INFORMATYKA

Lab1. *Podstawowe działania w R*

1. Oblicz: $\sin(2\pi)$; $\cos\left(\frac{3}{4}\right)$; $\mathrm{tg}(\pi)$; $\log(100)$; $\ln(15)$; $\log_7\left(\frac{1}{7}\right)$; e^3; $\sqrt[3]{64}$.

2. Utwórz wektor z wartościami od 1 do 10. Oblicz sumę składowych tego wektora.

3. Utwórz wektor **x** składający się z parzystych liczb od 2 do 20.
   a) Wyznacz najmniejszą i największą składową wektora **x**.
   b) Wyznacz indeks najmniejszej i największej składowej wektora **x**.
   c) Wyznacz składowe wektora **x** większe od 12.
   d) Wyznacz indeksy składowych wektora **x** większych od 12.
   e) Wyznacz indeks składowej wektora **x** równej 8.
   f) Odwróć kolejność elementów wektora **x**.
   g) Oblicz liczbę składowych wektora **x**.
   h) Sprawdź wynik działania **x*x** oraz **x**^2.
   i) Oblicz długość wektora **x** (Uwaga! Długość wektora to pierwiastek kwadratowy z sumy kwadratów składowych wektora).

4. Pod nazwą **a1** zapisz wektor o składowych: 3, 1, 6, 2, 7, 4, 4.
   a) Zwiększ każdą składową wektora **a1** o 3.
   b) Pod nazwą **a2** zapisz wektor **a1** bez składowych numer cztery i pięć.
   c) Pod nazwą **a3** zapisz sumę wektorów **a1** oraz **a2**. Wyświetl **a3**.

5. Utwórz wektor **a4** o składowych: 0, 1, 2, 3, 5, 7, 8.
   a) Utwórz bazę danych (za pomocą polecenia data.frame), w której wektory **a4** i **a1** przedstawiają liczbę par butów letnich i zimowych odpowiednio w grupie uczniów pewnej szkoły językowej i zapisać ją pod nazwą **X**.
   b) Korzystając z symbolu $ z poszczególnych kolumn bazy danych **X** utwórz dwa wektory o nazwach odpowiednio *lato* i *zima*.
   c) Ile w tej grupie jest butów zimowych, a ile letnich?
   d) Przedstaw dane z bazy na wykresie punktowym (polecenie plot). Nadaj wykresowi tytuł, a osiom układu nowe nazwy. Poeksperymentuj z kolorami (col), znacznikami (pch) i typami (type) wykresów.

6. Utwórz bazę danych składającą się z trzech wektorów pięcioelementowych o dowolnych naturalnych wartościach. Niech pierwszy wektor oznacza liczbę osób w poszczególnych klasach grających w koszykówkę, drugi – w piłkę nożną, a trzeci – w siatkówkę. Zapisz bazę pod nazwą **Y**.
   a) Który wymieniony sport najchętniej uprawiają uczniowie?
   b) Ile jest klas, gdzie na zajęcia z koszykówki chodzi więcej niż pięć osób? Jaka jest sytuacja z piłką nożną i siatkówką?
   c) Przedstaw zależność między liczbą uczniów grających w piłkę nożną i siatkówkę na wykresie punktowym.

7. W dwóch wektorach **b1**=(4, 7, 9, 2, 1, 5)’ oraz **b2**=(7, 1, 1, 6, 3, 6)’ przedstawiono w setkach liczbę długopisów złożonych przez każdego pracownika na dziennej (**b1**) i nocnej (**b2**) zmianie. Każdy pracownik na jednej zmienia powinien złożyć co najmniej 500 długopisów.
   a) Utwórz bazę danych, w której wektory **b1** i **b2** przedstawiają liczbę długopisów złożonych przez pracowników dziennej i nocnej zmiany. Zapisać ją pod nazwą **D**.
   b) Z poszczególnych kolumn bazy danych utwórz dwa wektory o nazwach odpowiednio *dzien* i *noc*.
   c) Oblicz, ilu jest pracowników, którzy nie wykonali dziennego limitu liczby złożonych długopisów.
   d) Która zmiana, dzienna czy nocna, złożyła więcej długopisów?
   e) Na którym stanowisku pracownik z dziennej zmiany złożył najwięcej długopisów? (pierwszy element wektora oznacza pierwsze stanowisko itd.)

8. Narysuj wykres funkcji $f(x) = x^2+3x-5$ na przedziale (-3, 4) (polecenie curve). Przetestuj działanie polecenia dla wykresów innych funkcji.

