ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Лекция 8.

Основные изучаемые вопросы:

- **1. Точечные оценки параметров генеральной совокупности.**
- **2. Ошибка выборочных наблюдений.**
- **3. Распределение Стьюдента (Госсета).**
- **4. Построение интервальных оценок.**
- **5. Интервальные оценки генеральной средней (математического ожидания).**

- ***В качестве точечных оценок параметров генеральной совокупности используются соответствующие выборочные характеристики.***
- ***Выборочная средняя является точечной оценкой генеральной средней***, т.е.

$$\overline{X}_{выб} = \overline{X}.$$

- Генеральная дисперсия имеет две точечные оценки:

$\sigma^2_{выб}$ - ***выборочная дисперсия***, исчисляется при $n \geq 30$

$$\sigma^2_{выб} = \frac{\sum_{i=1}^{k}(X_i - \overline{X}_{выб})^2 m_i}{n},$$

S^2 - ***исправленная выборочная дисперсия***, при $n < 30$

$$S^2 = \frac{n}{n-1}\sigma^2_{выб}.$$

- ***При больших объемах выборки $\sigma^2_{выб}$ и S^2 практически совпадают.***
- Для того чтобы статистики служили хорошими оценками параметров генеральной совокупности, они должны обладать рядом свойств: ***несмещенности, эффективности, состоятельности, достаточности.***
- ***Оценка называется несмещенной, если математическое ожидание выборки равно оцениваемому параметру.***
- ***Оценка называется эффективной, если при заданном объеме выборки она обеспечивает наименьшую дисперсию среди всех возможных несмещенных оценок.***

Оценка называется состоятельной, если она удовлетворяет закону больших чисел, т.е при увеличении объема выборки ($n \to \infty$) сходится по вероятности к оцениваемому параметру

$$\lim_{n\to\infty} P(|\overline{X}_{выб} - \overline{X}_{ген}) < \varepsilon) = 1$$

Оценка называется достаточной, если она содержит всю информацию об оцениваемом параметре.

1. Всем указанным свойствам отвечает <u>выборочная средняя</u>.

$$\overline{X}_{выб} = \overline{X}.$$

2. Выборочная дисперсия $\sigma^2_{выб}$ - <u>смещенная оценка</u>. Для устранения смещения при малых выборках вводится поправка

$$\frac{n}{n-1}.$$

ОШИБКА ВЫБОРОЧНЫХ НАБЛЮДЕНИЙ

- ***Разность между генеральными характеристиками и соответствующими выборочными статистиками*** называется ***ошибкой выборки,*** или ***ошибкой репрезентативности***.
- Статистические методы позволяют оценить эту разность, которая зависит как от характеристик выборки, так и от ее объема. В процессе выборочного исследования параметры генеральной совокупности определяются ***в виде интервала***, построенного вокруг выборочной статистики. Из теоремы Чебышева следует, что

$$P(\bar{X}_{выб} - \Delta < \bar{X}_{ген} < \bar{X}_{выб} + \Delta) = 2\Phi_0(t) = \Phi(t) = \gamma.$$

- Таким образом определяется ***интервальная оценка генеральной средней, которая представляет собой доверительный интервал,*** содержащий оцениваемый параметр генеральной совокупности:

$$\bar{X}_{выб} - \Delta < \bar{X}_{ген} < \bar{X}_{выб} + \Delta,$$

где Δ - ***предельная ошибка выборки***.

- ***Интервальной оценкой*** называют оценку, которая определяется двумя числами - ***концами интервала, который с определенной вероятностью накрывает неизвестный параметр генеральной совокупности.***
- Для определения доверительного интервала необходимо вычислить ***предельную ошибку выборки*** *Δ*, позволяющую установить предельные границы, в которых с заданной вероятностью (надежностью) должен находиться параметр генеральной совокупности.
- ***Предельная ошибка выборки равна t-кратному числу средних ошибок выборки.***
- ***Коэффициент t позволяет установить, насколько надежно высказывание о том, что заданный интервал содержит параметр генеральной совокупности***.

