ПРИЛОЖЕНИЕ 6

Ошибаться человеку свойственно,
но окончательно все запутать может только компьютер.
Мерфология, Пятый закон ненадежности

Проверка выборки на нормальность (метод моментов)

В случае выборок небольшого объема для проверки гипотезы о нормальности закона распределения можно использовать простые критерии, основанные на сравнении генеральных параметров распределения и их оценок, полученных по выборке. В качестве оцениваемых параметров удобнее всего брать моменты распределения – асимметрию и эксцесс.

Эксцесс выборочной совокупности можно вычислить по формуле:

$$E = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{\sigma} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)},$$

где n – численность выборки, x_i, i=1,2…,n – значения вариант выборки, $\overline{x}$ – выборочное среднее значение, σ – стандартное отклонение.

Дисперсию эксцесса можно определить формулой

$$D_E = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}.$$

Коэффициент A асимметрии выборочной совокупности и его дисперсия вычисляется по формулам:

$$A = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{\sigma} \right)^3, \qquad D_A = \frac{6(n-1)}{(n+1)(n+3)}.$$

Зная дисперсии D_A и D_E, можно оценить, значимо ли выборочные коэффициенты асимметрии и эксцесса отличаются от нуля. Если $|E| \leq 5\sqrt{D_E}$, а $|A| \leq 3\sqrt{D_A}$, то наблюдаемое распределение можно считать нормальным.

Пример П6. Проверить выборки X и Y (данные выделены цветом на рис. П.6.1) на нормальность и равенство дисперсий. Использовать уровень значимости α = 0,05.

При проверке двух выборок X и Y на нормальность и равенство дисперсий последовательность вычислений в среде Excel {Calc} следующая (рис. П6.1).

1. В диапазон A4:A13 и B4:B13 заносятся исходные данные по выборкам X и Y, в ячейку E2 – величина уровня значимости. Диапазонам данных выборок присваиваются имена X и Y соответственно.
2. В ячейке E3 формулой =СЧЁТ(X) {=COUNT(X)} определяются объемы выборок (в данном случае они одинаковы). Ячейке E3 присваивается имя n.
3. В ячейки E5:E9 для выборки X заносятся значения коэффициента асимметрии (формула =СКОС(X) {=SKEW(X)}), дисперсии асимметрии (формула =6*(n-1)/(n+1)/(n+3)), коэффициента эксцесса (формула =ЭКСЦЕСС(X) {=KURT(X)}) и его дисперсии (формула =24*n*(n-2) /(n+1)^2 *(n-3)/ (n+3)/(n+5)), а также выборочной дисперсии (формула =ДИСП(A4:A13) {=VAR(A4:A13)} (см. рис. П6.1).

	A	B	C	D	E	F	G	H
1	**Проверка нормальности и однородности выборок X**							
2				α=	0,05			**и Y**
3	*X*	*Y*		*n*=	10			
4	14	18				=СКОС(X)		
5	20	19		*Ax*=	-0,144	=6*(n-1)/(n+1)/(n+3)		
6	15	22		*D* **A** *x*=	0,378	=ЭКСЦЕСС(X)		
7	11	17		*Ex*=	-0,067			
8	16	24		*D* **E** *x*=	0,570	=24*n*(n-2)/(n+1)^2 *(n-3)/(n+3)/(n+5)		
9	13	21		σ_x^2 =	11,067	=ДИСП(X)		
10	16	25						
11	19	26		*Ay*=	-0,291	=СКОС(Y)		
12	15	24		*D* **Ay**=	0,378			
13	9	15		*Ey*=	-1,247	=ЭКСЦЕСС(Y)		
14				*D* **E** *y*=	0,570			
15				σ_y^2 =	13,878	=ДИСП(Y)		
16					=МАКС(E9;E15)/МИН(E9;E15)			
17			*F*=	1,254				
18			*Fкр*=	3,179	=FРАСПОБР(E2;n-1;n-1)			
20				ΙEΙ	$5\sqrt{D_E}$	ΙAΙ	$3\sqrt{D_A}$	
21			X	0,067	3,774	0,144	1,844	
22			Y	1,247	3,774	0,291	1,844	
24	Поскольку *F*< *Fкр*, то				*выборки X и Y однородны*			
26	Поскольку для обоих выборок выполняется							
27	$\lvert E\rvert \le 5\sqrt{D_E}$ и $\lvert A\rvert \le 3\sqrt{D_A}$							
28					*то выборки X и Y*			
29		*имеют нормальный закон распределения*						

Рис. П6.1. Скриншот листа MS Excel Проверка выборки на нормальность методом моментов

4. В ячейки E11:E15 аналогично п.3 заносятся соответствующие данные по выборке Y.

5. В ячейке D17 определяется значение критерия Фишера $F_{эмп} = \frac{\sigma_y^2}{\sigma_x^2} \geq 1 \quad \left(\sigma_y^2 \geq \sigma_x^2\right)$ формулой =МАКС(E9;E15) /МИН(E9;E15) {=MAX(E9;E15)/MIN(E9;E15)}. В ячейку D18 при помощи формулы =FРАСПОБР(E2; n-1; n-1) {=FINV(E2; n-1; n-1)} заносится соответствующее критическое значение $F_{кр}$.

6. Для удобства анализа в ячейки D21:G21 заносятся абсолютная величина коэффициента эксцесса (формула =ABS(E7)) и его пороговое значение $5\sqrt{D_E}$ (формула =5*КОРЕНЬ(E8) {=5*SQRT(E8)}), абсолютная величина коэффициента асимметрии (формула =ABS(E5)) и его пороговое значение $3\sqrt{D_A}$ (формула =3*КОРЕНЬ(E6) {=3*SQRT(E6)}) для данных выборки X. В ячейки D22:G22 заносятся аналогичные данные для выборки Y.

7. На основании расчетных данных и построенной таблицы делаются выводы:

- ✓ Поскольку $F < F_{кр}$, то <u>выборки X и Y однородны по дисперсии</u>.
- ✓ Из $|E_X| \leq 5\sqrt{D_{E_X}}$ и $|A_X| \leq 3\sqrt{D_{A_X}}$ следует, что <u>выборка X относится к нормальному распределению</u>.
- ✓ Из $|E_Y| \leq 5\sqrt{D_{E_Y}}$ и $|A_Y| \leq 3\sqrt{D_{A_Y}}$ следует, что <u>выборка Y</u> также <u>относится к нормальному распределению</u>.

Таким образом, в частности, для сравнения средних значений данных выборок можно использовать критерий Стьюдента.