

# Inżynierskie zastosowania statystyki

Ćwiczenia Semestr zimowy 2020/21

# ĆWICZENIA 4

**Testowanie hipotez** 

dr inż. Agata Kirjanów-Błażej



Wydział Elektroniki Katedra Systemów i Sieci Komputerowych



# Testy dla dwóch frakcji

#### Przykład 2.

Pewien importer owoców cytrusowych twierdzi, że owoce zawijane w papierki mniej się psują w transporcie od owoców, które importuje się starą metodą bez zawijania. Jednak wprowadzenie nowej metody wiąże się ze zwiększeniem kosztów. Dlatego importer przeprowadził eksperyment, który miał udowodnić, że owoce zawijane w papierki mniej się psują od nie zawijanych. Pobrał próbę losową 200 owoców zawijanych w papierki, z których uległo zepsuciu 85, oraz 150 owoców nie zawijanych w papierki, w których znaleziono 60 owoców zepsutych. Na poziomie istotności 0.05 oceń czy badania importera potwierdzają jego twierdzenie.

Testujemy hipotezę: H0: p1 = p2 H1: p1 < p2

Estymatory dla frakcji owoców popsutych w partii zawijanej i nie zawijanej w papierki są postaci:

Zbiór krytyczny jest postaci [1.96, ∞)

Nie mamy podstaw do odrzucenia hipotezy o równości frakcji. Można przyjąć, że frakcje zepsutych owoców w obu metodach są takie same. Czyli zawijanie owoców w papierki nie zmienia ich podatności na psucie.



### Testy frakcji

#### Zadania:

- 1. Wysunięto przypuszczenie, że jakość produkcji pewnego wyrobu po wprowadzeniu nowej, tańszej technologii nie uległa zmianie. Wylosowano próbę 120 sztuk tego wyrobu spośród wyprodukowanych starą technologią i otrzymano 12 sztuk złych. Wśród 160 wylosowanych sztuk wyprodukowanych nową technologią było 20 sztuk wadliwych. Czy wysunięte przypuszczenie można w świetle uzyskanych wyników uznać za uzasadnione?
- 2. W losowej próbie 700 mieszkańców pewnego rejonu będących w wieku produkcyjnym znalazło się 122 bezrobotnych. Czy na poziomie istotności 0,05 można stwierdzić, że stopa bezrobocia w tym rejonie jest większa od 20%?
- 3. Zbadano n=140 wylosowanych gospodarstw domowych w pewnym mieście ze względu na wysokość miesięcznych opłat za energię elektryczną. Spośród nich 84 gospodarstwa domowe płaciły miesięcznie za energię co najmniej 80 zł. Czy na poziomie istotności α=0,05 można stwierdzić, że % gospodarstw domowych, których miesięczne opłaty za energię elektryczną wynosiły co najmniej 80 zł jest mniejszy niż 70%?



- ightharpoonup Jeżeli:  $H_1:...<....$  to obszar krytyczny testu jest lewostronny
- ► Jeżeli: H<sub>1</sub>:...>.... to obszar krytyczny testu jest prawostronny
- ▶ Jeżeli:  $H_1: \dots \neq \dots$  to obszar krytyczny testu jest dwustronny

# α/2 α/2

### Testy dla jednej średniej

Populacja ma rozkład normalny

Populacja ma inny rozkład

odchylenie standardowe (σ)

Jest znane

odchylenie standardowe (σ) Nie jest znane

Próba duża

test z

$$z = \frac{\overline{X} - \mu}{\sigma} \sqrt{n}$$

Próba mała (n≤30)

Test t

Próba duża (n>30)

$$z = \frac{\overline{X} - \mu}{s} \sqrt{n-1}$$
  $z = \frac{\overline{X} - \mu}{s} \sqrt{s}$ 

Test z

 $z = \frac{X - \mu}{s} \sqrt{n}$ 



# Testowanie hipotez: etapy

- 1. Sformułowanie tezy rzeczowej i ustaleniu hipotez H<sub>0</sub> i H<sub>a</sub>;
- 2. Wyboru właściwej funkcji testowej (statystyki z próby);
- 3. Przyjęciu stosownego poziomu istotności  $\alpha$ ;
- 4. Odczytaniu wartości krytycznych w tablicach dystrybuanty właściwego rozkładu i ustaleniu obszaru krytycznego;
- 5. Odrzuceniu hipotezy zerowej na korzyść hipotezy alternatywnej, gdy funkcja testowa obliczona z próby znajduje się w obszarze krytycznym i nie odrzucenie jej, gdy funkcja testowa jest poza obszarem krytycznym.

