



Zintegrowany  
Program  
Rozwoju  
Politechniki  
Lubelskiej

Biuro Projektu:  
ul. Nadbystrzycka 38H  
20-618 Lublin

# **METODYKA BADAŃ DOŚWIADCZALNYCH**

*Workbook*

Autor: dr Marcin Bogucki

Lublin, 2020 rok

# 1. Rozkłady statystyki z próby. Centralne Twierdzenie Graniczne. Właściwości rozkładu normalnego

## Zagadnienia

- Treść i interpretacja Centralnego Twierdzenia Granicznego;
- Pojęcie i zastosowanie rozkładu statystyki z próby na przykładzie rozkładu wartości średniej;
- Błąd oszacowania wartości średniej populacji;
- Doświadczalny rozkład częstości wystąpień obserwacji a ciągły rozkład prawdopodobieństwa obserwacji (na przykładzie rozkładu normalnego);
- Zmienna standaryzowana rozkładu normalnego;
- Wykres kwantyl – kwantyl i jego zastosowania;

## Wprowadzenie

Wnioskowanie statystyczne pozwala na pozyskiwanie informacji o parametrach rozkładu badanej populacji na podstawie pobranej próby. Kluczowa koncepcja, na podstawie której przeprowadza się wnioskowanie statystyczne (estymację parametrów rozkładu lub testowanie hipotez statystycznych), bazuje na rozkładach statystyk z próby.

W typowych zastosowaniach metod statystycznych, **rozkład statystyki z próby** najczęściej odnosi się do rozkładu prawdopodobieństwa wartości średniej, wariancji lub ilorazu wariancji szacowanych na podstawie wyników obserwacji (z próby o licznosci  $n$ ). Podstawowe rozkłady statystyk z próby są przedstawione w wielu pozycjach literatury i mają znane postacie analityczne (patrz: witryna internetowa przedmiotu). Dla przykładu rozkładem wartości średniej z próby jest rozkład normalny, rozkładem wariancji z próby jest rozkład Chi kwadrat –  $\chi^2$ , a rozkładem ilorazu dwóch wariancji jest rozkład F-Snedecora. Dysponując generatorem liczb pseudolosowych - oprogramowaniem symulującym próbę losową z populacji o zadanym rozkładzie prawdopodobieństwa np. funkcjami programu Statistica lub OpenOffice - oraz posługując się techniką histogramu, możliwe jest doświadczalne, przybliżone odtworzenie wymienionych rozkładów statystyk z próby. Taki eksperyment jest częścią niniejszego opracowania. **Zakłada się, że w trakcie pozyskiwania próby, parametry populacji (rozkładu obserwacji indywidualnych) nie ulegają zmianie.**

Drugim zagadnieniem omawianym w ćwiczeniu są praktyczne ustalenia jakie wynikają z treści Centralnego Twierdzenia Granicznego (CTW):

**Jeżeli  $x_1, x_2, \dots, x_n$  są niezależnymi zmiennymi losowymi o wartościach średnich  $\mu_1, \mu_2, \dots, \mu_n$  oraz wariancjach  $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$  i jeżeli  $y = x_1 + x_2 + \dots + x_n$ , wówczas rozkład zmiennej losowej:**

$$\frac{y - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \quad (1)$$

**dąży do zmiennej standaryzowanej  $Z = N(0,1)$ , gdy  $n$  dąży do nieskończoności.**

Twierdzenie to pierwotnie odnosi się do sumy obserwacji (zmiennych losowych) z próby. Okazuje się, że dla dostatecznie wysokiej licznosci  $n$  próby pobieranej z populacji obserwacji indywidualnych rozkład zmiennej losowej  $y$  z próby jest zgodny z rozkładem normalnym. Efekt ten jest tym bardziej

zauważalny, im rozkład obserwacji indywidualnych bardziej przypomina swym kształtem rozkład normalny i im większa jest liczność próby  $n$ .

Treść CTW dotyczy sumy niezależnych zmiennych losowych, a każda ze zmiennych może pochodzić z odmiennego rozkładu prawdopodobieństwa. W szczególnym przypadku, gdy każda ze zmiennych  $x_i$  podlega temu samemu rozkładowi prawdopodobieństwa, z CTG wiemy, że parametry wynikowego rozkładu prawdopodobieństwa: wartość średnia i odchylenie standardowe ze średnich z prób wynoszą odpowiednio:

*Jeżeli niezależne zmienne losowe:*

$$x_1, x_2, \dots, x_n \quad \mu_{\bar{x}} = \mu \quad (2)$$

*pochodzą z tego samego rozkładu prawdopodobieństwa o wartości średniej  $\mu$  i wariancji  $\sigma^2$ , wówczas wartość średnia rozkładu tych zmiennych jest równa  $\mu$ , a odchylenie standardowe rozkładu średnich, tzw. błąd standardowy wynosi:  $\frac{\sigma}{\sqrt{n}}$*

