MUNI
ECON

χ^2 chí-kvadrát test

Povinná literatura: Mann (2016), Kapitola 11

Z čeho studovat desátou lekci?

Povinná literatura:	Mann (2016), kap. 11
Příprava na cvičení:	Leaflet 12 Koncepty a procedury, cv. 12, kap. 11
Příprava na zkoušku:	Mann (2016), kap. 11 Leaflet 12 Sbírka úloh, kap. 11 Koncepty a procedury, cv. 12, kap. 11

Proč je χ^2 rozdělení důležité pro ekonomy?

Testy dobré shody

- Chcete ověřit, zda vaše data odpovídají předpokladům trhu? χ^2 test to zvládne.
- Příklad: Sleduje spotřebitelské preference, které mají odpovídat teoretickému modelu.

Test nezávislosti a test homogenity

- Odhaluje vazby mezi kategoriemi, jako je vzdělání vs. pohlaví.
- Pomáhá zodpovědět otázky typu: „Ovlivňuje místo pobytu naše příjmy?“

Testy o populačním rozptylu

- Nezbytné pro měření rizik: Volatilita na akciovém trhu, fluktuace cen nemovitostí.

Obsah

χ^2 rozdělení

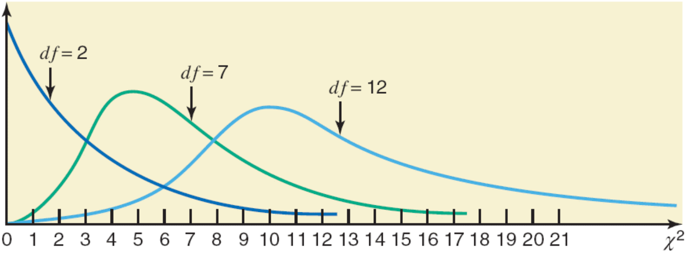
Testy dobré shody

Test nezávislosti, test homogenity

Testy o populačním rozptylu

Rozdělení chí-kvadrát

Rozdělení chí-kvadrát má pouze jeden parametr nazývaný stupně volnosti. Tvar křivky rozdělení chí-kvadrát je pro malé stupně volnosti zešikmené doprava a pro velké stupně volnosti se stává symetrickým. Celá křivka rozdělení chí-kvadrát leží napravo od vertikální osy. Rozdělení chí-kvadrát předpokládá pouze nezáporné hodnoty, které jsou označeny symbolem χ^2 (čteno jako “chí-kvadrát”).



Příklad 1

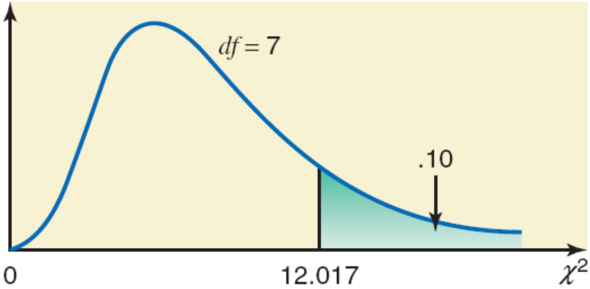
Zadání: Najděte hodnotu χ^2 pro 7 stupňů volnosti a plochu 0.10 v pravém ocase křivky rozdělení chí-kvadrát.

Řešení:

	Area in the Right Tail Under the Chi-Square Distribution Curve				
df	**.995**	...	**.100**	...	**.005**
1	0.000	...	2.706	...	7.879
2	0.010	...	4.605	...	10.597
.	...	...	...	...	...
.	...	...	...	...	...
.	...	...	...	...	...
7	0.989	...	12.017	...	20.278
.	...	...	...	...	...
.	...	...	...	...	...
.	...	...	...	...	...
100	67.328	...	118.498	...	140.169

Required value of χ^2

Příklad 1: Řešení



Příklad 2

Zadání: Najděte hodnotu χ^2 pro 12 stupňů volnosti a plochu 0.05 v levém ocase křivky rozdělení chí-kvadrát.

