

Zestaw 13 - Weryfikacja hipotez statystycznych

Artur Fortuna
AGH

Weryfikacja hipotez statystycznych

Weryfikacja hipotez statycznych polega na przeprowadzeniu pewnego testu na posiadanych danych. Testu czyli pewnej procedury.

Weryfikacja hipotez statystycznych

Weryfikacja hipotez statycznych polega na przeprowadzeniu pewnego testu na posiadanych danych. Testu czyli pewnej procedury. Zaczynamy oczywiście od zgromadzenia danych - wyników jakiś pomiarów, próbek. Zasada prosta, więcej to lepiej.

Weryfikacja hipotez statystycznych

Weryfikacja hipotez statycznych polega na przeprowadzeniu pewnego testu na posiadanych danych. Testu czyli pewnej procedury. Zaczynamy oczywiście od zgromadzenia danych - wyników jakiś pomiarów, próbek. Zasada prosta, więcej to lepiej. Następnie stawiamy **hipotezę zerową**.

Weryfikacja hipotez statystycznych

Weryfikacja hipotez statycznych polega na przeprowadzeniu pewnego testu na posiadanych danych. Testu czyli pewnej procedury. Zaczynamy oczywiście od zgromadzenia danych - wyników jakiś pomiarów, próbek. Zasada prosta, więcej to lepiej. Następnie stawiamy **hipotezę zerową**. To jak ona brzmi zależy od tego co konkretnie chcemy sprawdzić. Np hipoteza o wartości oczekiwanej mogłaby brzmieć tak: $H_0 : m = 13$.

Weryfikacja hipotez statystycznych

Weryfikacja hipotez statycznych polega na przeprowadzeniu pewnego testu na posiadanych danych. Testu czyli pewnej procedury. Zaczynamy oczywiście od zgromadzenia danych - wyników jakiś pomiarów, próbek. Zasada prosta, więcej to lepiej. Następnie stawiamy **hipotezę zerową**. To jak ona brzmi zależy od tego co konkretnie chcemy sprawdzić. Np hipoteza o wartości oczekiwanej mogłaby brzmieć tak: $H_0 : m = 13$. Hipotezie zerowej towarzyszyć musi **hipoteza alternatywna**. Najczęściej jest to zaprzeczenie hipotezy zerowej np $H_0 : m \neq 13$.

Weryfikacja hipotez statystycznych

Weryfikacja hipotez statycznych polega na przeprowadzeniu pewnego testu na posiadanych danych. Testu czyli pewnej procedury. Zaczynamy oczywiście od zgromadzenia danych - wyników jakiś pomiarów, próbek. Zasada prosta, więcej to lepiej. Następnie stawiamy **hipotezę zerową**. To jak ona brzmi zależy od tego co konkretnie chcemy sprawdzić. Np hipoteza o wartości oczekiwanej mogłaby brzmieć tak: $H_0 : m = 13$. Hipotezie zerowej towarzyszyć musi **hipoteza alternatywna**. Najczęściej jest to zaprzeczenie hipotezy zerowej np $H_0 : m \neq 13$. Ale z różnych powodów nie musi to być negacja H_0 , grunt, żeby była przeciwna.

Kolejny krok to wybór **poziomu istotności** oznaczanego standardowo α .

Weryfikacja hipotez statystycznych

Kolejny krok to wybór **poziomu istotności** oznaczanego standardowo α . Typową wartością jest $\alpha = 0,05$ czyli będziemy mieć 95% pewności. Ale nie ma przeszkód by wybrać inną.

Weryfikacja hipotez statystycznych

Kolejny krok to wybór **poziomu istotności** oznaczanego standardowo α . Typową wartością jest $\alpha = 0,05$ czyli będziemy mieć 95% pewności. Ale nie ma przeszkód by wybrać inną. Dalej teoretycznie jest prosto. Liczymy pewien współczynnik (liczbę) noszący nazwę **statystyki**.

Weryfikacja hipotez statystycznych

Kolejny krok to wybór **poziomu istotności** oznaczanego standardowo α . Typową wartością jest $\alpha = 0,05$ czyli będziemy mieć 95% pewności. Ale nie ma przeszkód by wybrać inną. Dalej teoretycznie jest prosto. Liczymy pewien współczynnik (liczbę) noszący nazwę **statystyki**. Stosuje się tu różne oznaczenia skojarzone z rozkładami - dla naszych potrzeb niech to będzie T .