- Если мы выберем коэффициент таким, что высказывание в 97 % случаев окажется правильным и только в 3 % - неправильным, то мы говорим - со статистической надежностью в 97 % доверительный интервал выборочной статистики содержит параметр генеральной совокупности. ***Статистической надежности*** в 97 % соответствует ***доверительная вероятность*** $\gamma = 0{,}97$.
- Если в 5 % случаев утверждение ***«параметр принадлежит доверительному интервалу»*** будет неверным, то 5 % задает ***уровень значимости*** (a = 0,05 - вероятность ошибки). Обычно в статистике уровень значимости выбирают таким, чтобы он не превысил 5 % (a < 0,05).
- ***Доверительная вероятность и уровень значимости дополняют друг друга до 1*** (или 100 %) и определяют надежность статистического высказывания.
- ***Имеет место соотношение***:

$$a = 1 - \gamma.$$

- Из теоремы Чебышева следует, что с вероятностью, сколь угодно близкой к единице, можно утверждать, что ***при достаточно большом объеме выборки и ограниченной дисперсии генеральной совокупности разность между выборочной средней и генеральной средней будет сколь угодно мала***

$$P\left\{\left|\overline{X}_{выб} - \overline{X}_{ген}\right| < \frac{t \cdot \sigma_{ген}}{\sqrt{n}}\right\} > 1 - \frac{1}{t^2},$$

где $\overline{X}_{выб}$ - средняя по совокупности выбранных единиц;
$\overline{X}_{ген}$ - средняя по генеральной совокупности;
$\sigma_{ген}$ - среднее квадратическое отклонение в генеральной совокупности;

n - объем выборочной совокупности.

Итак, ***о величине расхождения между параметром и статистикой можно судить лишь с определенной вероятностью, от которой зависит величина t.***

- Согласно центральной предельной теореме Ляпунова ***выборочные распределения статистик при $n > 30$***

  будут иметь <u>нормальное распределение</u> <u>независимо от того, какое распределение имеет генеральная совокупность</u>.
- ***В случае, если объем выборочной совокупности***

  $n < 30$, то <u>при определении величины t используют распределение Стьюдента</u>.
- Распределение Стьюдента приводится в таблицах. ***Величину t определяют, задаваясь***

  - уровнем значимости a;

  - числом степеней свободы $k = n - 1$,

  где n – ***объем выборочной совокупности***.

Значения функции Лапласа $\Phi(t) = P\ (|T| \leqslant (t_{табл});\ \ \Phi(-t) = -\Phi(t);\ \ \Phi(t) = \frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}\,dx$

Целые и десятые доли t	Сотые доли t									
	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	0,7994	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
1,7	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	0,9399	0,9412
1,9	0,9426	0,9439	0,9451	0,9464	0,9476	0,9488	0,9500	0,9512	0,9523	0,9534
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	0,9931	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
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	0,9990	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	0,999936	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
4,5	0,999994	–	–	–	–	–	–	–	–	–
5,0	0,99999994	–	–	–	–	–	–	–	–	–

Распределение Стьюдента (Госсета)

- Случайная величина *T* имеет распределение Стьюдента с *k* степенями свободы (под *k* обычно понимают размер выборки без единицы), если она определена на интервале (-∞, +∞) и имеет следующую плотность вероятности

$$f(t) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\ \Gamma(\frac{k}{2})}(1+\frac{t^2}{k})^{-\frac{k+1}{2}}, \quad t \in (-\infty, +\infty).$$

- График плотности вероятности распределения Стьюдента имеет вид, напоминающий нормальное распределение, однако спад значений *f(t)* более пологий, а максимум функции расположен ниже, чем у соответствующего нормального распределения.

- При стремлении k к бесконечности (уже при $k > 30$) распределение Стьюдента стремится к нормальному распределению.
- Математическое ожидание распределения Стьюдента равно 0 (оно является центрированным), а дисперсия равна

$$D(T) = \frac{k}{k-2}, \;\; k > 2.$$

- ***Использование распределения Стьюдента в математической статистике основано на следующей интерпретации.***
- Пусть Z и V – независимые случайные величины, причем Z распределена по нормированному нормальному закону с нулевым матожиданием и единичной дисперсией, а V имеет χ^2 – распределение с k степенями свободы.

- Тогда случайная величина

$$T = \frac{Z}{\sqrt{\frac{V}{k}}}$$

распределена по закону Стьюдента с k степенями свободы.