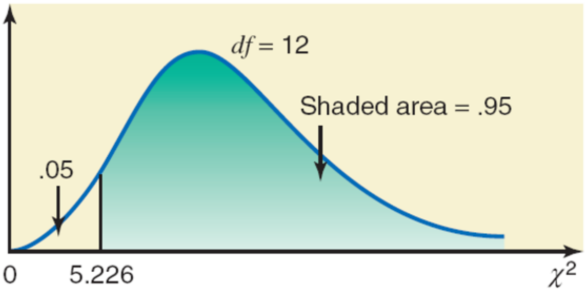
Řešení:

Plocha v pravém ocase = 1 - plocha v levém ocase = 1 - 0.05 = 0.95

	Area in the Right Tail Under the Chi-Square Distribution Curve				
df	**.995**	...	**.950**	...	**.005**
1	0.000	...	0.004	...	7.879
2	0.010	...	0.103	...	10.597
.	...	...	...	...	...
.	...	...	...	...	...
.	...	...	...	...	...
12	3.074	...	**5.226**	...	28.300
.	...	...	...	...	...
.	...	...	...	...	...
.	...	...	...	...	...
100	67.328	...	77.929	...	140.169

Required value of χ^2

Příklad 2: Řešení



Obsah

χ^2 rozdělení

Testy dobré shody

Test nezávislosti, test homogenity

Testy o populačním rozptylu

Testy dobré shody

Experiment s následujícími charakteristikami se nazývá **multinomický experiment**.

1. Experiment se skládá z n identických pokusů (opakování).
2. Každý pokus má jeden z k možných výsledků (nebo kategorií), kde $k > 2$.
3. Pokusy jsou nezávislé.
4. Pravděpodobnosti různých výsledků zůstávají pro každý pokus konstantní.

Pozorované a očekávané frekvence

Frekvence získané z provedení experimentu se nazývají **pozorované frekvence** a jsou označeny jako O. **Očekávané frekvence**, označené jako E, jsou frekvence, které očekáváme, že získáme, pokud platí nulová hypotéza. Očekávaná frekvence pro kategorii se vypočítá jako

$$E = np$$

kde n je velikost vzorku a p je pravděpodobnost, že prvek patří do této kategorie, pokud platí nulová hypotéza.

Stupně volnosti pro testy dobré shody

V testu dobré shody jsou **stupně volnosti** definovány jako

$$df = k - 1$$

kde k označuje počet možných výsledků (nebo kategorií) experimentu.

Testová statistika pro testy dobré shody

Testová statistika pro test dobré shody je označována χ^2 a její hodnota je spočítána jako

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

kde

O = pozorovaná frekvence pro kategorii

E = očekávaná frekvence pro kategorii = np

Poznámka: Pamatujte, že test dobré shody pomocí χ^2 je vždy jednostranný test.

Příklad 3

Zadání: V jedné bance je umístěn bankomat, který je zákazníkům dostupný pouze v pracovní dny od pondělí do pátku od 7:00 do 18:00.

Manažerka banky chtěla zjistit, zda je procento transakcí provedených na tomto bankomatu stejné pro každý z 5 pracovních dnů v týdnu. Náhodně vybrala jeden týden a spočítala počet transakcí provedených na tomto bankomatu každý den během tohoto týdne. Informace, které získala, jsou uvedeny v následující tabulce, kde počet uživatelů představuje počet transakcí na tomto bankomatu v jednotlivých dnech. Pro jednoduchost budeme tyto transakce označovat jako „lidé“ nebo „uživatelé“.

Příklad 3

Day	Monday	Tuesday	Wednesday	Thursday	Friday
Number of users	253	197	204	279	267

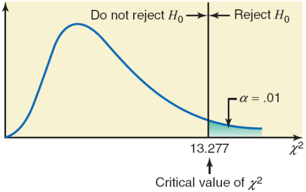
Zadání: Můžeme, na hladině významnosti 1 %, zamítnout nulovou hypotézu, že počet lidí, kteří využívají tento bankomat, je stejný pro každý z 5 dnů v týdnu? Předpokládejme, že tento týden je reprezentativní pro všechny týdny, pokud jde o využívání tohoto bankomatu.

