Historia i plan ćwiczeń

Uaktualnienie: 2 lutego 2020

Stanisław Jaworski

KEiS
SGGW
Nowoursynowska 159, PL-02-787 Warszawa
e-mail: stanislaw_jaworski@sggw.pl

Spis treści

1 Statystyka matematyczna - stacjonarne **4**
1.1 Ćwiczenia - wstęp 4
1.2 Ćwiczenia – rozkład warunkowy (statystyka dostateczna), model statystyczny 5
1.3 Ćwiczenia – rozkład warunkowy (statystyka dostateczna), model statystyczny 6
1.4 Ćwiczenia – $ENMW$ 6
1.5 Ćwiczenia – ENW 7
1.6 Ćwiczenia – ENW 7
1.7 Ćwiczenia – Przedziały ufności 8
1.8 Ćwiczenia – Przedziały ufności 9
1.9 Ćwiczenia – Hipotezy proste 10
1.10 Ćwiczenia – Hipotezy złożone 11
1.11 Ćwiczenia – Powtórzenie 12
1.12 Ćwiczenia – Omówienie zadań ze sprawdzianu 13
1.13 Uwagi 16

2 Rachunek prawdopodobieństwa i statystyka matematyczna - niestacjonarne, informatyka **16**
2.1 Ćwiczenia - temat 1 oraz temat 2 16
2.2 Ćwiczenia - temat 3 19
2.3 Ćwiczenia - temat 4 22
2.4 Ćwiczenia - temat 5 oraz temat 6 23
2.5 Ćwiczenia - temat 7 oraz temat 8 25
2.6 Ćwiczenia - temat 9 oraz temat 10 25
2.7 Ćwiczenia - temat 11 oraz temat 12 25
2.8 Ćwiczenia - Zaliczenie 25

3 Statystyka matematyczna - niestacjonarne, eis **25**
3.1 Ćwiczenia - Tematy: 1 25
3.2 Ćwiczenia - Tematy: 2 27
3.3 Ćwiczenia - Tematy: 3 27
3.4 Ćwiczenia - Tematy: 4 29
3.5 Ćwiczenia - Tematy: 5 30
3.6 Ćwiczenia - Tematy: 6 32

3.7 Ćwiczenia - Tematy: Zaliczenie 32

1. Statystyka matematyczna - stacjonarne

1.1 Ćwiczenia - wstęp

Przykłady na ćwiczeniach.

Przykład 1.
Niech $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} D(\theta)$.

- Wyznaczyć przestrzeń próby $\mathcal{X}$.
- Wyznaczyć rozkład prawdopodobieństwa $p_\theta(x_1, x_2, \ldots, x_n)$ wektora losowego $(X_1, X_2, \ldots, X_n)$.
- Rozkład prawdopodobieństwa $p_\theta(\cdot)$ zapisać w postaci $g_\theta(T(\cdot))h(\cdot)$.
- Rozkład prawdopodobieństwa $p_\theta(\cdot)$ zapisać w postaci wykładniczej

$$\exp\{c(\theta)T(\cdot) - b(\theta)\}h(\cdot)$$

  i sprawdzić, czy zbiór $c(\theta) : \theta \in \Theta = (0, 1)$ jest zbiorem jednowymiarowym.
- Zapisać model statystyczny.

Zadania domowe.

Zadanie 1.
Niech $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} Bin(\theta)$. Wyznaczyć rozkład prawdopodobieństwa zmiennej losowej X_1 pod warunkiem $\sum_{i=1}^n X_i$.

Zadanie 2.
Niech $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} E(\theta)$. Wyznaczyć rozkład prawdopodobieństwa zmiennej losowej X_1 pod warunkiem $\sum_{i=1}^n X_i$.

Zadanie 3.
Niech $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} Po(\theta)$. Wyznaczyć rozkład prawdopodobieństwa zmiennej losowej X_1 pod warunkiem $\sum_{i=1}^n X_i$.

Zadanie 4.
Niech $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} N(\mu, 1)$. Wyznaczyć rozkład prawdopodobieństwa zmiennej losowej X_1 pod warunkiem $\sum_{i=1}^n X_i$.

Zadanie 5.
Punkt a) w zadaniach 17.1-17.13 (zbiór zadań). Sprawdzić, czy rozkład prawdopodobieństwa w tych zadaniach da się zapisać w postaci $g_\theta(T(\cdot))h(\cdot)$ lub w postaci wykładniczej.

1.2 Ćwiczenia – rozkład warunkowy (statystyka dostateczna), model statystyczny

Przykłady

Przykład 2.
Niech $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} Bin(n\theta)$.

- Wyznaczyć przestrzeń próby $\mathcal{X}$.
- Wyznaczyć rozkład prawdopodobieństwa $p_\theta(x_1, x_2, \ldots, x_n)$ wektora losowego $(X_1, X_2, \ldots, X_n)$.
- Rozkład prawdopodobieństwa $p_\theta(\cdot)$ zapisać w postaci $g_\theta(T(\cdot))h(\cdot)$.
- Rozkład prawdopodobieństwa $p_\theta(\cdot)$ zapisać w postaci wykładniczej
$$\exp\{c(\theta)T(\cdot) - b(\theta)\}h(\cdot)$$
i sprawdzić, czy zbiór $c(\theta) : \theta \in \Theta = (0, 1)$ jest zbiorem jednowymiarowym.
- Zapisać model statystyczny.

Przykład 3. (Wskazówki)
Niech $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} E(\theta)$. Wyznaczyć rozkład prawdopodobieństwa zmiennej losowej X_1 pod warunkiem $\sum_{i=1}^n X_i$.

Wskazówki:

- $f(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ – funkcja gęstości rozkładu warunkowego zmiennej losowej X pod warunkiem $Y = y$.
- $f_{X,Y}(x, y) = f_{X,Y-X}(x, y - x)$
- Ogólnie: Niech $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ oraz $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.
$$f_{\mathbf{X}}(\mathbf{x}) = f_{\phi(\mathbf{X})}(\phi(\mathbf{x}))|J_\phi(\mathbf{x})|$$
Dla $\phi(\mathbf{x}) = A\mathbf{x}$ zachodzi $J_\phi(\mathbf{x}) = det(A)$
- $\sum_{i=1}^n X_i \sim G(n, \theta)$

1.3 Ćwiczenia – rozkład warunkowy (statystyka dostateczna), model statystyczny

Przykłady

Przykład 4. (Zrobione przez studena)
Niech $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} Po(\theta)$.

- Wyznaczyć przestrzeń próby $\mathcal{X}$.
- Wyznaczyć rozkład prawdopodobieństwa $p_\theta(x_1, x_2, \ldots, x_n)$ wektora losowego $(X_1, X_2, \ldots, X_n)$.
- Rozkład prawdopodobieństwa $p_\theta(\cdot)$ zapisać w postaci $g_\theta(T(\cdot))h(\cdot)$.
- Zapisać model statystyczny.
- Wyznaczyć model statystyczny dla statystyki dostatecznej.

Przykład 5. (Zrobione przeze mnie)
Niech $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} U(0, \theta)$.

- Wyznaczyć przestrzeń próby $\mathcal{X}$.
- Wyznaczyć gęstość rozkładu prawdopodobieństwa $p_\theta(x_1, x_2, \ldots, x_n)$ wektora losowego $(X_1, X_2, \ldots, X_n)$.
- Zapisać model statystyczny dla wektora $(X_1, X_2, \ldots, X_n)$.
- Wyznaczyć model statystyczny dla statystyki dostatecznej $T = X_{n:n}$.

Zadania domowe.

Zadanie 6.
Przerobić zadania 17.1–17.13.

Uwaga 1.
Studenci nadal nie robią zadań domowych.

