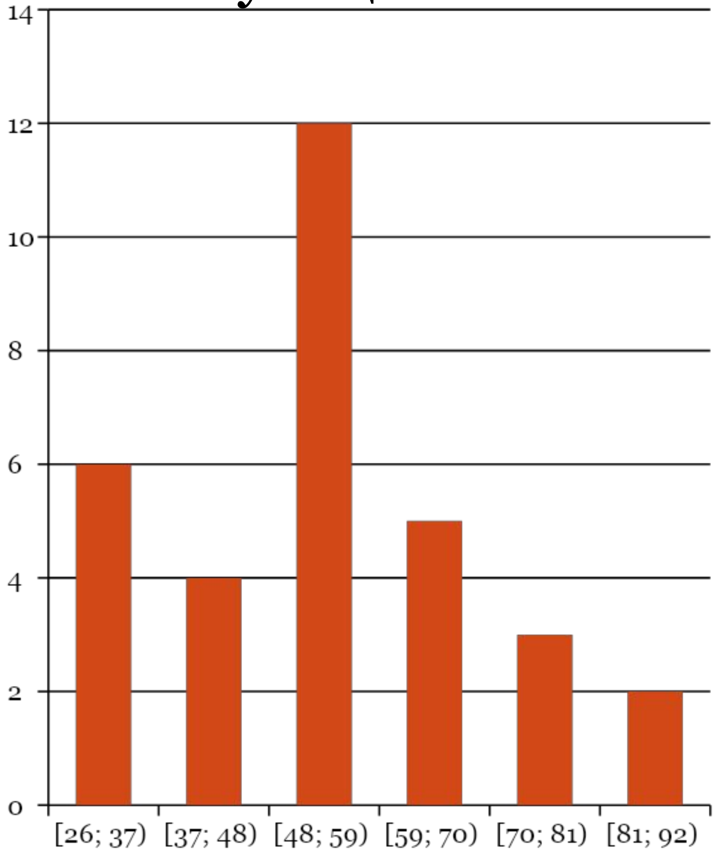
§2. Вариационные ряды

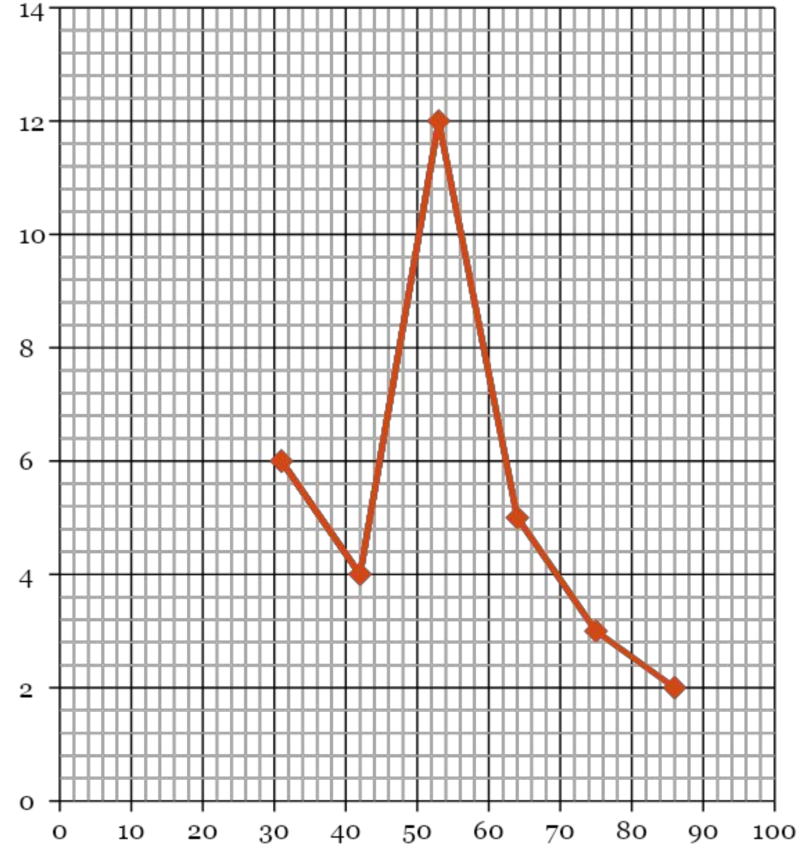
Иногда на практике интервальный вариационный ряд преобразуют в дискретный, заменяя каждый частичный интервал его серединой.