INFORMATYKA

Lab2. *Statystyka opisowa*

1. Wczytaj plik *loty.csv* zawierający dane dotyczące liczby pasażerów (w tysiącach) pewnej linii lotniczej w kolejnych miesiącach i latach, a następnie wykonaj polecenia:
   a) sprawdź, jakie wartości zawiera plik i jaki jest typ danych (class(dane));
   b) narysuj histogramy liczebności dla danych z kolejnych lat; zautomatyzuj rysowanie za pomocą pętli „for"; zadeklaruj tytuły kolejnych histogramów odwołując się do etykiet danych; przedstaw wszystkie wykresy w jednym oknie;
   c) wyznacz i zinterpretuj podstawowe miary statystyczne (średnia, mediana, pierwszy i trzeci kwartyl, odchylenie standardowe);
   d) porównaj dane z kolejnych lat za pomocą wykresów pudełkowych.

2. Wczytaj plik *oceny.csv* zawierający oceny końcowe studentów z czterech grup i wykonaj następujące polecenia:
   a) sprawdź typ danych wczytanych z pliku i zwróć uwagę na długości kolejnych zmiennych oraz sposób zapisu;
   b) wczytaj dane ponownie, zamieniając przecinki na kropki (opcja dec=",");
   c) sporządź szeregi rozdzielcze punktowe ocen w poszczególnych grupach (table);
   d) narysuj diagramy odcinkowe dla danych z kolejnych grup; zautomatyzuj rysowanie za pomocą pętli „for"; zadeklaruj tytuły kolejnych histogramów odwołując się do etykiet danych; wszystkie wykresy umieść w jednym oknie;

      UWAGA! Polecenie discrete.histogram należy do pakietu „arm".
   e) przedstaw dane z szeregów rozdzielczych na wykresach kołowych;
   f) wyznacz i zinterpretuj podstawowe miary statystyczne; w przypadku niepełnej długości danych posłuż się funkcją na.omit;
   g) porównaj dane z kolejnych lat za pomocą wykresów pudełkowych;

3. Wczytaj plik *truskawki.csv* zawierający wielkości plonów (w kg) truskawek z kilkudziesięciu plantacji w dwóch latach i wykonaj następujące polecenia:
   a) sprawdź typ danych wczytanych z pliku i zwrócić uwagę na długości kolejnych zmiennych; wyświetl dane zwracając uwagę na brakujące pomiary;
   b) sporządź szeregi rozdzielcze przedziałowe plonów w poszczególnych latach (cut);
   c) przedstaw dane z szeregów rozdzielczych na wykresach kołowych;
   d) narysuj histogramy probabilistyczne (freq=FALSE) dla plonów z kolejnych lat wykorzystując szeregi rozdzielcze z punktu (c); zautomatyzuj rysowanie za pomocą pętli „for"; zadeklaruj tytuły kolejnych histogramów odwołując się do etykiet danych; wszystkie wykresy przedstaw w jednym oknie;
   e) wyznacz i zinterpretuj podstawowe miary statystyczne; w przypadku danych „plon2010" wykorzystaj funkcję na.omit;
   f) porównaj dane z kolejnych lat za pomocą wykresów pudełkowych.

INFORMATYKA

Lab3. *Zmienna losowa*

1. Badania zanieczyszczeń wody pitnej z prywatnych studni zwykle wykazują, że 30% studni w pewnej okolicy jest skażona bakteriologicznie. Wylosowano 5 studni w tej okolicy. Niech zmienna losowa S oznacza liczbę skażonych studni wśród wylosowanych. Podaj rozkład prawdopodobieństwa zmiennej losowej S. Zilustruj rozkład zmiennej losowej S na wykresie odcinkowym.

2. Prawdopodobieństwo, że żarówka danego typu świeci przez przynajmniej 500 godzin wynosi 0,9. Niech B oznacza liczbę żarówek wśród 8 wylosowanych, których żywotność przekracza 500 godzin. Oblicz:
   (a) $P(B=8)$; (b) $P(B=7)$; (c) $P(B>5)$; (d) $E(B)$; (e) $SD(B)$.

3. Czas między awariami ogniw elektrycznych w systemie jest zmienną losową o rozkładzie wykładniczym z parametrem $\lambda = 0{,}01$. Obecnie w systemie pracują tylko dwa ogniwa, zaprojektowane równolegle, działające niezależnie od siebie. System może funkcjonować, jeśli przynajmniej jedno ogniwo dostarcza energii. Narysuj wykres funkcji gęstości (wykorzystaj funkcję `curve`). Wyznacz prawdopodobieństwo, że pojedyncze ogniwo:
   (a) przetrwa co najmniej 200 dni; (b) popsuje się przed upływem 100 dni.