1.4 Ćwiczenia – $ENMW$

Przykład 6.
Wyznaczyć $ENMW(\theta(1-\theta))$ z próby zmiennych losowych o rozkładzie $D(\theta)$.

1.5 Ćwiczenia – ENW

Przykład 7.
Przyjmujemy, że liczby wypadków $N_1, \ldots, N_k$ zgłoszonych w kolejnych k latach są niezależnymi zmiennymi losowymi. Zakładamy, że zmienna losowa N_i ma rozkład $Po(\lambda m_i)$, gdzie m_i jest znaną liczbą samochodów ubezpieczonych w $i-$tym roku, zaś λ nieznanym parametrem. Wyznaczyć $ENW(\lambda.)$

Przykład 8.
Zmienna losowa N ma rozkład $Po(\lambda)$ z parametrem λ, który chcemy oszacować. Niestety możemy obserwować jedynie zmienną losową M, która przyjmuje wartość zero, jeśli N równa się zero, a wartość jeden, jeśli N jest większa od zera. Średnią arytmetyczną z próbki niezależnych obserwacji zmiennej M oznaczmy przez $\bar{m}$. Wyznaczyć $ENW(\lambda)$.

Przykład 9.
Pobieramy próbę o rozkładzie $Po(\lambda)$. Niestety nasz sposób obserwacji uniemożliwia odnotowanie realizacji o wartości zero. Pobieranie próby kończymy w momencie, gdy liczebność odnotowanych realizacji wynosi T. Tak więc każda z naszych kolejnych odnotowanych realizacji $k_1, \ldots, k_T$ wynosi co najmniej 1, i nic nie wiemy o tym, ile w międzyczasie pojawiło się (i umknęło z naszego pola widzenia) obserwacji zerowych. Wyznaczyć $ENW(\lambda)$.

Zadanie 7. (do domu)
Zanalizować przykład ze zbioru zadań: wyznaczenie $ENW(\mu, \sigma^2)$ i zastosowanie do $ENW(\sigma/\mu)$

Uwaga 2.

- Ze względu na nadchodzący sprawdzian dokładnie przerobiłem wszystkie zamieszczone przykłady.
- Biorąc pod uwagę wyniki sprawdzianu, 70% grupy nie przerobiła przed sprawdzianem zadań z tych ćwiczeń i nie zrobiła zadania domowego.

1.6 Ćwiczenia – ENW

Przykład 10.
W pewnej populacji prawdopodobieństwo tego, że osobnik przeżyje rok jest równe $1-\theta$. Jeżeli osobnik przeżył rok, to (warunkowe) prawdopodobieństwo tego,że przeżyje następny rok też jest równe $1-\theta$. W próbce liczącej n osobników z tej populacji zanotowano n_0 przypadków, kiedy osobnik nie przeżył roku, n_1 przypadków, kiedy osobnik przeżył rok, ale nie przeżył drugiego oraz n_2 przypadków, kiedy osobnik przeżył dwa lata. Wyznaczyć $ENW(\theta)$.

1. Model statystyczny: generowany przez wektor (N_0, N_1, N_2) o rozkładzie trójmianowym o prawdopodobieństwach p_0, p_1, p_2.
2. Wyznaczanie $p_0 = 1-\theta, p_1 = (1-\theta)^2,\ p_2 = 1 - p_0 - p_1$.
3. Modyfikacja przestrzeni parametrów Θ zgodnie z warunkami: $\theta \in (0,1) \ \wedge \ p_2 \in (0,1)$.
4. Warunek na ekstremum.

Zadanie 8. (do domu)
Dla przykładu nr 10 podałem konkretne wartości n_0, n_1, n_2. Dla podanych wartości poprosiłem o numeryczne oraz graficzne wyznaczenie wartości $ENW(\theta)$.

1.7 Ćwiczenia – Przedziały ufności

Przykład 11.
Niech x będzie realizacją zmiennej losowej X z rozkładu $E(\lambda)$. Na zadanym poziomie ufności $(1-\alpha)$ skonstruować przedział ufności dla parametru λ.

Plan działania

1. Wyznaczenie rozkładu zmiennej X/λ.
2. Wykazanie, że $t(x, \lambda)$ jest funkcją centralną
3. Rozpoczęcie konstrukcji przedziału w oparciu o funkcję centralną.
4. Pokazanie, że przedziałów może być nieskończenie wiele.
5. Wyznaczenie liczbowe przedziału dla $x = 10$ oraz $1-\alpha = 0.95$.
6. Wyznaczenie przedziału w oparciu o kryterium równego prawdopodobieństwa przeszacowania i niedoszacowania parametru.
7. Wyznaczenie przedziału w oparciu o postulat, że przedział ufności ma być najkrótszy.

Zadanie 9. (do domu)
Dla przykładu 11 rozwiązać numerycznie równanie dla zagadnienia najkrótszego przedziału ufności (przyjmując $1 - \alpha = 0.95$).

Uwaga 3.
Studenci nie zrobili zadania domowego. Zapowiedziałem, że na następne zajęcia mają przynieść oba zadania, w przeciwnym przypadku wyciągnę wiadome dla nich konsekwencje.

1.8 Ćwiczenia – Przedziały ufności

Przykład 12.
Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu $E(\lambda)$. Na zadanym poziomie ufności $(1 - \alpha)$ skonstruować przedział ufności dla parametru λ zgodnie z postulatem

- prawdopodobieństwa przeszacowania i niedoszacowania mają być równe
- przedział ufności ma być najkrótszy.

Zadanie 10. (do domu)
Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu $N(0, \sigma^2)$. Na zadanym poziomie ufności $(1-\alpha)$ skonstruować przedział ufności dla parametru σ^2 zgodnie z postulatem, że prawdopodobieństwa przeszacowania i niedoszacowania tego parametru mają być równe.

Zadanie 11. (do domu)
Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu $N(0, \sigma^2)$. Na zadanym poziomie ufności $(1-\alpha)$ skonstruować przedział ufności dla parametru σ^2 zgodnie z postulatem, że prawdopodobieństwa przeszacowania i niedoszacowania tego parametru mają być równe.

Zadanie 12. (do domu)
Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu $N(\mu, \sigma^2)$. Na zadanym poziomie ufności $(1-\alpha)$ skonstruować przedział ufności dla parametru σ^2 zgodnie z postulatem, że prawdopodobieństwa przeszacowania i niedoszacowania tego parametru mają być równe.

Zadanie 13. (do domu)

Niech

$$X_1, X_2, \ldots, X_n$$

będzie próbą z rozkładu $N(\mu, 1)$ oraz

$$X_{n+1}, X_{n+2}, \ldots, X_{n+m}$$

próbą z rozkładu $N(\mu, 4)$. Przy czym zmienne losowe $X_1, X_2, \ldots, X_{n+m}$ są niezależne. Na zadanym poziomie ufności $(1-\alpha)$ skonstruować przedział ufności dla parametru μ zgodnie z postulatem, że prawdopodobieństwa przeszacowania i niedoszacowania tego parametru mają być równe.

Wskazówka. Przydatne fakty:

1. $X \sim N(\mu, \sigma^2) \Longrightarrow \dfrac{X-\mu}{\sigma} \sim N(0,1)$
2. $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} N(0,1) \Longrightarrow \sum_{i=1}^n X_i^2 \sim \chi(n)$
3. $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} N(\mu, \sigma^2) \Longrightarrow \dfrac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\sigma^2} \sim \chi(n-1)$
4. $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} N(\mu, \sigma^2) \Longrightarrow \dfrac{\bar{X}_n - \mu}{S}\sqrt{n} \sim t(n-1)$, gdzie $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$

1.9 Ćwiczenia – Hipotezy proste

Przykład 13. (zrobiłem bardzo dokładnie)
Niech $X \sim E(\lambda)$. Skonstruować test najmocniejszy do weryfikacji hipotezy $H0 : \lambda = 5$ przeciwko hipotezie $H_1 : \lambda = 10$ na poziomie istotności $\alpha = 0.05$. Wyznaczyć moc tego testu.