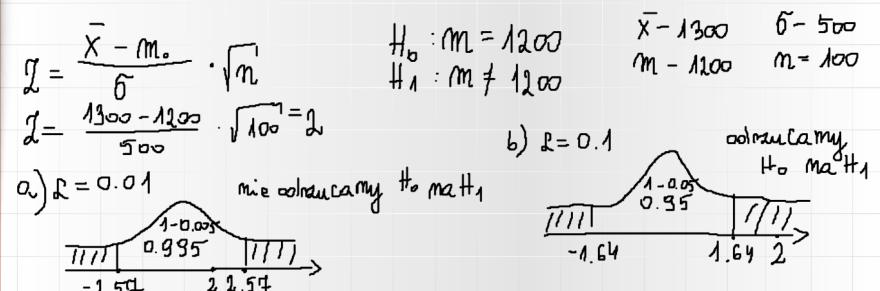


(rozkład normalny, znana wariancja)

#### Przykład:

W Politechnice Wrocławskiej pracownicy mogą brać pożyczki w Dziale Socjalnym. Rozkład wysokości tych pożyczek jest rozkładem normalnym z odchyleniem standardowym 500 zł. Rektor dąży do tego, żeby średni dług pracowników był równy 1200 zl. Główna księgowa sprawdziła zadłużenie 100 wybranych pracowników i policzyła średnią jego wartość, która wynosi 1300 zł.

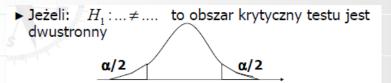
Czy średnie zadłużanie pracowników różni się od zakładanego? Sprawdź tę hipotezę na poziomie istotności 0,01; na poziomie istotności 0,1.





### Przykład:

2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916	
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936	
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952	_
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964	
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974	
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981	
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986	





### Przykład:

▶ Jeżeli:  $H_1:...≠....$  to obszar krytyczny testu jest dwustronny  $\alpha/2$ 

-		1									
	u	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
	1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
	1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
	1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
	1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
	1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
	1,5	0.9332	0.9345	0.9357	0,9370	0,9382	0.9394	0,9406	0,9418	0.9429	0.9441
	1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
r	1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
	1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
	1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767



#### Przykład:

Stawiamy hipotezę, że średnio 15 osób robi zakupy w kiosku w ciągu godziny. Na podstawie próby o liczebności 81 wyliczyliśmy m = 12, s = 4. Na poziomie istotności  $\alpha$  = 0.1 zweryfikuj, mniejszą ilość kupujących. Populacja ma rozkład normalny.



Pewien zakład bukmacherski zapewnia, że przeciętna stopa zwrotu z akcji w branży sportowej wynosi 11,5%.

Inwestor chce sprawdzić tę opinie. Pobiera próbę złożoną z akcji 50 spółki należących do branży. Na podstawie zebranych danych z próby stwierdza, że średnia stopa zwrotu z akcji wynosi 10,4%, przy odchyleniu standardowym 3,4%.

Czy inwestor ma dostateczne podstawy do odrzucenia zapewnień zakładu bukmacherskiego na poziomie istotności 0,05?

Nich m oznacza średnią stopę zwrotu z akcji dla ogółu danej branży.

Formułujemy hipotezę zerową H0, iż średnia ta równa jest wartości podanej przez firmę bukmacherską, tj. 11,5%, przeciwko hipotezie alternatywnej H1, iż średnia różni się od tej wartości.

 $H_0$ : m= 11,5%,

 $H_1$ : m  $\neq$  11,5%



Ponieważ dysponujemy dużą próbą, więc do weryfikacji hipotezy H0 możemy skorzystać z testy Z dla jednej średniej. W tym celu obliczamy wartość z statystyki testu Z.

I = -1.15

Budujemy dwustronny obszar odrzucenia dla  $\alpha$ = 0,05.

Jest nim suma przedziałów:  $(-\infty, -1,64]$  U  $[1,64,\infty)$ . -1,96

7,96

Odrewcamy to naty

၂, ၅၄ ၂, ၅ ၆ ၂ Wartość z statystyki testu leży w obszarze odrzucenia. Odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej. Stwierdzamy tym samym, że zapewnienia zakładu bukmacherskiego nie są prawdziwe. Ryzyko tego, że nasz wniosek nie jest słuszny, jest małe i wynosi  $\alpha$  (tutaj 0,05).