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (3)$$

co wynika z przekształcenia równania (1):

$$Z = \frac{y - \sum_{i=1}^n \mu_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} = \frac{y - n\mu}{\sqrt{n\sigma^2}} = \frac{y/n - \mu}{\frac{\sigma\sqrt{n}}{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad (4)$$

Posługując się wynikiem (4), widzimy, że wartość średnia zmiennej losowej  $y$  podlega rozkładowi normalnemu:

$$\bar{X} = \mu + Z \frac{\sigma}{\sqrt{n}} = \mu_{\bar{x}} + Z\sigma_{\bar{x}}$$

gdzie  $Z$  jest zmienną standaryzowaną rozkładu normalnego  $N(\mu, \sigma^2) = N(0,1)$ . Wyraz  $se = \frac{\sigma}{\sqrt{n}}$  jest określany mianem błędu standardowego wartości średniej (od ang. se – standard error).

**UWAGA:** Literatura przedmiotu stanowi, że liczność  $n$  powinna być większa niż 30 (w aspekcie technicznym - czytaj: liczność pomiarów badanej cechy w próbie np. skurczu wtórnego, wytrzymałości, wymiarów geometrycznych itp. na podstawie których jest liczona statystyka). W przypadku gdy rozkład obserwacji indywidualnych jest rozkładem normalnym ten warunek nie jest tak restrykcyjny i liczność może być dowolna. Dodatkowo pamiętajmy, że w czasie prób doświadczalnych parametry populacji obserwacji indywidualnych  $\mu$  i  $\sigma$  nie mogą ulegać zmianie. Innymi słowy, proces z którego pobieramy obserwacje powinien znajdować się w stanie kontroli statystycznej (powinien podlegać jedynie zakłóceniom losowym), a każda obserwacja  $x_i$  nie powinna mieć wpływu na wynik kolejnej obserwacji  $x_{i+1}$  (niezależność zmiennych).

Rozkłady statystyk z próby oraz Centralne Twierdzenie Graniczne są podstawą wielu metod statystycznych. Zrozumienie tych idei jest kluczowe w sprawnym posługiwaniu się rachunkiem statystycznym, interpretacją wyników badań doświadczalnych oraz metodami użytecznymi w doskonaleniu jakości produktów i procesów np. opracowaniu wyników pomiarów, konstrukcji kart kontrolnych, planowaniu doświadczeń i optymalizacji doświadczalnej (modelowanie empiryczne), walidacji przyrządów pomiarowych.

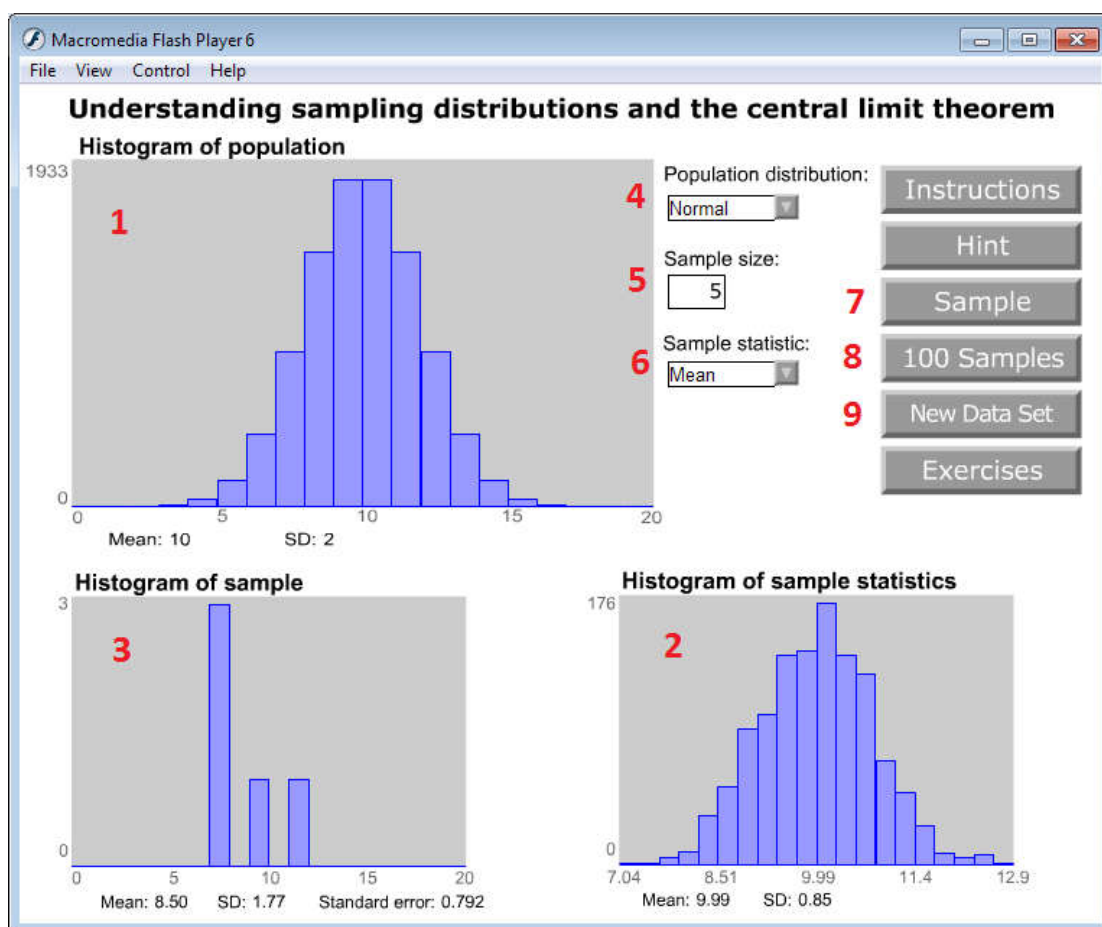
## Ćwiczenie 1.1 Interpretacja treści Centralnego Twierdzenia Granicznego