Przykład 14. (do domu; podałem wskazówki dotyczące postaci testu oraz wartości krytycznej)
Niech $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} E(\lambda)$. Skonstruować test najmocniejszy do weryfikacji hipotezy $H0 : \lambda = 5$ przeciwko hipotezie $H_1 : \lambda = 10$ na poziomie istotności $\alpha = 0.05$. Wyznaczyć moc tego testu przy $n = 10$.

Uwaga 4.
Tylko dwójka studentów z mojej grupy była na wykładzie z weryfikacji hipotez statystycznych. Zrobiłem kartkówkę. Wynik zdecydowanie niesatysfakcjonujący. Pytania:

- Co to jest hipoteza statystyczna?
- Na czym polega weryfikacja hipotezy statystycznej?
- Co to jest poziom istotności testu statystycznego?
- Do czego służy test statystyczny?
- Wymień rodzaje błędów, które można popełnić przy weryfikacji hipotez statystycznych.
- Czy byłaś/byłeś na dzisiejszym wykładzie ze statystyki matematycznej?

1.10 Ćwiczenia – Hipotezy złożone

Przykład 15. (zrobiłem bardzo dokładnie)
Niech $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} N(\mu, 1)$. Skonstruować test do weryfikacji hipotezy $H_0 : \mu \leq \mu_0$ przeciwko hipotezie $H_1 : \mu > \mu_0$, (przy znanej wartości μ_0) na poziomie istotności $\alpha = 0.05$.

Przykład 16. (do domu)
Niech $X_1, X_2, \ldots, X_n \overset{\texttt{iid}}{\sim} E(\lambda)$. Skonstruować test do weryfikacji hipotezy $H0 : \lambda \leq \lambda_0$ przeciwko hipotezie $H_1 : \lambda > \lambda_0$ na poziomie istotności $\alpha = 0.05$.

Uwaga 5.

- Na początku zajęć zrobiłem kartkówkę o treści: *Sformułować lemat Neymana-Pearsona oraz wyjaśnić pojęcie testu najmocniejszego.* Jeden student otrzymał 0.5 punktu. Pozostali otrzymali 0 punktów. Wniosek: studenci nie zanalizowali treści lematu pomimo tego, że był przedmiotem poprzednich zajęć.
- Wyjaśniłem, co będzie robione na ćwiczeniach: konstrukcja testu opartego na ilorazie wiarogodności. Zaznaczyłem, że test ten nie musi być jednostajnie najmocniejszy. Wynik kartkówki pokazał, że moje wyjaśnienia nie mogły być należycie zrozumiane, ponieważ studenci nie dysponowali odpowiednią wiedzą.
- Zauważyłem, że studenci mają problemy z robieniem notatek. Studenci nie notują ważnych uwag. Przypuszczam, że nakładają się na to dwie rzeczy: studenci nie przemyśleli poprzednich tematów oraz mają braki z analizy matematycznej (przebieg zmienności funkcji).

– Część studentów (ok. 4 osób) zaczęła robić zadania domowe.

1.11 Ćwiczenia – Powtórzenie

Cel zajęć: Ponieważ na ostatnich dwóch zajęciach omawiałem temat weryfikacji hipotez, na tych ćwiczeniach postanowiłem położyć nacisk na starszy temat dotyczący konstrukcji przedziałów ufności.

Przykład 17.
Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu o gęstości

$$f_a(x) = (\gamma - 1)a^{\gamma-1}x^{-\gamma}\mathbb{1}_{(a,\infty)}(x),$$

gdzie $a > 0$ jest nieznane, zaś $\gamma > 1$ jest znaną wielkością. Wykazać, że X_{min}/a jest funkcją centralną dla parametru a i na tej podstawie, na poziomie ufności $1 - \alpha$, skonstruować dla tego parametru przedział ufności.

Plan działania

1. Uwagi na temat statystyki X_{min}: a) X_{min} jest statystyką dostateczną (nawiązałem do twierdzenia o faktoryzacji oraz przykładu 5, w którym technika pokazania tego faktu jest taka sama), b) X_{min} jest estymatorem obciążonym dla parametru a (na podstawie analizy postaci funkcji gęstości, z której wynika, że $P_a(X_{min} > a) = 1$), c) X_{min} jest $ENW(a)$.
2. Przypomnienie definicji funkcji centralnej.
3. Wyznaczenie rozkładu prawdopodobieństwa zmiennej losowej X_{min}/a.
4. Wyznaczenie przedziału ufności dla a zgodnie z kryterium, że przeszacowanie i niedoszacowanie parametru ma być jednakowo-prawdopodobne (dokładne omówienie tego, co oznacza przyjęte kryterium, a co za tym idzie, dlaczego musimy wyznaczyć kwantyl rzędu $\alpha/2$ oraz $1-\alpha/2$).
5. Uwagi końcowe na temat ewentualnej konstrukcji testu dla $H_0 : a \leq a_0$ przeciwko $H_1 : a > a_0$ opartego na statystyce X_{min}. Objaśnienie (intuicyjne, w oparciu o uwagi z punktu 1.), dlaczego obszar krytyczny testu powinien mieć postać $\{(x_1, x_2, \ldots, x_n) : x_{min} > k\}$.

Uwaga 6.
Wybrany przykład ma tę zaletę, że kwantyle dają się wyznaczyć analitycznie. Sposób wyznaczenia rozkładu minimum został przypomniany (było na rachunku prawdopodobieństwa) i może przydać się na sprawdzianie.

W ramach przygotowania do sprawdzianu zasugerowałem studentom przerobienie wszystkich przykładów omówionych na ćwiczeniach (niezbędne minimum).
Dwie panie oddały pracę domową. Przy konstrukcji testu poprawnie zastosowały technikę opartą na monotonicznym ilorazie wiarogodności. Technika ta nie jest omawiana na wykładzie.

1.12 Ćwiczenia – Omówienie zadań ze sprawdzianu

Cel zajęć: Omówienie zadań z jednego z dwóch zestawów. Podanie szkiców rozwiązań. Zwrócenie uwagi na elementy, które sprawiły największe trudności.
Realizacja: Omówiono 3 zadania. Najbardziej złożone zadanie ze sprawdzianu znajduje się w przykładzie 18.

Przykład 18.
Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu o gęstości

$$f_\theta(x) = \begin{cases} \frac{\gamma}{\theta} x^{\gamma-1} \exp\{-\frac{x^\gamma}{\theta}\}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

gdzie $\theta > 0$ jest nieznane, zaś $\gamma > 0$ jest znaną wielkością. Niech $\theta_0 > 0$ będzie daną liczbą. Na poziomie istotności α skonstruować test hipotezy $H_0: \ \theta \geq \theta_0$ przeciw alternatywie $H_1: \ \theta < \theta_0$.