Номер интервала	Интервал / Середина интервала	Частота	Относительная частота
i	**[xi;xi+1) / xi***	**ni**	**wi**
1	**[26;37) / 31,5**	6	0,19
2	**[37;48) / 42,5**	4	0,12
3	**[48;59) / 53,5**	12	0,38
4	**[59;70) / 64,5**	5	0,16
5	**[70;81) / 75,5**	3	0,09
6	**[81;92) / 86,5**	2	0,06

§2. Вариационные ряды

В этом случае вместо гистограммы вариационного ряда для его визуализации используется соответствующий полигон.





§2. Вариационные ряды

Вариационный ряд является статистическим аналогом распределения признака X, а полигон или гистограмма играют роль кривой распределения.

Эмпирической функцией распределения называется относительная частота (частость) того, что признак X примет значение, меньшее заданного значения x, т.е. представляет собой накопленную частость варианты:

$$F_n\left(x\right) = w\left(X < x\right) = w_x^{\text{нак}} = \frac{n_x^{\text{нак}}}{n}. \qquad (3)$$

Эмпирическая функция распределения являются статистическим аналогом функции распределения случайной величины.

§3. Меры усреднения данных

Вариационный ряд содержит полную информацию об изменчивости признака X. Однако часто бывает достаточно информации лишь о сводных характеристиках выборки: средних величинах и показателях изменчивости. Расчёт таких характеристик и представляет собой процедуру обработки данных наблюдений.

Средние величины «демонстрируют» значения признака вокруг которых наблюдения некоторым образом «концентрируются», т.е. средние величины характеризуют так называемую центральную тенденцию.

§3. Меры усреднения данных

Мода вариационного ряда – это значение признака с наибольшей частотой.

$$Mo = 4.$$

Номер варианты	i	*1*	*2*	*3*
Варианта	**xi**	3	$\boxed{4}$	5
Частота варианты	**ni**	10	$\boxed{15}$	7
Относительная частота варианты	**wi**	0,31	0,47	0,22

§3. Меры усреднения данных

Выборочная средняя – среднее арифметическое всех вариант.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i \cdot n_i}{n}, \qquad \bar{x} = \frac{3\cdot 10 + 4\cdot 15 + 5\cdot 7}{32} = \frac{30+60+35}{32} = \frac{125}{32} \approx 3,9.$$

Номер варианты	**i**	***1***	***2***	***3***
Варианта	**xi**	3	4	5
Частота варианты	**ni**	10	15	7
Относительная частота варианты	**wi**	0,31	0,47	0,22

§3. Меры усреднения данных

Медиана вариационного ряда – это значение признака, которое делит ранжированный ряд данных на две равные по объёму части. Если ряд содержит чётное количество вариант, то медиана равна среднему арифметическому двух вариант, стоящих в середине.