4. Wielkość trzęsień ziemi zarejestrowanych w rejonie Ameryki Północnej jest zmienną losową o rozkładzie wykładniczym ze średnią 2,4 stopnia w skali Richtera. Wyznaczyć prawdopodobieństwo, że następne trzęsienie ziemi w tym regionie
   (a) przekroczy 3 stopnie w skali Richtera;
   (b) będzie miało między 2 a 3 stopnie w skali Richtera.

5. Przewody elektryczne wyprodukowane do zastosowania w pewnym systemie komputerowym powinny mieć opór (rezystancję) pomiędzy 0,12 i 0,14 oma. Rezystancja przewodów produkowanych przez firmę A jest zmienną losową o rozkładzie normalnym ze średnią 0,13 oma i odchyleniem standardowym 0,005 oma. Narysuj wykres funkcji gęstości rezystancji. Jakie jest prawdopodobieństwo, że losowo wybrany przewód produkowany przez firmę A spełnia wymagania stawiane przez system?

6. Czas schnięcia farby pewnego typu jest zmienną losową o rozkładzie normalnym z wartością oczekiwaną 2 godziny i odchyleniem standardowym 15 minut. Narysuj wykres funkcji gęstości czasu schnięcia badanej farby. Wyznacz prawdopodobieństwo, że farba schnie między 1h 51min i 2h 15 min.

7. Załóżmy, że 25% wszystkich studentów dużej uczelni publicznej otrzymuje stypendium. Niech X będzie liczbą studentów w losowej próbie o rozmiarze 100, którzy otrzymują stypendium. Oblicz dokładne i przybliżone prawdopodobieństwo, że pomoc finansową otrzyma co najwyżej 15 uczniów.

8. Średnia rezystancja przewodników danego typu wynosi 200 omów, z odchyleniem standardowym 10 omów. W obwodzie użytych zostało 25 przewodników. Wyznacz prawdopodobieństwo, że
   (a) <u>średnia</u> rezystancja wszystkich 25 przewodników zawiera się między 199 i 202 omy;
   (b) <u>całkowita</u> rezystancja wszystkich 25 przewodników nie przekracza 5100 omów.

9. Poziom cholesterolu we krwi pracowników pewnej firmy jest zmienną losową, dla której średnia to 202, a odchylenie standardowe to 14. Jakie jest prawdopodobieństwo, że <u>średni</u> poziom cholesterolu 64 wylosowanych do badania pracowników będzie zawierał się w przedziale między 198 a 206?

10. Wytrzymałość nici jest zmienną losową o średniej 0,5 kg i odchyleniu standardowym 0,2 kg. Załóżmy, że wytrzymałość liny jest sumą wytrzymałości nici w linie. Oblicz prawdopodobieństwo, że lina składająca się ze 100 nitek utrzyma 47 kg.

INFORMATYKA

Lab4. *Estymacja*

UWAGA! Większość danych pojawiających się w poniższych zadaniach dostępna jest w pliku *dane_est_hip.csv*

1. W celu oceny nowego procesu produkcji syntetycznych diamentów sprawdzono wagę [karaty] diamentów wyprodukowanych tą metodą uzyskując następujące wyniki:

   0,46 0,61 0,52 0,48 0,57 0,54 0,47 0,63 0,51 0,49 0,58 0,55.

   Przyjmijmy, że badana zmienna ma rozkład normalny. Wykonaj poniższe polecenia, przy czym do zrealizowania punktów (c)–(e) samodzielnie zdefiniuj funkcje wyznaczające oceny przedziałowe parametrów.

   (a) Określ populację, próbę i badaną zmienną.
   (b) Wyznacz oceny punktowe średniej, wariancji i odchylenia standardowego wagi diamentów produkowanych tą metodą. 0,534167; 0,00308; 0,0555073
   (c) Oceń metodą przedziałową z ufnością 0,95 średnią badanej populacji. Uzasadnij wybór metody budowy przedziału ufności. (0,498; 0,57)
   (d) Zwiększ ufność z jaką chcemy wnioskować i porównaj długości uzyskanych przedziałów ufności.
   (e) Wybierz współczynnik ufności i oceń za pomocą przedziału ufności wariancję oraz odchylenie standardowe badanej populacji.

