

Lista siódma*
Metody probabilistyczne i statystyka
kierunek: Informatyka, studia I°

dr Jarosław Kotowicz
wersja z dnia 6 maja 2020

Spis treści

| | | |
|---|-----------------------------------|---|
| 1 | Testowanie hipotez | 1 |
| 2 | Niezależność. Testy niezależności | 5 |
| 3 | Regresja | 7 |
| 4 | Zadania prof. L.Uby | 8 |

1 Testowanie hipotez

Zadanie 1. Cecha X ma w zbiorowości generalnej rozkład $\mathcal{N}(m, 4)$. Na podstawie czteroelementowej próby należy zweryfikować hipotezę H_0 , że wartość przeciętna jest równa 10, wobec hipotezy H_1 , że wartość przeciętna wynosi 15. Przyjmujemy taką regułę postępowania, że hipotezę H_0 odrzucamy, gdy \bar{X} obliczona na podstawie próby przyjmuje wartość większą od 13. Oblicz prawdopodobieństwo popełnienia błędu I i II rodzaju.

Zadanie 2. Hipotezę, że wariancja w rozkładzie normalnym wynosi 80, odrzuca się, gdy wartość wariancji policzona na podstawie trzejelementowej próby jest mniejsza od 4 lub większa od 240. Oblicz prawdopodobieństwa popełnienia błędu I rodzaju.

Zadanie 3. Czas pracy pewnego rodzaju baterii ma rozkład $\mathcal{N}(m, 70)$. Na poziomie istotności $\alpha = 0,05$ zweryfikować hipotezę, że przeciętny czas pracy tego typu baterii wynosi ponad 500 godz., jeśli dla 16 losowo wybranych baterii otrzymano $\bar{X} = 560$ godz.

Zadanie 4. Zakłada się, że „długość życia” opon samochodowych ma rozkład normalny. Producent twierdzi, że wartość przeciętna tej charakterystyki jest równa 50 tys. km. Na podstawie 100 losowo wybranych opon otrzymano $\bar{X} = 45$ tys. km i $s = 8$ tys. km. Czy na poziomie istotności $\alpha = 0,05$ można uważać, że producent ma rację?

Zadanie 5. W zakładzie A dla losowo wybranych 10 pracowników otrzymano średni wiek 32 lata i odchylenie standardowe $s = 4$ lata. Czy można uważać (przy założeniu, że wiek pracowników ma rozkład normalny) że przeciętny wiek pracownika w tym zakładzie jest istotnie wyższy od 30 lat? Poziom istotności $\alpha = 0,05$.

Zadanie 6. W celu sprawdzenia, czy po dokonaniu usprawnienia w silniku samochodowym zmalało zużycie paliwa, przeprowadzono jazdy próbne i otrzymano następujące wyniki:
przed usprawnieniem: 5.7, 6.5, 6.1, 5.5, 5.0, 6.1, 6.2, 5.9 (l/100 km);
po usprawnieniu: 4.9, 5.0, 4.7, 5.0, 5.0, 4.0 (l/100 km).

Zalóżmy, że zużycia paliwa są niezależnymi zmiennymi losowymi o rozkładach normalnych o równych wariancjach. Na poziomie istotności $\alpha = 0,1$ zweryfikować hipotezę o jednakowym średnim zużyciu paliwa przed i po zmianie, przeciwko hipotezie mówiącej o mniejszym zużyciu paliwa po przeróbkach.

Zadanie 7. Przyjęto, że stopy zwrotu z inwestycji A i B są niezależnymi zmiennymi losowymi o rozkładzie normalnym z tą samą wariancją. Poprzednie stopy zwrotu były równe (w %):

A: 10, 15, 0, 1, 12, 7, 7, 8, 11, 12, 11, 14, 11, 8, 3, -2, 2.

B: 20, -10, 5, 9, -3, -12, -5, 1, 1, 6, 5, 16.

Na poziomie istotności $\alpha = 0.05$, zweryfikować hipotezę, że oczekiwane stopy zwrotu z obu inwestycji są jednakowe przeciwko hipotezie, że oczekiwana stopa zwrotu inwestycji B jest większa.

*©J.Kotowicz

Zadanie 8. Miesięczne wydatki na żywność w przeliczeniu na jedną osobę w gospodarstwach pracowniczych (w PLN) mają rozkład normalny. Na podstawie badania 400 losowo wybranych gospodarstw stwierdzono, że średnie wydatki w tej grupie wynoszą 250 PLN ze współczynnikiem zmienności równym 40%.

1. Czy na podstawie poniższych danych, zakładając prawdopodobieństwo popełnienia błędu I rodzaju na poziomie 0.05, można uznać, że średnie wydatki na żywność ogółu gospodarstw przekraczają 240 PLN?
2. Przy jakim poziomie istotności podjęta decyzja weryfikacyjna ulegnie zmianie?

