Федеральное агентство по образованию

ГОУ ВПО «Уральский государственный технический университет – УПИ

имени первого Президента России Б.Н. Ельцина»

# Проверка гипотезы о нормальном распределении генеральной совокупности по критерию Пирсона и проверка гипотезы о виде распределения.

Екатеринбург 2016

Таблица для расчета показателей.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| xi | Кол-во, fi | xi \* fi | Накопленная частота, S | |x - xср|\*f | (x - xср)2\*f | Частота, fi/n |
| 72 | 5 | 360 | 5 | 0.88 | 0.15 | 0.00503 |
| 71 | 10 | 710 | 15 | 11.75 | 13.81 | 0.0101 |
| 70 | 15 | 1050 | 30 | 32.63 | 70.96 | 0.0151 |
| 73 | 20 | 1460 | 50 | 16.5 | 13.61 | 0.0201 |
| 74 | 25 | 1850 | 75 | 45.62 | 83.26 | 0.0252 |
| 75 | 30 | 2250 | 105 | 84.75 | 239.41 | 0.0302 |
| 76 | 35 | 2660 | 140 | 133.87 | 512.06 | 0.0352 |
| 61 | 40 | 2440 | 180 | 447 | 4995.27 | 0.0402 |
| 62 | 45 | 2790 | 225 | 457.88 | 4658.92 | 0.0453 |
| 63 | 50 | 3150 | 275 | 458.75 | 4209.08 | 0.0503 |
| 64 | 60 | 3840 | 335 | 490.5 | 4009.89 | 0.0604 |
| 65 | 65 | 4225 | 400 | 466.38 | 3346.29 | 0.0654 |
| 66 | 70 | 4620 | 470 | 432.25 | 2669.19 | 0.0704 |
| 67 | 75 | 5025 | 545 | 388.13 | 2008.59 | 0.0755 |
| 68 | 80 | 5440 | 625 | 334 | 1394.48 | 0.0805 |
| 69 | 85 | 5865 | 710 | 269.88 | 856.88 | 0.0855 |
| 96 | 90 | 8640 | 800 | 2144.25 | 51086.54 | 0.0905 |
| 88 | 95 | 8360 | 895 | 1503.37 | 23790.76 | 0.0956 |
| 77 | 55 | 4235 | 950 | 265.37 | 1280.41 | 0.0553 |
| 63 | 44 | 2772 | 994 | 403.7 | 3703.99 | 0.0443 |
| Итого | 994 | 71742 |  | 8387.46 | 108943.54 | 1 |

Для оценки ряда распределения найдем следующие показатели:  
**Показатели центра распределения**.  
*Средняя взвешенная* (выборочная средняя)  
https://chart.googleapis.com/chart?cht=tx&chl=\overline%7bx%7d%20=%20\frac%7b%20\sum%7bx\cdot%20f%7d%7d%7b\sum%7bf%7d%7d%20=%20\frac%7b71742%7d%7b994%7d%20=%2072.18  
**Показатели вариации**.  
*Абсолютные показатели вариации*.  
Размах вариации - разность между максимальным и минимальным значениями признака первичного ряда.  
R = Xmax - Xmin = 96 - 61 = 35  
*Дисперсия* - характеризует меру разброса около ее среднего значения (мера рассеивания, т.е. отклонения от среднего).  
https://chart.googleapis.com/chart?cht=tx&chl=D%20=%20\frac%7b\sum%7b(x_%7bi%7d%20-%20\overline%7bx%7d)%5e%7b2%7d%20f%7d%7d%7b\sum%7bf%7d%7d%20=%20\frac%7b108943.54%7d%7b994%7d%20=%20109.6  
*Несмещенная оценка дисперсии* - состоятельная оценка дисперсии (исправленная дисперсия).  
https://chart.googleapis.com/chart?cht=tx&chl=S%5e%7b2%7d%20%20=%20\frac%7b\sum%7b(x_%7bi%7d%20-%20\overline%7bx%7d)%5e%7b2%7d%20f%7d%7d%7b\sum%7bf%7d-1%7d%20%20=%20\frac%7b108943.54%7d%7b993%7d%20=%20109.71  
*Среднее квадратическое отклонение* (средняя ошибка выборки).  
https://chart.googleapis.com/chart?cht=tx&chl=\sigma%20%20=%20\sqrt%7bD%7d%20=%20\sqrt%7b109.601%7d%20=%2010.47  
Каждое значение ряда отличается от среднего значения 72.18 в среднем на 10.47  
*Оценка среднеквадратического отклонения*.  
https://chart.googleapis.com/chart?cht=tx&chl=s%20=%20\sqrt%7bS%5e%7b2%7d%20%7d%20=%20\sqrt%7b109.71%7d%20=%2010.47