Weryfikacja hipotez statystycznych

Kolejny krok to wybór **poziomu istotności** oznaczanego standardowo α . Typową wartością jest $\alpha = 0,05$ czyli będziemy mieć 95% pewności. Ale nie ma przeszkód by wybrać inną. Dalej teoretycznie jest prosto. Liczymy pewien współczynnik (liczbę) noszący nazwę **statystyki**. Stosuje się tu różne oznaczenia skojarzone z rozkładami - dla naszych potrzeb niech to będzie T . Jak się liczy ową statystykę to już zależy od hipotezy. Są różne testy, w zależności co badamy i różne wzory na statystykę. Inne dla testów o wartości oczekiwanej, inne dla testów losowości, niezależności itp.

Następnie należy policzyć **obszar krytyczny**.

Weryfikacja hipotez statystycznych

Następnie należy policzyć **obszar krytyczny**. Tu znów, to jak się go liczy zależy od testu. Dostaniemy pewien przedział lub sumę przedziałów. Oznaczmy go sobie jako K .

Weryfikacja hipotez statystycznych

Następnie należy policzyć **obszar krytyczny**. Tu znów, to jak się go liczy zależy od testu. Dostaniemy pewien przedział lub sumę przedziałów. Oznaczmy go sobie jako K . Czas na wynik.

Weryfikacja hipotez statystycznych

Następnie należy policzyć **obszar krytyczny**. Tu znów, to jak się go liczy zależy od testu. Dostaniemy pewien przedział lub sumę przedziałów. Oznaczmy go sobie jako K . Czas na wynik.

- $T \in K \Rightarrow$ na poziomie istotności α odrzucamy H_0 , przyjmujemy H_1
- $T \notin K \Rightarrow$ na poziomie istotności α nie ma podstaw do odrzucenia H_0

Weryfikacja hipotez statystycznych

Następnie należy policzyć **obszar krytyczny**. Tu znów, to jak się go liczy zależy od testu. Dostaniemy pewien przedział lub sumę przedziałów. Oznaczmy go sobie jako K . Czas na wynik.

- $T \in K \Rightarrow$ na poziomie istotności α odrzucamy H_0 , przyjmujemy H_1
- $T \notin K \Rightarrow$ na poziomie istotności α nie ma podstaw do odrzucenia H_0

No właśnie, nie napisałem "przyjmujemy H_0 ". Tak dobrze to nie ma. Jeśli hipoteza zerowa brzmiała $m = 13$ i wyszło, że nie możemy jej odrzucić to przecież nie znaczy, że m to naprawdę 13, a nie np. 13,01.

Weryfikacja hipotez statystycznych

Następnie należy policzyć **obszar krytyczny**. Tu znów, to jak się go liczy zależy od testu. Dostaniemy pewien przedział lub sumę przedziałów. Oznaczmy go sobie jako K . Czas na wynik.

- $T \in K \Rightarrow$ na poziomie istotności α odrzucamy H_0 , przyjmujemy H_1
- $T \notin K \Rightarrow$ na poziomie istotności α nie ma podstaw do odrzucenia H_0

No właśnie, nie napisałem "przyjmujemy H_0 ". Tak dobrze to nie ma. Jeśli hipoteza zerowa brzmiała $m = 13$ i wyszło, że nie możemy jej odrzucić to przecież nie znaczy, że m to naprawdę 13, a nie np. 13,01. No i jest jeszcze α szans, że wynik jest błędny.

Test losowości (test serii)

Test losowości (test serii) pozwala sprawdzić czy nasze dane są losowe (przypadkowe) czy jednak powstały w jakimś porządku.

Test losowości (test serii)

Test losowości (test serii) pozwala sprawdzić czy nasze dane są losowe (przypadkowe) czy jednak powstały w jakimś porządku.

H_0 : dane są losowe

H_1 : dane nie są losowe

Test losowości (test serii)

Test losowości (test serii) pozwala sprawdzić czy nasze dane są losowe (przypadkowe) czy jednak powstały w jakimś porządku.

H_0 : dane są losowe

H_1 : dane nie są losowe

Dane zamieniamy na zera i jedynki. Czasem dane w naturalny sposób są dwuwartościowe np. płeć.

Test losowości (test serii)

Test losowości (test serii) pozwala sprawdzić czy nasze dane są losowe (przypadkowe) czy jednak powstały w jakimś porządku.