#### Zadanie 1:

Istnieje opinia, ze pasażerowie linii lotniczych maj tendencję do zabierania coraz większego bagażu podręcznego. Kabiny w samolotach pewnej linii lotniczej umożliwiają przechowywanie bagażu podręcznego o nominalnej wadze 20 kg.

Aby właściwie przeprojektować kabiny, zbadano wagę bagażu podręcznego dla losowej próby 150 pasażerów, uzyskując średnią wagę bagażu równą 22 kg, przy odchyleniu standardowym 6 kg.

Na poziomie istotności 0,05 zweryfikować przypuszczenie, że średnia waga bagażu podręcznego wśród pasażerów linii lotniczych nie różni się od wagi nominalnej, przeciwko hipotezie, iż ją przekracza.

ODP. Obszar odrzucenia: [1,64, ∞). Odrzucamy hipotezę, zerową na rzecz hipotezy alternatywnej.



#### Zadanie 2:

Na opakowaniu pewnego towaru widnieje napis: "przeciętna waga wynosi 200 g". Do Stowarzyszenia Konsumentów napływają jednak skargi klientów, iż producent zaniża wagę produktu.

W celu sprawdzenia prawdziwości informacji podanej przez producenta, zważono zawartość 100 losowo wybranych opakowań danego produktu. Uzyskano średnią wagę równą 199,5 g, przy odchyleniu standardowym 6 g.

Czy na podstawie uzyskanych wyników można sądzić, że informacja na opakowaniu nie jest prawdziwa, przyjmując poziom istotności 0,05?

Odp. Obszar odrzucenia, który dla  $\alpha$ = 0, 05 jest postaci: (-  $\infty$  , - 1, 64].

Nie ma podstaw do odrzucenia hipotezy zerowej.



(rozkład normalny, parametry nieznane)

Drugi test dla jednej średniej (test Studenta), stosowany jest przy założeniu, że cecha ma rozkład normalny  $N(\mu, \sigma)$ , przy czym parametry  $\mu$  i  $\sigma$  są nieznane.

Statystyka t ma rozkład Studenta z n-1 stopniami swobody, nie zależy od parametru σ ale od parametru S, S jest odchyleniem standardowym obliczonym z próby.



#### (rozkład normalny, parametry nieznane)

Odchylenie standardowe populacji nie jest znane oraz n<30 (mała próba)

#### Przykład:

W zakładzie produkującym czekoladę, dla losowo wybranych 26 (n=26) pracowników, otrzymano:

średni wiek X=38 lat, s=4 lata.

Na poziomie istotności α=0,01 należy zweryfikować hipotezę, że przeciętny wiek pracowników w tym zakładzie jest wyższy niż 35 lat?

$$T = \frac{38 - 35}{4} \cdot \sqrt{26 - 1} \quad \text{Hom} = 35$$

$$T = \frac{3}{4} \cdot 5 = 3.75 \quad \text{Hom} = 35$$

$$T = \frac{3}{4} \cdot 5 = 3.75 \quad \text{Hom} = 35$$

$$T = \frac{3}{4} \cdot 5 = 3.75 \quad \text{Hom} = 35$$

promedriat dorsan krytycamego (1,787, +0)

oolp. Odrzucarny Ho ma H1



(rozkład normalny, parametry nieznane)

#### Przykład:

Kierownictwo pewnej firmy ubezpieczeniowej wysunęło przypuszczenie, że średnie wypłaty ponoszone z tytułu odszkodowań powodziowych przekraczają kwotę 2 mln zł.

Przeanalizowano dane dotyczące wysokości odszkodowań poniesionych przez tę firmę podczas 5 kolejnych powodzi. Ustalono, że łączne kwoty odszkodowań powodziowych wypłaconych w rozważanych okresach wynosiły odpowiednio (w mln zł): 1,9; 3,7; 2,9; 2,0; 3,3. Czy można przyjąć, że kierownictwo firmy ma rację?

Zweryfikować odpowiednią hipotezę na poziomie istotności  $\alpha$  = 0, 05, zakładając, że rozkład wysokości odszkodowań powodziowych jest normalny.



