Хи-квадрат статистикасына негізделген тексеру

Веб-тестілеу көбінесе A / B тестілеуінен асып түседі және бірден бірнеше нұсқаны тексереді. Статистикаға негізделген тексеру β-квадрат (2 β , хи-квадрат) сандық мәліметтермен бірге олардың күтілетін үлестірімге қаншалықты сәйкес келетінін анықтау үшін қолданылады.Статистикалық тәжірибеде айнымалылар арасында Тәуелсіздік туралы нөлдік гипотезаның негіздері бар-жоғын анықтау үшін RC × конъюгация кестелерімен бірге 2-ші статистиканы қолдану жиі кездеседі.2 - ші статистикаға негізделген тексеруді бастапқыда Карл Пир-Сонг 1900 жылы жасаған. "Хи" термині Пирсон өз мақаласында қолданған грек әрпінен шыққан.

Негізгі терминдер

2 статистикасы

(chi-square statistic) бақыланатын деректердің күтуден шегіну дәрежесін өлшейтін метрикалық көрсеткіш.

Күту немесе күту (expectation or expected) деректердің мінез-құлқы, біздің күткеніміз бойынша, қандай да бір ережеге сәйкес, әдетте нөлдік гипотеза болып табылады.

d. f.еркіндік дәрежелері.

Сынау χ2: қайта іріктеуге негізделген тәсіл

Сіз жарнаманың үш түрлі тақырыбын — A, B және C — тексеріп жатырсыз және олардың әрқайсысын 1000 келушіге тексересіз делік. Нәтижелер кестеде келтірілген. 3.4.

Таблица 3.4. *Результаты веб-тестирования трех разных заголовков*

	Заголовок *A*	**Заголовок *B***	**Заголовок *C***
Нажатия	14	8	12
Нет нажатий	986	992	988

Кесте 3.4. Үш түрлі тақырыптағы веб-тестілеу нәтижелері

Тақырыптар сөзсіз әр түрлі. A тақырыбы B тақырыбына қарағанда 2 есе көп түртуге мүмкіндік береді. Қайта таңдау процедурасы басу пайызы

кездейсоқтық тудыруы мүмкін деңгейден әлдеқайда жоғары екенін тексере алады. Мұндай тексеру үшін бізде "күтілетін" басу үлестірімі болуы керек, бұл жағдайда тексеру нөлдік гипотезаның болжамына сүйене отырып жүзеге асырылады, барлық үш тақырып 34/3000 - ға тең басудың жалпы пайызымен бірдей басу пайызына ие болады. Осы болжамға сүйене отырып, біздің кесте біріктірілген-жаңалықтар кестеде көрсетілгендей болады. 3.5.

Таблица 3.5. *Ожидаемое значение, если все три заголовка имеют одинаковый процент нажатий (нулевая гипотеза)*

	Заголовок *A*	**Заголовок *B***	**Заголовок *C***
Нажатия	11,33	11,33	11,33
Нет нажатий	988,67	988,67	988,67

Кесте 3.5. Егер барлық үш тақырыпта бірдей басу пайызы болса, күтілетін мән (нөлдік гипотеза)

Пирсонның қалдығы келесі формуламен берілген:

$$R = \frac{\text{наблюдаемое} - \text{ожидаемое}}{\sqrt{\text{ожидаемое}}}$$

және нақты шамалардың олардың күтілетін мөлшерінен қаншалықты ерекшеленетінін көрсетеді (кесте. 3.6).

Таблица 3.6. *Остатки Пирсона*

	Заголовок *A*	**Заголовок *B***	**Заголовок *C***
Нажатия	0,792	–0,990	0,198
Нет нажатий	–0,085	0,106	–0,021

Статистика 2 χ Пирсонның квадрат қалдықтарының қосындысы ретінде берілген:

$$\chi^2 = \sum_{i}^{r} \sum_{j}^{c} R^2,$$

мұндағы r және c сәйкесінше жолдар мен бағандар саны. Берілген мысал үшін 2- χ Статистика 1,666 құрайды.Бұл шынымен кездейсоқ модельде орын алуы мүмкін емес пе?