H_0 : dane są losowe

H_1 : dane nie są losowe

Dane zamieniamy na zera i jedynki. Czasem dane w naturalny sposób są dwuwartościowe np. płeć. Jeśli są to liczby to np. te większe od mediany zamieniamy na 1, a mniejsze na 0 (a te równe medianie usuwamy z danych).

Test losowości (test serii)

Test losowości (test serii) pozwala sprawdzić czy nasze dane są losowe (przypadkowe) czy jednak powstały w jakimś porządku.

H_0 : dane są losowe

H_1 : dane nie są losowe

Dane zamieniamy na zera i jedyne. Czasem dane w naturalny sposób są dwuwartościowe np. płeć. Jeśli są to liczby to np. te większe od mediany zamieniamy na 1, a mniejsze na 0 (a te równe medianie usuwamy z danych). Liczymy kilka parametrów.

- n_0 - liczba zer
- n_1 - liczba jedynek
- n - liczba danych ($n = n_0 + n_1$)
- k - liczba serii

Test losowości (test serii)

Test losowości (test serii) pozwala sprawdzić czy nasze dane są losowe (przypadkowe) czy jednak powstały w jakimś porządku.

H_0 : dane są losowe

H_1 : dane nie są losowe

Dane zamieniamy na zera i jedynki. Czasem dane w naturalny sposób są dwuwartościowe np. płeć. Jeśli są to liczby to np. te większe od mediany zamieniamy na 1, a mniejsze na 0 (a te równe medianie usuwamy z danych). Liczymy kilka parametrów.

- n_0 - liczba zer
- n_1 - liczba jedynek
- n - liczba danych ($n = n_0 + n_1$)
- k - liczba serii

To ostatnie wymaga wyjaśnienia. Np. w ciągu 0001100011110 jest 5 serii.

Test losowości (test serii)

Dla małych danych są specjalne tablice, dla dużych danych ($n_0, n_1 > 20$) potrzebujemy jeszcze dwóch parametrów.

Test losowości (test serii)

Dla małych danych są specjalne tablice, dla dużych danych ($n_0, n_1 > 20$) potrzebujemy jeszcze dwóch parametrów.

- $m = \frac{2n_0n_1}{n} + 1$
- $s = \sqrt{\frac{2n_0n_1(2n_0n_1 - n)}{n^2(n-1)}}$

Test losowości (test serii)

Dla małych danych są specjalne tablice, dla dużych danych ($n_0, n_1 > 20$) potrzebujemy jeszcze dwóch parametrów.

- $m = \frac{2n_0n_1}{n} + 1$
- $s = \sqrt{\frac{2n_0n_1(2n_0n_1-n)}{n^2(n-1)}}$

No i kluczowe wartości:

- statystyka $Z = \frac{k-m}{s}$
- obszar krytyczny $K = (-\infty, \Phi^{-1}(\frac{\alpha}{2})] \cup [\Phi^{-1}(1 - \frac{\alpha}{2}), +\infty)$,
gdzie Φ to dystrybuanta rozkładu $\mathcal{N}(0, 1)$.

Test niezależności pozwala sprawdzić czy dwie serie danych (oznaczymy je X i Y) są niezależne.

Test niezależności pozwala sprawdzić czy dwie serie danych (oznaczymy je X i Y) są niezależne. Używamy go gdy przynajmniej jedno dane nie są liczbami (mogą być zakodowane liczbami, ale nimi de facto nie są) i dają się zgrupować w kilka typów.

Test niezależności

Test niezależności pozwala sprawdzić czy dwie serie danych (oznaczymy je X i Y) są niezależne. Używamy go gdy przynajmniej jedno dane nie są liczbami (mogą być zakodowane liczbami, ale nimi de facto nie są) i dają się zgrupować w kilka typów. Dla danych typowo liczbowych prościej jest policzyć estymator współczynnika korelacji (potrafią to wszystkie arkusze kalkulacyjne nazywając tę funkcję po prostu "współczynnik korelacji") i z niego wnioskować. Dla zmiennych niezależnych $\rho = 0$ czyli jeśli wyjdzie bliski zeru to traktujemy je jak niezależne.

Test niezależności pozwala sprawdzić czy dwie serie danych (oznaczymy je X i Y) są niezależne. Używamy go gdy przynajmniej jedno dane nie są liczbami (mogą być zakodowane liczbami, ale nimi de facto nie są) i dają się zgrupować w kilka typów. Dla danych typowo liczbowych prościej jest policzyć estymator współczynnika korelacji (potrafią to wszystkie arkusze kalkulacyjne nazywając tę funkcję po prostu "współczynnik korelacji") i z niego wnioskować. Dla zmiennych niezależnych $\rho = 0$ czyli jeśli wyjdzie bliski zeru to traktujemy je jak niezależne.