2. Aby oszacować średnią zawartość nikotyny w nowej marce papierosów, wybrano 15 paczek papierosów i zbadano w nich zawartość nikotyny otrzymując dane:

   1,87 2,28 1,77 2,13 1,43 1,64 2,38 1,39 1,94 2,68 1,95 0,86 1,98 1,69 1,15

   Z wcześniejszych badań wiadomo, że odchylenie standardowe zawartości nikotyny w papierosach nowej marki jest równe 0,7 miligrama. Oceń przedziałowo z ufnością 95% średnią zawartości nikotyny we wszystkich papierosach? (1,455; 2,164)

3. Badacz zajmujący się możliwością zastosowania wodorostów do karmienia zwierząt badał zawartość białka w wodorostach. Wyniki 18 pomiarów z 50-kilogramowych próbek wodorostów przedstawiają się następująco:

   4,28 3,3 4,22 2,77 2,75 2,93 3,86 3,05 4,12 2,88 3,94 4,99 2,08 4,35 2,7 4,09 2,81 2,82

   Przyjmijmy, że zawartość białka w wodorostach ma rozkład normalny, ale parametry nie są znane. Wykonaj poniższe polecenia. Do wykonania zdań (b) i (c) wykorzystaj gotowe funkcje pakietu R.

   (a) Podaj wartość estymatora punktowego średniej i wariancji populacji. 3,44; 0,63
   (b) Oceń metodą przedziałową prawdziwą średnią zawartość białka w 50-kilogramowych porcjach wodorostów (przyjmij współczynnik ufności 0,90). (3,115; 3,767)
   (c) Zbuduj 90% przedział ufności dla wariancji zawartości białka. (0,388; 1,235)

4. Zużycie wody w fabryce podlega losowym wahaniom w kolejnych dniach roku. Na podstawie 365 obserwacji stwierdzono, że średnie dzienne zużycie wynosi 102 hl, a wariancja 81 hl^2.

   (a) Przyjmując współczynnik ufności 0,98 oceń przedziałowo średnie dzienne zużycie wody w fabryce. (100,9;103,1)
   (b) W następnym roku cena wody ma wzrosnąć. Produkcja będzie musiała być ograniczona, jeżeli średnie dzienne zużycie wyniesie co najmniej 122 hl. Czy na podstawie uzyskanego wyniku jest to prawdopodobna sytuacja? nie

Lab4. *Estymacja*

5. W pewnej firmie badano liczbę wypijanych dziennie filiżanek kawy. Wśród przebadanych 120 pracowników uzyskano następujące odpowiedzi:

Liczba filiżanek kawy	0	1	2	3	4
Liczba osób	14	28	36	28	14

(a) Budując samodzielnie odpowiednią funkcję i przyjmując współczynnik ufności 0,95 wyznacz ocenę przedziałową frakcji osób pracujących w tej firmie, które:
    (i) wypijają do 2 filiżanek kawy dziennie (włącznie); binom: (55,7%; 73,5%)
    (ii) wypijają do 3 filiżanek kawy dziennie (włącznie). binom: (81,2%; 93.5%)

(b) Oceń metodą przedziałową z ufnością 0,95 ile średnio filiżanek kawy wypija pracownik w badanej firmie. (1,78; 2,22)

6. Przeprowadzano obserwacje dotyczące opóźnień pociągów. Stwierdzono, że spośród 1000 losowo wybranych 160 przyjechało z opóźnieniem. Zakładając, że opóźnienia pociągów są niezależne od siebie i jednakowo prawdopodobne, oceń przedziałowo prawdopodobieństwo opóźnienia pociągu (z ufnością 0,9). (0,14; 0,18); (0,141; 0,181); (0,141; 0,181)

   Pani K zna ten wynik i pewnego dnia tak się złożyło, że wyszła zbyt późno i nie wie czy zdąży na pociąg. Postanawia zaryzykować i spróbować. Czy podjęła słuszną decyzję?