Zadanie 9. Rozkład tygodniowego czasu poświęconego na naukę poza uczelnią studentów I roku studiów dziennych SGH jest rozkładem $\mathcal{N}(m, 5)$, natomiast w rozkładzie normalnym tygodniowego czasu studentów II roku odchylenie standardowe wynosi 6 godz. Pobrano niezależnie 10-elementową próbę studentów I roku oraz 24-elementową studentów II roku; średnie w tych próbach wynosiły odpowiednio: 20 godz. oraz 15 godz.

1. Czy na poziomie istotności 0.1, można przyjąć, iż średni czas nauki poza uczelnią ogółu studentów I roku jest wyższy niż na roku II?
2. Do jakiego przedziału liczbowego powinny należeć wartości odpowiedniej statystyki, aby nie było podstaw do odrzucenia weryfikowanej hipotezy?

Zadanie 10. W wyniku badania zmian poziomu płac pracowników firmy Intraco w Warszawie w latach 1993-1994 otrzymano następujące dane dla 50 losowo wybranych pracowników w każdym roku:

$$\begin{aligned} 1993: \bar{X} &= 398 \text{ PLN}, & s &= 118,7 \text{ PLN}, & n &= 50, \\ 1994: \bar{X} &= 654 \text{ PLN}, & s &= 213,5 \text{ PLN}, & n &= 50. \end{aligned}$$

Czy na podstawie powyższych wyników można mówić o wzroście poziomu płac ogółu pracowników Intraco w 1994 r. W porównaniu z rokiem 1993?

Zadanie 11. Wiadomo, że rozkład wyników pomiaru głębokości morza w pewnym rejonie jest normalny z odchyleniem standardowym 5 m. Dokonano 5 niezależnych pomiarów głębokości morza w pewnym rejonie i otrzymano następujące wyniki (w m): 862, 870, 876, 866, 871.

1. Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę, że średnia głębokość morza w tym rejonie równa jest 870 m.
2. W innym rejonie rozkład wyników pomiaru głębokości morza jest $\mathcal{N}(m, 10)$. Dla 10 niezależnych pomiarów otrzymano $\bar{X} = 865$ m. Czy można uważać, że przeciętna głębokość morza w obu rejonach jest jednakowa? Poziom istotności $\alpha = 0.01$.

Zadanie 12. Stalowe obręcze produkowane są na dwóch maszynach: A i B. Kontroler jakości uważa, że obręcze produkowane przez maszynę A mają średnicę istotnie większą od średnicy obręczy produkowanych przez maszynę B. Zakładamy, że rozkłady średnic obręczy dla maszyn A i B są: $\mathcal{N}(m_1, \sigma_1)$ i $\mathcal{N}(m_2, \sigma_2)$ oraz $\sigma_1 = \sigma_2$. Sprawdzić, czy kontroler jakości ma rację, jeśli dla 10 losowo wybranych obręczy z maszyny A otrzymano $\bar{X}_1 = 1.051$ i $s_1^2 = 0.000397$, a dla 15 obręczy z maszyny B mamy $\bar{X}_2 = 1.036$ i $s_2^2 = 0.00021$. Poziom istotności $\alpha = 0.01$.

Zadanie 13. W zakładzie A otrzymano następujące informacje o 16 pracownikach:

| Wiek pracowników | 20-24 | 24-28 | 28-32 | 32-36 |
|--------------------|-------|-------|-------|-------|
| Liczba pracowników | 4 | 6 | 4 | 2 |

Czy można uważać (zakładając, że wiek ma rozkład normalny), że wariancja wieku jest większa niż 10, na poziomie istotności $\alpha = 0,05$?

Zadanie 14. Tygodniowe wydatki na żywność mają rozkład normalny. Uważa się, że wartość przeciętna tych wydatków jest wyższa niż 40 zł.

1. Zweryfikować prawdziwość tego sądu na poziomie istotności $\alpha = 0,01$, jeśli dla 10 losowo wybranych rodzin otrzymano $\bar{X} = 48$ zł i $s = 10,8$ zł.
2. Czy na poziomie istotności $\alpha = 0,05$ można uważać, że odchylenie standardowe wydatków wynosi 9 zł?

Zadanie 15. W dwóch firmach przewozowych badano odległości przejazdów i otrzymano:

dla firmy A: wielkość próby - 15 przewozów i odchylenie standardowe z próby - 158 km,

dla firmy B: wielkość próby - 10 przewozów i odchylenie standardowe z próby - 283 km.

Czy można na poziomie istotności $\alpha = 0,05$ uważać, że wariancje odległości przewozów w obu firmach są takie same?

Zadanie 16. Badano zawartość nikotyny w dwóch gatunkach papierosów. W próbie liczącej 50 papierosów gatunku A zaobserwowano średnią arytmetyczną zawartości nikotyny $\bar{X}_1 = 23,8$ mg przy odchyleniu standardowym $s_1 = 1,2$ mg. W próbie liczącej 40 papierosów gatunku B zaobserwowano: $\bar{X}_2 = 24,1$ mg, $s_2 = 1,4$ mg.