(rozkład normalny, parametry nieznane)

#### Rozwiązanie:

Ponieważ rozkład badanej cechy jest normalny, więc korzystamy z testu Studenta.

$$\bar{X}$$
 - 1,76  $m = 5$  Ho :  $m = 2$   $f = 0.05$   
 $S$  - 0.43  $T = \frac{2.46 - 1}{0.49} - \sqrt{5 - 1}$  Shopnie swobody - 4  
 $T = 1,92$ 

[1,446, +50)



#### Zadanie 3:

Według normy technicznej wykonanie obróbki mechanicznej jednego pierścienia stalowego powinno zajmować szlifierzowi 22 minuty. Wylosowano 16 stanowisk roboczych, dla których średni czas obróbki wynosił 24 minuty. Jednocześnie z przeprowadzonego badania generalnego wiadomo, że odchylenie standardowe σ czasu obróbki wynosi 4 minuty.

Zakładając, że czas obróbki ma rozkład normalny, zweryfikuj na poziomie istotności  $\alpha = 0$ , 05 hipotezę H0 : a = 22 wobec hipotezy alternatywnej H1 : a  $\neq$ 22

Odp. T=2, 4- stopnie swobody, odrzucamy H0 na korzyść H1.



(Próby niezależne, znane odchylenia standardowe populacji lub duże próby.)

**Próby niezależne** – np. kobiety i mężczyźni, mieszkańcy miasta A i miasta B. **Próby zależne** – najczęściej "przed i po", np. ciśnienie przed i po wzięciu leku – badanie dotyczy tej samej próbki, ale w dwóch momentach czasu.

#### Model 1

Próby niezależne, znane odchylenia standardowe populacji lub duże próby.

$$u_{obl}=rac{\overline{x}_1-\overline{x}_2}{\sqrt{\left(rac{s_1^2}{n_1}+rac{s_2^2}{n_2}
ight)}}$$

Obszar krytyczny znajdujemy z tablic rozkładu normalnego.

#### Model 2

Próby niezależne, nieznane odchylenie standardowe populacji i małe próby.

$$t = rac{\overline{x}_1 - \overline{x}_2}{\sqrt{rac{n_1 \cdot s_1^2 + n_2 \cdot s_2^2}{n_1 + n_2 - 2} \left(rac{1}{n_1} + rac{1}{n_2}
ight)}}$$

Obszar krytyczny znajdujemy z tablic rozkładu t-Studenta dla n1+n2-2n1+n2-2 stopni swobody.



(Próby niezależne, znane odchylenia standardowe populacji lub duże próby.)

**Próby niezależne** – np. kobiety i mężczyźni, mieszkańcy miasta A i miasta B. **Próby zależne** – najczęściej "przed i po", np. ciśnienie przed i po wzięciu leku – badanie dotyczy tej samej próbki, ale w dwóch momentach czasu.

#### Model 3

Próby zależne (inaczej: test dla różnicy średnich).

Testujemy hipotezę  $H_0:\mu_z=0$ , gdzie  $\mu_z$  to średnia różnic (czyli testujemy hipotezę, że różnica między badanymi próbkami jest równa 0).

$$t=rac{\overline{z}}{s_z}\sqrt{n-1}$$

Obszar krytyczny znajdujemy z tablic rozkładu t-Studenta dla n−1 stopni.



(Próby niezależne, znane odchylenia standardowe populacji lub duże próby.)

#### Przykład:

Istnieje powszechne przekonanie, że pracownicy z wykształceniem wyższym zarabiają przeciętnie więcej niż pracownicy z niższym poziomem wykształcenia.

Zbadano wysokość zarobków w dwóch losowych próbach pracowników: z wykształceniem wyższym i z wykształceniem co najwyżej gimnazjalnym.

W pierwszej próbie, liczącej 60 osób, średnia wysokość miesięcznych zarobków wynosiła 3 tys. zł, przy odchyleniu standardowym 0,9 tys. zł. W drugiej próbie, liczącej 100 osób, średnia zarobków wynosiła 2,5 tys. zł, przy odchyleniu standardowym 0,5 tys. zł.

Zweryfikować odpowiednią hipotezę, przyjmując  $\alpha = 0,02$ .



(Próby niezależne, znane odchylenia standardowe populacji lub duże próby.)

Przykład:  $\mu_1$  M Niech  $\mu_1$  i  $\mu_2$  oznaczają średnie zarobki pracowników z wykształceniem wyższym i co najwyżej gimnazjalnym.