Szkic rozwiązania. Ze względu na proponowaną postać testu ilorazu wiarogodności, przy ustalonych $x_1, \ldots, x_n \in R_+$, należy wyznaczyć

$$\sup_{\theta>0} \prod_{i=1}^n f_\theta(x_i) \text{ oraz } \sup_{\theta\geq\theta_0} \prod_{i=1}^n f_\theta(x_i)$$

Zauważmy, że

$$\begin{aligned} \ln\left(\prod_{i=1}^n f_\theta(x_i)\right) &= \sum_{i=1}^n \ln\left(\frac{\gamma}{\theta} x_i^{\gamma-1} \exp\{-\frac{x_i^\gamma}{\theta}\}\right) \\ &= \sum_{i=1}^n \left(\ln\gamma - \ln\theta + (\gamma-1)\ln(x_i) - \frac{x_i^\gamma}{\theta}\right) \end{aligned}$$

Zatem

$$\frac{\partial}{\partial\theta}\ln\left(\prod_{i=1}^{n} f_\theta(x_i)\right) = \sum_{i=1}^{n}\left(-\frac{1}{\theta}+\frac{x_i^\gamma}{\theta^2}\right) > 0 \Leftrightarrow \theta < \frac{1}{n}\sum_{i=1}^{n} x_i^\gamma$$

Oznaczmy $\frac{1}{n}\sum_{i=1}^{n} x_i^\gamma$ przez $\hat{\theta}$. Pokazano zatem, że funkcja

$$h(\theta; x_1, \ldots, x_n) \stackrel{\texttt{def}}{=} \prod_{i=1}^{n} f_\theta(x_i)$$

jest rosnąca na przedziale $(0, \hat{\theta})$ i malejąca na $(\hat{\theta}, \infty)$. Stąd

$$\sup_{\theta>0} h(\theta; x_1, \ldots, x_n) = h(\hat{\theta}; x_1, \ldots, x_n)$$

oraz

$$\sup_{\theta\geq\theta_0} h(\theta; x_1, \ldots, x_n) = \begin{cases} h(\hat{\theta}; x_1, \ldots, x_n), & \text{dla } \theta_0 \leq \hat{\theta} \\ h(\theta_0; x_1, \ldots, x_n), & \text{dla } \theta_0 > \hat{\theta} \end{cases}$$

Stąd wynika, że

$$\frac{\sup_{\theta>0} h(\theta; x_1, \ldots, x_n)}{\sup_{\theta\geq\theta_0} h(\theta; x_1, \ldots, x_n)} = \begin{cases} 1, & \text{dla } \theta_0 \leq \hat{\theta} \\ \frac{h(\hat{\theta}; x_1, \ldots, x_n)}{h(\theta_0; x_1, \ldots, x_n)}, & \text{dla } \theta_0 > \hat{\theta} \end{cases}$$

Funkcja $h(\theta; x_1, \ldots, x_n)$ jest postaci

$$\begin{aligned} \prod_{i=1}^{n} f_\theta(x_i) &= \frac{\gamma^n}{\theta^n}\left(\prod_{i=1}^{n} x_i\right)^{1-\gamma} \exp\left\{-\frac{\sum_{i=1}^{n} x_i^\gamma}{\theta}\right\} \\ &= \frac{\gamma^n}{\theta^n}\left(\prod_{i=1}^{n} x_i\right)^{1-\gamma} \exp\left\{-\frac{n\hat{\theta}}{\theta}\right\} \end{aligned}$$

Zatem przy danym θ_0 funkcja

$$\begin{aligned}
\lambda(x_1,\ldots,x_n) &\stackrel{\texttt{def}}{=} \\
&= \frac{\sup_{\theta>0} h(\theta; x_1,\ldots,x_n)}{\sup_{\theta\geq\theta_0} h(\theta; x_1,\ldots,x_n)} \\
&= \begin{cases} 1, & \text{dla } \theta_0 \leq \hat{\theta} \\ \left(\frac{\theta_0}{\hat{\theta}}\right)^n \exp\left\{-n + \frac{n\hat{\theta}}{\theta_0}\right\}, & \text{dla } \theta_0 > \hat{\theta} \end{cases} \\
&= \begin{cases} 1, & \text{dla } \theta_0 \leq \hat{\theta} \\ \left(\frac{\theta_0}{\hat{\theta}} \exp\left\{-1 + \frac{\hat{\theta}}{\theta_0}\right\}\right)^n, & \text{dla } \theta_0 > \hat{\theta} \end{cases}
\end{aligned}$$

zależy od $(x_1,\ldots,x_n)$ poprzez funkcję $\hat{\theta} = \hat{\theta}(x_1,\ldots,x_n) = \frac{1}{n}\sum_{i=1}^n x_i^\gamma$. Dla $\hat{\theta} \in \langle\theta_0,\infty)$ jest stała i równa jeden, a dla $\hat{\theta} \in (0,\theta_0)$ jest, ze względu na $\hat{\theta}$, funkcją malejącą, bo jak łatwo sprawdzić

$$\left(\frac{1}{x}\exp\left(-1+x\right)\right)' = \frac{e^{-1+x}}{x}\left(1-\frac{1}{x}\right) < 0 \text{ dla } x \in (0,1)$$

Stąd dla danego $k >$ mamy takie t, że

$$\lambda(x_1,\ldots,x_n) > k \Leftrightarrow \sum_{i=1}^n x_i^\gamma < t.$$

Z powyższych rozważań wynika, że postać testu opartego na ilorazie wiarogodności jest następująca:

$$\varphi(x) = \begin{cases} 1, & \text{dla } \sum_{i=1}^n x_i^\gamma < t \\ 0, & \text{dla } \sum_{i=1}^n x_i^\gamma \geq t \end{cases}$$

Łatwo sprawdzić, że $X_i^\gamma \sim E(\theta)$. Zatem $\sum_{i=1}^n X_i^\gamma \sim G(m,\theta)$. Stąd t wyznaczamy z równania

$$\sup_{\theta\geq\theta_0} F_\theta(t) = \alpha,$$

gdzie F_θ jest dystrybuantą rozkładu $G(n,\theta)$. Intuicyjnie im mniejsze θ, tym chętniej zmienna losowa $\sum_{i=1}^n X_i^\gamma$ przyjmuje małe wartości (wartości mniejsze lub równe t). Zatem $\sup_{\theta\geq\theta_0} F_\theta(t) = F_{\theta_0}(t)$, co daje $t = F_{\theta_0}^{-1}(\alpha)$.

1.13 Uwagi

Patrz uwagi: 1, 2, 3, 4, 5, 6.

2. Rachunek prawdopodobieństwa i statystyka matematyczna - niestacjonarne, informatyka

Tematy

1. Prawdopodobieństwo klasyczne i geometryczne.
2. Własności prawdopodobieństwa: wzór na prawdopodobieństwo całkowite, wzór Bayesa, wzór łańcuchowy.
3. Interpretacja rozkładów dyskretnych. Związki między rozkładami dyskretnymi. Wyznaczanie parametrów tych rozkładów
4. Rozkłady ciągłe: przykłady, charakterystyka, wyznaczanie parametrów
5. Wyznaczanie rozkładów funkcji od zmiennych losowych
6. Wektory losowe: wyznaczanie rozkładów brzegowych i ich parametrów
7. Wektory losowe: sprawdzanie niezależności, wyznaczanie rozkładów warunkowych i ich parametrów
8. Twierdzenia graniczne. Przybliżanie prawdopodobieństw za pomocą rozkładu normalnego oraz rozkładu Poissona.
9. Estymacja przedziałowa: analiza jednej populacji
10. Weryfikacja hipotez statystycznych: analiza jednej populacji
11. Estymacja przedziałowa: analiza dwóch populacji
12. Weryfikacja hipotez statystycznych: analiza dwóch populacji
13. Analiza korelacji i regresji
14. Test zgodności i niezależności

2.1 Ćwiczenia - temat 1 oraz temat 2

Przykłady na ćwiczeniach.

Przykład 19.
Tarcza strzelecka składa się z trzech koncentrycznych kół o promieniach odpowiednio 1, 2 i 3. Za trafienie w środkowe koło zdobywa się trzy punkty, za trafienie w kolejne pierścienie (licząc od środka koła) odpowiednio dwa i

jeden punkt. Jakie jest prawdopodobieństwo uzyskania co najmniej trzech punktów w dwóch strzałach? (Zakładamy, że każdy strzał trafia w tarczę.)