Біз мұны келесі қайта таңдау алгоритмімен тексере аламыз:

1.Қорапқа 34 бірлік (басу) және 2966 нөл (басу жоқ) салыңыз.

2.Араластыру, 1000 элементтен тұратын үш бөлек үлгіні алып тастау және әрқайсысында басуды санау.

3.Аралас шамалар мен күтілетін шамалар арасындағы квадраттық айырмашылықтарды тауып, оларды қорытындылаңыз.

4. 2 және 3-қадамдарды 1000 рет қайталаңыз. Квадраттық ауытқулардың қайта тексерілген қосындысы байқалғандардан қаншалықты жиі асып түседі? Бұл p мәні. Chisq функциясы.test қайта таңдалған статистиканы есептеу үшін пайдаланылуы мүмкін χ2 . Басу деректері үшін χ2-ші тексеру келесідей болады:

```
> chisq.test(clicks, simulate.p.value=TRUE)

  Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: clicks
X-squared = 1.6659, df = NA, p-value = 0.4853
```

Тексеру бұл нәтижені ерік-жігердің арқасында оңай алуға болатындығын көрсетеді.

Сынау χ2: статистикалық теория

Асимптотикалық статистикалық теория статистиканың таралуы χ2-ге жуықтауға болатындығын көрсетеді . Сәйкес стандартты үлестіру 2 χ еркіндік дәрежелерімен анықталады (бөлімді қараңыз."Бостандық дәрежелері" осы тараудың басында). Конъюгация кестесі үшін еркіндік дәрежелері жолдар (r) және бағандар (c) санымен келесідей байланысты:

$$\text{Бостандық дәрежелері} = (r-1)(c-1)$$

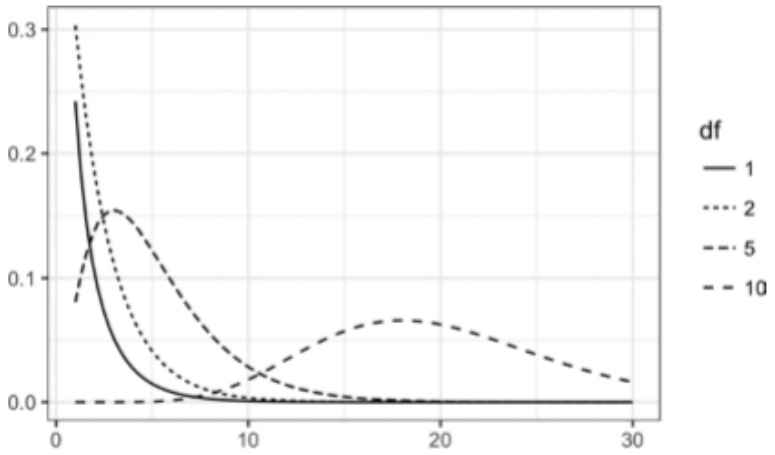
χ2-ші таралу, әдетте, ұзын құйрығы оңға қарай қиғаш болатын асимметриялық көрініске ие (суретті қараңыз. 3.7 1, 2, 5 және 10 дәрежелі еркіндікпен бөлуге қатысты).Бақыланатын статистика 2 χ үлестірімінде

неғұрлым алыс болса, p мәні соғұрлым төмен болады.Chisq функциясы.test эталон ретінде 2 χ үлестірімін қолдана отырып, p мәнін есептеу үшін қолданыла алады:

```
> chisq.test(cticks, simulate.p.value=FALSE)
Pearson's Chi-squared test test

data:  clicks
X-squared = 1.6659, df = 2, p-value = 0.4348
```

Берілген p мәні қайта іріктеудің p мәнінен сәл аз: бұл хи-квадрат үлестірімі статистиканың нақты үлестірімінің жуықтауы болып табылады.



Сурет. 3.7. Әр түрлі еркіндік дәрежелері бар χ2 таралуы (ықтималдық-у осінде, статистиканың χ2 мәні — x осінде)

Фишерді дәл тексеру.

