

Wykład nr

--

Nr indeksu

--	--	--	--	--

Nazwisko, imię (studenta)

Temat A

**Egzamin ze statystyki
Studia Licencjackie Stacjonarne
ZESTAW PRZYKŁADOWY NR 1**

Zad		Pkt
1	a	
	b	
	c	
2	a	
	b	
	c	
3	a	
	b	
	c	
4	a	
	b	
	c	
5	a	
	b	
	c	
6	a	
	b	
	c	
Test		
Razem		

Przy rozwiązywaniu zadań – jeśli to konieczne – należy przyjąć poziom istotności 0,05 oraz współczynnik ufności 0,90.

Zadanie 1

Zbadano 200 emerytów i 200 osób aktywnych zawodowo pod względem miesięcznych wydatków na komunikację miejską (w zł) i stwierdzono, że:

- w grupie emerytów dokładnie 25% osób wydawało na komunikację miejską co najwyżej 30 złotych, 60% co najwyżej 60 złotych, a 80% co najwyżej 90 złotych. Pozostała część emerytów wydawała na komunikację miejską co najwyżej 120 złotych;

- w grupie osób aktywnych zawodowo (pracownicy) strukturę wydatków obrazuje następujący szereg rozdzielczy:

(Wydatki>	Liczba pracowników
0-30	10
30-60	20
60-90	55
90-120	80
120-150	35
Ogółem	200

Na podstawie tego szeregu obliczono:

$$\sum \dot{x}_i \cdot n_i = 18300;$$

$$\sum (\dot{x}_i - \bar{x})^2 \cdot n_i = 197550;$$

Korzystając z powyższych danych należy:

- Zbudować szereg rozdzielczy miesięcznych wydatków na komunikację miejską w grupie emerytów oraz określić wartość mediany w tej grupie. [2]
- Przedstawić graficzny sposób wyznaczenia miary obliczonej w punkcie 1a). [1]
- Za pomocą miar pozycyjnych ocenić i porównać zróżnicowanie obu badanych grup ze względu na wydatki na komunikację miejską. Wiadomo, że kwartył pierwszy, mediana i kwartył trzeci w rozkładzie dla osób aktywnych zawodowo wyniosły odpowiednio: 71,9; 96,6; 115,4. [4]

Zadanie 2

Korzystając z danych z zadania 1 proszę:

- oszacować punktowo i przedziałowo średnie wydatki na komunikację miejską w populacji osób aktywnych zawodowo [4]
- Zweryfikować hipotezę mówiącą o tym, iż brak jest istotnych różnic między frakcjami emerytów i osób aktywnych zawodowo, którzy wydają miesięcznie od 60 do 90 złotych na komunikację miejską. [3]
- Dla jakiego minimalnego poziomu istotności decyzja weryfikacyjna z punktu 2b) mogłaby ulec zmianie? [1]

Zadanie 3

Przyjmuje się, że czas „żywności” pewnego rodzaju żarówki można opisać rozkładem normalnym o parametrach 4,11 tys. godz. oraz 0,1 tys. godz.

- Ile wynosi prawdopodobieństwo, że czas „żywności” losowo wybranej żarówki będzie różnić się nie więcej niż o 0,2 tys. godz. od średniego czasu założonego dla tego rodzaju żarówek? [1]
- Na wykresie dystrybucyjnym rozkładu teoretycznego czasu „żywności” tego rodzaju żarówki przedstawić graficzną prezentację prawdopodobieństwa z punktu 3a. [1]

- c. Z partii wyprodukowanych żarówek tego typu w sposób losowy wybrano 144 sztuki. Ile wynosi prawdopodobieństwo, że średni czas „żywności” żarówek w tej próbie będzie dłuższy niż 4,10 tys. godz.? [2]