Uruchom program o nazwie „Centralne Twierdzenie Graniczne.exe” dołączony do skryptu. Rysunek 1 przedstawia panel programu wraz z opisem funkcji zamieszczonym w legendzie.

Ćwiczenie nawiązuje do treści zamieszczonych we wprowadzeniu, a zadaniem studenta jest odtworzenie rozkładów prawdopodobieństwa statystyk z próby z wykorzystaniem techniki histogramu oraz doświadczalne odtworzenie treści Centralnego Twierdzenia Granicznego.

W menu *Population Distribution* student ma do dyspozycji cztery funkcje programu odpowiedzialne za generowanie/losowanie obserwacji pochodzących z następujących rozkładów prawdopodobieństwa:

- rozkładu normalnego (ang. Normal distribution);
- rozkład Poisson’a;
- rozkład wykładniczy (ang. Exponential distribution);
- dowolny rozkład, zdefiniowany przez użytkownika (ang. custom distribution),



Rysunek 1 Panel programu „Understanding Sampling Distribution And Central Limit Theorem” udostępnionego wraz z podręcznikiem D. Montgomery: „Applied Statistics and Probability for Engineers” wydanie V. Legenda: (1) Histogram obserwacji indywidualnych z rozkładu wraz z wartością średnią  $\mu$  i odch. stand  $\sigma$ ; (2) Histogram rozkładu statystyki z próby wraz z aktualnymi obliczeniami  $\mu_S$  i  $\sigma_S$  (w tym przykładzie:  $\mu_{\bar{x}}$  i  $\sigma_{\bar{x}}$ ). (3) Histogram z próby; (4) Lista wyboru rozkładu obserwacji indywidualnych; (5) Liczność próby  $n$  (do 5 do 100); (6) Wybór statystyki obliczanej na podstawie próby; (7) Wygenerowanie pojedynczej próby o licznosci  $n$ , obliczenie wartości statystyki i dodanie jej do puli skumulowanych wyników wraz z aktualizacją histogramów: 2 i 3; (8) Podobnie do funkcja 7, lecz generowanych jest jednocześnie 100 prób; wynik wartości średniej i odchylenia standardowego jest aktualizowany na bieżąco pod rysunkiem histogramu; (9) Usunięcie wszystkich wyników z prób i obliczeń statystyk;

Obserwacje mogą być generowane wyłącznie w ramach prób o liczebności  $n$  (liczebność próby wynosi od 5 do 100 obserwacji – patrz funkcja (5)). Nawiązując do wprowadzania, każda z wygenerowanych/wylosowanych obserwacji w próbie jest realizacją doświadczenia i jest traktowana jako niezależna zmienna losowa. Student ma także do wyboru statystykę z próby (6), która będzie obliczona na podstawie wylosowanych  $n$  obserwacji.

Po dokonaniem wyborze rozkładu prawdopodobieństwa (4), statystyki z próby (6) oraz liczebności  $n$  (5) należy wygenerować próbę z populacji wybierając przycisk „Sample”. W efekcie, na lewym dolnym wykresie programu, przedstawiony zostanie: histogram statystyki z próby o  $n$  obserwacjach wraz z obliczoną wartością średnią, odchyleniem standardowym, oraz błędem standardowym z próby. Po ponownej aktywacji przycisku „Sample” (7) lub wybraniu funkcji „100 Samples” (8) do łącznej puli wyników zostanie wylosowana i dołączona kolejna próba lub kolejnych 100 prób, a następnie zostaną obliczone (zaktualizowane) wartości statystyk: średniej oraz odchylenia standardowego statystyki ze wszystkich dotychczas uzyskanych wyników prób. Wynik tych obliczeń bazuje na rozkładzie/histogramie z ogółu statystyk z prób i jest przedstawiony na prawym dolnym wykresie. Dla przykładu, na rysunku 1, w oknie (2) przedstawiono histogram z wartości średnich z wielu prób wraz z oszacowaniem wartości średniej i odchylenia standardowego rozkładu/histogramu (wyniki obliczeń: poniżej okna histogramu).