Příklad 3: Řešení

Krok 1 $H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = 0.20$
H_1 : Alespoň dva z pěti poměrů nejsou rovny 0.20

Krok 2 Existuje 5 kategorií - 5 dnů, ve kterých je bankomat používán (multinomický experiment), pro test používáme rozdělení χ^2.

Krok 3



$\alpha = 0.01$
k = počet kategorií = 5
$df = k - 1 = 5 - 1 = 4$
Kritická hodnota $\chi^2 = 13.277$

Příklad 3: Řešení

Krok 4 Všechny potřebné výpočty k určení hodnoty testové statistiky χ^2 jsou uvedeny v tabulce:

Category (Day)	Observed Frequency O	p	Expected Frequency $E = np$	$(O - E)$	$(O - E)^2$	$\frac{(O-E)^2}{E}$
Monday	253	.20	1200(.20) = 240	13	169	.704
Tuesday	197	.20	1200(.20) = 240	−43	1849	7.704
Wednesday	204	.20	1200(.20) = 240	−36	1296	5.400
Thursday	279	.20	1200(.20) = 240	39	1521	6.338
Friday	267	.20	1200(.20) = 240	27	729	3.038
	$n = 1200$					Sum = 23.184

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 23.184$$

Příklad 3: Řešení

Krok 5 Hodnota testové statistiky $\chi^2 = 23.184$ je větší než kritická hodnota $\chi^2 = 13.277$, tedy spadá do oblasti zamítnutí. Proto zamítáme nulovou hypotézu.

Tvrdíme, že počet osob, které používají tento bankomat, se v průběhu 5 dnů v týdnu liší.

Příklad 4

Zadání: V průzkumu provedeném 3.–6. dubna 2023 byli dotazováni občané Evropy ve věku 18 a více let, zda si myslí, že lidé s vyššími příjmy „platí svůj spravedlivý podíl na daních, platí příliš mnoho nebo platí příliš málo."

Z respondentů 61 % uvedlo, že platí příliš málo, 24 % uvedlo spravedlivý podíl, 13 % uvedlo příliš mnoho a 2 % nemělo názor. Předpokládejme, že tato procenta odrážejí názor populace Evropanů ve věku 18 a více let v roce 2023. Nedávno bylo na stejnou otázkou dotázáno 1000 náhodně vybraných Evropanů ve věku 18 a více let. Následující tabulka uvádí počet Evropanů v tomto vzorku, kteří patřili do každé kategorie odpovědí.

Příklad 4

Response	Too Little	Fair Share	Too Much	No Opinion
Frequency	581	256	138	25

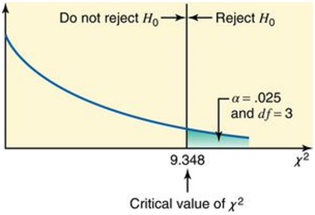
Zadání: Na hladině významnosti 2.5 % testujte, zda se současné rozložení názorů od roku 2023 liší.

Příklad 4: Řešení

Krok 1 H_0 : Současné procentní rozložení názorů je stejné jako v 2023.
H_1 : Současné procentní rozložení názorů je rozdílné oproti 2023.

Krok 2 Existují 4 kategorie - multinomický experiment; pro test používáme rozdělení χ^2.

