Estymacja parametryczna

Estymacja to dział wnioskowania statystycznego, który zajmuje się szacowaniem wartości parametrów oraz postaci rozkładu w populacji generalnej na podstawie obserwacji uzyskanych w próbie losowej.

Estymacja parametryczna określa metody znajdowania nieznanych wartości parametrów. Z kolei wnioskowaniem o postaci rozkładu w populacji generalnej zajmuje się **estymacja nieparametryczna**.

Proces estymacji

Punktem wyjściowym w estymacji jest wylosowanie z populacji n-elementowej próby i wyznaczenie na jej podstawie wartości estymatora nieznanego parametru.

Estymatorem parametru Θ rozkładu populacji generalnej jest funkcja wyznaczona na podstawie próby losowej, służąca do oceny wartości tego parametru.

Główne własności estymatorów

- **Nieobciążoność**
  Estymator nazywamy **nieobciążonym**, gdy $E(\hat{\Theta}) = \Theta$, jeśli $E(\hat{\Theta}) - \Theta = f(\Theta)$ to estymator nazywamy **obciążonym**, a funkcję $f(\hat{\Theta})$ nazywamy **obciążeniem** estymatora.

- **Zgodność**
  Estymator nazywamy zgodnym, jeśli jest stochastycznie zbieżny do szacowanego parametru

$$\bigwedge_{\varepsilon>0} \lim_{n\to\infty} P\left\{|\hat{\theta} - \theta| < \epsilon\right\} = 1$$

- **Efektywność**
  Spośród zbioru wszystkich nieobciążonych estymatorów $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_r$ najefektywniejszym nazywamy estymator o najmniejszej wariancji.

Podstawowe estymatory

- Estymatorem wartości oczekiwanej populacji jest wartość średnia $\bar{x}$ z próby losowej wyrażona wzorem

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

  Jest to estymator nieobciążony, zgodny i najefektywniejszy.

- Estymatorem wariancji populacji σ^2 jest wariancja z próby prostej wyrażona wzorem

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

  Jest to estymator zgodny i nieobciążony.

Rodzaje estymacji parametrycznej

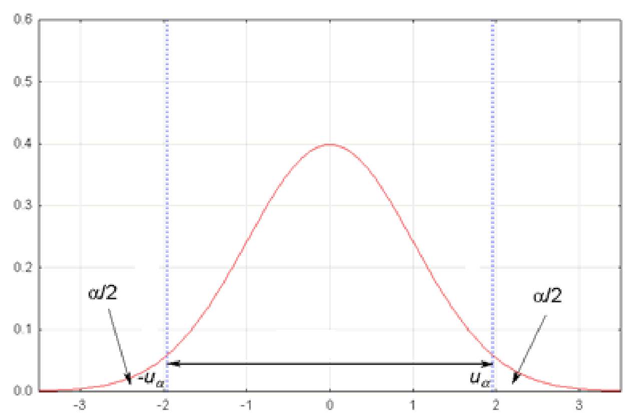
- **Estymacja punktowa**
  Za parametr populacji przyjmuje się wartość estymatora otrzymaną z danej, n-elementowej próby losowej. Estymacja punktowa nie daje oszacowania nieznanego parametru Θ rozkładu populacji. Prawdopodobieństwo, że estymator przyjmie wartość równą wartości szacowanego parametru, jest równe 0. Z tego wynika, że przy stosowaniu estymacji punktowej prawdopodobieństwo popełnienia błędu w ocenie parametru populacji jest równe 1.

- **Estymacja punktowa**
  Błąd oceny parametru populacji Θ za pomocą jego estymatora Q nie powinien przekraczać odpowiednio małej wartości ε z przyjętym dużym prawdopodobieństwem $1-\alpha$, czyli musi być spełnione równanie:
  $$P(|\Theta - Q| < \varepsilon) = 1 - \alpha$$
  Przedział liczbowy $(Q - \varepsilon, Q + \varepsilon)$, który z określonym z góry, dużym (bliskim jedności) prawdopodobieństwem będzie zawierał nieznaną wartość parametru zbiorowości generalnej, jest nazywany przedziałem ufności, a prawdopodobieństwo $1-\alpha$ współczynnikiem ufności. Do wyznaczenia wartości ε potrzebna jest znajomość rozkładu estymatora Q. Procedura wyznaczania przedziału ufności jest nazywana estymacją przedziałową.