1. Czy można uważać, na poziomie istotności 0,05, że przeciętna zawartość nikotyny w papierosach gatunku A jest niższa niż w papierosach gatunku B?
2. Na poziomie istotności $\alpha = 0,01$ zweryfikować hipotezę, że wariancje zawartości nikotyny w obu gatunkach papierosów są jednakowe.

Zadanie 17. Badając odruchy warunkowe u psa otrzymano następujące ilości śliny wydzielającej się przy bodźcu: 0,76; 0,54; 0,65; 0,40; 0,27; 0,65; 0,16, natomiast przy drugim bodźcu otrzymano: 0,32; 0,40; 0,20; 0,09; 0,38; 0,50; 0,15; 0,28. Na poziomie istotności 0,05 zweryfikować hipotezę, że przy drugim bodźcu ilość wydzielającej się śliny psa jest mniejsza zakładając, że badana cecha ma rozkład normalny. Jaką hipotezę należy najpierw zweryfikować?

Zadanie 18. Pomiarzy prędkości samochodów osobowych na pewnym odcinku autostrady dały wyniki:

| Prędkość w $\frac{km}{godz.}$ | Liczba samochodów |
|-------------------------------|-------------------|
| poniżej 80 | 7 |
| 80 - 90 | 30 |
| 90 - 100 | 40 |
| 100 - 110 | 69 |
| 110 - 120 | 48 |
| powyżej 120 | 6 |

Na poziomie istotności $\alpha = 0,01$ zweryfikowano hipotezę, że ponad 55% samochodów na tym odcinku jedzie z prędkością przynajmniej $100 \frac{km}{godz.}$. Jaka będzie decyzja, gdy $\alpha = 0,1$?

Zadanie 19. W ciągu 100 dni obserwowano liczbę awarii w sieci wodno-kanalizacyjnej w pewnym rejonie miasta i otrzymano:

| Liczba awarii | 0 | 1 | 2 | 3 | 4 | 5 |
|---------------|----|----|----|----|----|---|
| Liczba dni | 13 | 30 | 25 | 15 | 10 | 7 |

Czy można uważać, że mniej niż 60% obserwacji to takie dni, kiedy zarejestrowano przynajmniej 2 awarie? Poziom istotności $\alpha = 0,1$.

Zadanie 20. Zbadano wiek pracowników w dwóch przedsiębiorstwach A i B i otrzymano:

| Wiek pracownika | 20-30 | 30-40 | 40-50 | 50-60 |
|------------------------|-------|-------|-------|-------|
| Liczba pracowników w A | 10 | 20 | 40 | 30 |
| Liczba pracowników w B | 10 | 30 | 45 | 15 |

Na poziomie istotności $\alpha = 0,05$ zweryfikować hipotezę, że odsetek pracowników w wieku 40-50 lat w obydwu przedsiębiorstwach jest taki sam.

Zadanie 21. Poniższe zestawienie zawiera informację o łącznej liczbie punktów uzyskanych przez studentów studiów dziennych SGH z czterech kolejnych prac kontrolnych ze statystyki:

| Liczba pkt. | 65< | 65-70 | 71-75 | 76-80 | 81-85 | >85 |
|--------------|-----|-------|-------|-------|-------|-----|
| Liczba stud. | 10 | 20 | 40 | 50 | 40 | 40 |

1. Korzystając z testu λ -Kolmogorowa zweryfikować hipotezę, że powyższy rozkład jest zgodny z rozkładem normalnym o wartości oczekiwanej i odchyleniu standardowym równym odpowiednio 78 oraz 9 punktów.
2. Do jakiego przedziału liczbowego powinny należeć wartości statystyki λ , aby przy poziomie istotności równym 0,01 nie było podstaw do odrzucenia hipotezy zerowej?

Zadanie 22. Ewidencja liczby awarii urządzeń technicznych w zakładzie produkcyjnym w ciągu kolejnych 100 dni roboczych dostarczyła następujących informacji:

| Liczba awarii | Liczba dni |
|---------------|------------|
| 0 | 20 |
| poniżej 2 | 65 |
| poniżej 3 | 95 |
| poniżej 4 | 100 |

Czy zakładając prawdopodobieństwo popełnienia błędu I rodzaju na poziomie 0.05, można uznać powyższy rozkład za zgodny z rozkładem Poissona?

Zadanie 23. Zakłada się, że rozkład wagi noworodków (w kg) jest rozkładem normalnym o wartości średniej równej 3,5 kg oraz odchyleniu standardowym 0,5 kg. Na podstawie losowej próby 200 noworodków ustalono, co następuje:

| | | | | | | | | |
|--------------------|----|----|----|-----|-----|----|----|--------|
| Numer przedziału | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Ogółem |
| Liczby teoretyczne | 10 | 15 | 50 | ... | ... | 20 | 18 | 200 |

1. Oblicz i zinterpretuj liczebności teoretyczne w czwartym i piątym przedziale, wiedząc, że $x_{0.4} = 3,0$ oraz $x_{1.4} = 3,5$.
2. Z jakiego przedziału liczbowego pochodzi obliczona wartość statystyki chi-kwadrat, jeśli przy poziomie istotności równym 0,1 hipotezę zerową należało odrzucić?