**Проверка гипотез о виде распределения**.  
1. Проверим гипотезу о том, что Х распределено по *нормальному закону* с помощью критерия согласия Пирсона.  
https://chart.googleapis.com/chart?cht=tx&chl=K%20=%20\sum%7b\frac%7b(n_%7bi%7d%20-%20n\cdot%20_%7bi%7d)%5e%7b2%7d%7d%7bn\cdot%20_%7bi%7d%7d%7d  
где n\*i - теоретические частоты:  
https://chart.googleapis.com/chart?cht=tx&chl=n\cdot%20_%7bi%7d%20=%20\frac%7bn\cdot%20h%7d%7b\sigma%20%7d\phi%20_%7bi%7d  
Вычислим теоретические частоты, учитывая, что:  
n = 994, h=-1 (ширина интервала), σ = 10.47, xср = 72.18  
https://chart.googleapis.com/chart?cht=tx&chl=n\cdot%20_%7bi%7d%20=%20\frac%7b994\cdot%20-1%7d%7b10.47%7d\phi%20_%7bi%7d%20=%20-94.95\phi%20_%7bi%7d

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| i | xi | ui | φi | n\*i |
| 1 | 72 | -0.0167 | 0,3989 | -37.87 |
| 2 | 71 | -0.11 | 0,3961 | -37.61 |
| 3 | 70 | -0.21 | 0,3902 | -37.05 |
| 4 | 73 | 0.0788 | 0,3977 | -37.76 |
| 5 | 74 | 0.17 | 0,3925 | -37.27 |
| 6 | 75 | 0.27 | 0,3847 | -36.53 |
| 7 | 76 | 0.37 | 0,3725 | -35.37 |
| 8 | 61 | -1.07 | 0,2251 | -21.37 |
| 9 | 62 | -0.97 | 0,2468 | -23.43 |
| 10 | 63 | -0.88 | 0,2709 | -25.72 |
| 11 | 64 | -0.78 | 0,292 | -27.72 |
| 12 | 65 | -0.69 | 0,3144 | -29.85 |
| 13 | 66 | -0.59 | 0,3352 | -31.83 |
| 14 | 67 | -0.49 | 0,3521 | -33.43 |
| 15 | 68 | -0.4 | 0,3683 | -34.97 |
| 16 | 69 | -0.3 | 0,3802 | -36.1 |
| 17 | 96 | 2.28 | 0,0297 | -2.82 |
| 18 | 88 | 1.51 | 0,1257 | -11.93 |
| 19 | 77 | 0.46 | 0,3572 | -33.91 |
| 20 | 63 | -0.88 | 0,2709 | -25.72 |

Сравним эмпирические и теоретические частоты. Составим расчетную таблицу, из которой найдем наблюдаемое значение критерия:  
https://chart.googleapis.com/chart?cht=tx&chl=\chi%5e%7b2%7d%20=%20\sum%7b\frac%7b(n_%7bi%7d%20-%20n\cdot%20_%7bi%7d)%5e%7b2%7d%7d%7bn\cdot%20_%7bi%7d%7d%7d