Przedział ufności to losowy przedział wyznaczony za pomocą rozkładu estymatora, mający tę własność, że z dużym, z góry zadanym prawdopodobieństwem pokrywa wartość szacowanego parametru. Zapisujemy go zwykle w postaci: $P(a < \Theta < b) = 1 - \alpha$. Liczby a i b są nazywane dolną i górną granicą przedziału ufności. Współczynnik ufności $1 - \alpha$ jest miarą zaufania do prawidłowego szacunku. Najczęściej ma on wartość $0,99, 0,95$ lub $0,90$.

Przedział ufności dla wariancji

Przedział ufności dla wartości oczekiwanej $E(X)$ populacji o rozkładzie normalnym $N(m, \sigma)$ jest wyznaczany według wzoru:
$$P\left(\bar{x} - u_\alpha \frac{\sigma}{\sqrt{n}} < m < \bar{x} + u_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$
gdzie u_α jest taką wartością rozkładu standardowego, że pole pod krzywą gęstości w przedziale $(-u_\alpha, u_\alpha)$ wynosi $1 - \alpha$, a pole na lewo od $-u_\alpha$ i na prawo od u_α wynosi $\frac{\alpha}{2}$.



Z tego wynika zależność
$$\Phi(u_\alpha) = 1 - \frac{\alpha}{2}$$

Z kolei dla małej ilości prób ($n < 30$) przedział ufności dla wariancji wylicza się z rozkładu t-Studenta za pomoca następującego wzoru

$$P\left(\bar{x} - t_\alpha \frac{s}{\sqrt{n-1}} < m < \bar{x} + t_\alpha \frac{s}{\sqrt{n-1}}\right) = 1 - \alpha$$

gdzie t_α jest wartością zmiennej losowej o rozkładzie t-Studenta o $n-1$ stopniach swobody wyznaczana z relacji $P(-t_\alpha < t < t_\alpha) = 1 - \alpha$

Wyliczenie przedziału ufności w programie STATISTICA
Wyliczymy przedział ufności pomiarów ołowiu dla danych zaimportowanych z pliku `Dane_Estymacja.xls` z arkusza `Ołów (Dużo prób)` dla $\alpha = 0,95$.

I) **Ręcznie przy pomocy kalkulatora prawdopodobieństwa**
Liczba pomiarów wynosi 52, zatem użyjemy pierwszego ze wzorów dla przedziałów ufności. Zauważmy, że