Tabela 1. Wyniki symulacji ćwiczenia: Centralne Twierdzenie Graniczne

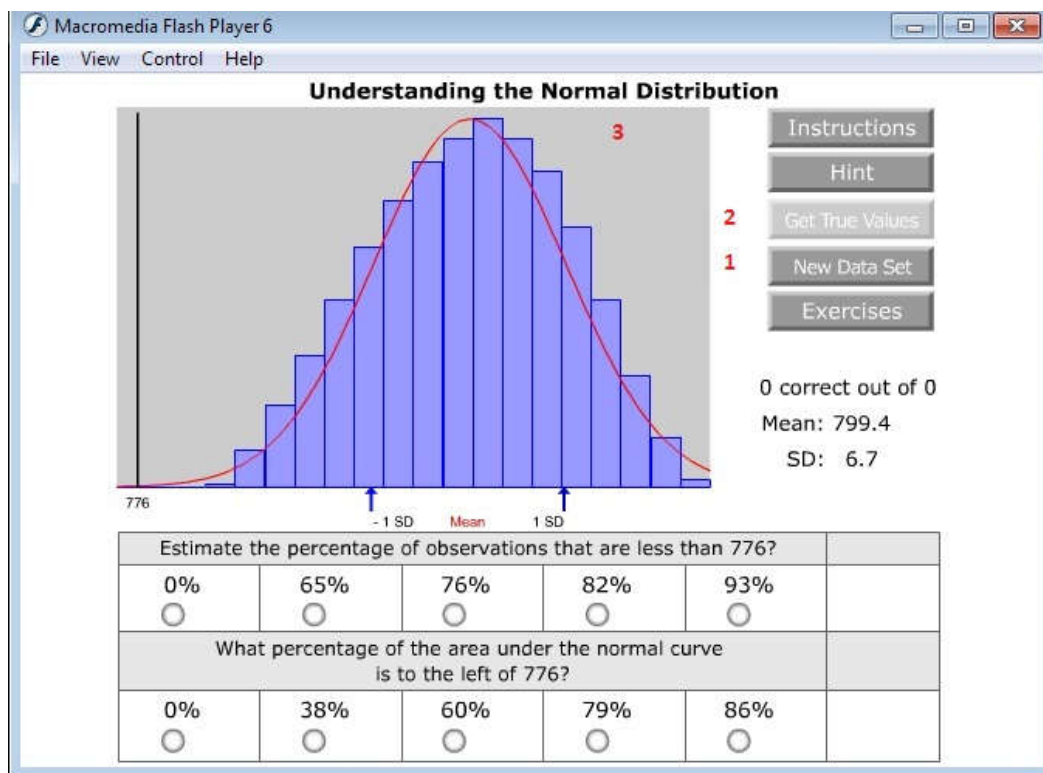
Rozkład populacji	liczebność próby, $n$	Wartość średnia populacji $\mu$	Odchylenie standardowe populacji $\sigma$	Wartość średnia rozkładu statystyki z próby $\mu_{\bar{x}}$	Odchylenie standardowe rozkładu statystyki z próby $\sigma_{\bar{x}}$
normalny	9	10	2		
	16	10	2		
	64	10	2		
	100	10	2		
wykładniczy	9	10	10		
	16	10	10		
	64	10	10		
	100	10	10		

Dysponując dużą liczbą obserwacji z prób, student ma możliwość poznania rozkładu/histogramu innych statystyk z próby (np. wariancji, czy odchylenia standardowego) przez wybór z pozycji listy „Sample Statistic”.

1. Wybierz rozkład normalny z listy "Population distribution" (4). Za liczbę pomiarów w próbie przyjmij wstępnie wartość:  $n = 9$ , a za statystykę z próby wybierz: **wartość średnią** (lista (6) „Sample Statistic” – pozycja „mean”). Następnie wygeneruj dużą liczbę prób (np. ok. 1500, przez wielokrotny wybór funkcji (8)).
  - a. Zapisz wyniki wartości średniej ze średnich oraz błąd standardowy średniej przedstawiony na prawym dolnym rysunku histogramu (2) do tabeli 2 (są to tzw. parametry rozkładu statystyki z próby)?
  - b. Czy zaobserwowane wartości średniej ze średnich oraz błędu standardowego są zbliżone z wartościami wynikającymi z CTG (patrz: równania 2 i 3)?