Formulujemy hipotezy:  $H_0: \stackrel{\mathcal{M}}{\mu_1} = \stackrel{\mathcal{M}}{\mu_2}, \qquad H_1: \stackrel{\mathcal{M}}{\mu_1} > \stackrel{\mathcal{M}}{\mu_2}.$ 

Ponieważ dysponujemy dużymi próbami, więc korzystamy z testu <u>U</u> dla dwóch średnich. Wartość statystyki testu

wynosi:

$$J = \sqrt{\frac{(0.9)^2}{60} + \frac{(0.5)^{2}}{100}} = 3.95$$

Prawostronny obszar odrzucenia dla  $\alpha = 0,02$ : [2,05,  $\infty$ ).

Odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej. Osoby z wykształceniem wyższym zarabiają średnio więcej niż osoby z wykształceniem co najwyżej gimnazjalnym.



(Próby niezależne, znane odchylenia standardowe populacji lub dużę próby.)

#### Przykład:

Bank chce sprawdzić, która metoda pozyskiwania pieniędzy – ze źródeł publicznych czy prywatnych – prowadzi do pozyskania większego funduszu.

Bank pobrał losową próbę 12 firm, które zaciągnęły kredyt tylko ze źródeł publicznych, stwierdzając, że przeciętna wartość kredytu w tej próbie wynosiła 60 tys. zł, przy odchyleniu standardowym 10 tys. zł. W losowej próbie 18 firm, które zaciągnęły kredyt tylko ze źródeł prywatnych, średnia wysokość kredytu wynosiła 80 tys. zł, przy odchyleniu standardowym 15 tys. zł.

Czy można sądzić, że publiczne źródła finansowania udzielają, przeciętnie biorąc, mniejszych kredytów, zakładając, że wysokość kredytów prywatnych i publicznych ma rozkład normalny o tej samej wariancji? (przyjąć  $\alpha$ =0,01).



(Próby niezależne, znane odchylenia standardowe populacji lub dużę próby.)

#### Przykład:

Niech  $\mu_1^{\text{M}}$  i  $\mu_2^{\text{M}}$  oznaczają średnie wartości kredytów ze źródeł odpowiednio publicznych i prywatnych.

Formulujemy hipotezy:  $H_0: \mu_1 = \mu_2$ ,  $H_1: \mu_1 < \mu_2$ .

Korzystamy z testu Studenta dla dwóch średnich.

Wartość statystyki testu jest równa:

Budujemy lewostronny obszar odrzucenia dla  $\alpha = 0,01$  i 28 stopni swobody. Jest nim przedział:  $(-\infty, -2, 47]$ .

Odrzucamy hipotezę zerową na rzecz alternatywnej, co pozwala wyciągnąć wniosek, że kredyty udzielane ze źródeł publicznych są, średnio biorąc, niższe.



#### Zadanie 4:

Badano wielkość plonu z hektara dla upraw gruszek gatunku "A" i gatunku "B". Zmierzono wielkość plonu z dziesięciu 1-hektarowych pół obsianych gatunkiem "A" i z 10 obsianych gatunkiem "B". Otrzymano dla gatunku "A" średnią wartość plonu x1=6,65, a dla gatunku "B" x2=6,36. Wiadomo, że wariancja pomiaru wynosi dla gatunku "A" σ21=0,05, a dla gatunku "B" σ22=0,06.

Zakładamy, że wielkość plonu z hektara ma rozkład normalny.

Na poziomie istotności α=0,05 zweryfikować hipotezę, że wartości przeciętne plonu z hektara są dla obu gatunków jednakowe wobec hipotezy alternatywnej mówiącej, że są różne.

 $(-\infty;-1,96)\cup(1,96;+\infty)$ 

odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej,



#### Zadanie 5:

Zbadano dwie partie próbek 100 gramowych wątroby ze względu na zawartość węglowodanów. Zawartości węglowodanów dla I partii próbek (wg) były następujące: 15; 17; 16; 17; 18,5; 18; 17,5. Dla II partii liczącej n=10 próbek średnia zawartość węglowodanów wynosi x=15,5 g i odchylenie standardowe s=1,1g.

Na poziomie istotności  $\alpha$ =0,05 zweryfikować hipotezę, że badane partie próbek pochodzą z populacji generalnych o tej samej średniej zawartości węglowodanów.

```
Opp. T=2,5975
Tkryt=2,131
(-∞;-2,131)∪(2,131;+∞)
hipotezę zerową odrzucamy na korzyść hipotezy alternatywnej
```