Номер варианты	**i**	***1***	***2***	***3***
Варианта	**xi**	3	4	5
Частота варианты	**ni**	10	15	7
Относительная частота варианты	**wi**	0,31	0,47	0,22

$$n = 32 \Rightarrow Me = \frac{x_{16} + x_{17}}{2} = \frac{4+4}{2} = 4.$$

§3. Меры усреднения данных

$$Mo = 53, \quad Me = 53,$$

$$\bar{x} = \frac{31\cdot 6 + 42\cdot 4 + 53\cdot 12 + 64\cdot 5 + 75\cdot 3 + 86\cdot 2}{32} \approx 53,3.$$

Номер интервала	Интервал / Середина интервала	Частота	Относительная частота
i	[xi;xi+1) / xi*	ni	wi
1	[26;37) / 31	6	0,19
2	[37;48) / 42	4	0,12
3	[48;59) / 53	12	0,38
4	[59;70) / 64	5	0,16
5	[70;81) / 75	3	0,09
6	[81;92) / 86	2	0,06

§4. Меры разброса данных

Изменчивость признака отражают показатели вариации. Наибольший интерес представляют меры рассеяния наблюдений вокруг средних величин.

Размах варьирования – это разница между наибольшей и наименьшей вариантами.

Номер варианты	**i**	***1***	***2***	***3***
Варианта	**xi**	3	4	5
Частота варианты	**ni**	10	15	7
Относительная частота варианты	**wi**	0,31	0,47	0,22

$$R = x_{\max} - x_{\min}, \quad R = 5 - 3 = 2.$$

§4. Меры разброса данных

Выборочная дисперсия – это среднее арифметическое квадратов отклонений вариант от их выборочной средней:

$$S^2 = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \cdot n_i}{n}.$$

Номер варианты	i	*1*	*2*	*3*
Варианта	**xi**	3	4	5
Частота варианты	**ni**	10	15	7
Относительная частота варианты	**wi**	0,31	0,47	0,22

$$S^2 = \frac{(3-3,9)^2 \cdot 10 + (4-3,9)^2 \cdot 15 + (5-3,9)^2 \cdot 7}{32} \approx 0,52.$$

§4. Меры разброса данных

«Исправленная» выборочная дисперсия определяется по формуле:

$$\overline{S}^2 = \frac{n}{n-1} \cdot S^2.$$

Номер варианты	**i**	***1***	***2***	***3***
Варианта	**xi**	3	4	5
Частота варианты	**ni**	10	15	7
Относительная частота варианты	**wi**	0,31	0,47	0,22

$$\overline{S}^2 = \frac{32}{31} \cdot 0,52 \approx 0,53.$$

§4. Меры разброса данных

$$R = 86 - 31 = 55$$

$$S^2 = \frac{(31-53,3)^2 \cdot 6 + (42-53,3)^2 \cdot 4 + (53-53,3)^2 \cdot 12 + (64-53,3)^2 \cdot 5 + (75-53,3)^2 \cdot 3 + (86-53,3)^2 \cdot 2}{32} \approx 238$$

$$\overline{S}^2 = \frac{n}{n-1} \cdot S^2 = \frac{32}{31} \cdot 238 \approx 246$$

$$\overline{S} = \sqrt{246} \approx 16$$

Номер интервала	Интервал / Середина интервала	Частота	Относительная частота
i	[xi;xi+1) / xi*	ni	wi
1	[26;37) / 31	6	0,19
2	[37;48) / 42	4	0,12
3	[48;59) / 53	12	0,38
4	[59;70) / 64	5	0,16
5	[70;81) / 75	3	0,09
6	[81;92) / 86	2	0,06

§5. Статистическое оценивание

Пусть распределение признака X – генеральной совокупности – задаётся функцией, которая содержит неизвестный параметр распределения θ. Об этом параметре судят по выборке, рассматривая варианты x_1, x_2, ..., x_n как значения n независимых случайных величин X_1, X_2, ..., X_n, которые имеют такой же закон распределения, что и признак X.

Оценкой $\tilde{\theta}_n$ параметра θ называется всякая функция результатов наблюдений над случайной величиной X, с помощью которой судят о значении параметра θ :

$$\tilde{\theta}_n = \tilde{\theta}_n\left(X_1, X_2, \ldots, X_n\right).$$

Оценка параметра является случайной величиной, зависящей от распределения признака X и числа n.

