Statystyka w analizie i planowaniu eksperymentu

Paweł Błażej

6 marca 2013

Własności rozkładu normalnego

Rozkład normalny jest rozkładem najczęściej wykorzystywanym do modelowania zmienności pewnej cechy w zadanej populacji. Parametrami tego rozkładu są średnia μ (parametr położenia) oraz σ (parametr skali).

Własności rozkładu normalnego

Zapis

$$X \sim \mathcal{N}(\mu, \sigma).$$

Standardowy rozkład normalny

$$X \sim \mathcal{N}(0, 1), \; gdzie \; \mu = 0, \; \sigma = 1$$

Przekształcenia zmiennej losowej o rozkładzie normalnym

Niech

$$X \sim \mathcal{N}(\mu, \sigma),$$

wtedy

$$Y \sim \mathcal{N}(0, 1), \text{ gdzie } Y = \frac{(X - \mu)}{\sigma}.$$

Niech

$$X \sim \mathcal{N}(0, 1),$$

wtedy

$$Y \sim \mathcal{N}(\mu, \sigma) \text{ gdzie } Y = \sigma \cdot X + \mu$$

Gęstość rozkładu normalnego

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Podstawowe własności rozkładu normalnego

1. $E[X] = \mu$;

Podstawowe własności rozkładu normalnego

1. $E[X] = \mu$;
2. $Var[X] = \sigma^2$;

Podstawowe własności rozkładu normalnego

1. $E[X] = \mu$;
2. $Var[X] = \sigma^2$;
3. Jeżeli $X_1, X_2, \ldots, X_n$ są to niezależne zmienne losowe pochodzące z rozkładu normalnego $N(\mu, \sigma)$. to $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$;

Podstawowe własności rozkładu normalnego

1. $E[X] = \mu$;
2. $Var[X] = \sigma^2$;
3. Jeżeli $X_1, X_2, \ldots, X_n$ są to niezależne zmienne losowe pochodzące z rozkładu normalnego $N(\mu, \sigma)$. to $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$;
4. $Me = \mu$;

Podstawowe własności rozkładu normalnego

1. $E[X] = \mu$;
2. $Var[X] = \sigma^2$;
3. Jeżeli $X_1, X_2, \ldots, X_n$ są to niezależne zmienne losowe pochodzące z rozkładu normalnego $N(\mu, \sigma)$. to $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$;
4. $Me = \mu$;
5. $Mo = \mu$;

Dalsze własności rozkładu normalnego garść przydatnych definicji

Kurtoza - miara spłaszczenia rozkładu

$$K = \frac{E(X - E[X])^4}{\sigma^4} - 3$$

Uwaga

Dla rozkładu normalnego kurtoza wynosi 0.

Kurtoza - podstawowe własności

Rozkłady prawdopodobieństwa można podzielić ze względu na wartość kurtozy na rozkłady:

1. **mezokurtyczne** - wartość kurtozy wynosi 0, spłaszczenie rozkładu jest podobne do spłaszczenia rozkładu normalnego (dla którego kurtoza wynosi dokładnie 0);

Kurtoza - podstawowe własności

Rozkłady prawdopodobieństwa można podzielić ze względu na wartość kurtozy na rozkłady:

1. **mezokurtyczne** - wartość kurtozy wynosi 0, spłaszczenie rozkładu jest podobne do spłaszczenia rozkładu normalnego (dla którego kurtoza wynosi dokładnie 0);
2. **leptokurtyczne** - kurtoza jest dodatnia, wartości cechy bardziej skoncentrowane niż przy rozkładzie normalnym;

Kurtoza - podstawowe własności

Rozkłady prawdopodobieństwa można podzielić ze względu na wartość kurtozy na rozkłady:

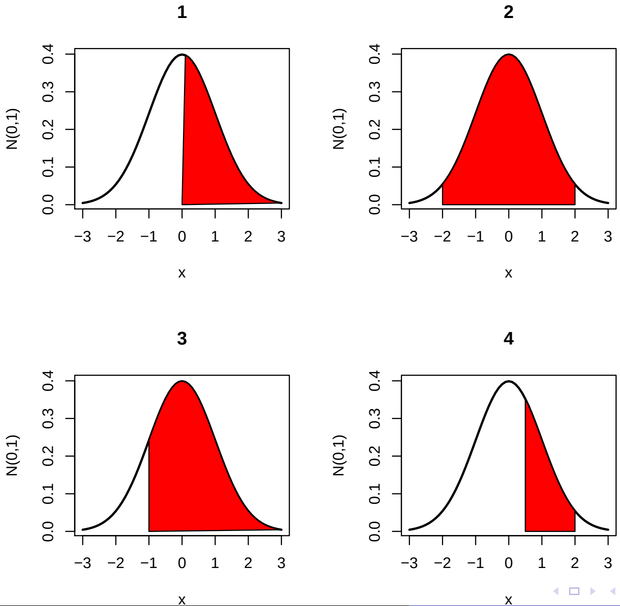
1. **mezokurtyczne** - wartość kurtozy wynosi 0, spłaszczenie rozkładu jest podobne do spłaszczenia rozkładu normalnego (dla którego kurtoza wynosi dokładnie 0);
2. **leptokurtyczne** - kurtoza jest dodatnia, wartości cechy bardziej skoncentrowane niż przy rozkładzie normalnym;
3. **platokurtyczne** - kurtoza jest ujemna, wartości cechy mniej skoncentrowane niż przy rozkładzie normalnym.

Praktyczne posługiwanie się rozkładem normalnym

Niech $X \sim \mathcal{N}(0, 1)$. Oblicz następujące prawdopodobieństwa:

1. $P(X > 0) = 0.5$;
2. $P(|X| < 2) = 0.954499736103642$;
3. $P(X > -1) = 0.841344746068543$;
4. $P(0.5 < X < 2) = 0.285787406777808$.

Praktyczne posługiwanie się rozkładem normalnym

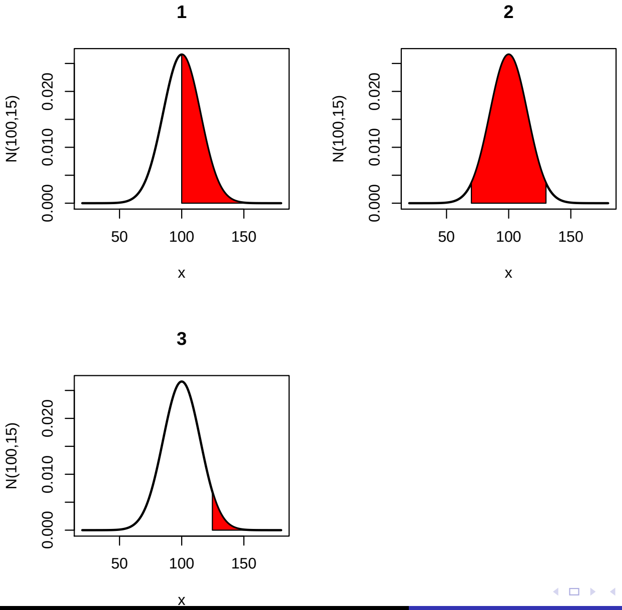


Praktyczne posługiwanie się rozkładem normalnym

Przyjmuje się, że współczynnik IQ ma w populacji rozkład normalny o średniej $\mu = 100$ i odchyleniu standardowym $\sigma = 15$

1. Ile osób ma większe IQ niż 100?;
2. Ile osób ma IQ w przedziale $70 - 130$?;
3. Jaki przedział przyjąć aby określić odsetek 0.05 osób o największym IQ?

Praktyczne posługiwanie się rozkładem normalnym



Praktyczne posługiwanie się rozkładem normalnym - odpowiedzi

1. $P(X > 100)$ =0.5;
2. $P(70 < X < 130)$ =0.954499736103642;
3. $P(X > 124.75) = 0.05$.

Załóżmy, że długość piór ogonowych pawia wynosi średnio 65 cm, z odchyleniem standardowym 5 cm. Oszacuj prawdopodobieństwo, że losowo wyjęte pióro ma długość:

1. mniejszą niż 54 cm;
2. większą niż 64 cm;
3. jeśli mieszkańcy Łobzowa zwykli nosić na czapkach pióra o długości od 70 do 75 cm, to jak często natrafiają na takie pióro?