- c. Co można sądzić o doświadczalnym rozkładzie statystyki z próby (2)? Czy histogram statystyki z próby (2) przypomina swym kształtem rozkład normalny? Uzasadnij odpowiedź.
  - d. Jak ma się wartość średnia rozkładu statystyki z próby (2) do odpowiadającej jej wartości parametru rozkładu obserwacji indywidualnych umieszczonych pod rysunkiem histogramu – patrz: wykres (1). Jak uzasadnisz/wytłumaczysz ten wynik? Czy zauważasz regułę?
  - e. Zmień statystykę z próby na wariancję (w liście (6) pozycja SD - ang. Standard Deviation). Zaobserwuj/zapisz wartości parametrów statystyki z próby (przedstawione pod histogramem (2)) i odnieś się ponownie do pkt. (c i d).
  - f. Wykonaj ponownie polecenie z pkt. 1 ćwiczenia zmieniając licznosci próby odpowiednio  $n = \{16, 64, 100\}$  za każdym razem generując dużą liczbę prób i odnieś się wyłącznie do pytań 1a i 1b postawionych w pkt. 1. Wyniki symulacji przedstaw w formie tabeli 1.
2. Powtórz polecenia na pytania 1a i 1b zmieniając funkcję rozkładu populacji na rozkład wykładniczy (lista (4) "Population distribution"). Wyniki symulacji zestaw w tabeli 1 (patrz: wiersze i licznosci odnoszone do rozkładu wykładniczego).

Student ma możliwość wykonania podobnych, pouczających doświadczeń dla dwóch pozostałych opcji z listy wyboru (4) rozkładu populacji indywidualnych („Custom” i „Poisson”) i sprawdzenia jak będą zachowywać się: histogramy (rozkłady doświadczalne) statystyk z prób: wartości średniej i wariancji wraz ze zmianą licznosci próby;



Rysunek 2 Panel programu „Understanding Normal Distribution” udostępnionego wraz z podręcznikiem D. Montgomery: „Applied Statistics and Probability for Engineers” wydanie V. Opis funkcji programu: (1) Generowanie nowego zestawu obserwacji, (2) Porównanie oszacowań (odpowiedzi) z prawidłowymi wynikami, (3) Panel programu wykorzystywany do prezentacji histogramu wraz z dopasowanym rozkładem prawdopodobieństwa;

## Ćwiczenie 1.2 Właściwości rozkładu normalnego

W wielu metodach doświadczalnych (np. kartach kontrolnych, weryfikacji hipotez statystycznych i innych) przyjmuje się założenie, że rozkład obserwacji przedstawiony na histogramie może być w zadawalający sposób przybliżony rozkładem normalnym. Ma to swoje ważne uzasadnienie

teoretyczne i będzie dyskutowane w dalszej części kursu. W takich też przypadkach, wartości częstości wystąpień obserwacji reprezentowane na rysunku histogramu w postaci prostokątów - przypominają swym kształtem symetryczną i charakterystyczną figurę w kształcie dzwonu. Symetrię i właściwości rozkładu normalnego wykorzystuje się do wykonania szybkich oszacowań prawdopodobieństwa interesujących nas zdarzeń losowych (patrz: pkt. 2 – oszacowanie zdolności procesu). Jakkolwiek, jeżeli kształt histogramu jest zgoła odmienny od kształtu rozkładu normalnego, oszacowanie prawdopodobieństwa zdarzeń losowych na podstawie powyższego założenia prowadzi do błędnych wyników. Zatem, w jaki sposób ocenić czy histogram przedstawia sobą rozkład normalny? To zagadnienie jest przedstawione w skrypcie w postaci dwóch ćwiczeń laboratoryjnych.

Uruchom dołączony do skryptu program o nazwie „Understanding Normal Distribution”, a następnie zapoznaj się z dalszą treścią zadania.

a) W nawiązaniu do wprowadzenia, w rozważanym ćwiczeniu rozważa się przypadki w których rozkład obserwacji (pomiarów) jest zgodny z rozkładem normalnym, jak i te, w których założenie o normalności rozkładu nie jest spełnione.

Główne okno programu zawiera przycisk „New Data Set” (1) służący do generowania pomiarów (tu: obserwacji). Z chwilą wygenerowania ciągu obserwacji, w oknie aplikacji zostanie przedstawiony ich histogram wraz z krzywą reprezentującą rozkład normalny (czerwona linia) (3). Ćwiczenie polega na oszacowaniu częstości wystąpień i prawdopodobieństwa obserwacji **mniejszych lub równych** od jednej, losowo wybranej wartości. Ta wartość jest oznaczona na wykresie czarną pionową linią. Dokonując oszacowania należy wykorzystać zarówno właściwości histogramu jak i krzywej rozkładu normalnego. W tym ćwiczeniu, częstość wystąpień obserwacji oceniona na podstawie histogramu jest porównywana z prawdopodobieństwem oszacowanym na podstawie krzywej rozkładu normalnego, bez względu na to czy kształt histogramu jest zgodny z kształtem rozkładu normalnego. **Co jest istotne, obydwa oszacowania należy wykonać niezależnie.**