§5. Статистическое оценивание

О качестве оценки $\tilde{\theta}_n$ можно судить по выборочному распределению её значений

Оценка $\tilde{\theta}_n$ параметра θ называется **несмещённой**, если её математическое ожидание равно оцениваемому параметру: $M\left(\tilde{\theta}_n\right) = \theta$.

В противном случае оценку называют **смещённой**.

Требование несмещённости оценки гарантирует отсутствие систематических ошибок (всегда только преувеличивающих или только преуменьшающих результат наблюдения) при оценивании.

§5. Статистическое оценивание

Оценка $\tilde{\theta}_n$ параметра θ называется **состоятельной**, если она удовлетворяет закону больших чисел, т.е. сходится по вероятности к оцениваемому параметру:

$$\lim_{n\to\infty} P\left(\left|\tilde{\theta}_n - \theta\right| < \varepsilon\right) = 1.$$

Если оценка $\tilde{\theta}_n$ параметра θ является несмещённой, а её дисперсия $\sigma^2_{\tilde{\theta}_n} \to 0$ при $n \to \infty$, то оценка $\tilde{\theta}_n$ является состоятельной.

Несмещённая оценка $\tilde{\theta}_n$ параметра θ называется **эффективной**, если она имеет наименьшую дисперсию среди всех возможных несмещённых оценок параметра θ, вычисленных по выборкам одного и того же объёма n.

§5. Статистическое оценивание

В качестве статистических оценок параметров генеральной совокупности желательно использовать оценки, которые являются одновременно *несмещёнными, состоятельными и эффективными*. На практике, однако, это трудно достижимо.

Оценки параметров генеральной совокупности одним числом называют **точечными**. Для выборок небольшого объёма точечные оценки даже будучи несмещёнными, состоятельными и эффективными могут существенно отличаться от оцениваемого параметра.

§5. Статистическое оценивание

Выборочная доля повторной и бесповторной выборки есть *несмещённая и состоятельная* оценка генеральной доли.

Выборочная средняя повторной и бесповторной выборки есть *несмещённая и состоятельная* оценка генеральной средней.

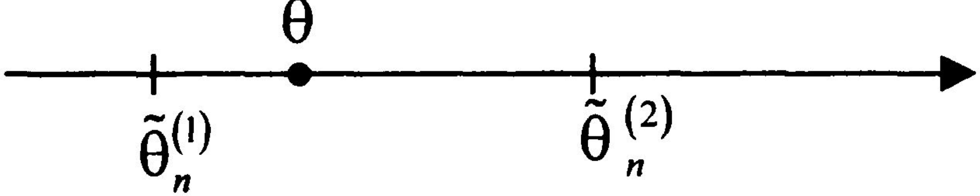
Выборочная дисперсия повторной и бесповторной выборки есть ***с**мещённая и состоятельная* оценка генеральной дисперсии.

«Исправленная» выборочная дисперсия повторной и бесповторной выборки есть *несмещённая и состоятельная* оценка генеральной дисперсии.

§6. Интервальное оценивание

Чтобы получить информацию о точности и надёжности оценки используют интервальное оценивание.

Интервальной оценкой параметра θ называется числовой интервал $\left(\tilde{\theta}_n^{(1)};\tilde{\theta}_n^{(2)}\right)$, который с заданной вероятностью γ накрывает неизвестное значение параметра θ.



Интервал $\left(\tilde{\theta}_n^{(1)};\tilde{\theta}_n^{(2)}\right)$ называется **доверительным**, а вероятность γ – доверительной вероятностью или **надежностью** оценки.

§6. Интервальное оценивание

На практике доверительный интервал параметра θ целесообразно выбирать симметричным относительно оценки $\tilde{\theta}_n$, т.е. в виде $\left(\tilde{\theta}_n - \Delta; \tilde{\theta}_n + \Delta\right)$. Положительное число Δ характеризует **точность** интервальной оценки параметра θ по выборке объёма n и называется **предельной ошибкой выборки**.