Zadanie 24. Badanie 200 losowo wybranych czteroosobowych gospodarstw domowych pod względem miesięcznych wydatków na kulturę dostarczyło następujących danych: $\bar{X} = 30$ PLN, $s = 6,5$ PLN: pozostałe dane zaprezentowano w następującym zestawieniu:

| | | | | | |
|---|-------|-------|-------|-------|-------|
| Miesięczne wydatki | 15-21 | 21-27 | 27-33 | 33-39 | 39-45 |
| Liczba gospodarstw | 20 | 45 | 70 | 50 | 15 |
| $\frac{(n_i - \hat{n}_i)^2}{\hat{n}_i}$ | 0,610 | 0,164 | 0,009 | 0,101 | ... |

Obliczając brakujące dane, na poziomie istotności 0,05 zweryfikować hipotezę, że wydatki na kulturę w czteroosobowych gospodarstwach domowych mają rozkład normalny.

Zadanie 25. Strukturę 60 losowo wybranych pracowników SGH według liczby posiadanych dzieci przedstawia poniższe zestawienie:

| | | | | | | |
|--------------------------------|-----|----|----|----|---|-----|
| Liczba dzieci | 0 | 1 | 2 | 3 | 4 | 5 |
| Liczba pracowników | 10 | 19 | 14 | 9 | 6 | 2 |
| Teoretyczna liczba pracowników | ... | 18 | 16 | 10 | 4 | ... |

1. Oblicz brakujące liczby.
2. Na poziomie istotności 0,10 zweryfikować hipotezę, że powyższy rozkład pochodzi z populacji o rozkładzie Poissona.

Zadanie 26. Teoretyczne prawdopodobieństwa p_k powstania k cząstek w wyniku pewnej reakcji jądrowej są równe: $p_0 = 7/16$, $p_1 = 1/4$, $p_2 = p_3 = 1/8$, $p_4 = 1/16$. Przeprowadzono 496 niezależnych powtórzeń eksperymentu i w 212 powtórzeniach nie pojawiła się ani jedna cząstka, w 123 powstała jedna cząstka, w 62 dwie cząstki, w 45 trzy oraz w 54 powtórzeniach powstały cztery cząstki. Testem chi-kwadrat na poziomie istotności $\alpha = 0.01$ zweryfikować hipotezę H , że teoria dobrze opisuje zjawisko zderzeń.

Zadanie 27. Postawiono hipotezę H , że czas czekania na kolejnego klienta w pewnym systemie obsługi ma rozkład jednostajny na przedziale $(0,1)$. W celu sprawdzenia hipotezy wykonano 50 pomiarów odstępu czasu między chwilami przybycia kolejnych klientów. Otrzymano następujące wyniki:

| | | | | | |
|-----------------|-----------|-------------|-------------|-------------|-----------|
| Czas | $(0,0.2)$ | $[0.2,0.4)$ | $[0.4,0.6)$ | $[0.6,0.8)$ | $[0.8,1)$ |
| Liczba klientów | 20 | 15 | 9 | 5 | 1 |

Zweryfikować hipotezę H na poziomie istotności $\alpha = 0.05$.

Zadanie 28. Niech X będzie liczba klientów, którzy zgłoszą się do pewnego systemu obsługi w ciągu godziny. Postawiono hipotezę H , że zmienna X ma rozkład Poissona z parametrem $\lambda = 1$. Aby zweryfikować tę hipotezę, liczono ilu klientów pojawiło się w systemie w każdej ze stu godzin i otrzymano następujące wyniki:

| | | | | | |
|-----------------|----|----|----|----|------------|
| Liczba klientów | 0 | 1 | 2 | 3 | 4 i więcej |
| Liczba godzin | 39 | 30 | 19 | 10 | 2 |

Testem chi-kwadrat zweryfikować hipotezę H na poziomie istotności $\alpha = 0.05$.

Zadanie 29. ch na terenie poszczególnych województw w latach 1980 i 1992 dostarczyło danych zawartych w poniższej tablicy:

| | | |
|------------------------------------|-------------------|------|
| Odsetek studni o złej jakości wody | Liczba województw | |
| | 1980 | 1992 |
| poniżej 30 | 3 | 2 |
| poniżej 50 | 13 | 15 |
| poniżej 70 | 29 | 34 |
| poniżej 90 | 42 | 45 |
| poniżej 100 | 49 | 49 |

Korzystając z testu Kolmogorowa-Smirnowa, zweryfikować hipotezę o identyczności rozkładu województw węglug jakości wody w studniach w obu zbadanych latach. Przyjąć $\alpha = 0,005$.