Jak zaznaczono, po wygenerowaniu nowych danych – przycisk „New Data Set”(1), oszacuj częstości wystąpień obserwacji na podstawie informacji przedstawionych na histogramie, a następnie wykonaj jeszcze raz to samo zadanie wykorzystując właściwości rozkładu normalnego. Zauważ, że u dołu rysunku (tuż pod krzywą rozkładu normalnego) znajdują się dwie małe strzałki oznaczające dystans jednego odchylenia standardowego liczony od wartości średniej (wartość „Mean”). Wykorzystaj tę informację do oszacowania wartości prawdopodobieństwa liczonego wg krzywej rozkładu normalnego. **Pamiętaj, że w przedziale ( $\mu \pm \sigma$ ) (na rysunku w miejsce symbolu  $\sigma$  użyto oznaczenia „SD”) znajduje się ok. 68% obserwacji, a w przedziałach ( $\mu \pm 2\sigma$ ) i ( $\mu \pm 3\sigma$ ), odpowiednio ok. 95% i 99.76% ogółu obserwacji, o ile faktycznie rozkład obserwacji może być traktowany jako normalny.** Tą właściwość rozkładu normalnego wykorzystuje się do szybkiego, przybliżonego oszacowania prawdopodobieństwa zajścia zdarzeń losowych. Spośród dwóch wierszy odpowiedzi, wybierz swoim zdaniem poprawne wartości obydwu oszacowań. Wybór „Get True Values”(2) zwraca prawidłowe odpowiedzi. Powtórz to zadanie dla kilku różnych, losowo wygenerowanych rozkładów obserwacji.

**Należy zwrócić uwagę**, że w przypadku zgodności dopasowania krzywej rozkładu do histogramu, obydwa wyniki oszacowania: częstości wystąpień (histogram) i prawdopodobieństwa wystąpień obserwacji (rozkład normalny) powinny być zbliżone. I przeciwnie, gdy ten warunek nie jest spełniony, wyniki mogą różnić się znacząco.

b) Druga część ćwiczenia dotyczy pojęcia **zmiennej standaryzowanej** rozkładu normalnego. Podobnie jak w „pkt. a” zadania, wygeneruj nowy ciąg danych („New Data Set”). Z zamieszczonego rysunku odczytaj wartość średnią rozkładu ( $\mu$  - etykieta: „Mean”) oraz odchylenie standardowe ( $\sigma$  - „Standard Deviation” – etykieta „SD”). Na podstawie tych dwóch wartości ( $\mu$  i  $\sigma$ ) oraz wartości wylosowanej obserwacji, oblicz wartość zmiennej standaryzowanej z rozkładu normalnego. Kolejno,

wykorzystując moduł programu STATISTICA - „kalkulator prawdopodobieństwa” – (Statystyka/Kalkulator prawdopodobieństwa/Rozkłady) (lub inny program np. arkusz kalkulacyjny, ew. tablice matematyczne) oceń wartość prawdopodobieństwa, o którym jest mowa w pkt. „a” ćwiczenia. Sprawdź poprawność otrzymanego wyniku prawdopodobieństwa z odpowiedzią programu („Get True Values”). W raporcie/sprawozdaniu należy umieścić rysunek (okno programu wraz z wygenerowanym rozkładem obserwacji) oraz wartość oszacowanego prawdopodobieństwa poparte stosownym rachunkiem.

### Ćwiczenie 1.3 Wykres kwantyl-kwantyl i jego zastosowania

Tabela 1 zestawia wyniki wielokrotnych pomiarów skurczu wzdłużnego (s, %) wyprasek, jednej z charakterystyk jakości procesu wtryskiwania termoplastów. **UWAGA:** Wyniki przedstawione w tabeli są dołączone do zadania w postaci arkusza programu Statistica (\*.sta) lub w pliku tekstowym (\*.csv).

Specyfikacje technologiczne nałożone na wartości skurczu wtórnego wynoszą odpowiednio: LSL = 2.2 %, USL = 3.8 % (oznaczenia LSL – dolna granica specyfikacji (ang. Lower Specification Limit); USL – górna granica specyfikacji (ang. Upper Specification Limit). Celem przypomnienia - z punktu widzenia przyjętego kryterium jakości - wartości pomiarów skurczu wzdłużnego, należące do przedziału (LSL , USL) są uznawane za prawidłowe. Pozostałe są traktowane jako wady.

Na podstawie obserwacji przedstawionych w tabeli 1 oceń zdolność procesu wtryskiwania termoplastów tzn. oszacuj odsetek wadliwie wytworzonych wyprasek i wyraż w jednostkach PPM (ang. Parts Per Millon – tu: liczba wadliwych egzemplarzy produktu na milion potencjalnie wyprodukowanych).