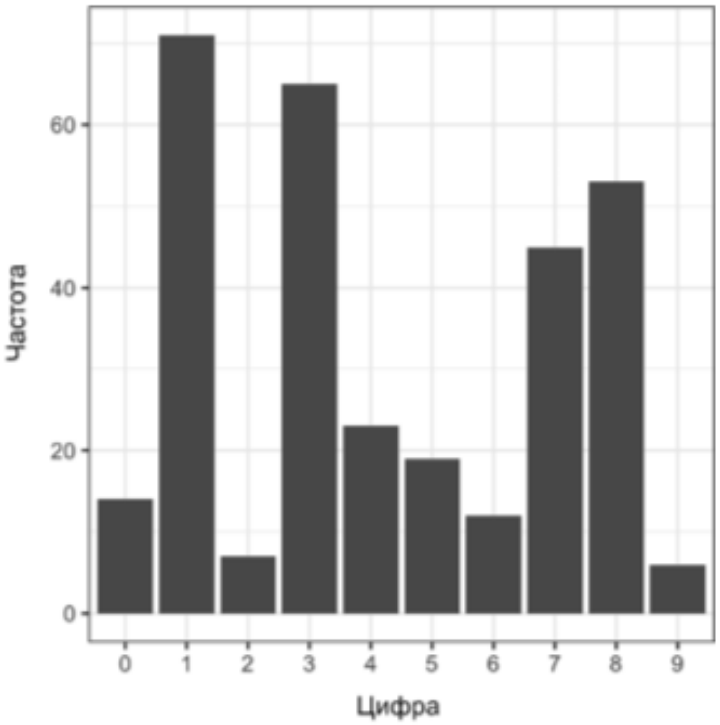
χ2 бөлу-бұл шамалар өте төмен болған жағдайларды қоспағанда, жаңадан сипатталған қайта тексерудің жақсы жуықтауы (бір разрядтар, әсіресе бес немесе одан аз).Мұндай жағдайларда қайта таңдау процедурасы дәлірек p мәндерін береді. Шын мәнінде, көптеген статистикалық бағдарламалық жүйелерде болуы мүмкін барлық мүмкін ауыстыруларды нақты тізімдеу процедурасы бар, ол олардың жиілігін кестеге келтіреді және бақыланатын нәтиженің қаншалықты шекті екенін анықтайды. Бұл процедура Фишердің ұлы статистик Р. А. Фишердің құрметіне нақты статистикалық тексеру деп аталады. Фишерді канондық түрде дәл тексеруге арналған R бағдарламасының коды өте қарапайым:

```
> fisher.test(cticks)
Fisher's Exact Test for Count Data

data:  clicks
p-value = 0.4824
alternative hypothesis: two.sided
```

p мәні қайта іріктеу әдісімен алынған 0,4853 p мәніне өте жақын.

Кейбір шамалар өте төмен, ал басқалары өте жоғары болған жағдайда (мысалы, айырбастау деңгейіндегі бөлгіш), мүмкін болатын барлық ауыстыруларды есептеудің қиындығына байланысты толық дәл тексерудің орнына аралас ауыстыру сынағын өткізу қажет болуы мүмкін. Бұрын айтылған R-функциясында осы жуықтауды қалай қолдануға болатындығы туралы бірнеше дәлелдер бар (simulate.p. value=TRUE немесе FALSE) қанша итерацияны қолдану керек (B=...) және есептеу шегін орнатыңыз (workspace=...) нақты нәтиже алу үшін есептеулер қаншалықты алыс болуы керек.



Сурет. 3.8. Иманиси - Кари зертханалық мәліметтеріне арналған жиілік гистограммасы

Деректер ғылымы үшін тексерулердің өзектілігі.