Zadanie 30. Losowa próba licząca $n = 200$ niezależnych obserwacji wagi noworodków (w kg) dała następujące wyniki:

| | | | | | |
|------------|---------|---------|---------|---------|---------|
| Waga | 1,0-1,4 | 1,4-1,8 | 1,8-2,2 | 2,2-2,6 | 2,6-3,0 |
| Liczebność | 15 | 45 | 70 | 50 | 20 |

Na poziomie istotności 0,05 zweryfikować hipotezę, że rozkład wagi noworodków jest rozkładem normalnym.

Zadanie 31. W pewnej miejscowości sprawdzono w 200 losowo wybranych chwilach czerwca stopień zachmurzenia nieba i otrzymano następujące wyniki:

| | | | | | | | | | |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Stopień | 0-1 | 1-2 | 2-3 | 3-4 | 4-5 | 5-6 | 6-7 | 7-8 | 8-9 |
| Liczba | 43 | 20 | 15 | 14 | 13 | 16 | 15 | 22 | 42 |

Testem λ -Kolmogorowa na poziomie istotności $\alpha = 0,05$ zweryfikować hipotezę, że stopień zachmurzenia w danym miesiącu w tej miejscowości ma rozkład normalny.

Zadanie 32. W pewnym doświadczeniu mierzy się czas określonego efektu świetlnego. Przeprowadzono 140 doświadczeń i otrzymano wyniki:

| | | | | | |
|--------|---------|---------|---------|---------|---------|
| Czas | 0,0-0,2 | 0,2-0,4 | 0,4-0,6 | 0,6-0,8 | 0,8-1,0 |
| Liczba | 10 | 30 | 45 | 34 | 21 |

1. Czy można uważać, że rozkład czasu efektu świetlnego jest rozkładem $\mathcal{N}(0,6; 0,2)$ na poziomie istotności $\alpha = 0,05$?
2. Jaki będzie wynik, gdy zastosujemy test λ -Kolmogorowa?

Zadanie 33. Rejestrując straty czasu na skutek przestoju maszyn i urządzeń otrzymano dla dwóch wydziałów pewnego zakładu następujące wyniki:

| | | | | |
|---------------------------------|------|-------|-------|-------|
| Straty czasu (w min) | 0-10 | 10-20 | 20-30 | 30-40 |
| Liczba stanowisk na wydziale I | 10 | 14 | 15 | 11 |
| Liczba stanowisk na wydziale II | 20 | 30 | 40 | 10 |

1. Zweryfikować hipotezę, że rozkład strat czasu na obydwu wydziałach jest taki sam, $\alpha = 0,05$.
2. Czy można uważać na poziomie istotności 0,05, że rozkład strat czasu na wydziale II jest rozkładem $\mathcal{N}(20; 9)$?

2 Niezależność. Testy niezależności

Zadanie 34. W pewnym badaniu ankietowym przeprowadzonym wśród studentów uczelni warszawskich zanotowano m.in. wysokość miesięcznych wydatków na utrzymanie i miejsce pochodzenia studenta. Dla wylosowanej grupy 1000 studentów otrzymano następujące dane:

| | | | | |
|----------|------|-------------------------------|------------------------------------|-------|
| | Wieś | Miasta do 50 tys. mieszkańców | Miasta powyżej 50 tys. mieszkańców | n_i |
| 200-400 | 40 | 100 | 20 | 160 |
| 400-600 | 90 | 250 | 30 | 370 |
| 600-800 | 100 | 120 | 150 | 370 |
| 800-1000 | 20 | 30 | 50 | 100 |
| n_j | 250 | 500 | 250 | 1000 |

Sprawdzić, czy występuje zależność wydatków na cele kulturalne (X) od miejsca pochodzenia (Y). Oblicz wartość współczynnika V Cramera.

Zadanie 35. W przedsiębiorstwie A wprowadzono eksperymentalnie do produkcji nowe urządzenie i obserwowano jego pracę przez 100 dni, dokonując pomiarów liczby braków produkowanych detali (x_i - liczba sztuk/dzień) i liczby awarii automatu (y_j - liczba awarii/dzień). Uzyskano następujące wyniki:

| | 0-5 | 6-11 | Razem |
|-------|-----|------|-------|
| 0-4 | 7 | 12 | 19 |
| 5-9 | 12 | 44 | 56 |
| 10-14 | 17 | 8 | 25 |
| Razem | 36 | 64 | 100 |

Określić siłę zależności liczby braków (X) i liczby awarii automatu (Y), stosując współczynnik zbieżności V Cramera. Jaka jest jego interpretacja?

