Статистические закономерности малых выборок. Распределение Стьюдента.

Лекция 4

План лекции:

- Этапы обработки малой выборки с использованием распределения Стьюдента
- Определение достоверности различия двух зависимых выборочных совокупностей (разностный метод)
- Определение достоверности различия двух независимых выборочных совокупностей
- Сводка основных формул

Этапы обработки малой выборки с использованием распределения Стьюдента

Стьюдент – псевдоним английского математика Госсета.

Обозначения для **генеральной совокупности**:

μ – среднее арифметическое;

D – дисперсия **(σ^2);**

σ– среднее квадратическое отклонение.

Обозначения для **выборочной совокупности**:

$\bar{x}$ – среднее арифметическое;

s^2 – дисперсия

s – среднее квадратическое отклонение.

- Пусть дан ряд значений пульса (ЧСС) у больных: 95 130 83 115 120

1. Найдем среднее арифметическое значение выборки:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + ... + x_5}{n} = \frac{95 + 130 + 83 + 115 + 120}{5} = 108,6$$

2. Вычислим дисперсию (рассеивание ряда)

$$D(x) = s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$$

где df = n-1

число степеней свободы

$$D(x) = \frac{(95 - 108,6)^2 + (130 - 108,6)^2 + ... + (120 - 108,6)^2}{4} = 367,3$$

3. Среднее квадратическое отклонение выборки:

$$s = \sqrt{D} = \sqrt{367{,}3} = 19{,}6$$

- Это - **точечные** (т.е. выраженные одним значением) параметры малой выборки.
- Результат записывается в виде:

ЧСС= $\mathbf{\bar{x} \pm s} = 108{,}6 \pm 19{,}16\,\mathbf{(уд\,/\,мин)}$

- Оценим генеральную совокупность по нашей выборке.

4. Определим среднюю величину расхождения между параметрами выборки и генеральной совокупности. Эту величину называют средней квадратической ошибкой (или средней ошибкой, ошибкой выборочности, стандартной ошибкой) s_x:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{19{,}16}{\sqrt{5}} = 8{,}57$$

Если объекты отобраны в выборку случайным образом, то чем **больше** ее размеры, тем **меньше** стандартная ошибка, а значит, **меньше** расхождения в выборочной и генеральной совокупностях.

- Критерий Стьюдента для малой выборки:

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$$

- Доверительный интервал имеет вид:

$$\mu = \bar{x} \pm t \cdot s_{\bar{x}}$$

- Распределение значений t отличается от нормального тем **сильнее**, чем **меньше** n. По мере увеличения **n**, t – распределение Стьюдента приближается к нормальному. При **n** $\geq$ **30** разница между ними практически **исчезает**.

- Определим доверительный интервал для уровня P=0,95 в нашем примере. По таблицам Стьюдента находим t для доверительной вероятности 0,95 и числа степеней свободы n-1=4: t=2,78, следовательно:

$$\mu = 108{,}6 \pm 2{,}78 \cdot 8{,}57 = 108{,}6 \pm 23{,}82$$

$$\text{или} \quad 84{,}78 \le \mu \le 132{,}42$$

Определение достоверности различия двух зависимых выборочных совокупностей (разностный метод)

- Исследовалась урожайность некоторой культуры на контрольной и опытной (внесено удобрение) делянах.

N	$X_К$(ц/га)	$X_Э$(ц/га)	Δd	$Δd^2$
1	31,6	31,6	0	0
2	24	24,2	0,2	0,04
3	24,6	24,8	0,2	0,04
4	28,6	29,1	0,5	0,25
5	29,1	29,9	0,8	0,64
6	30,1	31	0,9	0,81