Итак, $P\left(\left|\tilde{\theta}_n - \theta\right| < \Delta\right) = \gamma$, т.е. с вероятностью γ выполняется неравенство $\tilde{\theta}_n - \Delta < \theta < \tilde{\theta}_n + \Delta.$

§6. Интервальное оценивание

Для построения доверительных интервалов для генеральной средней a и генеральной доли p используют точечные оценки: $\overline{x}$ – выборочную среднюю и ω – выборочную долю.

Пусть N и n – объёмы генеральной и выборочной совокупностей соответственно;

$\overline{S}^2$ – исправленная выборочная дисперсия;

γ – надёжность оценки;

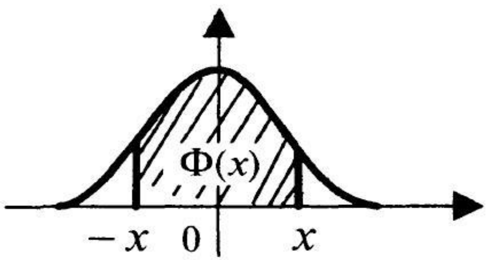
t – аргумент функции Лапласа $\Phi(t)$ и $\Phi(t) = \gamma$;

ξ – случайная величина, имеющая распределение Стьюдента с n-1 степенями свободы и $P\left(\left|\xi\right| < t_{n-1}\right) = \gamma$.

§6. Интервальное оценивание

Значения функции Лапласа

$$\Phi(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$$

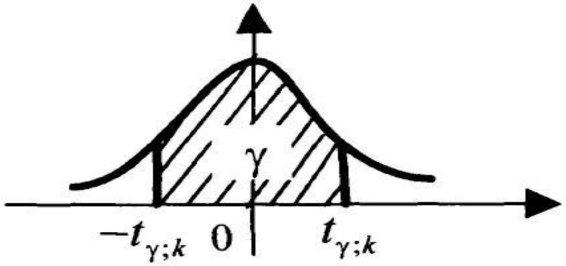


Целые и деся-тые доли x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0,0797	0,0876	0,0955	0,1034	0,1113	0,1192	0,1271	0,1350	0,1428	0,1507
0,2	0,1585	0,1663	0,1741	0,1819	0,1897	0,1974	0,2051	0,2128	0,2205	0,2282
0,3	0,2358	0,2434	0,2510	0,2586	0,2661	0,2737	0,2812	0,2886	0,2960	0,3035
0,4	0,3108	0,3182	0,3255	0,3328	0,3401	0,3473	0,3545	0,3616	0,3688	0,3759
0,5	0,3829	0,3899	0,3969	0,4039	0,4108	0,4177	0,4245	0,4313	0,4381	0,4448
0,6	0,4515	0,4581	0,4647	0,4713	0,4778	0,4843	0,4907	0,4971	0,5035	0,5098
0,7	0,5161	0,5223	0,5285	0,5346	0,5407	0,5467	0,5527	0,5587	0,5646	0,5705
0,8	0,5763	0,5821	0,5878	0,5935	0,5991	0,6047	0,6102	0,6157	0,6211	0,6265
0,9	0,6319	0,6372	0,6424	0,6476	0,6528	0,6579	0,6629	0,6679	0,6729	0,6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	0,7287	0,7330	0,7373	0,7415	0,7457	0,7499	0,7540	0,7580	0,7620	0,7660
1,2	0,7699	0,7737	0,7775	0,7813	0,7850	0,7887	0,7923	0,7959	0,7984	0,8029
1,3	0,8064	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	0,9616	0,9625	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	0,9331	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
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	0,9981	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986
3,2	0,9986	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,3	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
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	0,9996	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
3,7	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
4,0	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