7. Kontrola w browarze ma na celu sprawdzenie, czy proporcja niedopełnianych puszek z napojem, p, nie jest zbyt duża. Otwarcie wszystkich puszek jest oczywiście niemożliwe (spowodowałoby to zniszczenie produktu), dlatego wybrano próbę 100 puszek i zmierzono objętość znajdującego się w nich napoju. Okazało się, że 4 opakowania były niepełne. Oceń z ufnością 95% prawdziwą proporcję niedopełnianych puszek. (0,15%; 7,85%); (1,1%; 10%); (1,2%; 10,6%)

8. Próba 1000 magistrów zarządzania wykazała, że w chwili obecnej 122 jest bezrobotnych. Oceń przedziałowo z ufnością 90% prawdziwą proporcję wszystkich bezrobotnych magistrów zarządzania. (10,4%; 14%); (10,5%; 14,1%); (10,5%; 14,1%)

INFORMATYKA

Lab5. *Testowanie hipotez*

UWAGA! Większość danych pojawiających się w poniższych zadaniach dostępna jest w pliku *dane_est_hip.csv*

1. Aby mogły pracować urządzenia prądotwórcze elektrowni wiatrowej <u>średnia</u> prędkość wiatru powinna przekraczać 4 m/s. W celu stwierdzenia czy sensowna jest budowa elektrowni wiatrowej mierzono przez okres roku każdego miesiąca przeciętną prędkość wiatru w okolicach Darłowa uzyskując wyniki (w m/s):

   5,9 4,4 5,4 3,8 4,0 4,2 3,4 3,6 4,6 6,5 5,6 4,8.

   Zakładając, że prędkość wiatru jest zmienną losową o rozkładzie normalnym oraz przyjmując poziom istotności $\alpha = 0{,}05$ sprawdź, czy okolice Darłowa nadają się do budowy elektrowni wiatrowych. W tym celu skonstruuj odpowiednią procedurę testującą. Porównaj otrzymane wyniki z uzyskanymi po zastosowaniu odpowiedniej funkcji z pakietu R. *odrzucamy H_0*

2. Przyjęto, że współczynnik efektywności pompy cieplnej *COP* jest zadawalający, gdy jego <u>średnia</u> wartość wynosi co najmniej 3,5 (co oznacza, że ponad 70% dostarczonego przez pompę ciepła pochodzi z naturalnego źródła ciepła, a reszta pochodzi z pracy sprężarki). Nabywca pompy ma wątpliwości co do jakości zakupu i wysunął przepuszczenie, że współczynnik efektywności pompy cieplnej w jego gospodarstwie domowym jest znacznie mniejszy niż 3,5. Przez pewien okres mierzono współczynnik *COP* w tym gospodarstwie otrzymując następujące wyniki:

   3,5 3,2 3,6 3,0 3,3 3,8 2,5 3,0 3,7 3,9.

   Zakładając, że zmienna opisująca wartości współczynnika *COP* jest zmienną losową o rozkładzie normalnym i na podstawie powyższych wyników (przyjmując poziom istotności $\alpha=0{,}01$) sprawdź, czy wątpliwości nabywcy są słuszne. *brak podstaw do odrzucenia H_0*

3. Podejrzewa się, że w związku z ociepleniem klimatu <u>średnia</u> głębokość morza w pewnym rejonie uległa zmianie na przestrzeni lat. W celu weryfikacji tej hipotezy, dokonano 5 niezależnych pomiarów głębokości morza w tym rejonie i otrzymano następujące wyniki (w m):

   862, 870, 876, 866, 871.

   Wiedząc, że rozkład głębokości morza jest normalny z odchyleniem standardowym 5 m oraz że średnia głębokość wynosiła w 1950 roku 870 m, przetestuj wspomnianą hipotezę. Przyjmij poziom istotności $\alpha = 0{,}05$. *brak podstaw do odrzucenia H_0*

4. Automat produkuje blaszki określonych wymiarów o nominalnej grubości 0,04 mm. Inżynier podejrzewa, że automat uległ rozregulowaniu i produkowane blaszki są grubsze niż nominalny wymiar. W tym celu wylosował próbę 40 blaszek otrzymując następujące wyniki:

   0,048 0,028 0,037 0,033 0,054 0,046 0,041 0,043 0,044 0,05 0,047 0,052 0,053 0,048
   0,027 0,056 0,058 0,039 0,026 0,034 0,043 0,042 0,047 0,022 0,046 0,04 0,036
   0,043 0,041 0,044 0,043 0,044 0,038 0,046 0,041 0,038 0,047 0,03 0,041 0,049.

   Przyjmując poziom istotności $\alpha = 0{,}02$ sprawdź poprawność przypuszczenia inżyniera.
   brak podstaw do odrzucenia H_0

5. W próbce laboratoryjnej mleka spożywczego wykonano 10 oznaczeń (w %) zawartości tłuszczu i uzyskano:

   1,5 1,8 1,5 1,7 1,6 1,6 1,8 1,6 1,7 1,6.