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| i | ni | n\*i | ni-n\*i | (ni-n\*i)2 | (ni-n\*i)2/n\*i |
| 1 | 5 | -37.87 | -42.87 | 1838.19 | -48.53 |
| 2 | 10 | -37.61 | -47.61 | 2266.55 | -60.27 |
| 3 | 15 | -37.05 | -52.05 | 2709.01 | -73.12 |
| 4 | 20 | -37.76 | -57.76 | 3336.24 | -88.35 |
| 5 | 25 | -37.27 | -62.27 | 3877.12 | -104.04 |
| 6 | 30 | -36.53 | -66.53 | 4425.7 | -121.17 |
| 7 | 35 | -35.37 | -70.37 | 4951.59 | -140 |
| 8 | 40 | -21.37 | -61.37 | 3766.58 | -176.24 |
| 9 | 45 | -23.43 | -68.43 | 4683.05 | -199.85 |
| 10 | 50 | -25.72 | -75.72 | 5733.67 | -222.92 |
| 11 | 60 | -27.72 | -87.72 | 7695.57 | -277.57 |
| 12 | 65 | -29.85 | -94.85 | 8996.74 | -301.39 |
| 13 | 70 | -31.83 | -101.83 | 10368.55 | -325.79 |
| 14 | 75 | -33.43 | -108.43 | 11757.21 | -351.69 |
| 15 | 80 | -34.97 | -114.97 | 13217.82 | -377.99 |
| 16 | 85 | -36.1 | -121.1 | 14664.88 | -406.24 |
| 17 | 90 | -2.82 | -92.82 | 8615.54 | -3055.25 |
| 18 | 95 | -11.93 | -106.93 | 11435.05 | -958.13 |
| 19 | 55 | -33.91 | -88.91 | 7905.86 | -233.11 |
| 20 | 44 | -25.72 | -69.72 | 4861.02 | -188.99 |
| ∑ | 994 | 994 |  |  | -7710.64 |