§6. Интервальное оценивание

Значения $t_{\gamma,k}$-критерия Стьюдента

Число степеней свободы k	Вероятность γ											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99
1	0,16	0,32	0,51	0,73	1,00	1,38	1,96	3,08	6,31	12,71	31,82	63,66
2	14	29	44	62	0,82	06	34	1,89	2,92	4,30	6,96	9,92
3	14	28	42	58	76	0,98	25	64	35	3,18	4,54	5,84
4	13	27	41	57	74	94	19	53	13	2,78	3,75	4,60
5	13	27	41	56	73	92	16	48	01	57	36	03
6	0,13	0,26	0,40	0,55	1,72	1,91	1,13	1,44	1,94	2,45	3,14	3,71
7	13	26	40	55	71	90	12	41	89	36	00	50
8	13	26	40	55	70	89	11	40	86	31	2,90	35
9	13	26	40	54	70	88	10	38	83	26	82	25
10	13	26	40	54	70	88	09	37	81	23	76	17
11	0,13	0,26	0,40	0,54	0,70	0,88	1,09	1,36	1,80	2,20	2,72	3,11
12	13	26	39	54	69	87	08	36	78	18	68	05
13	13	26	39	54	69	87	08	35	77	16	65	01
14	13	26	39	54	69	87	08	34	76	14	62	2,98
15	13	26	39	54	69	87	07	34	75	13	60	95
16	0,13	0,26	0,39	0,53	0,69	0,86	1,07	1,34	1,75	2,12	2,58	2,92
17	13	26	39	53	69	86	07	33	74	11	57	90
18	13	26	39	53	69	86	07	33	73	10	55	88
19	13	26	39	53	69	86	07	33	73	09	54	86
20	13	26	39	53	69	86	06	32	72	09	53	84
21	0,13	0,26	0,39	0,53	0,69	0,86	1,06	1,32	1,72	2,08	2,52	2,83
22	13	26	39	53	69	86	06	32	72	07	51	82
23	13	26	39	53	68	86	06	32	71	07	50	81
24	13	26	39	53	68	86	06	32	71	06	49	80
25	13	26	39	53	68	86	06	32	71	06	48	79
26	0,13	0,26	0,39	0,53	0,68	0,86	1,06	1,31	1,71	2,06	2,48	2,78
27	13	26	39	53	68	85	06	31	70	05	47	77
28	13	26	39	53	68	85	06	31	70	05	47	76
29	13	26	39	53	68	85	05	31	70	04	46	76
30	13	26	39	53	68	85	05	31	70	04	46	75
40	0,13	0,25	0,39	0,53	0,68	0,85	1,05	1,30	1,68	2,02	2,42	2,70
60	13	25	39	53	68	85	05	30	67	00	39	66
120	0,13	0,25	0,39	0,53	0,68	0,84	1,04	1,29	1,66	1,98	2,36	2,62
∞	13	25	38	52	67	84	04	28	64	96	33	58



§6. Интервальное оценивание

Доверительные интервалы для генеральной средней a и генеральной доли p для выборок небольшого объёма $n \le 30$ строятся **только** для *нормальной* генеральной совокупности. При n>30 распределение Стьюдента можно приближённо заменить на стандартное нормальное распределение.

Параметр	Оценка	Предельная ошибка выборки			
		Повторная выборка		Бесповторная выборка	
		$n > 30$	$n \le 30$	$n > 30$	$n \le 30$
a	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\Delta = t\frac{\bar{S}}{\sqrt{n}}$	$\Delta = t_{n-1}\frac{\bar{S}}{\sqrt{n}}$	$\Delta = t\frac{\bar{S}}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$	$\Delta = \frac{t_{n-1}\bar{S}}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$
p	$\omega = \frac{m}{n}$	$\Delta = t\sqrt{\frac{\omega(1-\omega)}{n}}$		$\Delta = t\sqrt{\frac{\omega(1-\omega)}{n}}\sqrt{1-\frac{n}{N}}$	

§6. Интервальное оценивание

Уильям Сили Госсет
(1876-1937)

t-критерий Стьюдента
был разработан британским учёным У.Госсетом для оценки качества пива в компании «Гиннесс». В связи с обязательствами перед компанией по неразглашению коммерческой тайны (руководство считало таковой использование в своей работе статистического аппарата), в 1908г. статья вышла в журнале «Биометрика» под псевдонимом «Student».