Zadanie 4

Badanie zależności miesięcznych dochodów (x - tys. zł) w losowej próbie 100 gospodarstw domowych w małych miastach (liczących do 10 tys. mieszkańców) i miesięcznych wydatków ogółem w tych gospodarstwach (y – tys. zł) dostarczyło informacji:

$$\bar{x} = 3,19; \quad \bar{y} = 2,66;$$

$$\sum(\hat{y}_i - \bar{y})^2 = 104,51; \quad \sum(y_i - \hat{y}_i)^2 = 108,66;$$

$$s^2(x) = 2,79; \quad s^2(y) = 2,16; \quad c(x, y) = c(y, x) = 1,72;$$

(s^2 - wariancje nieobciążone)

Korzystając z powyższych danych należy:

- a) Zapisać funkcję regresji liniowej wydatków względem dochodów w tych gospodarstwach. Podać interpretację współczynnika regresji [3]
- b) Obliczyć odpowiednią miarę i ocenić w jakim stopniu zróżnicowanie wydatków może być wyjaśniane przez regresję liniową dochodów? [2]
- c) Dokonać prognozy miesięcznych wydatków w gospodarstwie domowym, którego dochody wynoszą 5 tys. zł. Ile wynosi standardowy błąd dla tej prognozy? [2]

Zadanie 5

Na podstawie informacji z zadania 4 dokonano analizy występowania zależności między wysokością dochodów gospodarstwa i płcią głowy gospodarstwa domowego. Dane charakteryzujące dochody zostały pogrupowane w przedziały i zbudowano tablicę korelacyjną:

Dochody gospodarstwa	Mężczyzna	Kobieta
do 2 tys.	10	17
2-4 tys.	28	13
powyżej 4 tys.	23	9

Obliczono:

$$\sum_i \sum_j \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} = 9,024$$

- a) Zweryfikować hipotezę dotyczącą niezależności dochodów od płci głowy gospodarstwa domowego w populacji mieszkańców małych miast. [2]
- b) Obliczyć i zinterpretować wartość \hat{n}_{21} [1]

Zadanie 6

Dane dotyczące półrocznych wydatków na leki (w zł) w 100 zbadanych gospodarstwach domowych w małych miastach w latach 2011-2014 przedstawia poniższe zestawienie:

Lata	2011		2012		2013		2014	
Półrocze	I	II	I	II	I	II	I	II
Wydatki (w zł)	200	230	210	250	210	260	220	270
Średnie ruchome		217,5	...	230	232,5	237,5	242,5	

a) Na podstawie tych danych dokonano analizy sezonowości i wyznaczono absolutny (addytywny) surowy wskaźnik sezonowości dla drugich półroczy $S'_2 = 18,333$. Uzupełnić brakującą średnią ruchomą a następnie obliczyć i zinterpretować poziom wahań sezonowych (oczyszczony) dla pierwszych półroczy analizowanych lat. [3]

b) Czy roczne wydatki na leki w latach 2012-2014 średnio rosły czy malały i w jakim tempie? (Podać odpowiedź na podstawie odpowiedniej miary dynamiki zmian). [2]

c) Wydatki na leki w badanych gospodarstwach analizowano według ich struktury (na receptę i bez recepty). Na podstawie indeksów agregatowych ustalono, że gdyby w 2014 roku ilość i struktura zakupionych leków była taka sama jak w 2012 roku, to tylko pod wpływem zmiany cen, ogólne wydatki na leki wzrosłyby w 2014 r w stosunku do 2012 roku o 4,5%. Na podstawie odpowiedniego indeksu ocenić wpływ zmian w ilości zakupionych leków na ogólne wydatki gospodarstw w tych latach. [2]

Część testowa

Punktacja: poprawna odpowiedź 1 pkt; brak odpowiedzi 0 pkt; odpowiedź błędna -1 pkt. *Jeżeli całkowita suma punktów z części testowej będzie ujemna, jako wynik części testowej zostanie przyjęte 0 pkt.*