   Przyjmując, że rozkład zawartości tłuszczu w mleku spożywczym jest normalny, odpowiedź na poniższe pytania (przyjmij poziom istotności $\alpha = 0{,}05$):

   (a) Czy obserwacje potwierdzają, że <u>średnia</u> zawartość tłuszczu w mleku jest istotnie różna od 1,7 %?
   brak podstaw do odrzucenia H_0

   (b) Czy można twierdzić, że <u>wariancja</u> zawartości tłuszczu w mleku jest mniejsza niż $0{,}02\ (\%)^2$.
   brak podstaw do odrzucenia H_0

6. Kukułki podrzucają swoje jaja różnym ptakom, między innymi małym strzyżykom. Obserwacje przyrodników wskazują, że kukułka potrafi znieść jajo wielkości podobnej do jaj „adopcyjnych rodziców". Zmierzono długość [w mm] 21 jaj podrzuconych strzyżykom otrzymując wyniki:

17,93	18,52	19,66	14,30	17,52	20,76	20,26
19,82	21,40	16,54	18,64	17,62	20,79	19,14
16,74	14,93	18,56	15,43	15,19	21,05	20,79.

Wiadomo, że średnia długość jaj strzyżyka wynosi 17 mm, a odchylenie standardowe 2,5 mm. Zakładamy, że badana cecha ma rozkład normalny.

(a) Na poziomie istotności 0,05 zweryfikuj, czy zebrane dane przeczą stawianej przez przyrodników hipotezie dotyczącej:
(i) średniej długości podrzuconych jaj; *odrzucamy H_0*
(ii) wariancji długości podrzuconych jaj. *brak podstaw do odrzucenia H_0*

(b) Zbuduj 95% przedział ufności dla średniej długości jaj podrzucanych strzyżykom. Jaki jest związek między skonstruowanym przedziałem ufności a decyzją podjętą przy testowaniu hipotez? *(17,36;19,37)*

7. Sondaż opinii publicznej na temat frekwencji oczekiwanej na wyborach wykazał, że w losowo wybranej grupie 2500 osób 1600 zamierza uczestniczyć w głosowaniu. Czy na poziomie istotności równym 0,05 próba przeczy twierdzeniu, że 60% ogółu osób zamierza wziąć udział w wyborach? Skonstruuj odpowiednią procedurę testującą. Porównaj otrzymane wyniki z uzyskanymi po zastosowaniu testu dokładnego dostępnego w pakiecie R. *odrzucamy H_0*

8. Przeprowadzono badanie jakości jaj kurzych pochodzących z pewnej fermy. Zakłada się z góry, że 2% jaj jest złej jakości. Wylosowano 1200 jaj do zbadania i wśród nich 16 okazało się złej jakości. Czy obserwacje przeczą przyjętemu założeniu i potwierdzają, że frakcja ta w badanej fermie jest mniejsza? Wnioskuj na poziomie istotności 0,05. W tym celu skonstruuj odpowiednią procedurę testującą. Porównaj otrzymane wyniki z uzyskanymi po zastosowaniu testu dokładnego dostępnego w pakiecie R. *brak podstaw do odrzucenia H_0*

9. W sondażu przeprowadzonym przez pracownię badania opinii społecznych wśród 1100 dorosłych Polaków, 1000 ankietowanych odpowiedziało, że w ubiegłym miesiącu nie przeczytało żadnej książki. Pozostali twierdzili, że przeczytali przynajmniej jedną książkę. Na podstawie tych danych, na poziomie 0,05, zweryfikuj, czy opinia, że procent Polaków, którzy nie przeczytali żadnej książki jest większy niż 90, jest uzasadniona? *brak podstaw do odrzucenia H_0*

INFORMATYKA

Lab6. *Analiza wariancji (ANOVA)*

UWAGA! Przedstawione dane dostępne są (zgodnie z kolejnością) w plikach csv:

anova_cisnienie, anova_kopalnie, anova_mikrometr, anova_sportowcy, anova_chomiki

1. Inżynier chemiczny chce sprawdzić, czy różne warunki ciśnieniowe mają wpływ na średnią produkcję pewnego wyrobu z danego typu surowca. Poniższa tabela zawiera wyniki przeprowadzonego eksperymentu (w g/l):

Ciśnienie	Replikacje									
Niskie	28	26	29	30	28	31	26	32	25	29
Średnie	30	29	30	30	28	32	29	32	28	30
Silne	31	29	33	33	29	33	28	32	27	32
Bardzo silne	29	27	30	31	27	32	27	32	27	30

Czy ciśnienie ma wpływ na wielkość produkcji? Wnioskuj na poziomie istotności 0,05. Analizę wariancji poprzedź testowaniem założeń o normalności (np. lillie.test) oraz jednorodności wariancji (np. bartlett.test).