Krok 3



$\alpha = 0.025$
k = počet kategorií = 4
df = $k - 1$ = $4 - 1 = 3$
Kritická hodnota $\chi^2 = 9.348$

Příklad 4: Řešení

Krok 4 Všechny potřebné výpočty k určení hodnoty testové statistiky χ^2 jsou uvedeny v tabulce:

Category (Response)	Observed Frequency O	p	Expected Frequency $E = np$	$(O - E)$	$(O - E)^2$	$\frac{(O-E)^2}{E}$
Too little	581	.61	1000(.61) = 610	−29	841	1.379
Fair share	256	.24	1000(.24) = 240	16	256	1.067
Too much	138	.13	1000(.13) = 130	8	64	.492
No opinion	25	.02	1000(.02) = 20	5	25	1.250
	$n = 1000$					Sum = 4.188

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 4.188$$

Příklad 4: Řešení

Krok 5 Hodnota testové statistiky $\chi^2 = 4.188$ je menší než kritická hodnota $\chi^2 = 9.348$, tedy spadá do oblasti nezamítnutí. Proto nezamítáme nulovou hypotézu.

Tvrdíme, že současné procentuální rozložení názorů se neliší od rozložení v roce 2023.

Obsah

χ^2 rozdělení

Testy dobré shody

Test nezávislosti, test homogenity

Testy o populačním rozptylu

Test nezávislosti, test homogenity

Často máme informace o více než jedné proměnné pro každý prvek. Takové informace mohou být shrnuty a prezentovány pomocí tabulky dvojrozměrné klasifikace, která se nazývá **kontingenční tabulka** nebo **křížová tabulka**.

	Full-Time	**Part-Time**
Male	6768	2615
Female	7658	3717

Students who are male and enrolled part-time

Test nezávislosti

Test nezávislosti zahrnuje test nulové hypotézy, že dva atributy populace nejsou vzájemně závislé.

Hodnota **testové statistiky χ^2 pro test nezávislosti** je vypočítána jako

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

kde O a E jsou pozorované a očekávané frekvence pro danou buňku.

Stupně volnosti pro test nezávislosti jsou

$$df = (R - 1)(C - 1)$$

kde R a C jsou počet řádků a počet sloupců v dané kontingenční tabulce.

Příklad 5

Zadání: Nedostatek kázně se stal hlavním problémem ve školách v USA. Byl vybrán náhodný vzorek 300 dospělých, kteří byli dotázáni, zda podporují větší svobodu pro učitele při trestání studentů za nedostatek kázně. Dvojrozměrná klasifikace odpovědí těchto dospělých je uvedena v následující tabulce.

	In Favor (F)	Against (A)	No Opinion (N)
Men (M)	93	70	12
Women (W)	87	32	6

Spočítejte očekávané frekvence pro tuto tabulku za předpokladu, že tyto dva atributy, pohlaví a názory na danou otázku, jsou nezávislé.

Příklad 5: Řešení

$$\text{Očekávaná frekvence E pro buňku} = \frac{\text{(počet v řádku)(počet v sloupci)}}{\text{velikost vzorku}}$$

	In Favor (F)	Against (A)	No Opinion (N)	Row Totals
Men (M)	93 (105.00)	70 (59.50)	12 (10.50)	175
Women (W)	87 (75.00)	32 (42.50)	6 (7.50)	125
Column Totals	180	102	18	300

Příklad 6

Zadání: Znovu zvažte klasifikační tabulku uvedenou v příkladu 11-5. V tomto příkladu byla náhodně vybrána skupina 300 dospělých, kteří odpovídali, zda podporují poskytnutí větší svobody učitelům k trestání studentů za nedisciplinovanost. Na základě výsledků průzkumu byla připravena a prezentována klasifikační tabulka v příkladu 11-5. Poskytuje vzorek dostatečné informace k závěru, že tyto dva atributy, pohlaví a názory dospělých, jsou závislé? Při testování použijte hladinu významnosti 1 %.

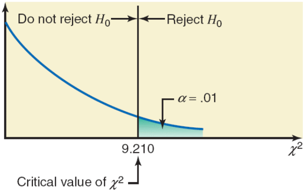
	In Favor **(F)**	**Against** **(A)**	**No Opinion** **(N)**
Men (M)	93	70	12
Women (W)	87	32	6

Příklad 6: Řešení

Krok 1 H_0 : Pohlaví a názory dospělých jsou nezávislé.
H_1 : Pohlaví a názory dospělých jsou závislé.