Przykład 20.
W grupie studenckiej jest 20 osób. Na ćwiczeniach Student do odpowiedzi losowany jest na podstawie wyniku rzutu kostką dwudziestościenną. Jakie jest prawdopodobieństwo, że ten sam Student zostanie wyrwany do odpowiedzi trzykrotnie z rzędu?

Przykład 21.
Troje dzieci: Ania, Basia i Czesio zmywają szklanki. Najstarsza Ania zmywa dwa razy częściej niż młodsza Basia, zaś Basia trzy razy częściej niż najmłodszy Czesio. Wiadomo, że prawdopodobieństwo zbicia szklanki w czasie zmywania wynosi dla Ani 0.01, dla Basi wynosi 0.04 natomiast dla Czesia 0.5. Jakie jest prawdopodobieństwo, że w czasie zmywania zostanie zbita jedna szklanka? Pewnego dnia po powrocie z pracy mama zauważyła, że jedna ze szklanek jest zbita, a żadne z dzieci nie chce się przyznać do zniszczenia szklanki. Które z dzieci najprawdopodobniej zmywało tego dnia?

Przykład 22.
Przedsiębiorstwo zawarło umowy z zakładami Z_1, Z_2 oraz Z_3 na dostawę podzespołów. Zakład Z_1 dostarcza 50%, zakład Z_2 dostarcza 35% natomiast zakład Z_3 dostarcza 15% potrzebnych podzespołów. Wiadomo, że 95% dostaw zakładu Z_1, 80% dostaw zakładu Z_2 oraz 85% dostaw zakładu Z_3 odpowiada wymaganiom technicznym. Jakie jest prawdopodobieństwo, że jeden wylosowany podzespół odpowiada wymaganiom technicznym? Do punktu serwisowego zgłasza się klient z urządzeniem, w którym uszkodzony jest podzespół. Jakie jest prawdopodobieństwo, że producentem zepsutego podzespołu był zakład Z_1?

Zadania domowe.

Zadanie 1.
Autobus przyjeżdża na przystanek co piętnaście minut. Jakie jest prawdopodobieństwo tego, że przychodząc na przystanek w losowym momencie będziemy czekać na autobus nie dłużej niż pięć minut?

Zadanie 2.
Pan Roztargniony zapomniał ostatniej cyfry telefonu do znajomego. W związku z tym wykręcając numer telefonu ostatnią cyfrę wybiera losowo. Jakie jest prawdopodobieństwo tego, że dodzwoni się, jeżeli ma do dyspozycji cztery żetony telefoniczne?

Zadanie 3.
Na egzamin przygotowanych jest 100 pytań. Student zna odpowiedź na 80 z nich. Egzaminator przerywa egzamin w chwili, gdy Student nie umie odpowiedzieć na pytanie, lecz nie później niż po piątym pytaniu. Ocena końcowa jest równa liczbie pytań, na które odpowiedział Student. Jakie jest prawdopodobieństwo tego, że Student otrzyma ocenę co najmniej dobrą?

Zadanie 4.
Na wspólnej klasówce z matematyki spotkali się Studenci I roku z dwóch grup. W pierwszej grupie jest 15 pań oraz 10 panów, zaś w drugiej jest 12 panów i 13 pań. Prawdopodobieństwo, że pani z grupy pierwszej rozwiąże zadanie na klasówce wynosi 0.8, natomiast prawdopodobieństwo to dla pana wynosi 0.7. W drugiej grupie prawdopodobieństwa te kształtują się odpowiednio 0.9 oraz 0.85. Jak duży odsetek wszystkich Studentów rozwiąże zadanie na klasówce? Przy sprawdzaniu prac okazało się, że ktoś przygotował ściągawkę. Określić, kim najprawdopodobniej był autor ściągawki (tzn. określić płeć i grupę autora).

Zadanie 5.
Wśród 300 zdających egzamin wstępny z matematyki jest 200 absolwentów klas matematyczno–fizycznych, 75 absolwentów klas ogólnokształcących oraz 25 absolwentów klas humanistycznych. Prawdopodobieństwo zdania egzaminu przez absolwenta klasy matematyczno–fizycznej wynosi 0.9, klasy ogólnokształcącej wynosi 0.25, zaś klasy humanistycznej 0.1. Jakie jest prawdopodobieństwo, że losowo wybrany przystępujący do egzaminu zda go pomyślnie? Jaki jest odsetek absolwentów klas matematyczno–fizycznych, klas humanistycznych oraz klas ogólnokształcących wśród osób, które zdały egzamin?

Zadanie 6.
Sklep jest zaopatrywany w żarówki pochodzące z trzech fabryk, przy czym 20% żarówek pochodzi z pierwszej fabryki, 30% z drugiej, 50% z trzeciej. Pro-

dukcja pierwszej fabryki zawiera 1% żarówek wadliwych, produkcja drugiej — 5% żarówek wadliwych, a produkcja trzeciej fabryki 10% żarówek wadliwych. Obliczyć prawdopodobieństwo, że losowo wybrana żarówka będzie wadliwa.

Zadanie 7.
W szpitalu na oddziale wewnętrznym przebywa rocznie średnio 2000 chorych. Wśród leczonych było 800 cierpiących na chorobę K_1, 600 cierpiących na chorobę K_2, 400 cierpiących na chorobę K_3 oraz 200 cierpiących na chorobę K_4. Prawdopodobieństwo pełnego wyleczenia z chorób wynosiło odpowiednio 0.9, 0.8, 0.7 oraz 0.5. Obliczyć prawdopodobieństwo, że losowo wypisany pacjent jest całkowicie wyleczony.

Zadanie 8.
W magazynie znajdują się opony do samochodów osobowych. Pochodzą one w 20% z fabryki F_1, w 30% z fabryki F_2 i w 50% z fabryki F_3. Wiadomo, że produkcja wadliwych opon w poszczególnych fabrykach wynosi odpowiednio 5%, 4%, 3%. Obliczyć prawdopodobieństwo, że losowo wybrana opona będzie wadliwa.

2.2 Ćwiczenia - temat 3

Przykłady na ćwiczeniach.

Przykład 23.
Partia dostarczanych detali ma wadliwość 5%. Niech zmienną losową będzie liczba dobrych detali spośród czterech dostarczonych. Obliczyć prawdopodobieństwo, że dostarczono co najmniej dwa detale dobre.

Przykład 24.
Z partii nasion o sile kiełkowania 75% losujemy osiem nasion. Obliczyć prawdopodobieństwo, że nie wykiełkuje co najmniej pięć nasion.

Przykład 25.
Obliczyć prawdopodobieństwo, że na siedem rzutów kostką co najwyżej trzy razy wypadnie liczba oczek nie mniejsza niż 4.

Przykład 26.

Z partii dostarczanych detali o wadliwości 5% losujemy detale do momentu uzyskania trzech detali wadliwych. Wyznaczyć prawdopodobieństwo, że wśród wylosowanych detali całkowita liczba dobrych detali nie przekroczyła dziesięciu.

Przykład 27.
Popularność pewnej partii wynosi 45%. Z populacji wyborców losujemy kolejne osoby do momentu, aż wylosujemy czwartego wyborcę, który popiera tę partię. Wyznaczyć prawdopodobieństwo, że nie wylosujemy więcej niż 20 osób.

Przykład 28.
Jasiek rzuca piłką do kosza do momentu, aż trafi do niego po raz trzeci. Szansa trafienia przez Jaśka do kosza wynosi 0.8 przy pojedynczym rzucie. Ile wynosi prawdopodobieństwo, że Jasiek będzie musiał rzucać przynajmniej 5 razy?