Определим границу критической области. Так как статистика Пирсона измеряет разницу между эмпирическим и теоретическим распределениями, то чем больше ее наблюдаемое значение Kнабл, тем сильнее довод против основной гипотезы.  
Поэтому критическая область для этой статистики всегда правосторонняя: [Kkp;+∞).  
Её границу Kkp = χ2(k-r-1;α) находим по таблицам распределения χ2 и заданным значениям σ, k = 20, r=2 (параметры xcp и σ оценены по выборке.  
Kkp(0.05;17) = 27.58711; Kнабл = -7710.64  
Наблюдаемое значение статистики Пирсона не попадает в критическую область: Кнабл < Kkp, поэтому нет оснований отвергать основную гипотезу. Справедливо предположение о том, что данные выборки имеют **нормальное распределение**.  
4. Проверка гипотезы о *равномерном распределении* генеральной совокупности.  
Для того чтобы проверить гипотезу о равномерном распределении X,т.е. по закону: f(x) = 1/(b-a) в интервале (a,b)  
надо:  
1. Оценить параметры a и b - концы интервала, в котором наблюдались возможные значения X, по формулам (через знак \* обозначены оценки параметров):  
https://chart.googleapis.com/chart?cht=tx&chl=a%5e%7b\cdot%20%7d%20=%20\overline%7bx%7d%20-%20\sqrt%7b3%7d\sigma%20,%20b%5e%7b\cdot%20%7d%20=%20\overline%7bx%7d%20%2B%20\sqrt%7b3%7d\sigma  
2. Найти плотность вероятности предполагаемого распределения f(x) = 1/(b\* - a\*)  
3. Найти теоретические частоты:  
n1 = nP1 = n[f(x)\*(x1 - a\*)] = n\*1/(b\* - a\*)\*(x1 - a\*)  
n2 = n3 = ... = ns-1 = n\*1/(b\* - a\*)\*(xi - xi-1)  
ns = n\*1/(b\* - a\*)\*(b\* - xs-1)  
4. Сравнить эмпирические и теоретические частоты с помощью критерия Пирсона, приняв число степеней свободы k = s-3, где s - число первоначальных интервалов выборки; если же было произведено объединение малочисленных частот, следовательно, и самих интервалов, то s - число интервалов, оставшихся после объединения.  
Решение:  
1. Найдем оценки параметров a\* и b\* равномерного распределения по формулам:  
https://chart.googleapis.com/chart?cht=tx&chl=a%5e%7b\cdot%20%7d%20=%20\overline%7bx%7d%20-%20\sqrt%7b3%7d\sigma%20,%20b%5e%7b\cdot%20%7d%20=%20\overline%7bx%7d%20%2B%20\sqrt%7b3%7d\sigma  
https://chart.googleapis.com/chart?cht=tx&chl=a%5e%7b\cdot%20%7d%20=%2072.18%20-%20\sqrt%7b3%7d\cdot%2010.47%20=%2054.04,%20b%5e%7b\cdot%20%7d%20=%2072.18%20%2B%20\sqrt%7b3%7d\cdot%2010.47%20=%2090.31  
2. Найдем плотность предполагаемого равномерного распределения:  
f(x) = 1/(b\* - a\*) = 1/(90.31 - 54.04) = 0.0276  
3. Найдем теоретические частоты:  
n1 = n\*f(x)(x1 - a\*) = 994 \* 0.0276(72-54.04) = 492.2  
n20 = n\*f(x)(b\* - x19) = 994 \* 0.0276(90.31-63) = 748.48  
Остальные ns будут равны:  
ns = n\*f(x)(xi - xi-1)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| i | ni | n\*i | ni - n\*i | (ni - n\*i)2 | (ni - n\*i)2/n\*i |
| 1 | 5 | 492.2 | -487.2 | 237365.88 | 482.25 |
| 2 | 10 | -27.41 | 37.41 | 1399.41 | 0 |
| 3 | 15 | -27.41 | 42.41 | 1798.5 | 0 |
| 4 | 20 | 82.23 | -62.23 | 3872.08 | 47.09 |
| 5 | 25 | 27.41 | -2.41 | 5.8 | 0.21 |
| 6 | 30 | 27.41 | 2.59 | 6.71 | 0.24 |
| 7 | 35 | 27.41 | 7.59 | 57.63 | 2.1 |
| 8 | 40 | -411.13 | 451.13 | 203518.53 | 0 |
| 9 | 45 | 27.41 | 17.59 | 309.45 | 11.29 |
| 10 | 50 | 27.41 | 22.59 | 510.37 | 18.62 |
| 11 | 60 | 27.41 | 32.59 | 1062.19 | 38.75 |
| 12 | 65 | 27.41 | 37.59 | 1413.11 | 51.56 |
| 13 | 70 | 27.41 | 42.59 | 1814.02 | 66.18 |
| 14 | 75 | 27.41 | 47.59 | 2264.93 | 82.64 |
| 15 | 80 | 27.41 | 52.59 | 2765.85 | 100.91 |
| 16 | 85 | 27.41 | 57.59 | 3316.76 | 121.01 |
| 17 | 90 | 740.03 | -650.03 | 422544.85 | 570.98 |
| 18 | 95 | -219.27 | 314.27 | 98765.31 | 0 |
| 19 | 55 | -301.5 | 356.5 | 127089.07 | 0 |
| 20 | 44 | 748.48 | -704.48 | 496286.52 | 663.06 |
| Итого | 994 |  |  |  | 2256.91 |

Определим границу критической области. Так как статистика Пирсона измеряет разницу между эмпирическим и теоретическим распределениями, то чем больше ее наблюдаемое значение Kнабл, тем сильнее довод против основной гипотезы.  
Поэтому критическая область для этой статистики всегда правосторонняя: [Kkp;+∞).  
Её границу Kkp = χ2(k-r-1;α) находим по таблицам распределения χ2 и заданным значениям s, k (число интервалов), r=2 (параметры *a* и *b*).  
Kkp(17,0.05) = 27.58711; Kнабл = 2256.91  
Наблюдаемое значение статистики Пирсона попадает в критическую область: Кнабл > Kkp, поэтому есть основания отвергать основную гипотезу. Данные выборки распределены **не по равномерному закону**.