Zadanie 36. Istnieje przypuszczenie, że częste infekcje górnych dróg oddechowych zależą od liczby wypalanych dziennie papierosów. W pewnym mieście przeprowadzono odpowiednie badania, otrzymując poniższe dane:

| | 0 | 1-5 | 6-10 | 11-20 | 21 i więcej | Razem |
|----------------|----|-----|------|-------|-------------|-------|
| Nie choruje | 14 | 3 | - | - | - | 17 |
| Choruje rzadko | 20 | 42 | 9 | 2 | 1 | 74 |
| Choruje często | 6 | 15 | 39 | 49 | 57 | 166 |
| Razem | 40 | 60 | 48 | 51 | 58 | 257 |

Sprawdzić, czy prawdziwe jest powyższe przypuszczenie, stosując test niezależności chi-kwadrat i przyjmując poziom istotności 0,02. Jaka jest siła tego związku?

Zadanie 37. Na I roku studiów dziennych wyniki egzaminów w sesji letniej kształtowały się następująco:

| | Ndst | Dst | Dobry | B.dobry |
|-----------|------|-----|-------|---------|
| Kobiety | 28 | 22 | 32 | 18 |
| Mężczyźni | 34 | 30 | 36 | 10 |

Zbadać, czy rozkład stopni zależy od płci. Założyć poziom istotności $\alpha = 0,01$.

Zadanie 38. Poniższa tablica przedstawia dwuwymiarowy rozkład czasu pisania pracy kontrolnej i ocen uzyskanych z tej pracy w grupie 25 studentów:

| | 50-60 | 60-70 | 70-80 | 80-90 | n_i |
|-------|-------|-------|-------|-------|-------|
| 2 | 2 | 0 | 0 | 2 | 4 |
| 2,5 | 0 | 1 | 0 | 2 | 3 |
| 3 | 0 | 0 | 1 | 6 | 7 |
| 3,5 | 0 | 1 | 2 | 1 | 4 |
| 4 | 1 | 0 | 1 | 2 | 4 |
| 4,5 | 0 | 0 | 1 | 0 | 1 |
| 5 | 0 | 0 | 0 | 2 | 2 |
| n_j | 3 | 2 | 5 | 15 | 25 |

Wyznaczyć:

- średnie i wariancje dla rozkładów brzegowych zmiennych X i Y ,
- średnie i wariancje dla rozkładów warunkowych zmiennych X i Y ,
- kowariancję dla rozkładu łącznego,
- współczynnik korelacji ocen uzyskanych z pracy i czasu pisania pracy,
- wskaźnik krzywoliniowości zależności ocen od czasu pisania pracy,
- wskaźnik korelacyjny ocen z pracy względem czasu jej pisania,
- określić kształt zależności ocen uzyskanych przez badanych studentów od czasu pisania pracy w oparciu o empiryczną krzywą regresji ocen względem czasu pisania.

Zadanie 39. Badanie płacy (x_1 - w zł) oraz stażu pracy (x_2 - w latach) w próbie 14 robotników dało następujące notowania:

| | | | | | | | | | | | | | | |
|-------|-----|-----|------|------|-----|------|------|-----|-----|------|------|------|------|------|
| x_1 | 800 | 950 | 1110 | 1450 | 900 | 1320 | 1200 | 860 | 920 | 1500 | 1000 | 1600 | 1390 | 1100 |
| x_2 | 5 | 8 | 9 | 10 | 6 | 8 | 8 | 6 | 7 | 10 | 7 | 10 | 9 | 9 |

Określić charakter i natężenie związku między powyższymi zmiennymi za pomocą współczynnika korelacji Pearsona oraz ocenić jego statystyczną istotność wiedząc, iż krytyczna wartość odpowiedniego testu wynosi 2,1788.

Zadanie 40. Zbadać zależność między czasem przeznaczonym na reklamę telewizora marki Philips (w min/miesiąc) a miesięczną ich sprzedaż. W tym celu zebrano dane za 7 miesięcy 1996 roku:

| Miesiąc | Styczeń | Luty | Marzec | Kwiecień | Maj | Czerwiec | Lipiec |
|--|---------|------|--------|----------|-----|----------|--------|
| Czas w mediach przeznaczony na reklamy w min | 10 | 18 | 13 | 14 | 20 | 15 | 8 |
| Liczba sprzedanych telewizorów w tys. | 2,5 | 4,6 | 5,2 | 4,0 | 5,6 | 3,2 | 1,5 |

Do określenia stopnia zależności wykorzystać:

- współczynnik korelacji liniowej Pearsona,
- współczynnik korelacji rang Spearmana.