H_0 : X i Y są niezależne

H_1 : X i Y są zależne

Tworzymy **tabelę krzyżową** dla danych z próby wpisując ile par zmiennych ma daną cechę np. dla zmyślonego badania o ulubiony kolor:

	czarny	czerwony	niebieski	zielony	inne
kobiety	150	120	15	15	120
mężczyźni	160	10	100	100	10

Oznaczmy wartości w tej tabelce jako n_{ij} .

Test niezależności

Dodajemy sumy.

	czarny	czerwony	niebieski	zielony	inne	suma
kobiety	150	120	15	15	120	420
mężczyźni	160	10	100	100	10	380
suma	310	130	115	115	130	800

Test niezależności

Ta sama tabelka przeliczona na procenty:

	czarny	czerwony	niebieski	zielony	inne	suma
K	18,75%	15%	1,875%	1,875%	15%	52,5%
M	20%	1,25%	12,5%	12,5%	1,25%	47,5%
suma	38,75%	16,25%	14,375%	14,375%	16,25%	100%

Tworzymy taką samą tabelkę dla danych teoretycznych. Jak?
Skoro w naszym badaniu brało udział 800 osób w tym 52,2% kobiet, a czarny lubi 38,75% wszystkich badanych na skrzyżowaniu pozycji kobiety-czarny powinno figurować
 $800 \cdot 0,525 \cdot 0,3875 = 162,75$ osób

Tworzymy taką samą tabelkę dla danych teoretycznych. Jak?
Skoro w naszym badaniu brało udział 800 osób w tym 52,2% kobiet, a czarny lubi 38,75% wszystkich badanych na skrzyżowaniu pozycji kobiety-czarny powinno figurować
 $800 \cdot 0,525 \cdot 0,3875 = 162,75$ osób

	czarny	czerwony	niebieski	zielony	inne
kobiety	162,75	68,25	60,375	60,375	68,25
mężczyźni	147,25	61,75	54,625	54,625	61,75

Oznaczmy wartości w tej tabelce jako $\hat{n}_{ij}$.

Oznaczenia

- g_1 - liczba grup dla danych X (u nas 2)
- g_2 - liczba grup dla danych Y (u nas 5)

Oznaczenia

- g_1 - liczba grup dla danych X (u nas 2)
- g_2 - liczba grup dla danych Y (u nas 5)

I kluczowe wartości

- statystyka $\chi^2 = \sum_{i=1}^{g_1} \sum_{j=1}^{g_2} \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$
- obszar krytyczny $K = [F^{-1}(1 - \alpha), +\infty)$, gdzie F to dystrybuanta rozkładu χ^2 o $(g_1 - 1)(g_2 - 1)$ stopniach swobody.

Oznaczenia

- g_1 - liczba grup dla danych X (u nas 2)
- g_2 - liczba grup dla danych Y (u nas 5)

I kluczowe wartości

- statystyka $\chi^2 = \sum_{i=1}^{g_1} \sum_{j=1}^{g_2} \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}$
- obszar krytyczny $K = [F^{-1}(1 - \alpha), +\infty)$, gdzie F to dystrybuanta rozkładu χ^2 o $(g_1 - 1)(g_2 - 1)$ stopniach swobody.

Uwaga, jak korzystamy z gotowych tablic to często zawierają one nie dystrybuantę, a wartości krytyczne i wtedy lewy kraniec tego obszaru szukamy jako $\chi^2_{\alpha, (g_1-1)(g_2-1)}$.

Test zgodności Pearsona

Test zgodności Pearsona (test zgodności chi-kwadrat) pozwala nam sprawdzić czy rozkład danych w próbce jest zgodny z rozkładem, z którym uważamy, że zgodny powinien być.

Test zgodności Pearsona

Test zgodności Pearsona (test zgodności chi-kwadrat) pozwala nam sprawdzić czy rozkład danych w próbie jest zgodny z rozkładem, z którym uważamy, że zgodny powinien być. Np. prowadzimy badanie statystyczne i chcemy wiedzieć czy rozkład liczby lat wśród naszych badanych jest reprezentatywny czyli zgodny z rozkładem wieku mieszkańców Polski.

Test zgodności Pearsona

Test zgodności Pearsona (test zgodności chi-kwadrat) pozwala nam sprawdzić czy rozkład danych w próbie jest zgodny z rozkładem, z którym uważamy, że zgodny powinien być.