Przykład 29.
Jedna na dziesięć kobiet ma naturalnie rude włosy. Losowo spotykamy kobiety. Wyznaczyć prawdopodobieństwo, że liczba spotkanych kobiet do momentu trafienia na rudą po raz trzeci nie przekroczy sześciu.

Zadania domowe.

Zadanie 9.
Załóżmy, że prawdziwa jest hipoteza Mendla, iż dla krzyżówki grochu w drugim pokoleniu stosunek nasion żółtych do zielonych jest jak $3:1$. Wylosowano dziesięć nasion. Obliczyć prawdopodobieństwo, że będą co najwyżej cztery nasiona żółte.

Zadanie 10.
Załóżmy, że prawdziwa jest hipoteza Mendla, iż dla krzyżówki grochu w drugim pokoleniu stosunek nasion żółtych do zielonych jest jak $3:1$. Losowano kolejno nasiona do momentu uzyskania dwóch zielonych nasion. Wyznaczyć prawdopodobieństwo, że w sumie wylosowano nie więcej niż 10 nasion.

Zadanie 11.

Siła kiełkowania nasion wynosi 0.75. Obliczyć prawdopodobieństwo, że na dziesięć wysianych nasion wzejdzie co najmniej osiem nasion.

Zadanie 12.
Środek owadobójczy zabija przeciętnie 90% owadów. Środek ten zastosowano na dziesięciu owadach. Obliczyć prawdopodobieństwo, że co najwyżej dwa osobniki przeżyją.

Zadanie 13.
Wadliwość procesu produkcyjnego wynosi 10%. Obliczyć prawdopodobieństwo, że na osiem wylosowanych produktów będą co najwyżej dwa złe.

Zadanie 14.
Widomo, że 30% pewnej populacji kobiet ma nadciśnienie. Z populacji tej kolejno losowano kobiety do momentu natrafienia na drugą kobietę z nadciśnieniem. Wyznaczyć prawdopodobieństwo, że wylosowano nie więcej niż dziesięć kobiet bez nadciśnienia.

Zadanie 15.
W pewnym gatunku zwierząt prawdopodobieństwo urodzenia osobnika płci męskiej wynosi 0.6. Obliczyć prawdopodobieństwo, że w miocie, w którym urodziło się pięcioro młodych będą co najmniej cztery osobniki męskie.

Zadanie 16.
W stawie hodowlanym są dwa gatunki ryb w proporcji 8 : 2. Obliczyć prawdopodobieństwo, że wśród dziesięciu złowionych ryb będzie co najmniej siedem ryb liczniejszego gatunku.

Zadanie 17.
W jeziorze jest 1000 ryb, w tym 100 ryb zaobrączkowanych. Obliczyć prawdopodobieństwo, że wśród 10 złowionych ryb będzie co najmniej siedem ryb zaobrączkowanych.

Zadanie 18.
Przyjmując, że co czwarte wezwanie pogotowia jest nieuzasadnione określić prawdopodobieństwo, że na osiem wyjazdów co najmniej połowa z nich będzie uzasadniona.

Zadanie 19.
Załóżmy, że co czwarty zawał kończy się zejściem oraz, że prawdopodobieństwo przeżycia zawału nie zależy od tego, który to zawał z kolei. Jakie jest prawdopodobieństwo przeżycia trzech zawałów?

Zadanie 20.
Właściciel kurzej fermy stwierdził, że kogutków wykluwa się trzy razy więcej niż kurek. Obliczyć prawdopodobieństwo, że z pięciu losowo wybranych jajek wykluje się co najmniej jeden kogutek, ale nie mniej niż dwie kurki.

Zadanie 21.
Producent podaje, że w co czwartym jajku niespodziance znajduje się zajączek Ribbon. Jakie jest prawdopodobieństwo, że wśród dwudziestu kupionych jajek jest a) przynajmniej pięć jajek z zajączkiem Ribbon; b) nie więcej niż piętnaście jajek bez zajączka. Jaka jest najbardziej prawdopodobna ilość jajek z zajączkami?

Zadanie 22.
Poparcie dla polityka J wynosi 34%. Wylosowano 12 wyborców. Wyznaczyć prawdopodobieństwo, że wśród wylosowanych będzie co najmniej trzech popierających polityka J.

Zadanie 23.
Widomo, że 20% pewnej populacji mężczyzn ma nadciśnienie. Z populacji tej kolejno losowano mężczyzn do momentu natrafienia na trzeciego mężczyznę z nadciśnieniem. Wyznaczyć prawdopodobieństwo, że w sumie wylosowano nie więcej niż dziesięciu mężczyzn.

2.3 Ćwiczenia - temat 4

Przykłady na ćwiczeniach.

Przykład 30.
Niech X będzie zmienną losową o rozkładzie $N(-5, 100)$. Obliczyć prawdopodobieństwa

$$P\{X \in (-1, 5)\},\ P\{X \leq -9\},\ P\{X \in (-7, 1)\},\ P\{X \geq -7\},\ P\{|X+5| \leq 10\}.$$

Przykład 31.
Wzrost kobiety jest zmienną losową o rozkładzie $N(160, 100)$. Obliczyć jaki jest procent kobiet wyższych niż 170 lub niższych niż 150.

Przykład 32.
Automat tokarski produkuje nity, których średnica ma rozkład normalny z odchyleniem standardowym 0.04 mm. Wartość średnia tej zmiennej losowej może być dowolnie regulowana przez odpowiednie ustawienie automatu. Nit uważa się za dobry, jeżeli jego średnica mieści się w przedziale $(2.9, 3.1)$. Jakie jest prawdopodobieństwo wyprodukowania braku, gdy automat tokarski ustawiony jest tak, że średnia średnica jest równa 3.05 mm? Jak powinien być ustawiony automat, by wadliwość procesu produkcyjnego była najmniejsza?

Przykład 33.
Na pewnym terenie mazurki łączą się w stada z wróblami w proporcji 2:3. Rozkład wagi mazurka jest normalny o średniej 23 g i wariancji 9 g^2. Rozkład wagi wróbla też jest normalny. Chociaż oba gatunki są do siebie niezwykle podobne, wróble są cięższe od mazurków średnio o 7 g, a wariancja ich wagi wynosi $25g^2$.
Każdego dnia ze stada wróbli i mazurków łapiemy jednego ptaka do momentu, aż waga złapanego będzie większa od 25g. Wyznaczyć wartość oczekiwaną liczby złapanych ptaków.

Zadania domowe.

Zadanie 24.
W populacji królików rozkład wagi samiczek jest normalny o średniej 3kg i wariancji $0.25kg^2$, a samczyków wykładniczy o średniej 4kg. Odsetek samiczek w populacji wynosi 60%. Losujemy króliki do momentu, aż waga wylosowanego królika będzie większa od 5kg lub mniejsza od 2kg. Wyznaczyć wartość oczekiwaną liczby wylosowanych królików.

2.4 Ćwiczenia - temat 5 oraz temat 6

Przykłady na ćwiczeniach.

Zadanie 25.

Dwuwymiarowy rozkład pary zmiennych losowych ξ oraz η dany jest za pomocą tablicy

	η		
ξ	-1	0	1
0	0.10	0.09	0.08
1	0.27	0.05	0.20
2	0.05	0.03	0.13

Wyznaczyć

- rozkład prawdopodobieństwa zmiennej losowej η
- rozkład prawdopodobieństwa zmiennej losowej ξ
- rozkład prawdopodobieństwa zmiennej losowej $\xi + \eta$
- $E(\eta)$ oraz $E\xi)$
- $D^2(\eta)$ oraz $D^2(\xi)$
- $D^2(\xi + \eta)$
- $\varrho(\xi, \eta)$
- $E(\xi + \eta)$
- rozkład prawdopodobieństwa zmiennej losowej η^2 pod warunkiem $\xi > 0$
- $E(\eta^2 | \xi > 0)$
- $(a_o, b_o) = \underset{a,b \in \mathbb{R}}{\operatorname{argmin}} E(\eta - (a\xi + b))^2$

Zadania domowe.