Szacując zdolność procesu, wykonaj następujące polecenia:

Oceń czy wyniki pomiarów podlegają rozkładowi normalnemu; Wykorzystaj właściwości wykresu normalności (menu programu STATISTICA: Wykresy/Wykresy2W/Wykresy normalności);

Tabela 1 Wyniki pomiarów skurczu wzdłużnego wyprasek

3.3	3.9	3.9	3.4	3.6	3.2	2.9	3.8	3.8	3.4	3.2	3.3
3.2	3.4	3.6	3.4	3.4	3.4	3.5	3.5	3.9	3.2	3.6	3.5
3.7	3.6	3.4	2.8	3.9	3.8	3.2	2.6	3.3	3.7	3.9	3.9
3.6	3.0	3.4	3.1	3.3	3.6	3.3	3.4	3.8	3.3	3.7	3.4
3.4	3.3	2.9	4.0	3.1	3.0	3.1	3.6	3.5	3.5	3.3	3.4

Jeżeli wykres normalności wykazuje zgodność dopasowania (tu: obserwacje pomiarów skurczu wtórnego powinny układać się wzdłuż linii idealnego dopasowania), z wykresu odczytaj wartość średnią i odchylenie standardowe; [**Wskazówka:** Zauważ, że na mocy definicji zmiennej standaryzowanej z (oś rzędnych), wartości  $z = 0$  odpowiada wartość średnia zmiennej losowej (oś odciętych wykresu – tu: wyniki pomiarów skurczu - s), a wartości  $z = 1$  (lub  $z = -1$ ) odpowiada wartość zmiennej losowej (oś odciętych) oddalonej o jedno odchylenie standardowe względem wartości średniej (odczytanej wcześniej z tego wykresu). Różnica między tymi wartościami jest miarą odchylenia standardowego;] Zweryfikuj poprawność oszacowania obliczając wartość średnią i odchylenie standardowe z wykorzystaniem programu STATISTICA (Statystyka/Statystyki podstawowe i tabele/Statystyki opisowe); W przypadku gdy rozkład obserwacji faktycznie podlega rozkładowi normalnemu, obydwa oszacowania (na podstawie wykresu normalności i bezpośrednich obliczeń) powinny być zgodne;



Dysponując wartościami liczbowymi parametrów rozkładu normalnego: wartości średniej i odchylenia standardowego:  $\mu$  i  $\sigma$ , uruchom kalkulator prawdopodobieństwa (Statystyka/Kalkulator prawdopodobieństwa/Rozkłady) i wybierz z listy rozkładów rozkład Z (Normalny). Wstawiając w pola kalkulatora wartości parametrów rozkładu i wartość LSL (pole **X**), oblicz frakcję obserwacji znajdujących się poniżej dolnej granicy specyfikacji (pole **p**). Podobnie, oblicz frakcję obserwacji położonych powyżej górnej granicy specyfikacji. Suma obydwu wyników jest poszukiwanym odsetkiem niezgodności. Zdolność procesu wyraż w jednostkach PPM (iloczyn:  $p \cdot 10^6$ );

Wykorzystaj wykres normalności i kalkulator prawdopodobieństwa (lub tablic matematycznych) do oszacowania zdolności procesu wyłącznie wykorzystaniem rozkładu standaryzowanego  $Z = N(0,1)$  gdzie: 0 – wartość średnia, 1 – odchylenie standardowe. **Wskazówka:** podobnie jak w pkt. b, z wykresu normalności można odczytać wartości zmiennej standaryzowanej z odpowiadające wartościom dwóch granic specyfikacji LSL i USL. Odczytane wartości zmiennej standaryzowanej z, wstawione do pól kalkulatora prawdopodobieństwa pozwolą na oszacowanie frakcji niezgodności (za parametry rozkładu normalnego należy wówczas przyjąć: 0 - wartość średnia i 1 - odchylenie standardowe). Obliczając górną i dolną frakcję, stosuj się do informacji zawartych w pkt. c. (pkt. d jest alternatywą pkt. b i c);

**Zapamiętajmy etapy powyższej procedury.** W pierwszej kolejności należy wykazać, że rozkład obserwacji jest rozkładem normalnym. Jeżeli zostanie to pozytywnie zweryfikowane, szacujemy zdolność procesu wg zamieszczonych wskazówek. W przeciwnym bądź razie - **nie ma żadnych podstaw do wykonania tego rodzaju obliczeń.**

## Pytania kontrolne

1. Jakie są praktyczne wnioski z treści Centralnego Twierdzenia Granicznego?
2. Co to jest statystyka z próby? Przytocz kilka przykładów statystyk z próby i odpowiadających im rozkładów prawdopodobieństwa.
3. Co to jest błąd standardowy wartości średniej? Jaka jest jego interpretacja?
4. W jaki sposób przybliża się rozkład prawdopodobieństwa badanej zmiennej losowej?
5. Wymień właściwości rozkładu normalnego. Jak należy rozumieć, z czym należy kojarzyć wartości parametrów rozkładu normalnego?
6. Co to jest zmienna standaryzowana Z. Na czym polega standaryzacja zmiennej losowej?
7. Na czym polega konstrukcja wykresu kwartył-kwartył? Jakie są korzyści wynikające ze stosowania tego rodzaju wykresów?
8. Co to jest wykres normalności rozkładu?