Reguła 3σ

Dla rozkładu normalnego ma zastosowanie tzw. „ prawo 3σ ” mówi ono, że:

1. ok. 68% wszystkich wartości zmiennej odbiega od średniej oczekiwanej nie bardziej niż o jedno odchylenie standardowe;

Reguła 3σ

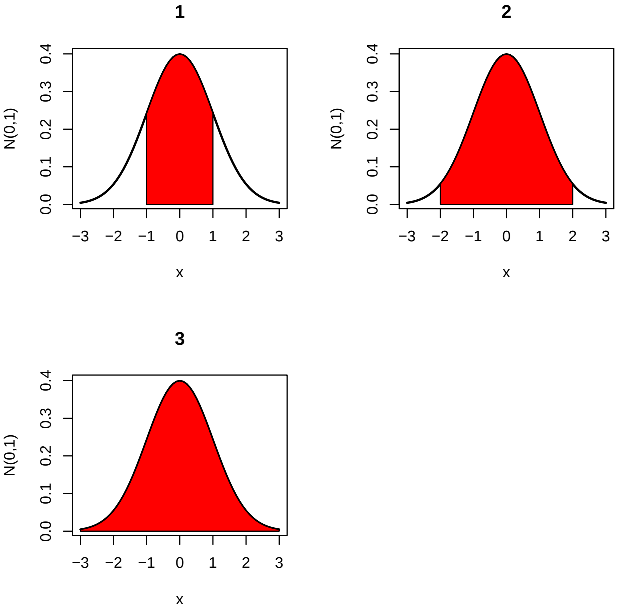
Dla rozkładu normalnego ma zastosowanie tzw. „ prawo 3σ ” mówi ono, że:

1. ok. 68% wszystkich wartości zmiennej odbiega od średniej oczekiwanej nie bardziej niż o jedno odchylenie standardowe;
2. ok. 95% wszystkich wartości nie bardziej niż o dwa odchylenia standardowe;

Reguła 3σ

Dla rozkładu normalnego ma zastosowanie tzw. „ prawo 3σ " mówi ono, że:

1. ok. 68% wszystkich wartości zmiennej odbiega od średniej oczekiwanej nie bardziej niż o jedno odchylenie standardowe;
2. ok. 95% wszystkich wartości nie bardziej niż o dwa odchylenia standardowe;
3. ok. 99.8% odbiega o nie więcej niż 3σ od wartości średniej.

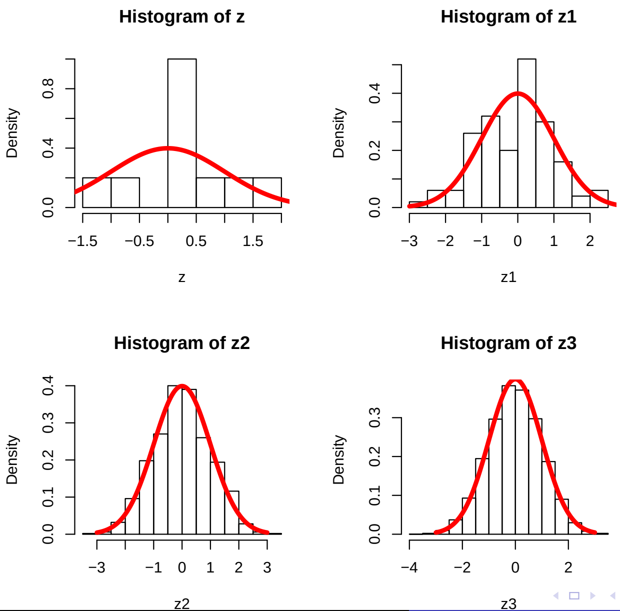

1
2
3
N(0,1)
0.0 0.1 0.2 0.3 0.4
−3 −2 −1 0 1 2 3
x

Rozkład normalny jest rozkładem granicznym dla standaryzowanych zmiennych losowych pochodzących z różnych rozkładów.

Centralne Twierdzenie Graniczne

Średnia n niezależnych ustandaryzowanych $\sqrt{n}\frac{\bar{X}-E[X]}{\sqrt{Var[X]}}$ zmiennych losowych pochodzących z porządnych rozkładów zbiega do rozkładu normalnego $\mathcal{N}(0,1)$.

Co to oznacza w praktyce?



Rzucamy 100 razy monetą jakie jest prawdopodobieństwo, że suma wyrzuconych orłów przekroczy 55?

Rzucamy 20 razy symetryczną kostką oblicz prawdopodobieństwo tego, że suma wyrzuconych oczek jest liczbą pomiędzy $60, 80$?

Przyjmijmy, że czas naszego codziennego dojazdu do szkoły jest zmienną losową o rozkładzie jednostajnym na odcinku $[0,5 godz., 1 godz.]$ i że czasy dojazdu w różne dni są niezależne. Ile w przybliżeniu wynosi prawdopodobieństwo zdarzenia, ze średni dzienny dojazd w ciągu 30 dni przekroczy $0,8$ godziny.