Zadanie 26.
Wygenerować 1000 realizacji wektora (ξ, η): $\{(\xi(\omega_i), \eta(\omega_i)) : i = 1, \ldots 1000\}$ oraz narysować dwa wykresy

1. $\{(\xi(\omega_i), \eta(\omega_i)) : i = 1, \ldots 1000\}$
2. $\{(a_o\xi(\omega_i) + b_o, \eta(\omega_i)) : i = 1, \ldots 1000\}$

i porównać.

Zadanie 27.
Zadania 3.34 – 3.37 ze zbioru
`http://wojtek.zielinski.statystyka.info/Innosci/SW_zbior.pdf`

2.5 Ćwiczenia - temat 7 oraz temat 8

Przykłady na ćwiczeniach.

Zadania domowe.

2.6 Ćwiczenia - temat 9 oraz temat 10

Przykłady na ćwiczeniach.

Zadania domowe.

2.7 Ćwiczenia - temat 11 oraz temat 12

Przykłady na ćwiczeniach.

Zadania domowe.

2.8 Ćwiczenia - Zaliczenie

3. Statystyka matematyczna - niestacjonarne, eis

Tematyka

1. Estymacja przedziałowa: analiza jednej populacji
2. Weryfikacja hipotez statystycznych: analiza jednej populacji
3. Estymacja przedziałowa: analiza dwóch populacji
4. Weryfikacja hipotez statystycznych: analiza dwóch populacji
5. Analiza korelacji i regresji
6. Analiza wariancji
7. Test zgodności i niezależności
8. Testy oparte na rangach

3.1 Ćwiczenia - Tematy: 1

Przykłady na ćwiczeniach.

Przykład 1.
Przeprowadzono anonimową ankietę wśród 26 uczniów najstarszych klas szkoły podstawowej. Celem ankiety było określenie ile uczniowie wydają tygodniowo na różnego rodzaju używki. Wyniki przedstawiają się następująco: $\sum x_i = 650$, $\sum x_i^2 = 16650$. Na podstawie wyników ankiety oszacować przeciętne tygodniowe wydatki uczniów na używki.

Przykład 2.
Wśród 17 losowo wybranych osób przeprowadzono ankietę na temat czasu dojazdu do pracy. Wyniki ankiety przedstawiały się następująco: $\sum x_i = 425$, $\sum x_i^2 = 10881$. Na podstawie wyników ankiety oszacować średni czas dojazdu do pracy.

Przykład 3.
Wyprodukowano pewien nowy środek owadobójczy. Środek ten zastosowano na tysiącu owadach, z których 852 padły. Oszacować skuteczność tego środka owadobójczego.

Przykład 4.
Oszacować wadliwość procesu produkcyjnego wiedząc, że na 50 przebadanych wyrobów stwierdzono dwa braki.

Zadania domowe.

Zadanie 1.
Badano dzienny czas poświęcany przez dzieci w wieku przedszkolnym na oglądaniu telewizji. Uzyskano następujące wyniki na próbie przedszkolaków (w minutach): 132, 114, 51, 97, 117, 119, 122, 65, 109, 84, 85, 134, 133, 107, 149. Czy zebrane wyniki potwierdzają przypuszczenia, że przedszkolaki spędzają na oglądaniu telewizji przeciętnie dwie godziny dziennie?

Zadanie 2.
Postanowiono oszacować parametr CCS (Czytelnictwo Cyklu Sezonowego) dla pewnego kwartalnika. Ponieważ jest to kwartalnik, za sezon przyjmuje się jeden rok. Zatem interesuje nas odsetek osób, które miało kontakt z danym kwartalnikiem w ciągu roku. Okazało się, że na tysiąc wylosowanych osób z grupy docelowej, 350 miało kontakt z tym kwartalnikiem. Oszacować

interesujący nas parametr CCS na poziomie ufności 95%. Zinterpretować wynik.

Zadanie 3.
Zadania 51 – 75 ze zbioru:
`http://sjaworski.wzim.sggw.pl/zadania/pdf/z_stat.pdf`

3.2 Ćwiczenia - Tematy: 2

Przykłady na ćwiczeniach.

Przykład 5.
Badano czy na zmianę wydajności pracowników wpływa przebyte przeszkolenie. Eksperyment przeprowadzono na wybranych losowo pracownikach. Zmiany ich wydajności po przeszkoleniu były następujące: -12, 13, -11, 14, 15, 9, -5, 5, 4, 8, 14, 6. Czy uzyskane wyniki potwierdzają przypuszczenie, że badany rodzaj szkolenia ma istotny wpływ na zmianę wydajności?

Przykład 6.
W biochemicznym doświadczeniu badano czas życia komórek w pewnym środowisku. Dokonano ośmiu pomiarów uzyskując wyniki (w godzinach): 4.7, 5.3, 4.0, 3.8, 6.2, 5.5, 4.5, 6.0 ($\bar{x} = 5,\ s = 0,891227083$). Czy można uznać, że średni czas życia komórek w badanym środowisku wynosi 4 godziny?

Zadania domowe.

Zadanie 4.
Zadania 76 – 92 ze zbioru:
`http://sjaworski.wzim.sggw.pl/zadania/pdf/z_stat.pdf`

3.3 Ćwiczenia - Tematy: 3

Przykłady na ćwiczeniach.

Przykład 7.
Z pewnego regionu Polski wylosowano do badań tysiąc osób, w tym dwieście osób z wyższym wykształceniem. Wśród osób z wyższym wykształceniem

zliczono dwanaście osób bezrobotnych, a wśród osób bez wyższego wykształcenia sześćdziesiąt osiem. Czy na podstawie uzyskanych rezultatów można uznać, że odsetek bezrobotnych zależy od poziomu wykształcenia?

Przykład 8.
W pewnym doświadczeniu farmakologicznym z podawaniem dwu preparatów badano potęgowanie narkozy. Dla preparatu A otrzymano następujące przedłużenia narkozy: 4, 3, 5, 2, 4, 6, 4, 5 ($\bar{x} = 4.125$, $s = 1.246423455$), a dla preparatu B: 6, 10, 8, 9, 9, 10, 8, 7 ($\bar{x} = 8.375$, $s = 1.407885953$). Czy można uznać, że preparaty w różnym stopniu przedłużają czas narkozy, jeżeli doświadczenie przeprowadzono na dobranych losowo osobach?

Przykład 9.
Dwóm grupom robotników zlecono wykonanie tej samej pracy z tym jednak, że robotnicy grupy pierwszej przeszli wcześniej przeszkolenie. Zaobserwowana wydajność pracy w pierwszej grupie kształtowała się następująco (w szt/h): 18.6, 17.9, 18.1, 17.0, 18.7, 18.3, podczas gdy w grupie drugiej zaobserwowano następujące wydajności: 17.3, 17.6, 17.1, 16.0, 17.8. Czy przeszkolenie zmieniło wydajność pracy robotnika?

Przykład 10.
Na sto losowo wybranych kobiet głosowało czterdzieści sześć, a na dwustu losowo wybranych mężczyzn głosowało stu sześćdziesięciu. Oszacować różnicę między odsetkiem głosujących kobiet a odsetkiem głosujących mężczyzn.

Zadania domowe.