- Установлено, что распределение Стьюдента имеет случайная величина, представляющая собой отношение точности оценки ε к дисперсии математического ожидания этой оценки:

$$t = \frac{\varepsilon}{\sqrt{\frac{\sigma^2}{n}}}$$

Распределение Стьюдента (*t*-распределение)

v	Вероятность $\alpha = St(t) = P(\lvert T\rvert > t_{табл})$												
	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,043	6,859
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,327	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,583
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,833
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,868	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,402	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

ПОСТРОЕНИЕ ИНТЕРВАЛЬНЫХ ОЦЕНОК

- Пусть ***найденная по данным выборки статистическая характеристика*** $\overline{X}_{выб}$ служит ***точечной оценкой*** неизвестного параметра $\overline{X}_{ген}$ генеральной совокупности.
- ***Доверительным интервалом*** $\Delta_{\gamma} = [\overline{X}_{выб} - \Delta; \overline{X}_{выб} + \Delta]$ для параметра X называет такой интервал, относительно которого можно утверждать с определенной, близкой к единице, вероятностью γ, что он содержит неизвестное значение параметра $\overline{X}_{ген}$.
- Величину γ называют ***доверительной вероятностью*** (надежностью) оценки параметра X. Величину Δ называют ***точностью оценки***. Нижняя и верхняя границы интервала равны:

$$X_{\min} = \overline{X}_{выб} - \Delta, \quad X_{\max} = \overline{X}_{выб} + \Delta.$$

- ***Ширина доверительного интервала***: $h = X_{\max} - X_{\min}$.
- Простейший способ построения интервальной оценки основан на использовании неравенства Чебышева.

- Пусть $\overline{X}_{выб}$ - несмещенная оценка параметра $\overline{X}_{ген}$, тогда

$$P\{|\overline{X}_{выб} - \overline{X}_{ген}| < \varepsilon\} > 1 - \frac{D_X}{\varepsilon^2}$$

(D_X предполагается существующей и известной), откуда доверительный интервал определяется как

$$h = [\overline{X}_{выб} - \varepsilon; \overline{X}_{выб} + \varepsilon].$$

- Итак, ***интервальное оценивание сводится к определению границ интервала, удовлетворяющему условию:***

$$P(X_{min} < \overline{X}_{ген} < X_{max}) = \gamma .$$

- Рассмотрим правила построения доверительных интервалов для параметров нормальной совокупности X на основании случайной выборки $x_1, x_2, \ldots , x_n$.

ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ГЕНЕРАЛЬНОЙ СРЕДНЕЙ (МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ)

- Пусть из генеральной совокупности *X*, имеющей нормальный закон распределения с математическим ожиданием μ и дисперсией σ^2 взята случайная выборка объемом *n*. В качестве основы интервальной оценки математического ожидания используется точечная оценка - ***среднее арифметическое*** $\bar{x}$, ***относительно которого строится симметричный интервал***.
- ***Правила построения доверительного интервала для математического ожидания зависят от того, известна или неизвестна дисперсия генеральной совокупности*** σ^2.

1. ***Доверительный интервал для*** μ ***при известной дисперсии*** σ^2.

- В этом случае ***полагают распределенной по нормальному закону величину***

$$t = \frac{\bar{x} - \mu}{\sigma}\sqrt{n}$$

Тогда $$P\left\{\left|\frac{\bar{x}-\mu}{\sigma}\sqrt{n}\right|<t_\gamma\right\}=P\{\bar{x}-t_\gamma\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+t_\gamma\frac{\sigma}{\sqrt{n}}\}=$$

$$=\frac{1}{2}[\Phi(t_\gamma)-\Phi(-t_\gamma)]=\Phi(t_\gamma)=\gamma,$$

где $\Phi(t)$ – интегральная функция Лапласа.

- Итак, ***построение доверительного интервала с заданной надежностью γ для генеральной средней при известной генеральной дисперсии осуществляется по формуле:***

$$P\{\bar{x}-t_\gamma\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+t_\gamma\frac{\sigma}{\sqrt{n}}\}=\Phi(t_\gamma)=\gamma.$$

где t_γ – значение стандартной нормальной величины, соответствующее надежности γ:

$$t_\gamma = \Phi^{-1}(\gamma).$$

- ***Точность оценки генеральной средней равна***

$$\delta = t_\gamma\frac{\sigma}{\sqrt{n}}.$$

Пример. Анализ доходности акций на основе случайной выборки за 16 дней показал, что средняя доходность составляет 10,37 %. Предполагая, что доходность акций подчиняется нормальному закону распределения:

- **А).** Определить ширину доверительного интервала для средней доходности с надежностью $\gamma = 0{,}97$, если известно, что $\sigma = 2$ %;
- **Б).** Найти доверительную вероятность того, что точность оценивания составит $\Delta = 0{,}98$;
- **В).** Определить минимальное число наблюдений, которое необходимо провести, чтобы с вероятностью $\gamma = 0{,}99$ можно было утверждать, что средняя доходность заключена в интервале шириной 3 %.

