Практическое занятие 9: Оценка вероятности биномиального распределения

Цель занятия: приобретение навыков оценки доверительного интервала при биномиальном распределении.

Оценка вероятности **биномиального распределения** – это одна из частных задач по теме статистических оценок.

Пусть проводятся **независимые испытания**, *в каждом из которых* некоторое событие ***A*** может наступить с вероятностью ***p***, причём эта вероятность **не известна**.

Пример. Игровой автомат, в котором разыгрываются призы. Неизвестна вероятность ***p*** выигрыша в каждой попытке. Но её реально *оценить*, и оценить весьма точно.

Совершено ***n***=300 испытаний и выиграно ***m***=75 призов. Тогда ***относительная частота*** ***w=m/n***=75/300=0,25 представляет собой **точечную оценку** неизвестной вероятности ***p***.

Теперь предположим, что совершена другая серия испытаний (не обязательно 300 раз). Какой будет результат? Почти наверняка будет выиграна иная *доля* призов, то есть, будет получена другая относительная частота. И, проводя многократные серии испытаний, будет получено множество *точечных оценок*, которые будут варьироваться вокруг точного значения ***p***.

Как отмечалось ранее, недостаток точечной оценки состоит в том, что она может оказаться далека от истины (особенно, при малом ***n***) и поэтому вероятность ***p*** выгодно **оценить интервалом**:

$$m/n - \delta < p < m/n + \delta,$$

который с заранее выбранной **доверительной вероятностью** γ накроет истинное значение ***p***.

Справка: δ («дельта») называется *точностью оценки* и вышесказанное можно записать компактнее:

$P\left(\left|\frac{m}{n} - p\right| < \delta\right) = \gamma$ – вероятность, того, что относительная частота ***w=m/n*** отклонится от вероятности ***p*** менее чем на δ.

Задание 1

Проводят независимые испытания с одинаковой, но неизвестной вероятностью $\boldsymbol{p}$ появления события $\boldsymbol{A}$ в каждом испытании. Найти доверительный интервал для оценки вероятности $\boldsymbol{p}$ с надёжностью $\boldsymbol{\gamma}$=0,95, если в $\boldsymbol{n}$=300 испытаниях событие $\boldsymbol{A}$ появилось $\boldsymbol{m}$=75 раз.

Методика выполнения

если количество испытаний $\boldsymbol{n}$ достаточно велико (порядка сотни и больше) и значение $\boldsymbol{p}$ не слишком мало, то требуемый доверительный интервал можно построить по следующей приближенной формуле:

$$w - t_y\sqrt{\frac{w(1-w)}{n}} < p < w + t_y\sqrt{\frac{w(1-w)}{n}},$$

где $\boldsymbol{w=m/n}$ – относительная частота, а $\boldsymbol{t_y}$ – *коэффициент доверия*, отыскиваемый из соотношения $2\Phi(\boldsymbol{t_y})=\boldsymbol{\gamma}$.

***Примечание**: при этих условиях* биномиальное распределение *близко к* нормальному.

Вычислим относительную частоту $\boldsymbol{w=m/n}$ =75/300 = 0,25 и точность оценки $t_y\sqrt{\frac{w(1-w)}{n}}$.

Коэффициент доверия найдём из соотношения $2\Phi(\boldsymbol{t_y})=\boldsymbol{\gamma}$.

В данном случае: $2\Phi(\boldsymbol{t_y})=0{,}95$, следовательно: $\Phi(\boldsymbol{t_y})=0{,}475$ и по таблице значений функции Лапласа или с помощью MS Excel определяем, что этому значению функции соответствует аргумент $\boldsymbol{t_y}=1{,}96$.

Таким образом, точность оценки:

$$\delta = t_y\sqrt{\frac{w(1-w)}{n}} = 1{,}96\sqrt{\frac{0{,}25*0{,}75}{300}} = 0{,}049$$

и искомый доверительный интервал: $\boldsymbol{w - \delta < p < w + \delta}$

*0,25-0,049<**p**< 0,25+0,049*

Ответ: 0,201<$\boldsymbol{p}$<0,299 – с вероятностью $\boldsymbol{\gamma}$=0,95 этот интервал накрывает истинную вероятность выигрыша в игре.

Оценка получилась неплохая, но её неплохо бы улучшить, т.е. уменьшить значение $\delta = t_y\sqrt{\frac{w(1-w)}{n}}$, сузив тем самым интервал. Очевидно, что для этого нужно увеличить количество $\boldsymbol{n}$ испытаний, что совершенно логично. Есть вариант уменьшить коэффициент доверия $\boldsymbol{t_y}$, но тогда упадёт и доверительная вероятность, поэтому это плохой вариант.