1. W pewnej grupie liczącej 25 osób średnia, mediana i dominanta wieku były równe po 22 lata. Do grupy dołączyło pięć osób - jedna mająca 18 lat, dwie po 22 lata i 2 po 24 lata. Czy wynika z tego, że:		
a) Średnia wzrosła	T	N
b) Dominanta nie zmieniła się	T	N
c) Mediana wzrosła i wynosi 24 lata	T	N
2. Współczynnik korelacji Pearsona:		
a) Może być zastosowany do pomiaru zależności dwóch cech niemierzalnych	T	N
b) Przyjmuje wartości z przedziału [0; 1]	T	N
c) Równy zero wskazuje na brak korelacji liniowej między zmiennymi	T	N
3. Z twierdzenia granicznego de Moivre'a-Laplace'a wynika, że:		
a) Ciąg zmiennych losowych $\{X_n\}$ o rozkładzie dwumianowym z parametrami n i p jest zbieżny do rozkładu normalnego o parametrach $m=1, \sigma=1$	T	N
b) Ciąg zmiennych losowych $\{X_n\}$ o rozkładzie dwumianowym z parametrami n i p jest zbieżny do standardowego rozkładu normalnego $N(0,1)$	T	N
c) Ciąg zmiennych losowych $\{X_n\}$ o rozkładzie dwumianowym z parametrami n i p jest zbieżny do rozkładu t-Studenta	T	N

EGZAMIN ZE STATYSTYKI: ZESTAW PRZYKŁADOWY NR 2

Uwaga. W rozwiązaniach – tam gdzie to konieczne - przyjąć poziom istotności 0,05 a współczynnik ufności 0,95

Zadanie 1.(7pkt)

Poniżej zaprezentowano dwa empiryczne szeregi rozdzielcze. Pierwszy dotyczy dziennych obrotów 130 spółek o najwyższej kapitalizacji WIG20, WIG40, WIG 80 (w mln zł), a drugi wartości jednostki udziałowej (w zł) 170 funduszy inwestycyjnych na giełdzie w wybranym dniu czerwca 2016 roku.

Spółki	
Obroty	Liczba spółek
0 - 10	5
10 - 20	117
20 - 30	2
30 - 40	3
40 - 50	2
50 - 60	1
Średnia ważona	16,0
Wariancja nieobciążona	22,51

Fundusze inwestycyjne	
Wartość jednostki	Liczba funduszy
5- 200	10
200 - 400	40
400 - 600	40
600 - 800	70
800-1000	10

- Oblicz i zinterpretuj wartość mediany obrotów spółek (1,5pkt).
- Wyznacz i zinterpretuj klasyczną względną miarę zróżnicowania dla obrotów (1 pkt)
- Wykonaj wykres dystrybuanty empirycznej i zaznacz (bez wykonania obliczeń) wartość trzeciego kwartyła w rozkładzie wartości jednostek udziałowych funduszy inwestycyjnych (2,5 pkt).
- Porównując odpowiednie miary (które zostały wyznaczone w poprzednich podpunktach) lub wykonując histogram scharakteryzuj kierunek asymetrii w obu rozkładach. Zinterpretuj wynik (2 pkt).

Zadanie 2 (5pkt)

Wiadomo, że rozkład wagi paczkowanego produktu spożywczego ma rozkład normalny z wartością oczekiwaną 20 dag i odchyleniem standardowym 6 dag.
Oblicz prawdopodobieństwo, że:

- waga losowo wybranej paczki przekroczy 25 dag. Wynik proszę zilustrować graficznie na rozkładzie funkcji gęstości wystandaryzowanego rozkładu normalnego. (2p)
- średnia waga losowo wybranych 16 paczek nie przekroczy 21 dag. Wynik proszę zilustrować na wykresie dystrybuanty wystandaryzowanego rozkładu normalnego (3p)

Zadanie 3. (4pkt)

W celu porównania wynagrodzeń kobiet i mężczyzn w październiku 2012 r. wykonano na próbach losowych badanie wynagrodzeń według płci i wieku. Sformułuj odpowiednie hipotezy i oceń, przy jakim granicznym poziomie istotności można stwierdzić, że średnie wynagrodzenie mężczyzn w najmłodszej grupie wieku w populacji jest wyższe niż średnie wynagrodzenie kobiet w tej grupie wieku w populacji. Wiadomo, że próby losowe w najmłodszej grupie wieku liczyły po

200 osób, a odchylenie standardowe (nieobciążone) stanowiło i dla kobiet, i dla mężczyzn w tej grupie po 60% średniego wynagrodzenia.