Np. prowadzimy badanie statystyczne i chcemy wiedzieć czy rozkład liczby lat wśród naszych badanych jest reprezentatywny czyli zgodny z rozkładem wieku mieszkańców Polski.

Albo mamy jakieś dane i chcemy sprawdzić czy pochodzą z rozkładu normalnego.

Test zgodności Pearsona

Test zgodności Pearsona (test zgodności chi-kwadrat) pozwala nam sprawdzić czy rozkład danych w próbkę jest zgodny z rozkładem, z którym uważamy, że zgodny powinien być.

Np. prowadzimy badanie statystyczne i chcemy wiedzieć czy rozkład liczby lat wśród naszych badanych jest reprezentatywny czyli zgodny z rozkładem wieku mieszkańców Polski.

Albo mamy jakieś dane i chcemy sprawdzić czy pochodzą z rozkładu normalnego.

H_0 : dane z próbki są zgodne z ...

H_1 : dane z próbki nie są zgodne z ...

Test zgodności Pearsona

Na początku zebrane dane o liczbie badanych trzeba pogrupować.

Test zgodności Pearsona

Na początku zebrane dane o liczbie badanych trzeba pogrupować. Czasem jest to naturalne. Tu dla przykładu fikcyjne liczby badanych i ich grupy krwi.

	0-	0+	A-	A+	B-	B+	AB-	AB+
n_i	49	293	67	302	18	145	21	105

Przez n_i oznaczamy licznosc każdej grupy.

Test zgodności Pearsona

A jeśli dane są ciągłe to trzeba je pogrupować w przedziały.

Test zgodności Pearsona

A jeśli dane są ciągłe to trzeba je pogrupować w przedziały.
Tu na przykład fikcyjne badanie prędkości sieci studentów w Mb/s.

	do 50	51-100	101-200	201-300	ponad 300
n_i	8	45	72	23	4

Test zgodności Pearsona

A jeśli dane są ciągłe to trzeba je pogrupować w przedziały.
Tu na przykład fikcyjne badanie prędkości sieci studentów w Mb/s.

	do 50	51-100	101-200	201-300	ponad 300
n_i	8	45	72	23	4

Podział zależy od intuicji badacza. Ważne! Nie wolno tworzyć przedziałów zbyt mało licznych. Przyjmuje się, że $n_i \geq 8$, ale zdarzają się też badania z warunkiem $n_i \geq 5$. Jeśli przedziały są za mało liczne trzeba je połączyć.

Test zgodności Pearsona

Dlatego musimy skorygować nasz przykład.

	do 50	51-100	101-200	201-300	ponad 300
n_i	8	45	72	23	4

Test zgodności Pearsona

Dlatego musimy skorygować nasz przykład.

	do 50	51-100	101-200	201-300	ponad 300
n_i	8	45	72	23	4

	do 50	51-100	101-200	ponad 200
n_i	8	45	72	27

Test zgodności Pearsona

Teraz trzeba policzyć jak często powinny występować dane, gdyby były zgodne z rozkładem z H_0 . Albo mamy takie informacje (np. jak często statystycznie występuje dana cecha w całej populacji). Albo liczymy prawdopodobieństwa przedziałów z dystrybuanty.

Test zgodności Pearsona

Teraz trzeba policzyć jak często powinny występować dane, gdyby były zgodne z rozkładem z H_0 . Albo mamy takie informacje (np. jak często statystycznie występuje dana cecha w całej populacji). Albo liczymy prawdopodobieństwa przedziałów z dystrybuanty. Dla badania z grupą krwi.

	0-	0+	A-	A+	B-	B+	AB-	AB+
n_i	49	293	67	302	18	145	21	105
p_i	6%	31%	6%	32%	2%	15%	1%	7%

Przez p_i oznaczamy prawdopodobieństwa teoretyczne dla każdej grupy.

Test zgodności Pearsona

Skoro wiemy z jakim prawdopodobieństwem grupa występuje i ile osób mamy łącznie w badaniu (u nas $n = 1000$) to możemy policzyć ile powinniśmy mieć osób w każdej grupie ($\hat{n}_i$).

Test zgodności Pearsona

Skoro wiemy z jakim prawdopodobieństwem grupa występuje i ile osób mamy łącznie w badaniu (u nas $n = 1000$) to możemy policzyć ile powinniśmy mieć osób w każdej grupie ($\hat{n}_i$). Dla badania z grupą krwi.