- Для контрольной группы:

$$\overline{x}_к = \frac{31,6 + 24 + \ldots + 30,1}{6} = 27,9$$

$$D(x) = \frac{(31,6 - 27,9)^2 + (24 - 27,9)^2 + \ldots + (30,1 - 27,9)^2}{5} = 9,31$$

$$s = \sqrt{9,31 =} 3,05 \qquad s_{\bar{x}} = \frac{3,05}{\sqrt{6}} = 1,24$$

- Для разности:

$$\overline{\Delta d} = \frac{0 + 0,2 + \ldots + 0,9}{6} = 0,43$$

$$D(x) = \frac{0 + 0,04 + \ldots + 0,81}{5} = 0,356$$

$$s = \sqrt{0,359 =} 0,59 \qquad s_d = \frac{0,59}{\sqrt{6}} = 0,242$$

- Определим, достоверно ли определена средняя арифметическая разности:

$$t_{эксп} = \frac{|d|}{s_d} = \frac{0,43}{0,242} = 1,47$$

$t_{таб}(0,95;5)=2,57$

$\mathbf{t_{эксп} < t_{таб}}$ **,** **недостоверно!**

Это означает, что нельзя утверждать, что изменение урожайности определяется внесением удобрения. Действуют другие причины, которые не были нами учтены в анализе.

Определение достоверности различия двух независимых выборочных совокупностей

- Нормированное отклонение:

$$t = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{d}{s_d}$$

- Для n<30, ошибка разницы s_d определяется по формуле:

$$s_d = \sqrt{\frac{\sum(x_{1i} - \bar{x}_1)^2 + \sum\left(x_{2i} - \bar{x}_2\right)^2}{(n_1 - 1) + (n_2 - 1)} \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}$$

- Изучался привес кроликов. В корм опытной группе добавлялся кобальт.

X_1(К)	X_2(Оп)	ΔX_1	$(\Delta X_1)^2$	ΔX_2	$(\Delta X_2)^2$
504	580	22	484	-58	3364
560	692	34	1156	54	2916
580	700	-54	2916	62	3844
600	621	74	5476	17	289
420	640	-106	11236	2	4
530	561	4	16	-77	5929
490	680	-36	1296	42	1764
580	630	54	2916	8	64
470		-56	3136		

$$s_d = \sqrt{\frac{28632 + 18174}{8+7} \cdot \frac{9+8}{9 \cdot 8}} = \sqrt{736{,}76} = 27{,}2$$

$$t_{эксп} = \frac{\left|\overline{x_1} - \overline{x_2}\right|}{s_d} = 4{,}11$$

df=(n1-1)+(n2-1)=15

$t_{эксп} > t_{таб}$, **нулевая гипотеза отвергается**

Добавка кобальта в рацион **достоверно** увеличила привес кроликов.

Сводка основных формул

- Средняя арифметическая выборки

$$\bar{\mathbf{x}} = \frac{\sum_{\mathbf{i}=1}^{\mathbf{n}} \mathbf{x}_{\mathbf{i}}}{\mathbf{n}} = \frac{\mathbf{x}_1 + \mathbf{x}_2 + \ldots + \mathbf{x}_{\mathbf{n}}}{\mathbf{n}}$$

- Дисперсия

$$D(x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

- Среднее квадратическое отклонение

$$s = \sqrt{D(x)}$$

- Средняя квадратическая ошибка

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

- Критерий нормированного отклонения (по Стьюденту)

$$t = \frac{\overline{x} - \mu}{s_{\overline{x}}}$$

- Доверительный интервал

$$t_{P,n} \cdot s_{\overline{x}}$$

- Критерий $t_{эксПер}$ для определения достоверности средней арифметической одной выборки

$$t_{эксПер} = \frac{\overline{x}}{s_{\overline{x}}}$$

- Критерий $t_{экспер}$ разности средних арифметических двух выборок

а) n≥30

$$t_{экспер} = \frac{\left|\overline{X_1} - \overline{X_2}\right|}{\sqrt{S_{x1}^2 + S_{x2}^2}}$$

б) *n<30*

$$t_{экспер} = \frac{\left|\overline{X_1} - \overline{X_2}\right|}{\sqrt{\frac{\sum(X_{1i} - \overline{X_1})^2 + (X_{2i} - \overline{X_2})^2}{(n_1 - 1) + (n_2 - 1)} \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}}$$

БЛАГОДАРЮ ЗА ВНИМАНИЕ