Варианты задания 1

№ вар.	γ	**n**	**m**
1	**0,94**	**309**	**76**
2	**0,93**	**332**	**80**
3	**0,95**	**288**	**78**
4	**0,86**	**279**	**71**
5	**0,83**	**340**	**97**
6	**0,89**	**287**	**90**
7	**0,93**	**286**	**96**
8	**0,85**	**315**	**84**
9	**0,83**	**272**	**85**
10	**0,91**	**360**	**99**
11	**0,86**	**302**	**86**
12	**0,80**	**355**	**71**
13	**0,91**	**265**	**71**

№ вар.	γ	**n**	**m**
14	**0,92**	**286**	**75**
15	**0,96**	**271**	**92**
16	**0,97**	**342**	**74**
17	**0,86**	**269**	**95**
18	**0,87**	**271**	**79**
19	**0,94**	**349**	**76**
20	**0,81**	**320**	**76**
21	**0,91**	**300**	**85**
22	**0,87**	**326**	**67**
23	**0,88**	**291**	**91**
24	**0,86**	**350**	**95**
25	**0,97**	**285**	**83**
26	**0,88**	**299**	**73**

Задание 2

Из 500 поступивших на сортировку шариков для подшипников 200 попало в первую группу. В предположении о биномиальном распределении, определить:

1) доверительную вероятность того, что найденная доля шариков отклонится от вероятности попадания шарика в первую группу, менее чем на 0,03.

2) доверительную вероятность того, что вероятность попадания шарика в 1-ю группу будет накрыта интервалом (0,35; 0,50).

Методика выполнения

вычислим относительную частоту ***w=m/n***=200/500=0,4 и обозначим через ***p*** неизвестную вероятность того, что шарик попадёт в 1-ю группу.

1) Используем формулу $\gamma = P\left(\left|\frac{m}{n} - p\right| < \delta\right) = 2\Phi\left(\frac{\delta\sqrt{n}}{\sqrt{w(1-w)}}\right)$.

В данном случае ***n***=50, ***w***=0,4, ***δ***=0,03, таким образом: $\gamma = P(|0,4 - p| < 0,03) = 2\Phi\left(\frac{0,03\sqrt{500}}{\sqrt{0,4*0,6}}\right) \approx 2\Phi(1,37) - 2*0,4147 \approx 0,83$ – вероятность того, что,

что значение p будет отличаться от $w=m/n$ менее чем на $\delta=0{,}03$. Иными словами, интервал $(w-\delta;\ w+\delta)$ = (0,37; 0,43) с вероятностью $\gamma=0{,}83$ накрывает истинное значение p.

2) Предложенный доверительный интервал (0,35; 0,50) не симметричен относительно относительной частоты $w=0{,}4$ и имеет вид: $(w-\delta_1;\ w+\delta_2)$, где $\delta_1=0{,}05$, $\delta_2=0{,}1$.

Запишем левостороннюю точность оценки: $\delta_1 = t_{y1}\sqrt{\dfrac{w(1-w)}{n}}$ и найдём соответствующий коэффициент доверия: $0{,}05 = t_{y1}\sqrt{\dfrac{0{,}4*0{,}6}{500}} \Rightarrow t_{y1} = 0{,}05\sqrt{\dfrac{500}{0{,}24}} \approx 2{,}28$.

По таблице значений функции Лапласа: $\gamma_1 = \Phi(t_{y1}) \approx \Phi(2{,}28) = 0{,}4887$ – левосторонняя доверительная вероятность.

Аналогично для правой стороны: $\delta_{21} = t_{y2}\sqrt{\dfrac{w(1-w)}{n}}$

$$0{,}1 = t_{y2}\sqrt{\frac{0{,}4*0{,}6}{500}} \Rightarrow t_{y2} = 0{,}1\sqrt{\frac{500}{0{,}24}} \approx 4{,}56$$

$\gamma_2=\Phi(t_{y2})\approx \Phi(4{,}56) = 0{,}50$ – правосторонняя доверительная вероятность.

Таким образом, двусторонняя доверительная вероятность составляет:

$\gamma=\gamma_1+\gamma_2=$ 0,4887+0,5 = 0,9887 – иными словами с такой вероятностью интервал $0{,}35<p<0{,}50$ накрывает истинное значение p.

Ответ: а) $\gamma=0{,}83$, б) $\gamma=0{,}9887$.