Krok 2 Pro provedení testu nezávislosti pro kontingenční tabulku používáme rozdělení χ^2.

Krok 3



$\alpha = 0.01$
$df = (R - 1)(C - 1)$
$= (2 - 1)(3 - 1) = 2$
Kritická hodnota $\chi^2 = 9.210$

Příklad 6: Řešení

Krok 4 Všechny potřebné výpočty k určení hodnoty testové statistiky χ^2 jsou uvedeny v Tabulce:

	In Favor (F)	Against (A)	No Opinion (N)	Row Totals
Men (M)	93 (105.00)	70 (59.50)	12 (10.50)	175
Women (W)	87 (75.00)	32 (42.50)	6 (7.50)	125
Column Totals	180	102	18	300

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(93-105)^2}{105} + \frac{(70-59.5)^2}{59.5} + \frac{(12-10.5)^2}{10.5} +$$
$$+ \frac{(87-75)^2}{75} + \frac{(32-42.5)^2}{42.5} + \frac{(6-7.5)^2}{7.5} =$$
$$= 1.371 + 1.853 + 0.214 + 1.92 + 2.594 + 0.3 =$$
$$= \mathbf{8.252}$$

Příklad 6: Řešení

Krok 5 Hodnota testové statistiky $\chi^2 = 8.252$ je menší než kritická hodnota $\chi^2 = 9.210$, a proto spadá do oblasti nezamítnutí. Proto nemůžeme zamítnout nulovou hypotézu.

Tvrdíme, že z dostupného vzorku nemáme dostatečný důkaz pro závěr, že pohlaví a názory dospělých jsou pro tuto otázku závislé.

Příklad 7

Zadání: Výzkumník chtěl studovat vztah mezi pohlavím a vlastnictvím smartphone v dospělé populaci, která vlastní mobilní telefony. Vzal vzorek 2000 dospělých a získal informace uvedené v následující tabulce.

	Own Smart Phones	Do Not Own Smart Phones
Men	640	450
Women	440	470

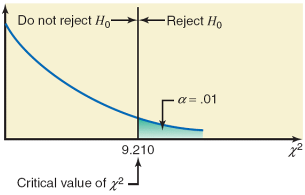
Můžete na hladině významnosti 5 % dospět k závěru, že pohlaví a vlastnictví chytrého telefonu jsou u všech dospělých provázané?

Příklad 7: Řešení

Krok 1 H_0 : Pohlaví a vlastnictví chytrého telefonu nejsou závislé.
H_1 : Pohlaví a vlastnictví chytrého telefonu jsou závislé.

Krok 2 Provádíme test nezávislosti. Používáme rozdělení χ^2 k provedení tohoto testu.

Krok 3



$\alpha = 0.05$
$df = (R-1)(C-1)$
$= (2-1)(2-1) = 1$
Kritická hodnota $\chi^2 = 3.841$

Příklad 7: Řešení

Krok 4 Hodnoty k určení testové statistiky χ^2 jsou uvedeny v Tabulce:

	Own Smart Phones (Y)	Do Not Own Smart Phones (N)	Row Totals
Men (*M*)	640 (588.60)	450 (501.40)	1090
Women (*W*)	440 (491.40)	470 (418.60)	910
Column Totals	1080	920	2000

$$
\begin{aligned}
\chi^2 = \sum \frac{(O-E)^2}{E} &= \frac{(640-588.60)^2}{588.60} + \frac{(450-501.40)^2}{501.40} + \\
&+ \frac{(440-491.40)^2}{491.40} + \frac{(470-418.60)^2}{481.60} = \\
&= 4.489 + 5.269 + 5.376 + 6.311 \\
&= \mathbf{21.445}
\end{aligned}
$$

Příklad 7: Řešení

Krok 5 Hodnota testové statistiky $\chi^2 = 21.445$ je větší než kritická hodnota $\chi^2 = 3.841$, a proto spadá do oblasti zamítnutí. Proto zamítáme nulovou hypotézu.