Решение.

- **А).** Так как дисперсия генеральной совокупности известна, то при построении доверительного интервала для генеральной средней будем исходить из формулы

$$P\{\bar{x}-t_\gamma \frac{\sigma}{\sqrt{n}} < \mu < \bar{x}+t_\gamma \frac{\sigma}{\sqrt{n}}\} = \Phi(t_\gamma) = \gamma .$$

- Для заданной надежности γ определим значение

$$t_\gamma = \Phi^{-1}(\gamma)$$

по таблице функции Лапласа $\Phi^{-1}(0{,}97) = 2{,}17$, ***откуда ширина доверительного интервала средней доходности***

$$h = 2\Delta = 2t_\gamma \frac{\sigma}{\sqrt{n}} = 2 \cdot 2{,}17 \cdot \frac{2}{\sqrt{16}} = 2{,}17\,\%.$$

Значения функции Лапласа $\Phi(t) = P(|T| < t_{табл})$; $\Phi(-t) = -\Phi(t)$; $\Phi(t) = \frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dx$

Целые и десятые доли t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
					....					
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

Б). Точность оценивания генеральной средней определяется как

$$\Delta = t_\gamma \frac{\sigma}{\sqrt{n}},$$

откуда

$$t_\gamma = \Delta \frac{\sqrt{n}}{\sigma}.$$

Следовательно, ***доверительная вероятность интервального оценивания генеральной средней при известной дисперсии равна***:

$$\gamma = \Phi(t_\gamma) = \Phi(\Delta \frac{\sqrt{n}}{\sigma}) = \Phi(0{,}98 \frac{\sqrt{16}}{2}) = \Phi(1{,}96).$$

По таблицам функции Лапласа $\gamma = \Phi(1{,}967) = 0{,}95.$

Значения функции Лапласа $\Phi(t) = P\,(|T| < (t_{табл});\ \Phi(-t) = -\Phi(t);\ \Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{x^2}{2}} dx$

Целые и десятые доли t	Сотые доли t									
	0	1	2	3	4	5	6	7	8	9
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	0,9399	0,9412
1,9	0,9426	0,9439	0,9451	0,9464	0,9476	0,9488	0,9500	0,9512	0,9523	0,9534
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

- В). Ширина доверительного интервала генеральной средней определяется выражением

$$h = 2\Delta = 2t_\gamma \frac{\sigma}{\sqrt{n}}.$$

- Отсюда

$$n = (2t_\gamma \frac{\sigma}{h})^2.$$

- Для заданной надежности γ определим значение

$$t_\gamma = \Phi^{-1}(\gamma)$$

по таблицам функции Лапласа, $t_\gamma = \Phi^{-1}(0,99) = 2,58$, откуда ***минимальное число наблюдений, которое необходимо провести, чтобы с вероятностью***

$\gamma = 0,99$ можно было утверждать, что средняя доходность заключена в интервале шириной 3 %, равно:

$$n = (2t_\gamma \frac{\sigma}{h})^2 = (2\cdot 2,58\frac{2}{3})^2 = 11,83.$$

- ***2. Доверительный интервал для*** μ ***<u>при неизвестной дисперсии</u>*** σ^2.
- ***В этом случае полагают величину*** t_a ***<u>распределенной по закону распределения Стьюдента</u>*** (t-распределение) с $k = n - 1$ степенями свободы:

$$P\left\{\left|\frac{\bar{x}-\mu}{S}\sqrt{n-1}\right|<t_a\right\} = P\{\bar{x}-t_a\frac{S}{\sqrt{n-1}} < \mu < \bar{x}+t_a\frac{S}{\sqrt{n-1}}\} = \gamma.$$

- ***Построение доверительного интервала с заданной надежностью*** γ ***для генеральной средней при неизвестной генеральной дисперсии осуществляется по формуле:***

$$P\{\bar{x}-t_a\frac{S}{\sqrt{n-1}} < \mu < \bar{x}+t_a\frac{S}{\sqrt{n-1}}\} = \gamma = 1-a,$$