Варианты задания 2

№ вар.	n	m	δ	Рнач	Рконеч	№ вар.	n	m	δ	Рнач	Ркон
1	258	523	0,021	0,4093	0,60671	14	218	528	0,038	0,348	0,508
2	254	514	0,023	0,3937	0,63825	15	222	469	0,021	0,296	0,667
3	174	466	0,038	0,4499	0,51296	16	253	574	0,029	0,322	0,657
4	222	435	0,022	0,3833	0,63662	17	195	509	0,023	0,372	0,595
5	243	555	0,033	0,3562	0,59329	18	212	581	0,022	0,46	0,616
6	165	423	0,021	0,3157	0,56486	19	198	541	0,022	0,399	0,551
7	187	460	0,037	0,3077	0,66407	20	231	549	0,023	0,449	0,51
8	213	444	0,026	0,375	0,49263	21	165	567	0,03	0,473	0,49
9	192	546	0,038	0,3712	0,6794	22	195	434	0,038	0,4	0,666
10	153	434	0,034	0,449	0,64117	23	202	489	0,037	0,383	0,542
11	159	567	0,025	0,408	0,49883	24	150	494	0,034	0,363	0,515
12	157	403	0,023	0,2869	0,63575	25	178	431	0,039	0,425	0,633
13	180	567	0,027	0,2951	0,58027	26	166	536	0,023	0,44	0,602

Задание 3

Проверив 1000 изделий, обнаружили, что 300 изделий высшего сорта. Сколько надо проверить изделий, чтобы с уверенностью 95% определить долю высшего сорта с точностью до 0,01?

Методика выполнения

Сразу вычислим относительную частоту ***w=m/n***=300/1000=0,3 и для исследовательского интереса найдём вероятность γ , с которой истинное значение ***p*** накрывается столь узким интервалом ***(w-δ; w+δ)*** = (0,29;0,31) .

Примечание: параметр ***p*** – есть вероятность того, что наугад извлечённое изделие окажется первосортным, его также называют **генеральной долей** (изделий высшего сорта).

Используем формулу $\gamma = P\left(\left|\frac{m}{n} - p\right| < \delta\right) = 2\Phi\left(\frac{\delta\sqrt{n}}{\sqrt{w(1-w)}}\right)$.

В данном случае:

$$\gamma = P(|0{,}3 - p| < 0{,}01) = 2\Phi\left(\frac{0{,}01\sqrt{1000}}{\sqrt{0{,}3*0{,}7}}\right) \approx 2\Phi(0{,}69) = 2*0{,}2549 \approx 0{,}51$$

– конечно, такое малое значение никуда не годится.

Поэтому в задаче и требуется обеспечить надёжность γ =0,95, и решение проводится по той же формуле:

$$\gamma = P\left(\left|\frac{m*}{n*} - p\right| < \delta\right) = 2\Phi\left(\frac{\delta\sqrt{n*}}{\sqrt{w(1-w)}}\right)$$

– откуда следует отыскать ***n****– объём выборки, обеспечивающий столь высокую доверительную вероятность.

В данном случае:

$$\gamma = P\left(\left|\frac{m*}{n*} - p\right| < 0{,}01\right) = 2\Phi\left(\frac{0{,}01\sqrt{n*}}{\sqrt{0{,}3*0{,}7}}\right) = 0{,}95$$

, следовательно:

$$2\Phi\left(\frac{0{,}01\sqrt{n*}}{\sqrt{0{,}3*0{,}7}}\right) = 0{,}95/2 = 0{,}475$$

– и по таблице значений функции Лапласа либо с помощью MS Excel выясняем, что этому значению функции соответствует аргумент 1,96: $\frac{0{,}01\sqrt{n*}}{\sqrt{0{,}3*0{,}7}}$ =1,96

теперь технически удобно возвести обе части в квадрат:

$$\left(\frac{0{,}01\sqrt{n*}}{\sqrt{0{,}3*0{,}7}}\right)^2 = 1{,}96^2$$

и найти искомый объём выборки:

$$\frac{0{,}0001n*}{0{,}21} \approx 3{,}8416$$

⇒ ***n****= 3,8416(0,21/0,0001) = 8067,36 ≈ 8068 – тут логично округлить в большую сторону.

Ответ: для того, чтобы с уверенностью 95% определить долю высшего сорта с точностью до 0,01, нужно проверить 8068 изделий.

Варианты задания 3

№ вар.	n	m	δ	p
1	1182	343	0,06	0,90
2	193	438	0,02	0,92
3	219	489	0,04	0,94
4	167	556	0,03	0,87
5	198	465	0,02	0,96
6	178	570	0,03	0,92
7	193	517	0,03	0,88
8	159	430	0,03	0,93
9	208	550	0,04	0,89
10	172	595	0,02	0,95
11	185	544	0,03	0,87
12	218	592	0,02	0,86
13	222	478	0,02	0,92

№ вар.	n	m	δ	p
14	212	574	0,03	0,95
15	190	598	0,04	0,96
16	205	475	0,03	0,93
17	253	450	0,02	0,89
18	160	437	0,03	0,96
19	229	591	0,02	0,92
20	236	439	0,03	0,90
21	162	540	0,03	0,90
22	159	515	0,04	0,94
23	172	542	0,03	0,87
24	233	419	0,04	0,92
25	241	410	0,03	0,93
26	226	543	0,03	0,96