Wiek	Mężczyźni	Kobiety
	średnie wynagrodzenie w tys. zł	
Ogółem	4,249	3,540
24 lata i mniej	2,627	2,332
25-34	3,908	3,339
35-44	4,887	3,793
45-54	4,379	3,579
55-59	4,155	3,674
60-64	4,682	4,947
65 lat i więcej	6,387	5,604

Zadanie 4 (5pkt)

Przeprowadzono badanie, którego celem była ocena zainteresowania bieżącą polityką i sprawami publicznymi w wyróżnionych grupach wieku. Respondentów podzielono na cztery grupy wg wieku po 7 osób w każdej grupie. Każdemu zadano po 10 pytań. Poniższa tabela zawiera liczby prawidłowych odpowiedzi uzyskane przez poszczególnych respondentów.

Liczba prawidłowych odpowiedzi w poszczególnych grupach			
Szkoła średnia	Młodzi dorośli	Osoby w wieku średnim	Osoby starsze
0	0	2	5
1	0	3	6
1	2	4	6
2	2	4	8
2	4	5	10
3	4	6	10
5	7	7	10

Ponadto obliczono średnie liczby poprawnych odpowiedzi dla każdej z grup, które wyniosły odpowiednio: 2; 2,71; 4,43; 7,86. Natomiast nieobciążona wariancja **w całej próbie** wyniosła 9.

a) Czy na podstawie badania można stwierdzić, że wiek istotnie różnicuje zainteresowanie polityką i sprawami bieżącymi w państwie? Zweryfikuj odpowiednią hipotezę. Wykorzystaj analizę wariancji (4p)

b) Jakie teoretyczne założenia są przyjmowane w analizie wariancji? (1p)

Zadanie 5 (10 pkt)

Poniższa tabela zawiera wybrane statystyki dotyczące przestępczości oraz możliwych determinant przestępczości dla 10 stanów USA. Dane o przestępczości przedstawiono, jako liczbę wyróżnionych rodzajów przestępstw przypadających na 10000 mieszkańców danego stanu.

Stan	Liczby przestępstw na 10000 mieszkańców:			Zmienne niezależne	
	morderstw	włamań	kradzieży samochodów	Gęstość zaludnienia (l. osób na milę kw.)	Stopa bezrobocia
Maine	1	27	96	42,7	5,4
Nowy Jork	4	161	145	412,8	5,5
Ohio	4	159	295	280,5	6,5

Iowa	1	44	164	53,7	4,0
Virginia	5	143	182	196,2	4,0
Kentucky	5	96	205	107,5	6,3
Teksas	6	162	293	92,9	4,8
Arizona	7	152	563	57,2	5,9
Waszyngton	3	94	382	98,4	5,3
Kalifornia	6	193	600	235,7	7,1
Średnia	4,2	123	293	158	5,48
Wariancja nieobciążona	4,18	3048	30277	14678	1,044

a) Naszkicuj wykres przedstawiający punkty reprezentujące obserwacje (stany) w dwuwymiarowej przestrzeni zmiennych: liczby kradzieży samochodów na 10000 mieszkańców i stopy bezrobocia. Czy można zauważyć liniową zależność między zmiennymi? Naszkicuj linię regresji. (1,5p)

b) Oblicz współczynnik korelacji w próbie i zweryfikuj hipotezę o dodatniej korelacji pomiędzy liczbą kradzieży samochodów na 10000 mieszkańców i stopą bezrobocia. Wiadomo, że kowariancja pomiędzy tymi zmiennymi wynosi 99,18. (3p)

c) Czy decyzja weryfikacyjna ulegnie zmianie, jeśli poziom istotności zmienimy do 0,1? Odpowiedź uzasadnij(1p)