§6. Интервальное оценивание

Для построения доверительного интервала генеральной дисперсии по выборке *нормальной* генеральной совокупности объёма n при неизвестных значениях генеральной средней a и генеральной дисперсии σ^2 используют статистику $Z = n\overline{S}^2 / \sigma^2$, которая имеет распределение «*хи-квадрат*» χ^2_{n-1}.

Доверительный интервал определяется условием:

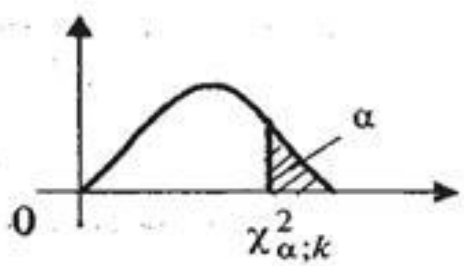
$$P\left(\frac{n\overline{S}^2}{z_2} < \sigma^2 < \frac{n\overline{S}^2}{z_1}\right) = \gamma,$$

а соотношения для выбора z_1 и z_2 по таблице распределения «*хи-квадрат*» χ^2_{n-1} имеют вид:

$$P\left(\chi^2_{n-1} > z_1\right) = \frac{1+\gamma}{2}, \quad P\left(\chi^2_{n-1} > z_2\right) = \frac{1-\gamma}{2}.$$