	0-	0+	A-	A+	B-	B+	AB-	AB+
n_i	49	293	67	302	18	145	21	105
p_i	6%	31%	6%	32%	2%	15%	1%	7%
$\hat{n}_i$	60	310	60	320	20	150	10	70

Test zgodności Pearsona

Teraz liczymy liczbę stopni swobody.

$d = g - 1 - p$ gdzie

- g - liczba grup
- p - liczba parametrów nieznanych i szacowanych

Test zgodności Pearsona

Teraz liczymy liczbę stopni swobody.

$d = g - 1 - p$ gdzie

- g - liczba grup
- p - liczba parametrów nieznanych i szacowanych

Wyjaśnijmy p . Załóżmy, że hipoteza zerowa brzmi dane są zgodne z rozkładem normalnym, to pytanie jakim? Ile wynoszą jego m i σ ?

Test zgodności Pearsona

Teraz liczymy liczbę stopni swobody.

$d = g - 1 - p$ gdzie

- g - liczba grup
- p - liczba parametrów nieznanych i szacowanych

Wyjaśnijmy p . Załóżmy, że hipoteza zerowa brzmi dane są zgodne z rozkładem normalnym, to pytanie jakim? Ile wynoszą jego m i σ ?

Jeśli ich nie znam to musiałem je estymować (m ze średniej arytmetycznej, a σ z estymatora odchylenia standardowego $\sqrt{S^2}$ jak w pomocy do zestawu 12) bo inaczej nie miałbym z czego policzyć p_i do tabeli. Czyli w tym przypadku $p = 2$.

Test zgodności Pearsona

Teraz liczymy liczbę stopni swobody.

$d = g - 1 - p$ gdzie

- g - liczba grup
- p - liczba parametrów nieznanych i szacowanych

Wyjaśnijmy p . Załóżmy, że hipoteza zerowa brzmi dane są zgodne z rozkładem normalnym, to pytanie jakim? Ile wynoszą jego m i σ ? Jeśli ich nie znam to musiałem je estymować (m ze średniej arytmetycznej, a σ z estymatora odchylenia standardowego $\sqrt{S^2}$ jak w pomocy do zestawu 12) bo inaczej nie miałbym z czego policzyć p_i do tabeli. Czyli w tym przypadku $p = 2$. Gdybym wiedział, ile wynosi m (np. w badaniu średniego napięcia w gniazdku wiem, że $m = 230 \text{ V}$), a nie znał σ to $p = 1$, gdybym znał obie to $p = 0$.

Test zgodności Pearsona

Teraz liczymy liczbę stopni swobody.

$d = g - 1 - p$ gdzie

- g - liczba grup
- p - liczba parametrów nieznanych i szacowanych

Wyjaśnijmy p . Załóżmy, że hipoteza zerowa brzmi dane są zgodne z rozkładem normalnym, to pytanie jakim? Ile wynoszą jego m i σ ? Jeśli ich nie znam to musiałem je estymować (m ze średniej arytmetycznej, a σ z estymatora odchylenia standardowego $\sqrt{S^2}$ jak w pomocy do zestawu 12) bo inaczej nie miałbym z czego policzyć p_i do tabeli. Czyli w tym przypadku $p = 2$. Gdybym wiedział, ile wynosi m (np. w badaniu średniego napięcia w gniazdku wiem, że $m = 230 \text{ V}$), a nie znał σ to $p = 1$, gdybym znał obie to $p = 0$. W zadaniu z grupami krwi $p = 0$ bo nie ma tam żadnych parametrów, które szacuję.

Test zgodności Pearsona

Kluczowe wartości

- statystyka $\chi^2 = \sum_{i=1}^g \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i}$
- obszar krytyczny $K = [F^{-1}(1 - \alpha), +\infty)$, gdzie F to dystrybuanta rozkładu χ^2 o d stopniach swobody.

Test zgodności Pearsona

Kluczowe wartości

- statystyka $\chi^2 = \sum_{i=1}^g \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i}$
- obszar krytyczny $K = [F^{-1}(1 - \alpha), +\infty)$, gdzie F to dystrybuanta rozkładu χ^2 o d stopniach swobody.

Uwaga, jak korzystamy z gotowych tablic to często zawierają one nie dystrybuantę, a wartości krytyczne i wtedy lewy kraniec tego obszaru szukamy jako $\chi^2_{\alpha, d}$.

Powodzenia!
Proszę pytać jeśli coś jest niejasne.