Zadanie 5.
W doświadczeniu Milgrama z 1961 roku jeden z uczestników eksperymentu („nauczyciel") aplikuje elektrowstrząsy drugiemu uczestnikowi („uczniowi") za każdą złą odpowiedź. Przy czym „nauczyciel" zwiększa napięcie po każdej złej odpowiedzi. „Nauczyciel" nie wie, że „uczeń"jest podstawiony i że faktycznie nie otrzymuje żadnych elektrowstrząsów, a jedynie udaje, że go to strasznie boli. Kiedy „nauczyciela" ogarniają wątpliwości, prowadzący eksperyment naciska go, aby kontynuował eksperyment. Okazuje się, że około 67% „nauczycieli" aplikuje dawki śmiertelne. Czy odsetek osób aplikujących dawki śmiertelne zależy od płci „nauczyciela", jeżeli na 1000 losowych ko-

biet dawkę śmiertelną zaaplikowało 680 kobiet, a na 1200 losowych mężczyzn dawkę śmiertelną zaaplikowało 780 mężczyzn?

Zadanie 6.
Pracownicy sklepów pewnej sieci handlowej charakteryzują się różną absencją. Wysunięto przypuszczenie, że struktura absencji zależy do płci. Zweryfikować to przypuszczenie na podstawie poniższych danych dotyczących czterystu wylosowanych kobiet oraz sześciuset wylosowanych mężczyzn.

Liczba dni nieobecności	Płeć	
	Kobiety	Mężczyźni
0–5	300	500
5 i więcej	100	100

Zadanie 7.
Ocenić różnicę między odsetkiem studentów, którzy zdali egzamin ze statystyki w pierwszym terminie jeżeli dla dwustu losowo wybranych studentów z wydziału A zdało 100, a dla 250 studentów wydziału B zdało taka sama liczba studentów jak na wydziale A. Oszacuj wielkość tej różnicy i odpowiedz na pytanie, który z wydziałów osiąga lepsze wyniki.

Zadanie 8.
Przypuszcza się, że popularność kawy rozpuszczalnej w SGGW zależy od bufetu. Zbadano, że na 1000 kaw wypitych w bufecie w budynku nr 37, 850 było rozpuszczalnych, a na 1200 wypitych w budynku nr 34, rozpuszczalnych było 875. Na podstawie oceny różnicy między odsetkami wypitych kaw rozpuszczalnych zweryfikować, czy przypuszczenia to można uznać za uzasadnione?

3.4 Ćwiczenia - Tematy: 4

Przykłady na ćwiczeniach.

Przykład 11.
W pewnym sklepie zważono jaja dostarczane przez dwóch różnych dostawców. Pobrano po dziesięć jaj od każdego dostawcy i otrzymano wyniki:
dostawca I: $\sum x_{1i} = 645$, $\sum x_{1i}^2 = 41615$; dostawca II: $\sum x_{2i} = 680$, $\sum(x_{2i} - \bar{x}_2)^2 = 10$.

Na podstawie uzyskanych wyników stwierdzić, czy średnie ciężary jaj dostarczane przez obu dostawców są takie same.

Przykład 12.
Badano zawartość tłuszczu w serach żółtych produkowanych zimą i latem. W każdym z dwóch okresów zbadano zawartość tłuszczu w dziesięciu serach i otrzymano wyniki:
zima: $\sum x_{1i} = 265$, $\sum x_{1i}^2 = 7034$; lato: $\sum x_{2i} = 276$, $\sum (x_{2i} - \bar{x}_2)^2 = 11$.
Na podstawie uzyskanych wyników stwierdzić, czy zawartość tłuszczu w serze żółtym zależy od pory roku.

Przykład 13.
W ramach badań dotyczących chorób alergiczych w Polsce stwierdzono, że w losowo wybranej próbie 4142osób z nieżytem nosa 1193 miało pleśń w domu, a w próbie 14031 osób, które nie miały nieżytu, pleśń w domu miało 3648 osób. Czy odsetek osób mających pleśń w domu jest inny w populacji osób z nieżytem nosa niż w populacji bez nieżytu? Jaka jest różnica między tymi odsetkami? Wyznaczyć przedział ufności dla tej różnicy.

Przykład 14.
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?

Zadania domowe.

Zadanie 9.
Zadania 93 – 108 ze zbioru:
`http://sjaworski.wzim.sggw.pl/zadania/pdf/z_stat.pdf`

3.5 Ćwiczenia - Tematy: 5

Przykłady na ćwiczeniach.

Przykład 15.
Istnieje podejrzenie że ludzie dzielą się na humanistów i matematyków tzn. jeśli ktoś jest dobry z przedmiotów humanistycznych to z matematyką może już mieć problemy. Wylosowano ośmiu uczniów z czwartej klasy liceum i obliczono dla nich średnie z ocen semestralnych z języka polskiego i z matematyki :

J. polski	3.4	2.9	4.3	3.8	3.3	4.7	3.6	3.9
Matematyka	4.6	4.4	3.6	3.2	3.8	3.5	4.6	3.1

$$\mathrm{cov}(x,y) = -1.435; \quad \mathrm{var}x = 2.29875; \quad \mathrm{var}y = 2.6.$$

Interesuje nas, czy oceny uczniów potwierdzają wspomniane twierdzenie. Jeśli tak, proszę wyznaczyć równanie regresji. Jaki jest przewidywana średnia ocena z matematyki dla ucznia którego średnia z języka polskiego wynosi 5.0?

Przykład 16.
W wylosowanej grupie studentów przeprowadzono dwa testy, matematyczny SATM oraz werbalny SATV. Na podstawie poniszych danych oceń, czy zdolności matematyczne i werbalne są względem siebie zależne? Jeżeli są, proszę opisać tę zależność przy pomocy liniowej funkcji regresji.

$SATM$	110	118	129	146	151	159	173	181	189	195
$SATV$	90	100	116	150	149	153	169	190	197	192

Przykład 17.
Zbadano w dziesięciu wylosowanych zakładach przemysłowych wielkość zaplanowanego i wykonanego funduszu na akcję socjalną. Zabadać, czy istnieje zależność między badanymi cechami. Jeżeli taka zależność istnieje, to opisać ją za pomocą liniowej funkcji regresji.

planowany	3.60	4.65	5.20	1.86	3.06	1.36	2.46	3.93	5.80	6.35
wykonany	3.56	4.59	5.13	1.84	3.02	1.35	2.43	3.89	5.72	6.26

Zadania domowe.

Zadanie 10.
Zadania 128 – 145 ze zbioru:
`http://sjaworski.wzim.sggw.pl/zadania/pdf/z_stat.pdf`

3.6 Ćwiczenia - Tematy: 6

Przykłady na ćwiczeniach.

Przykład 18.
W metodzie Kjeldahla oznaczania procentowej zawartości białka stosowany jest katalizator miedziowy. Podejrzewa się, że wynik oznaczenia zależy od ilości dodanego katalizatora. W tym celu przeprowadzono doświadczenie, w którym badano trzy różne ilości katalizatora. Czy poniższe dane udowadaniają zależność wyniku oznaczenia od ilości katalizatora miedziowego?

ilość katalizatora	średni wynik (z pięciu pomiarów)
1.0	18.48
2.0	18.44
5.0	18.02

$$s_e^2 = 6.408$$

Przykład 19.
Badano średnią zawartość (w ppm na kg s.m.) żelaza w roślinach łąkowych. Uzyskano następujące wyniki

dla traw	174.0	172.0	175.0	173.0	177.0	175.5
dla turzycowatych	134.0	135.0	137.0	138.0	135.0	135.2
dla motylkowych	117.0	118.0	116.0	119.0	116.3	116.5
dla ziół (chwastów)	111.0	112.0	113.0	111.5	112.7	114.2

Sformułować i zweryfikować odpowiednią hipotezę.

Zadania domowe.

3.7 Ćwiczenia - Tematy: Zaliczenie