Tvrdíme, že vzorek poskytuje silný důkaz pro závěr, že pohlaví a vlastnictví chytrých telefonů jsou u všech dospělých provázané.

Test homogenity

Test homogenity zahrnuje testování nulové hypotézy, že **podíly** prvků s určitými charakteristikami ve dvou nebo více různých populacích jsou stejné, oproti alternativní hypotéze, že tyto podíly nejsou stejné.

Příklad 8

Zadání: Zvažte data o rozdělení příjmů domácností v Kalifornii a Wisconsinu uvedená v Tabulce.

	California	Wisconsin	Row Totals
High income	70	34	104
Medium income	80	40	120
Low income	100	76	176
Column Totals	250	150	400

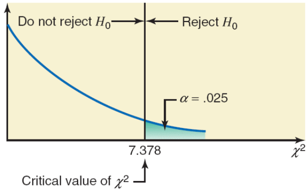
Na hladině významnosti 2.5 % otestujte, zda se rozdělení domácností s ohledem na úrovně příjmů mezi těmito dvěma státy liší (není homogenní).

Příklad 8: Řešení

Krok 1 H_0 : Podíly domácností, které patří do různých příjmových skupin, jsou v obou státech stejné.
H_1 : Podíly domácností, které patří do různých příjmových skupin, nejsou v obou státech stejné.

Krok 2 Používáme χ^2 rozdělení k provedení testu homogenity.

Krok 3



$\alpha = 0.025$
$df = (R-1)(C-1)$
$= (3-1)(2-1) = 2$
Kritická hodnota $\chi^2 = 7.378$

Příklad 8: Řešení

	California	Wisconsin	Row Totals
High income	70 (65)	34 (39)	104
Medium income	80 (75)	40 (45)	120
Low income	100 (110)	76 (66)	176
Column Totals	250	150	400

$$\begin{aligned}\chi^2 &= \sum \frac{(O-E)^2}{E} \\ &= \frac{(70-65)^2}{65} + \frac{(34-39)^2}{39} + \frac{(80-75)^2}{75} \\ &\quad + \frac{(40-45)^2}{45} + \frac{(100-110)^2}{110} + \frac{(76-66)^2}{66} \\ &= 0.385 + 0.641 + 0.333 + 0.566 + 0.909 + 1.515 = \mathbf{4.339}\end{aligned}$$

Příklad 8: Řešení

Krok 4 Hodnota testové statistiky $\chi^2 = 4.339$ je menší než kritická hodnota $\chi^2 = 7.378$, a proto nespadá do oblasti zamítnutí. Proto nezamítáme nulovou hypotézu.

Tvrdíme, že neexistuje důkaz, že se rozdělení domácností na základě příjmů mezi Kalifornií a Wisconsinem liší.

Obsah

χ^2 rozdělení

Testy dobré shody

Test nezávislosti, test homogenity

Testy o populačním rozptylu

Testy o populačním rozptylu

- Odhad rozptylu populace
- Testování hypotéz o rozptylu populace

Výběrové rozdělení $(n-1)s^2/\sigma^2$

Pokud je populace, ze které byl vzorek vybrán, (přibližně) normálně rozdělená, pak statistika

$$\frac{(n-1)s^2}{\sigma^2}$$

má chí-kvadrát rozdělení s $n-1$ stupni volnosti.

Interval spolehlivosti pro rozptyl populace σ^2

Za předpokladu, že populace, ze které byl vzorek získán, má (přibližně) normální rozdělení, můžeme získat $(1-\alpha)100\%$ **interval spolehlivosti pro rozptyl populace** σ^2 jako

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \quad \text{až} \quad \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

kde $\chi^2_{\alpha/2}$ a $\chi^2_{1-\alpha/2}$ jsou hodnoty získané z chí-kvadrát rozdělení pro plochy $\alpha/2$ a $1-\alpha/2$ v pravém chvostu křivky chí-kvadrát rozdělení při $n-1$ stupních volnosti. Interval spolehlivosti pro směrodatnou odchylku populace lze získat jako druhé odmocniny obou hranic intervalu spolehlivosti pro rozptyl populace.