$$1 - \alpha = 0,95$$

Zatem

$$\alpha = 0,05$$

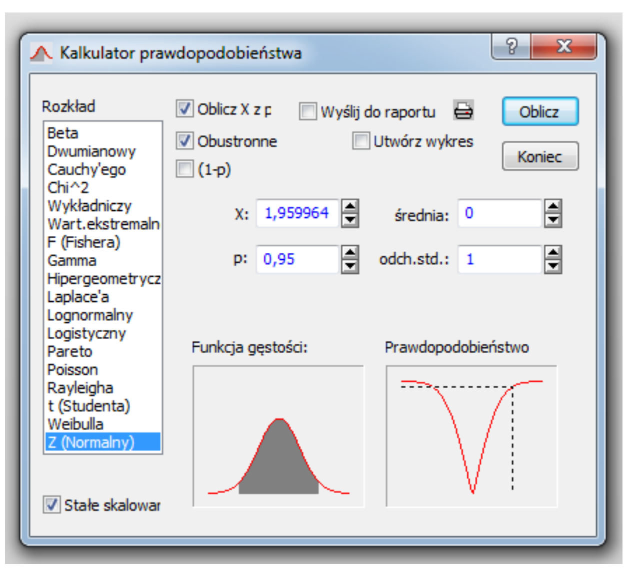
czyli

$$\frac{\alpha}{2} = 0,025$$

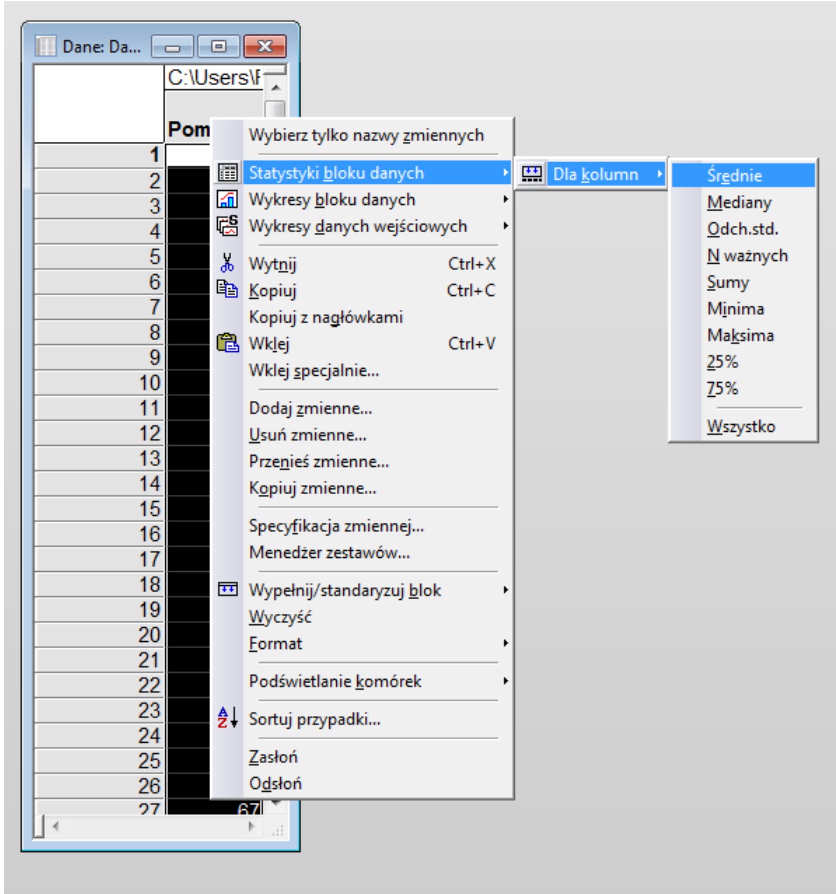
Musimy wyliczyć

$$\Phi(u_\alpha) = 1 - 0,025 = 0,975$$

W celu wyliczenia tego otwieramy **Statystyka** $\Longrightarrow$ **Kalkulatory** $\Longrightarrow$ **Rozkłady** $\Longrightarrow$ **Z(Normalny)** i wpisujemy albo $p = 0,975$ i dajemy **Oblicz** lub $p = 0,95$ i zaznaczamy kratkę **Obustronnie**, a następnie dajemy **Oblicz**.



Otrzymujemy wartość $u_\alpha = 1,96$. Potrzebne są jeszcze wartości średniej i odchylenia standardowego. W celu ich wyliczenia Klikamy na nagłówku kolumny **Pomiar** prawym klawiszem, następnie wybieramy **Statystyki bloku danych** $\Longrightarrow$ **Średnie / Odchylenie Standardowe**.

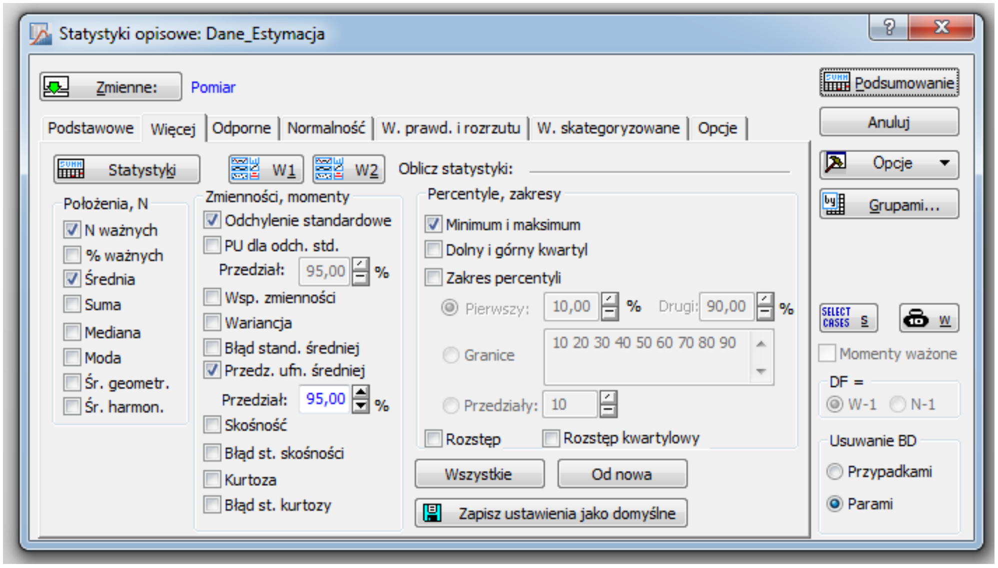


Otrzymujemy interesujące nas wartości $\bar{x} = 64,615$ i $\sigma = 3,13$. Mamy już wszystko co nas interesuje, zatem wstawmy dane, aby wyznaczyć przedział ufności:

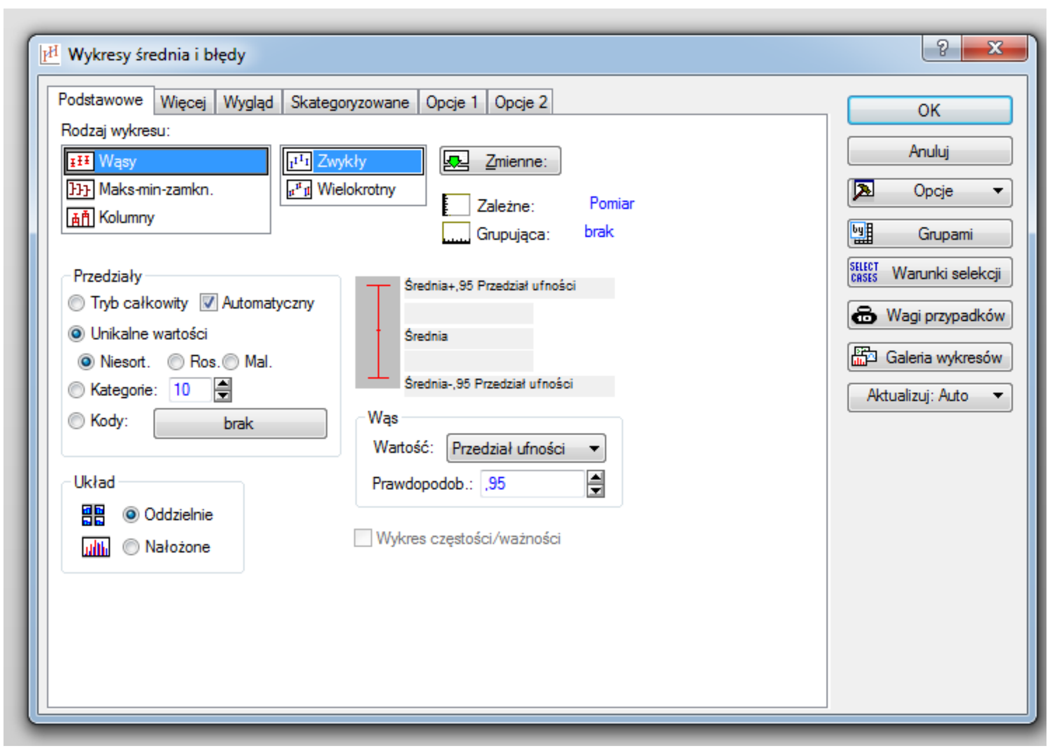
$$P\left(64,615 - 1,96 \cdot \frac{3,13}{\sqrt{52}} < m < 64,615 + 1,96 \cdot \frac{3,13}{\sqrt{52}}\right) = 0,95$$

Nasz przedział wynosi zatem $(63,765; 65,465)$

II) **Automatycznie**
Wybieramy **Statystyka** $\Longrightarrow$ **Statystyki Podstawowe** $\Longrightarrow$ **Statystyki Opisowe**, wybieramy zakładkę **Więcej**, zaznaczamy kratkę **Przedziały ufności średnie**, podajemy wartość 95% i klikamy **Podsumowanie**.



III) **Na wykresie ramka-wąsy** Wybieramy **Wykresy** $\Longrightarrow$ **Średnie**, następnie wpisujemy interesujący nas przedział ufności (domyślny to 95%), wybieramy zmienna zależną **Pomiar** i klikamy **Ok**.



Uwagi

Na koniec dwie istotne uwagi: w przypadku małych prób ($n < 30$) różnica polega na tym, że w kalkulatorze prawdopodobieństwa wybieramy rozkład t-Studenta, a nie normalny. Ponadto przy wyliczaniu automatycznym dla takiego przypadku program wylicza dzieląc nie przez $\sqrt{n-1}$, ale przez $\sqrt{n}$, ale dla $n-1$ stopni swobody. Można to sprawdzić licząc ręcznie i porównując wyniki.