F=2,2665; p-val=0,097; brak podstaw do odrzucenia H_0

2. W doświadczeniu badano zawartość popiołu (części niepalnych) dla ekogroszku wyprodukowanego na bazie węgla wysokogatunkowego pochodzącego z pięciu różnych kopalni. Otrzymano następujące wyniki:

Obiekty (kopalnie)	Replikacje			
K1	6,5	7,8	6,9	6,4
K2	7,2	8,5	7,3	7,0
K3	7,2	7,5	7,1	7,5
K4	7,1	7,0	7,1	7,2
K5	7,2	6,6	7,4	7,5

Czy średnie zawartości popiołu dla ekogroszku produkowanego w pięciu kopalniach można uznać za jednakowe? Wykonaj analizę wariancji na poziomie istotności 0,01.

F = 0,9563; p-val=0,459; brak podstaw do odrzucenia H_0

3. Każdym z trzech mikrometrów zmierzono kilkukrotnie ten sam przedmiot i uzyskano wyniki:

Mikrometr I:	4,5; 4,7; 4,8; 4,7
Mikrometr II:	4,7; 4,8; 4,5; 4,7; 4,4; 4,8
Mikrometr III:	4,8; 4,9; 4,8; 4,9; 4,8

Zakładając, że błędy pomiarów mają rozkłady normalne o takiej samej wariancji, na poziomie istotności 0,05 zweryfikuj hipotezę, że wybór mikrometru ma wpływ na uzyskane wyniki.

F = 3,3779; p-val=0,069; brak podstaw do odrzucenia H_0

4. Populacja sportowców została ostrzeżona, że palenie papierosów opóźnia rozwój. Jedną z miar wpływu palenia na rozwój jest badanie rytmu zatokowego serca. Przeprowadzono eksperyment, w którym zbadano rytm zatokowy serca u 24 sportowców po zadanym wysiłku fizycznym i 5-min. Otrzymano wyniki:

Niepalący	Lekko-palący	Średnio-palący	Dużo-palący
69	91	55	66
52	72	60	81
71	81	78	70
58	67	58	77
59	95	62	57
65	84	66	79

(a) Zakładając, że rytm serca u każdego rodzaju palaczy ma rozkład normalny, na poziomie istotności 0,01 zweryfikuj hipotezę, że palenie ma wpływ na rozwój. *$F = 6{,}12$; p-val=0,004; odrzucamy H_0*

(b) Zastosuj test uczciwych istotnych różnic (honest significant differences) Tukey'a do wyznaczenia grup jednorodnych porównywanych średnich obiektowych.

Grupy jednorodne: L-D, D-S, D-N, S-N => N-S-D, L-D

5. Badano masę tarczycy chomików w zależności od ich poziomu homozygotyczności (inbredu), wyróżniając cztery grupy (I – osobniki niezinbredowane, II – osobniki o poziomie inbredu z przedziału <0,01 – 0,10>, III – osobniki o poziomie inbredu z przedziału <0,11 – 0,20>, IV – osobniki o poziomie inbredu od 0,21). Uzyskano następujące wyniki:

I:	78	65	71	79	64	53	62
II:	67	88	76	86	76	79	
III:	86	90	78	76	89		
IV:	57	98	89	95	93		

(a) Sprawdź (przyjmując poziom istotności 0,05) czy słuszne jest przypuszczenie, że masa gruczołu tarczycowego zależy od poziomu inbredu. *$F = 3{,}9515$; p-val=0,024; odrzucamy H_0*

(b) Zastosuj test HSD Tukey'a do wyznaczenia grup jednorodnych porównywanych średnich obiektowych. *Grupy jednorodne: I-II, I-III, II-III => I-II-III*

II-III, II-IV, III-IV => II-III-IV

INFORMATYKA

Lab7. *Korelacja i analiza regresji*

UWAGA! Przedstawione dane dostępne są (zgodnie z kolejnością) w plikach csv:

reg_chemik, reg_urzadzenie, reg_arszenik

1. Poniższa tabela przedstawia wyniki eksperymentu, w którym inżynier chce wyznaczyć relację miedzy końcową wielkością produkcji środków chemicznych Y (w kg) w zależności od ilości zużytego surowca X (w litrach):

X	14	23	9	17	10	22	5	12	6	16
Y	68	105	40	79	81	95	31	72	45	93

(a) Narysuj wykres punktowy przedstawiający zależność wielkości produkcji od ilości zużytego surowca (scatter plot).