где t_a - значение функции распределения Стьюдента (t-распределения), соответствующее $k = n - 1$ степеням свободы и вероятности

$$\alpha = 1-\gamma\ ; \qquad t_a = St^{-1}\ (\alpha = 1 - \gamma;\ k = n - 1).$$

- ***Точность оценки генеральной средней равна***

$$\Delta = t_\alpha \frac{S}{\sqrt{n-1}}.$$

Пример. По данным предыдущего примера, при условии, что на основе случайной выборки за 16 дней получена оценка $S = 2{,}5$ %

- **А).** Определить верхнюю границу доверительного интервала для средней доходности с надежностью $\gamma = 0{,}9$;
- **Б).** Найти доверительную вероятность того, что средняя доходность заключена в интервале (10,35 %; 10,39 %).

Решение.

- **А).** Так как точное значение дисперсии генеральной совокупности неизвестно, то при построении доверительного интервала для генеральной средней будем исходить из формулы

$$P\{\bar{x} - t_a \frac{S}{\sqrt{n-1}} < \mu < \bar{x} + t_a \frac{S}{\sqrt{n-1}}\} = \gamma = 1 - a,$$

- Для заданной надежности γ определим значение

$$t_a = St^{-1}(\alpha = 1 - \gamma\ ;\ \nu = n - 1)$$

по таблице t-распределения Стьюдента

$$t_a = St^{-1}(1 - 0{,}9;\ 16 - 1) = St^{-1}(0{,}1;\ 15) = 1{,}753,$$

откуда верхняя граница доверительного интервала

$$\mu_{\max} = \bar{x} + \Delta = \bar{x} + t_a \frac{S}{\sqrt{n-1}} = 10{,}37 + 1{,}753 \frac{2{,}5}{\sqrt{16-1}} = 11{,}502\,\%.$$

Распределение Стьюдента (t-распределение)

ν	Вероятность $\alpha = St(t) = P(\lvert T\rvert > t_{табл})$											
	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,833

- Б. Поскольку интервал (10,35 %; 10,39 %) симметричен относительно точечной оценки математического ожидания (μ = 10,37 %), ***точность оценивания генеральной средней при неизвестной дисперсии определяется как***

$$\Delta = t_\alpha \frac{S}{\sqrt{n-1}},$$

откуда

$$t_a = \Delta \frac{\sqrt{n-1}}{S} = 2\frac{\sqrt{16-1}}{2,5} = 3,098.$$

Распределение Стьюдента (t-распределение)

v	Вероятность $\alpha = St(t) = P(\|T\| > t_{табл})$												
	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965

- Далее в таблице *t*-распределения Стьюдента для числа степеней свободы $k = n - 1 = 16 - 1 = 15$ берем ближайшее к полученному значению t и получаем ***приближенное значение надежности***:

$$\gamma = 1 - \alpha = 1 - St\,(t_{\alpha};\, n - 1) = 1 - St\,(3{,}098;\, 15) \approx$$
$$\approx 1 - St\,(2{,}947;\, 15) = 1 - 0{,}01 = 0{,}99.$$

- Чтобы получить более точное значение вероятности $\alpha = St(t_{\alpha};\, n - 1)$ и надежности γ , необходимо прибегнуть к ***методу линейной интерполяции*** в таблице *t*-распределения Стьюдента.

- 1. Производятся измерения размера детали с помощью штангенциркуля.
- 2. Генеральная совокупность измерений включает 20 результатов:

10,3 мм	10,1 мм	10,2 мм	10,3 мм	10,0 мм
10,1 мм	10,3 мм	10,2 мм	10,1 мм	10,3 мм
9,9 мм	9,7 мм	9,8 мм	10,2 мм	9,7 мм
9,7 мм	10,2 мм	9,9 мм	9,8 мм	9,9 мм

- 3. Определите ***математическое ожидание и дисперсию*** размера детали:
  - по всей генеральной совокупности;
  - по выборочной совокупности из серии, включающей первые десять измерений (две верхних строки);
  - по выборочной совокупности из серии, включающей вторую группу из десяти измерений (две нижних строки).

- 4. Постройте ***вариационный ряд и кумуляту*** с интервалами, равными 0,1 мм, для генеральной совокупности измерений.
- 5. Определите ***ширину доверительного интервала*** при доверительной вероятности 0,95 для первой выборки и для всей генеральной совокупности.