Zadanie 41. W finale jazdy figurowej na lodzie brało udział dwunastu zawodników: Z_1, Z_2, \dots, Z_{12} . Jazdę oceniali trzech sędziów, przy czym oprócz punktacji, każdy sędzia ustalał kolejność wszystkich zawodników; przedstawia ją tabela:

| Zawodnik | | Z_1 | Z_2 | Z_3 | Z_4 | Z_5 | Z_6 | Z_7 | Z_8 | Z_9 | Z_{10} | Z_{11} | Z_{12} |
|-------------------|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|
| Miejsce zawodnika | 1 | 3 | 4 | 5 | 6 | 1 | 12 | 2 | 9 | 8 | 10 | 11 | 7 |
| wg oceny | 2 | 2 | 5 | 4 | 6 | 1 | 9 | 3 | 8 | 10 | 11 | 12 | 7 |
| sędziego | 3 | 5 | 6 | 1 | 3 | 4 | 2 | 7 | 10 | 8 | 12 | 11 | 9 |

Wyznaczyć wartość współczynnika korelacji rang między ocenami wszystkich sędziów: $(1,2)$, $(1,3)$ i $(2,3)$. Na podstawie analizy otrzymanych wyników wytypować sędziego do odsunięcia od sędziowania zawodów za zbyt duże rozbieżności ze wskazaniami pozostałych sędziów.

3 Regresja

Zadanie 42. W pewnym zakładzie przemysłowym dokonano 10 pomiarów zużycia wody przy produkcji pewnego wyrobu i otrzymano następujące dane (X -wielkość produkcji w tys. sztuk, Y -zużycie wody w tys. m^3): $(1,8)$; $(2,15)$; $(3,8)$; $(4,10)$; $(5,22)$; $(6,14)$; $(7,17)$; $(8,28)$; $(9,22)$; $(10,26)$. Zakładając, że dla opisu zależności zużycia wody względem wielkości produkcji właściwy jest klasyczny model regresji liniowej

- oszacować parametry α i β odpowiedniej funkcji regresji,
- obliczyć wartość ocen parametrów σ^2 , $D(\hat{\alpha})$ i $D(\hat{\beta})$ w rozpatrywanym modelu regresji,
- obliczyć współczynnik determinacji dla oszacowanej funkcji,
- wyznaczyć przedziały ufności dla parametrów α i β ,
- ocenić istotność wpływu zmiennej zależnej ($\gamma = 0,05$).

Zadanie 43. W 1996 roku zebrano informacje w siedmiu krakowskich uczelniach o liczbie studentów $Y = [y_1, \dots, y_7]$ oraz o powierzchni (w m^2) sal dydaktycznych tych uczelni $X = [x_1, \dots, x_7]$:

| Uczelnia (i) | AR | AE | WSP | UJ | PK | ASP | AWF |
|-------------------------------|------|-------|------|-------|------|------|------|
| Liczba studiujących (y_i) | 4000 | 12000 | 6000 | 21000 | 9000 | 1000 | 3000 |
| Powierzchnia sal (x_i) | 4500 | 8800 | 4200 | 17000 | 6000 | 1800 | 2500 |

Należy:

- oszacować równanie regresji zmiennej Y względem zmiennej X ,
- oszacować równanie regresji zmiennej X względem zmiennej Y ,
- wyznaczyć wartość współczynnika korelacji r_{xy} ,
- określić szacunkową wielkość powierzchni sal wykładowych w jednej z krakowskich uczelni, jeśli ta zdecyduje się kształcić 15000 osób.
- oszacować wariancję resztową dla obu równań regresji.

Zadanie 44. Dane są:

$$\sum_{i=1}^n x_i y_i = 480, \quad S^2(X) = 10, \quad S^2(Y) = 40, \quad \bar{X} = 5, \quad \bar{Y} = 8, \quad n = 10.$$

Znaleźć:

1. równanie regresji zmiennej Y względem zmiennej X ,
2. równanie regresji zmiennej X względem zmiennej Y ,
3. wartość współczynnika korelacji liniowej między tymi zmiennymi.

Zadanie 45. Badając zależność między wielkością produkcji X pewnego wyrobu a zużyciem Y pewnego surowca wykorzystywanego w produkcji otrzymano dla losowej próby 7 obserwacji następujące wyniki (x_i w tys. sztuk, y_i w tonach):

| | | | | | | | |
|-------|---|----|----|----|----|----|----|
| x_i | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| y_i | 8 | 13 | 14 | 17 | 18 | 20 | 22 |

1. Wyznaczyć liniową funkcję regresji zużycia surowca względem wielkości produkcji.
2. Przy współczynniku ufności 0,95 oszacować metodą przedziałową współczynnik regresji zużycia surowca względem wielkości produkcji.
3. Na poziomie istotności 0,05 zweryfikować hipotezę, że współczynnik regresji w populacji generalnej ma wartość większą od 1,5.

4 Zadania prof. L.Uby

Zadanie 46. Badana cecha ma rozkład $N(\mu, \sigma)$ o znanej wariancji σ^2 . Weryfikujemy hipotezę $H : \mu = \mu_0$ na podstawie 25-elementowej próby, przy wykorzystaniu jako funkcji testowej statystyki U . Dla poziomu istotności $\alpha = 0.05$ wyznaczyć moc testu w przypadku gdy:

1. zbiorem krytycznym W jest przedział $[c, +\infty[$, gdzie c spełnia warunek $P(U \geq c | \mu_0)$.
2. zbiorem krytycznym W_1 jest suma przedziałów $]-\infty, -c_1] \cup [c_1, +\infty[$, gdzie $c_1 > 0$ spełnia warunek $P(U \leq -c_1 - \mu_0) + P(U \geq c_1 - \mu_0)$.