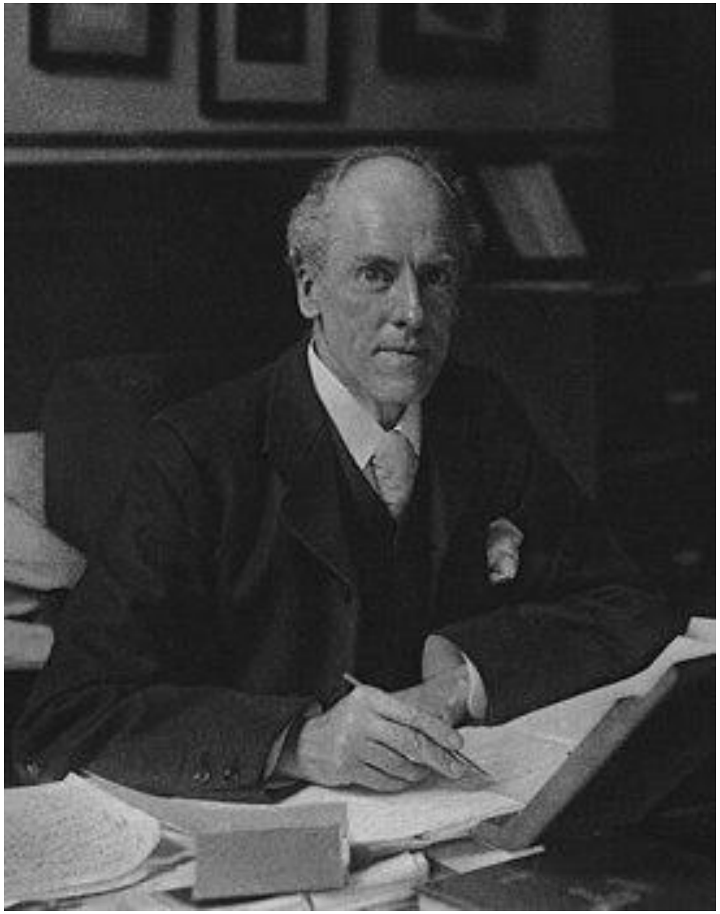
§6. Интервальное оценивание

Значения $\chi^2_{\alpha;k}$ критерия Пирсона



Число степеней свободы k	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
1	0,00	0,00	0,00	0,02	0,06	0,15	0,45	1,07	1,64	2,71	3,84	5,41	6,64
2	0,02	0,04	0,10	0,21	0,45	0,71	1,39	2,41	3,22	4,60	5,99	7,82	9,21
3	0,11	0,18	0,35	0,58	1,00	1,42	2,37	3,66	4,64	6,25	7,82	9,84	11,3
4	0,30	0,43	0,71	1,06	1,65	2,20	3,36	4,88	5,99	7,78	9,49	11,7	13,3
5	0,55	0,75	1,14	1,61	2,34	3,00	4,35	6,06	7,29	9,24	11,1	13,4	15,1
6	0,87	1,13	1,63	2,20	3,07	3,83	5,35	7,23	8,56	10,6	12,6	15,0	16,8
7	1,24	1,56	2,17	2,83	3,82	4,67	6,35	8,38	9,80	12,0	14,1	16,6	18,5
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,52	11,0	13,4	15,5	18,2	20,1
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,7	12,2	14,7	16,9	19,7	21,7
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,8	13,4	16,0	18,3	21,2	23,2
11	3,05	3,61	4,58	5,58	6,99	8,15	10,3	12,9	14,6	17,3	19,7	22,6	24,7
12	3,57	4,18	5,23	6,30	7,81	9,03	11,3	14,0	15,8	18,5	21,0	24,1	26,2
13	4,11	4,76	5,89	7,04	8,63	9,93	12,3	15,1	17,0	19,8	22,4	25,5	27,7
14	4,66	5,37	6,57	7,79	9,47	10,8	13,3	16,2	18,1	21,1	23,7	26,9	29,1
15	5,23	5,98	7,26	8,55	10,3	11,7	14,3	17,3	19,3	22,3	25,0	28,3	30,6
16	5,81	6,61	7,96	9,31	11,1	12,6	15,3	18,4	20,5	23,5	26,3	29,6	32,0
17	6,41	7,26	8,67	10,1	12,0	13,5	16,3	19,5	21,6	24,8	27,6	31,0	33,4
18	7,02	7,91	9,39	10,9	12,9	14,4	17,3	20,6	22,8	26,0	28,9	32,3	34,8
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
20	8,26	9,24	10,8	12,4	14,6	16,3	19,3	22,8	25,0	28,4	31,4	35,0	37,6
21	8,90	9,92	11,6	13,2	15,4	17,2	20,3	23,9	26,2	29,6	32,7	36,3	38,9
22	9,54	10,6	12,3	14,0	16,3	18,1	21,3	24,9	27,3	30,8	33,9	37,7	40,3
23	10,2	11,3	13,1	14,8	17,2	19,0	22,3	26,0	28,4	32,0	35,2	39,0	41,6
24	10,9	12,0	13,8	15,7	18,1	19,9	23,3	27,1	29,6	33,2	36,4	40,3	43,0
25	11,5	12,7	14,6	16,5	18,9	20,9	24,3	28,2	30,7	34,4	37,7	41,7	44,3
26	12,2	13,4	15,4	17,3	19,8	21,8	25,3	29,2	31,8	35,6	38,9	42,9	45,6
27	12,9	14,1	16,1	18,1	20,7	22,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0
28	13,6	14,8	16,9	18,9	21,6	23,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3
29	14,3	15,6	17,7	19,8	22,5	24,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6
30	14,9	16,3	18,5	20,6	23,4	25,5	29,3	33,5	36,2	40,3	43,8	48,0	50,9

§6. Интервальное оценивание



Карл П**и**рсон
(1857-1936)

Критерий согласия Пирсона или **критерий согласия χ^2** (хи – квадрат) был предложен английским математиком К. Пирсоном в 1900г.
Его работа рассматривается как фундамент современной математической статистики.
Это непараметрический метод, который позволяет оценить статистическую значимость различий двух или нескольких относительных показателей (частот, долей).

Продолжение следует…

математика всегда с вами!