Příklad 9

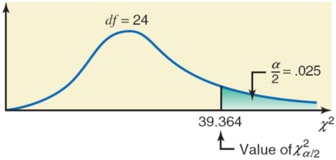
Zadání: Jedním z typů sušenek vyráběných společností **Danovy čočkové sušenky** jsou kakaové sušenky. Stroj, který plní balení sušenek, je nastaven tak, aby průměrná čistá hmotnost balení byla 32 uncí s rozptylem 0.015 čtvereční unce.

Kontrolor kvality občas vybere vzorek několika balení, vypočítá rozptyl jejich čisté hmotnosti a sestaví 95% interval spolehlivosti pro rozptyl populace. Pokud se jeden nebo oba konce tohoto intervalu nacházejí mimo rozmezí 0.008 až 0.030, je stroj zastaven a znovu nastaven.

Nedávno byl náhodným výběrem odebrán vzorek 25 balení z výrobní linky, u kterého byl zjištěn výběrový rozptyl 0.029 čtvereční unce. Myslíte, na základě těchto informací, že stroj potřebuje úpravu? Předpokládejte, že čistá hmotnost sušenek ve všech baleních má normální rozdělení.

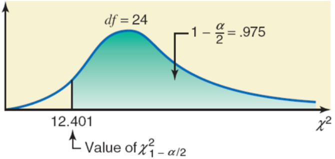
Příklad 9: Řešení

Krok 1 $n = 25$ a $s^2 = 0.029$



$$\alpha = 1 - 0.95 = 0.05$$
$$\alpha/2 = 0.05/2 = 0.025$$
$$1 - \alpha/2 = 1 - 0.025 = 0.975$$
$$\mathrm{df} = n - 1 = 25 - 1 = 24$$

Krok 2



Kritická hodnota χ^2 pro 24 stupňů volnosti a plochu 0.025 v pravém chvostu = 39.364
Kritická hodnota χ^2 pro 24 stupňů volnosti a plochu 0.975 v pravém chvostu = 12.401

Příklad 9: Řešení

Krok 3

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \quad \text{až} \quad \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

$$\frac{(25-1)(0.029)}{39.364} \quad \text{až} \quad \frac{(25-1)(0.029)}{12.401}$$

$$0.0177 \quad \text{až} \quad 0.0561$$

S 95% spolehlivostí tvrdíme, že rozptyl všech balení kakaových sušenek leží mezi **0.0177** a **0.0561** čtvereční unce.

Interval spolehlivosti pro směrodatnou odchylku populace σ můžeme získat jako kladné druhé odmocniny obou hranic výše uvedeného intervalu spolehlivosti pro rozptyl populace. Tedy 95% interval spolehlivosti pro směrodatnou odchylku populace je **0.133** až **0.237**.

Testová statistika pro test hypotézy o σ^2

Hodnota **testové statistiky** χ^2 se vypočítá jako

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

kde s^2 je výběrový rozptyl, σ^2 je předpokládaná hodnota rozptylu populace a $n-1$ představuje počet stupňů volnosti. Populace, ze které byl výběr proveden, se předpokládá jako (přibližně) normálně rozdělená.

Příklad 10

Zadání: Jedním z typů sušenek vyráběných společností **Haddad Food Company** jsou kakaové sušenky. Stroj, který plní balení těchto sušenek, je nastaven tak, aby průměrná čistá hmotnost balení byla 32 uncí s rozptylem 0.015 čtvereční unce.

Kontrolor kvality občas vybere vzorek několika těchto balení, vypočítá rozptyl jejich čisté hmotnosti a provede test hypotézy o rozptylu populace. Vždy používá hladinu významnosti $\alpha = 0.01$. Přijatelná hodnota rozptylu populace je 0.015 čtvereční unce nebo méně. Pokud závěr testu ukazuje, že rozptyl populace není v rámci přijatelných hodnot, je stroj zastaven a upraven.