χ2 Тексеруді немесе Фишерді дәл тексеруді стандартты қолданудың көпшілігі деректер ғылымы үшін аса маңызды емес.. Көптеген эксперименттерде A-B немесе ABC болсын.... мақсат тек статистикалық маңыздылықты анықтау емес, ең жақсы нұсқаны анықтау. Осы мақсатта көп қарулы қарақшылар (бөлімді қараңыз. "Көп қарулы қарақшы алгоритмі" бұдан әрі осы тарауда) неғұрлым толық шешім ұсынады. *χ2*тексерудің қосымшаларының бірі, әсіресе оның Фишердің нақты нұсқасы, деректер ғылымында веб-эксперименттер үшін тиісті үлгі өлшемдерін белгілеу болып табылады. Мұндай эксперименттер көбінесе өте төмен көрсеткіштерге ие және мыңдаған әсерлерге қарамастан, санның пайызы экспериментте категориялық қорытындылар алу үшін тым аз болуы мүмкін. Мұндай шайларда Фишерді дәл тексеру, *χ2*тексеру және басқа тексерулер қуат пен үлгі өлшемдерін есептеудің құрамдас бөлігі ретінде пайдалы болуы мүмкін (бөлімді қараңыз. "Қуат және үлгі өлшемі" осы тарауда). Тексерулер *χ2*сарапшылардың ақпарат жинауында статистикалық маңызды емес p - мәнін іздеуде кеңінен қолданылады, бұл жұмысты жариялауға мүмкіндік береді. *X2* тексерулер немесе ұқсас IP-қайта іріктеу модельдеулері деректер ғылымының қосымшаларында әсердің немесе белгінің маңыздылығын ресми тексеруден гөрі қосымша қарастыруға лайық екенін анықтау үшін сүзгі ретінде көбірек қолданылады. Мысалы, Олар геокеңістіктік статистика мен картографияда кеңістіктік деректердің көрсетілген нөлдік үлестірімге сәйкес келетіндігін анықтау үшін қолданылады. (Мысалы, Шынымен де қылмыстар белгілі бір салада шоғырланған, бұл кездейсоқ мүмкіндікке қарағанда көбірек?) Олар сондай - ақ барлық белгілердегі сыныптың таралуын анықтау және белгілі бір сыныптың таралуы әдеттен тыс жоғары немесе төмен, кездейсоқ вариациямен үйлеспейтін белгілерді анықтау үшін машиналық оқытудағы белгілерді автоматты түрде таңдауда қолданыла алады.

Статистикаға сүйене отырып тексеруге арналған негізгі идеялар χ2 •

Статистикадағы жалпы қабылданған процедура деректердің бақыланатын мөлшері Тәуелсіздік туралы болжамға сәйкес келетіндігін тексеруден тұрады (мысалы, бір немесе басқа затты сатып алу үрдісі жынысына байланысты емес). χ2 үлестіру-бұл анықтамалық үлестіру (ол тәуелсіздік туралы болжамды бейнелейді), оған сәйкес келуі керек бақылаумамыр есептелген статистика χ2 .

Қуат және үлгі өлшемі

Егер сіз веб-тест жүргізіп жатсаңыз, онда оның қанша уақытқа созылатындығын қалай шешесіз (яғни, бір нұсқаға қанша әсер қажет)? Желідегі көптеген веб-тестілеу нұсқаулықтарында оқуға болатын жалпы кеңестерге қарамастан, жақсы ұсыныстар жоқ-негізінен бәрі қалаған мақсатқа жету жиілігіне байланысты.

Негізгі терминдер

Эффект мөлшері (Effect size) статистикалық тексеру нәтижесінде табуға үміттенетін әсердің минималды мөлшері, атап айтқанда "басу санының 20 пайызға өсуі". Синоним: әсер мөлшері.

Қуат (қуат) берілген үлгі өлшемінде берілген әсер өлшемін анықтау ықтималдығы.

Маңыздылық деңгейі (significance level) тексеру жүргізілетін статистикалық маңыздылық деңгейі.

Синонимдер: Альфа, α.