Zadanie 47. Z populacji, w której badana cecha ma rozkład $N(\mu, 4)$ wylosowano próbkę złożoną z 9 obserwacji. Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę $H : \mu = 2$ przy alternatywie $K : \mu = \mu_1 < 2$, jeśli średnia z próbki wynosi $\bar{x} = 1.4$.

Zadanie 48. W celu ustalenia czy dotychczasowa norma okresu użytkowania ubrań ochronnych (wynoszącą 150 dni) nie jest zbyt wysoka, zbadano faktyczny okres użytkowania ich na przykładzie 65 losowo wybranych robotników pracujących w normalnych warunkach. Otrzymano średnia długość okresu użytkowania 139 dni z odchyleniem standardowym 9.8 dni. Zakładając, że czas użytkowania ubrań ma rozkład normalny, stwierdzić, na poziomie istotności $\alpha = 0.01$, czy uzyskane wyniki stanowią podstawę do zmiany normy.

Zadanie 49. Zmierzono długość 198 włókien bawełny, a wyniki pomiarów zgrupowano w następującym szeregu

| | | | | | | | |
|-------------------|---|----|----|----|----|----|----|
| numer klasy | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| środek przedziału | 8 | 13 | 18 | 23 | 28 | 33 | 38 |
| liczność | 4 | 9 | 18 | 70 | 75 | 19 | 3 |

Na poziomie istotności $\alpha = 0.01$ zweryfikować hipotezę, że średnia długość włókna dla całej badanej partii włókien jest równa $\mu = 24$, wobec hipotezy alternatywnej $K : \mu \neq 24$.

Zadanie 50. W celu oszacowania dokładności pomiarów wykonywanych pewnym przyrządem dokonano 8 pomiarów pewnej wielkości i otrzymano: 18.17, 18.21, 18.05, 18.14, 18.19, 18.22, 18.06, 18.08. Zweryfikować na poziomie istotności $\alpha = 0.05$ hipotezę $H : \sigma^2 = 0.06$ dotyczącą wariancji σ^2 wskazań przyrządu, wobec hipotezy alternatywnej $K : \sigma^2 \neq 0.06$.

Zadanie 51. Do tarczy oddano 50 strzałów, mierząc odległości trafień od środka tarczy. Okazało się, że wariancja tych odległości jest równa $s^2 = 107.3 \text{ cm}^2$. Zakładając, że te odległości mają rozkład normalny na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę H , że wariancja odległości trafień od środka tarczy przy tego rodzaju strzelaniu jest równa $\sigma^2 = 100 \text{ cm}^2$, jeśli hipoteza alternatywna jest hipoteza $K : \sigma^2 > 100 \text{ cm}^2$.

Zadanie 52. Z populacji pobrano 450-elementową próbkę i otrzymano $s^2 = 14.9$. Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę $H : \sigma^2 = 16$, jeśli hipoteza alternatywna jest $K : \sigma^2 < 16$.

Zadanie 53. Wylosowano 300 mieszkań w Łodzi, w 54 przypadkach mieszkanie te były wyposażone w telefon. Na poziomie istotności $\alpha = 0.05$ zweryfikować hipotezę H , że wskaźnik struktury (frakcja) θ mieszkań w Łodzi mających telefon jest równy 0.4 wobec hipotezy alternatywnej $K : \theta \neq 0.4$.

Zadanie 54. Przyjmując, że wynik pomiaru nieznannej rezystencji opornika jest zmienną losową o rozkładzie $\mathcal{N}(\mu, \sigma)$ należy przetestować na poziomie istotności $\alpha = 0.05$ hipotezę $H : \mu = 500\text{m}\Omega$ przeciw alternatywie $K : \mu \neq 500\text{m}\Omega$. Wykonano $n = 10$ pomiarów przyrządem pomiarowym, którego odchylenie standardowe błędu pomiaru jest znane i wynosi $4\text{m}\Omega$. Otrzymano następujące wyniki: 506, 502, 498, 501, 503, 504, 498, 501, 503, 504 m Ω .

Zadanie 55. Przyjmijmy, że pomiaru jest zmienną losową o rozkładzie $\mathcal{N}(\mu, \sigma)$. Jeżeli jest równe mierzonej wielkości, to mówimy, że pomiar jest nieobciążony. Odchylenie standardowe określa klasę dokładności urządzenia pomiarowego. Na poziomie istotności $\alpha = 0.05$ przetestuj hipotezę $H : \sigma^2 = 0.06$ wobec $K : \sigma^2 > 0.06$. W celu zweryfikowania hipotezy wykonano 8 pomiarów tej samej wielkości badanym przyrządem i otrzymano wyniki : 18.17, 18.21, 18.05, 18.14, 18.19, 18.22, 18.06, 18.08.