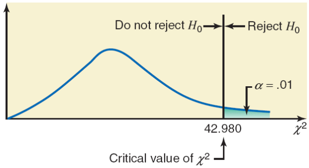
Nedávný náhodný vzorek 25 balení z výrobní linky ukázal výběrový rozptyl 0.029 čtvereční unce. Myslíte si, na základě těchto informací, že je potřeba stroj upravit? Předpokládejte, že čistá hmotnost sušenek ve všech baleních má normální rozdělení.

Příklad 10: Řešení

Krok 1 $H_0 : \sigma^2 \leq 0.015$ (Rozptyl populace je v rámci přijatelné hodnoty)
$H_1 : \sigma^2 > 0.015$ (Rozptyl pop. překračuje přijatelnou hodnotu)

Krok 2 Používáme χ^2 rozdělení k provedení testu rozptylu σ^2.

Krok 3



$\alpha = 0.01$
$df = n - 1 = 25 - 1 = 24$
Kritická hodnota $\chi^2 = 42.980$

Příklad 10: Řešení

Krok 4
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(25-1)(0.029)}{0.015} = 46.400$$

Krok 5 Hodnota testové statistiky $\chi^2 = 46.400$ je větší než kritická hodnota $\chi^2 = 42.980$, a proto spadá do oblasti zamítnutí. Proto zamítáme nulovou hypotézu.

Docházíme k závěru, že by stroj měl být zastaven a upraven, protože rozptyl populace není v přijatelných hodnotách.

Příklad 11

Zadání: Je známo, že rozptyl známek všech studentů na velké univerzitě byl v roce 2014 roven 0.24.

Profesorka chce zjistit, zda se rozptyl známek současných studentů na této univerzitě liší od hodnoty 0.24.

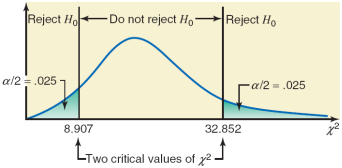
Odebrala náhodný vzorek 20 studentů a zjistila, že rozptyl jejich známek je 0.27. Můžeme na hladině významnosti 5 % dojít k závěru, že současný rozptyl známek studentů na této univerzitě je jiný než 0.24? Předpokládejte, že rozptyl známek všech současných studentů na této univerzitě má (přibližně) normální rozdělení.

Příklad 11: Řešení

Krok 1 $H_0 : \sigma^2 = 0.24$ (Rozptyl populace není rozdílný od 0.24)
$H_1 : \sigma^2 \neq 0.24$ (Rozptyl populace je rozdílný od 0.24)

Krok 2 Používáme χ^2 rozdělení k provedení testu rozptylu σ^2.

Krok 3



$\alpha = 0.05$
Plocha v chvostech $= 0.025$
$df = n - 1 = 20 - 1 = 19$
Kritické hodnoty:
$\chi^2 = 32.852$ a $\chi^2 = 8.907$

Příklad 11: Řešení

Krok 4

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(20-1)(0.27)}{0.24} = 21.375$$

Krok 5 Hodnota testové statistiky $\chi^2 = 21.375$ leží mezi kritickými hodnotami $\chi^2 = 8.907$ a $\chi^2 = 32.852$, a proto spadá do oblasti nezamítnutí. Proto nezamítáme nulovou hypotézu.

Docházíme k závěru, že rozptyl populace současných známek studentů na této univerzitě se zřejmě neliší od hodnoty 0.24.

Shrnutí přednášky:

χ^2 rozdělení

Testy dobré shody

Test nezávislosti, test homogenity

Testy o populačním rozptylu

Co si nastudovat na příští týden?

Příprava na cvičení: Leaflet 12
Koncepty a procedury, cv. 12, kap. 11

Povinná literatura: Mann (2016), Kapitola 12

Děkuji za pozornost!

X