d) Oszacuj i zinterpretuj oceny parametrów funkcji regresji, w której zmienną zależną będzie liczba kradzieży samochodów na 10000 mieszkańców, a zmienną niezależną stopa bezrobocia (2,5p)

e) Wiadomo, że w stanie Alabama stopa bezrobocia wyniosła 8,2, nie są natomiast znane statystyki dotyczące przestępczości. Wyznacz prognozę dla liczby kradzieży samochodów w tym stanie wykorzystując otrzymany model. Jaki będzie błąd standardowy tej prognozy, jeśli wiadomo, że odchylenie standardowe składnika resztowego $s(e)$ przyjęło wartość 153,17. (2p)

Zadanie 6 (5pkt).

W tablicy zestawiono dane na temat wskaźnika cen wybranych towarów konsumpcyjnych (zmiany z kwartału na kwartał)

Wyszczególnienie	2012				2013				2014			
	I kw.	II kw.	III kw.	IV kw.	I kw.	II kw.	III kw.	IV kw.	I kw.	II kw.	III kw.	IV kw.
	okres poprzedni = 100											
Ogółem	101,7	101,3	99,5	100,4	100,2	100,4	100,0	100,0	100,2	100,0	99,5	99,6
Żywność i napoje bezalkoholowe	103,5	102,0	98,0	100,8	101,9	100,8	99,0	99,9	101,8	98,8	97,5	99,3
Odzież i obuwie	95,2	102,2	95,4	102,7	94,8	102,6	95,3	102,7	95,1	102,4	95,2	102,7

a) proszę wyznaczyć, jak zmieniły się **cenę żywności i napojów bezalkoholowych** w czwartym kwartale 2014 r. w porównaniu do czwartego kwartału 2013 r. (2p)

b) wyznaczyć i zinterpretować średnie tempo zmian **cen żywności i napojów bezalkoholowych** w okresie I kwartał 2013 - III kwartał 2013.(2p)

c) czy uzasadniona byłaby – ze względów interpretacyjnych – ocena tempa zmian **cen odzieży i obuwia** w okresie IV kwartał 2012- IV kwartał 2013 za pomocą indeksu średniego tempa zmian. Odpowiedź proszę uzasadnić, niekoniecznie za pomocą obliczeń. (1p)

Zadanie 7 (testowe 9 pkt)**Otoczyć kółkiem w każdym przypadku odpowiedź T-tak lub N-nie***Punktacja w zadaniu: odpowiedź poprawna 1 pkt; brak odpowiedzi 0 pkt; odpowiedź błędna -1 pkt. Jeżeli całkowita suma punktów z części testowej będzie ujemna, jako wynik zadania testowego zostanie przyjęte 0 pkt.*

1	Przy poziomie ufności 0,95 wykonano estymację przedziałową wydatków na żywność w gospodarstwach domowych dwuosobowych.	
A	czy precyzja oszacowania zwiększyłaby się, jeśli współczynnik ufności przyjęty będzie na poziomie 0,99?	T N
B	gdyby w innym badaniu średnia była większa, czy przy pozostałych wartościach niezmiennych (liczebność próby, odchylenie standardowe, poziom ufności) długość przedziału ufności (rozpiętość) pozostałaby niezmienną?	T N
C	czy zwiększenie liczebności próby poprawi precyzję oszacowania?	T N
2	Test χ^2 wykorzystywany jest do:	
A	sprawdzania zgodności rozkładu z rozkładem normalnym	T N
B	badania niezależności zmiennych X,Y	T N
C	badania statystycznej istotności współczynnika regresji liniowej	T N
3	Roczny wskaźnik inflacji (agregatowy indeks cen dla ok. 1800 dóbr i usług konsumpcyjnych, obliczony formułą Laspeyresa) wyniósł 1,012. Oznacza to że:	
A	na pewno ceny wszystkich uwzględnionych dóbr i usług w badanym okresie wzrosły	T N
B	agregatowy indeks wartości jest na pewno mniejszy od 1,012	T N
C	agregatowy indeks ilości musi być mniejszy od jedności.	T N