Үлгі өлшеміне қатысты статистикалық есептеулердегі қадамдардың бірі: "статистикалық гипотезаны тексеру шын мәнінде А және В нұсқалары арасындағы айырмашылықты көрсете ме?"Гипотезаны тексерудің нәтижесі-р-мәні-А және В нұсқаларының арасындағы нақты айырмашылыққа байланысты, ол сонымен қатар таза кездейсоқтыққа байланысты - эксперимент үшін топтарға таңдалған адам. Бірақ А және В нұсқаларының арасындағы нақты айырмашылық неғұрлым көп болса, біздің эксперимент оны көрсету ықтималдығы соғұрлым жоғары болады деп болжау қисынды; айырмашылық неғұрлым аз болса, оны анықтау үшін соғұрлым көп деректер қажет болады. Бейсболдағы 0,350-ші соққыларды 0,200-ші соққылардан 2-ші соққылардан ажырату үшін соққыларға көп көзқарас қажет емес. Бірақ 0,300-ші және 0,280-ші жылдарды ажырату үшін көптеген тәсілдер қажет болады.

Қуат - бұл әсердің көрсетілген мөлшерін іріктеу сипаттамаларымен (мөлшері мен өзгергіштігі) анықтау ықтималдығы. Мысалы, (гипотетикалық) 25 биттік тәсілмен 0,330-шы және 0,200-ші соққыларды ажырату ықтималдығы 0,75 деп айтуға болады. Мұндағы әсер мөлшері-0,130 айырмашылығы. Ал " анықтау "гипотезаны тексеру" айырмашылық жоқ " деген жақсы гипотезаны қабылдамайтынын және нақты әсер бар деген қорытындыға келетіндігін білдіреді. Сонымен, әсер мөлшері 0,130 болатын екі соққы үшін 25 биттік тәсілмен (25 n =) тәжірибе (гипотетикалық) қуаты

0,75 немесе 75% құрайды. Мұнда бірнеше жылжымалы элементтер бар екенін және қажет болатын көптеген статистикалық болжамдар мен формулаларда оңай іске қосылатынын көруге болады (таңдамалы өзгергіштікті, әсер өлшемін, үлгі өлшемін, гипотезаны тексеру үшін α - деңгейді және т.б. көрсету және қуатты есептеу үшін). Шындығында, қуатты есептеу үшін арнайы мақсаттағы статистикалық бағдарламалық жүйелер бар. Көптеген деректер талдаушылары, мысалы, жарияланған жұмыста қуат туралы хабарлау үшін қажетті барлық ресми кезеңдерден өтудің қажеті жоқ. Дегенмен, олар A/B сынағы үшін біраз деректер жинау қажет болатын және деректерді жинау немесе өңдеу кейбір шығындарды талап ететін жағдайларға тап болуы мүмкін. Бұл жағдайда қанша деректерді жинау керектігі туралы шамамен ақпарат белгілі бір күш жұмсай отырып, деректерді жинаған кездегі жағдайдың алдын алуға көмектеседі және нәтижесінде нәтиже түпкілікті болмайды. Міне, өте интуитивті балама тәсіл: 1. Қорытындыға келетін нәтижелер туралы ең жақсы болжамды білдіретін кез келген гипотетикалық деректерден бастаңыз (мүмкін априорлық деректерге сүйене отырып) - мысалы, 20 бірлік және 80 нөл бар қорап, бұл жұмақ 0,200 - ші жылдарды білдіреді немесе "вебте өткізген уақыт" туралы бірнеше бақылаулары бар қорап.- сайтта". 2. Бірінші үлгіге қажетті эффект өлшемін қосу арқылы екінші үлгіні жасаңыз - мысалы, 33 бірлік және 67 нөл бар екінші қорап немесе әрбір бастапқы "веб — сайтта өткізілген уақытқа"25 секунд қосылған екінші қорап. 3. Әр қораптан n өлшемді жүктеу үлгісін алыңыз. 4. Екі жүктеу үлгісінде гипотезаны ауыстыру (немесе формулаға негізделген) тексеруді орындаңыз және олардың арасындағы айырмашылық статистикалық маңызды екенін жазыңыз. 5. Алдыңғы екі қадамды бірнеше рет қайталаңыз және айырмашылықтың қаншалықты жиі маңызды болғанын анықтаңыз — бұл болжамды қуат.