

Podstawy statystyki praktycznej

Lista nr 5

1. **Fakt.** Suma n niezależnych zmiennych losowych o rozkładzie wykładniczym z parametrem λ ma rozkład gamma z parametrem kształtu (shape) równym n i parametrem intensywności (rate) równym λ (parametr skali jest równy $1/\lambda$).

Narysuj gęstości rozkładów sumy 30, 50 i 100 niezależnych zmiennych losowych o rozkładzie wykładniczym z parametrem $\lambda = 1$ i porównaj z odpowiednio dopasowanymi (pod względem wartości oczekiwanej i wariancji) gęstościami rozkładów normalnych.

2. Narysuj "histogramy" rozkładów dwumianowych z parametrami

a) $p=0.5$, $n=20, 50, 100$

b) $p=0.1$, $n=20, 50, 100$

i porównaj z odpowiednio dopasowanymi (pod względem wartości oczekiwanej i wariancji) gęstościami rozkładów normalnych.

3. Wygeneruj 100 elementową próbę z rozkładu

a) $N(0,1)$

b) wykładniczego z parametrem $\lambda = 1$

Skonstruuj klasyczny (przy założeniu normalności) przedział ufności dla wartości oczekiwanej na poziomie ufności 95%.

i) Doświadczenie powtórz 1000 razy i oblicz jak często tak skonstruowane przedziały ufności zawierają rzeczywistą wartość oczekiwaną.

ii) Doświadczenie powtórz jeszcze raz zwiększając rozmiar próby do $n = 200$. Wyznacz p-stwa pokrycia rzeczywistej wartości oczekiwanej i porównaj średnią szerokość przedziałów ufności wyznaczonych w oparciu o próby 100 i 200 elementowe.

4. W tym zadaniu ponownie wykorzystamy zbiór danych *individuals.dat*.

a) Skonstruuj nową zmienną U będącą pierwiastkiem z dochodu. Narysuj histogram tej zmiennej i wyznacz średnią arytmetyczną μ_U z tej zmiennej w całym zbiorze danych. Poza tym wyznacz średni dochód μ_D oraz procenty osób z wykształceniem wyższym i średnim, p_W i p_S .

b) Pobierz 200 elementową próbę losową z tego zbioru danych i w oparciu o tą próbę wyznacz estymatory dla p_W , p_S , μ_U i μ_W . Skonstruuj klasyczne 95% przedziały ufności dla tych parametrów i sprawdź czy zawierają one rzeczywiste wartości.

- c) Punkt b) powtórz 200 razy i narysuj histogramy rozkładów powyższych estymatorów oraz wyznacz jak często przedziały ufności zawierały rzeczywistą wartość estymowanego parametru.

5. W tym zadaniu ponownie wykorzystamy zbiór danych *table1_6.txt*.

Zakładając, że zbiór danych jest prostą próbą losową z pewnej populacji skonstruuj przedziały ufności na średni iloraz inteligencji i średni wynik testu psychologicznego w tej populacji.

Zadania do policzenia ręcznego.

6. Wiadomo, że odchylenie standardowe wagi noworodków wynosi 500 g. Jaki powinien być rozmiar próby, żeby standardowe odchylenie średniej wagi noworodków w próbie było mniejsze niż 150 g.
7. Wysokość roślin kukurydzy w pewnej ich populacji ma rozkład normalny ze średnicą 145 cm i odchyleniem standardowym 22 cm.
- a) Jaki procent roślin ma wysokość w przedziale między 135 a 155 cm ?
 - b) Wybieramy losowo z populacji dużą liczbę prób, z których każda składa się z 16 roślin. Oszacuj w jakim procencie tych prób średnia wysokość (dla 16 roślin) będzie w przedziale między 135 a 155 cm.
 - c) Niech Y reprezentuje średnią wysokość w losowej próbie 16 roślin. Oblicz $P(135 < Y < 155)$.
 - d) Niech Y reprezentuje średnią wysokość w losowej próbie 36 roślin. Oblicz $P(135 < Y < 155)$.
8. Tygodniowe wypłaty z pewnego funduszu są niezależnymi zmiennymi losowymi o rozkładzie wykładniczym z tym samym parametrem $\lambda = \frac{1}{1000\text{zł}}$. Obliczyć prawdopodobieństwo, że łączna wypłata z tego funduszu w okresie roku, tzn. 52 tygodni, przekroczy 70 000 zł. Obliczenia wykonać wykorzystując dokładny rozkład gamma oraz stosując przybliżenie rozkładem normalnym. Porównać wyniki.
9. Prawdopodobieństwo uzyskania wygranej w pewnej grze liczbowej wynosi 0.1. Obliczyć prawdopodobieństwo, że spośród 500 grających osób wygra więcej niż 60 osób. Wykorzystać dokładny rozkład dwumianowy oraz przybliżenie rozkładem normalnym. Porównać wyniki.
10. W pewnej populacji mały 80% osobników jest zarażonych pewnym pasożytem. Biolog pobiera losową próbę 50 osobników. Stosując przybliżenie rozkładem normalnym oszacuj p-stwo, że mniej niż 35 spośród wybranych osobników będzie zakażonych. Wykonaj obliczenia dwa razy: bez poprawki na ciągłość i uwzględniając tę poprawkę. Porównaj z dokładną wartością tego p-stwa, wyznaczoną w oparciu o dystrybuantę rozkładu dwumianowego.
11. Dla danych -0.1, 0.15, 0.1, -0.05, oszacować na poziomie ufności $1 - \alpha = 0.9$ wartość oczekiwaną przyjmując, że rozkład jest normalny oraz $\sigma = 0.1$.

12. Oblicz średnią, odchylenie standardowe i błąd standardowy średniej w pięcioelementowej próbie : 10.0, 8.9, 9.1, 11.7 i 7.9. Skonstruuuj 90% przedział ufności dla μ , przy założeniu, że obserwacje pochodzą z rozkładu normalnego.

Małgorzata Bogdan