(b) Wyznacz i zinterpretuj kowariancję próbkową między ilością zużytego surowca a wielkością produkcji.
138,49

(c) Wyznacz i zinterpretuj współczynnik korelacji. 0,895

(d) Wyznacz ocenę prostej regresji między wielkością produkcji a ilością zużytego surowca. $y = 22{,}4 + 3{,}62x$

(e) Dodaj do wykresu punktowego prostą regresji.

(f) W jaki sposób zmieni się wielkość produkcji, jeśli ilość surowca wzrośnie o 1 litr? wzrośnie o 3,62 kg

(g) Jaka będzie wielkość produkcji, jeśli zużyjemy do produkcji 20 litrów surowca? 95

(h) Jaka będzie wielkość produkcji, jeśli zużyjemy do produkcji 15 litrów surowca? 77

(i) Oceń dopasowanie prostej regresji do danych. 80%

(j) Zweryfikuj test o istotności regresji. Przyjmij poziom istotności 5%. Zinterpretuj otrzymany wynik.
$F = 32{,}332$; $p\text{-}val=0{,}0005$; odrzucamy H_0

2. Żywotność pewnego urządzenia (w miesiącach) zależy od liczby wyprodukowanych przez to urządzenie elementów (efektywność urządzenia). Dla próby 9 urządzeń tego samego typu otrzymano następujące wyniki:

Efektywność (*X*)	18	20	18	17	15	15	14	12	10
Żywotność (*Y*)	2	3	3	4	5	6	7	11	9

(a) Narysuj wykres punktowy przedstawiający zależność żywotności od efektywności (scatter plot).

(b) Oblicz i zinterpretuj kowariancję między żywotnością i efektywnością. -8,65

(c) Oblicz i zinterpretuj współczynnik korelacji. -0,91

(d) Wyznacz ocenę prostej regresji żywotności urządzenia od jego efektywności. $y = 18{,}88 - 0{,}86x$

(e) Jak zmieni się żywotność urządzenia jeśli efektywność wzrośnie o 1 element? Zmaleje o 0,86 miesiąca

(f) Oszacuj żywotność urządzenia przy efektywności 11 elementów. 9,4

(g) Oszacuj żywotność urządzenia przy efektywności 19 elementów. 2,5

(h) Oceń dopasowanie prostej regresji. 83%

(i) Przeprowadź test istotności regresji. Przyjmij poziom istotności 1%. Zinterpretuj otrzymany wynik.
$F = 33{,}47$; $p\text{-}val=0{,}0007$; odrzucamy H_0

3. Przeprowadzono proces usuwania arszeniku z wód gruntowych. Poniższa tabela przedstawia procentowe ilości usuniętego przez proces arszeniku w zależności od zakwaszenia (pH) gleby:

pH	7,01	7,11	7,12	7,24	7,94	7,94	8,04	8,05	8,07
% ilość arszeniku	60	67	66	52	50	45	52	48	40
pH	8,90	8,94	8,95	8,97	8,98	9,85	9,86	9,86	9,87
% ilość arszeniku	23	20	40	31	26	9	22	13	7

(a) Narysuj diagram punktowy ilości usuniętego arszeniku w zależności od zakwaszenia gleby.

(b) Oblicz i zinterpretuj kowariancję i współczynnik korelacji między zakwaszeniem gleby a ilością usuniętego arszeniku. -18,32; -0,95

(c) Wyznacz prostą regresji zależności ilości usuniętego arszeniku od zakwaszenia gleby. $y = 190{,}27 - 18{,}03x$

(d) W jaki sposób zmieni się ilość usuniętego przez proces arszeniku jeśli pH gleby wzrośnie o 1? maleje o 18,03 %

(e) Ile arszeniku zostanie usunięte, jeśli pH gleby wyniesie 7,5? 55

(f) Ile arszeniku zostanie usunięte, jeśli pH gleby wyniesie 9? 28

(g) Jak dobra jest ocena liniowa regresji? 90%

(h) Przeprowadź test istotności regresji. Przyjmij poziom istotności 1%. Zinterpretuj otrzymany wynik $F = 149{,}7$; $p\text{-}val=0$; odrzucamy H_0