## 2. Przedział ufności wartości średniej i wariancji

### Zagadnienia

- Estymacja parametrów rozkładu prawdopodobieństwa populacji;
- Rozkład statystyki z próby na przykładzie rozkładu wartości średniej i wariancji;
- Błąd oszacowania wartości średniej populacji;
- Estymacja przedziałowa na przykładzie przedziałów ufności wartości średniej i wariancji;

### Wprowadzenie

Parametry rozkładu obserwacji indywidualnych (krótko: rozkładu populacji), których wartości szacuje się na podstawie próby doświadczalnej nazywa się mianem **wartości prawdziwych**. Zaznaczmy, że liczba szacowanych parametrów rozkładu populacji wynika z postaci analitycznej zakładanego modelu zjawisk losowych (funkcji rozkładu). W skrajnych przypadkach, by odtworzyć funkcję rozkładu prawdopodobieństwa szacujemy wartość tylko jednego parametru. Na przykład, jeżeli model badanego zjawiska losowego jest opisywany przez rozkład Poisson, wystarczy jedynie oszacowanie jednego parametru.

### Przykład

Zalóżmy, że inżynier stara się oszacować tzw. masowy wskaźnik szybkości płynięcia MFR [g/10 min] materiału termoplastycznego (wskaźnik charakteryzujący właściwości tworzywa). Inżynier posługuje się zaleceniami normy, dot. badań oraz próbą losową. Z doświadczenia, oraz z teorii (patrz skrypt: Centralne Twierdzenie Graniczne) wiemy, że jeżeli na badaną wielkość oddziałuje szereg czynników, a pomiary odbywają się w stałych warunkach, wówczas wynik pomiaru tej wielkości powinien podlegać zakłóceniom losowym opisywanym przez **rozkład normalny o parametrach  $\mu$  i  $\sigma$** . Wiedząc to, inżynier już przed realizacją badań dysponuje zakładanym, koncepcyjnym modelem zdarzeń losowych. W tym konkretnym przykładzie, z parametrem  $\mu$  inżynier kojarzy **wartość średnią** szybkości płynięcia materiału MFR, a z parametrem  $\sigma$  wiąże **rozrzut losowy obserwacji MFR** względem wartości  $\mu$ . **Problem inżyniera polega na jak najdokładniejszym oszacowaniu tych dwóch parametrów na podstawie pojedynczej próby losowej** tj. pobraniu kilku próbek materiału termoplastycznego, uplastycznieniu ich w aparaturze i pomiarze wskaźnika płynięcia. Na potrzeby dalszej dyskusji przyjmijmy, że znamy dokładne wartości MFR dla badanego tworzywa i wynoszą one:  $\mu = 7$  i  $\sigma = 1$  (wyrażone w jednostce: g/10min).

Spróbujmy uogólnić ten przykład na szereg innych praktyk doświadczalnych. Otóż zadaniem osoby prowadzącej badania doświadczalne jest wyznaczenie wartości parametrów rozkładu (czytaj: parametrów modelu matematycznego badanego procesu) na podstawie pojedynczej próby losowej o liczności  $n$ . Znając wartości tych parametrów oraz zakładając postać analityczną rozkładu prawdopodobieństwa wynikającą z charakteru zjawisk losowych, jesteśmy w stanie jednoznacznie odtworzyć poszukiwany rozkład obserwacji i wykorzystać go do rozwiązania szeregu praktycznych problemów.

Powyższy problem rozwiązuje się za pomocą dwóch metod statystycznych: **estymacji punktowej** i **estymacji przedziałowej** (od ang. słowa *estimate* – szacować, oceniać). Pierwsza technika, estymacja punktowa polega na zastosowaniu formuł matematycznych tzw. **estymatorów**, przybliżających wartości poszukiwanych parametrów rozkładu prawdopodobieństwa. Formuły te są wyznaczane na drodze rozważań teoretycznych (np. z zastosowaniem metod: momentów, największego prawdopodobieństwa, czy podejścia od strony metod Bayes'a lub metod numerycznych). Stąd wynika m.in. wyprowadzenie formuły wartości średniej arytmetycznej i wariancji (lub odchylenia standardowego) wykorzystywanych do szacowania wartości parametrów rozkładu  $\mu$  (położenia) i  $\sigma$  (rozrzutu) obserwacji  $\{X_1, X_2, \dots, X_n\}$  podlegających rozkładowi normalnemu:

Estymator średniej populacji  $\mu$  wyznaczony przy założeniu, że obserwacje podlegają rozkładowi normalnemu - formuła średniej arytmetycznej

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n}, \quad (1)$$

Estymator wartości rozrzutu obserwacji - wariancji wyznaczony przy założeniu, że obserwacje podlegają rozkładowi normalnemu

$$S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{n - 1}, \quad (2)$$

Estymatory/formuły są tak skonstruowane, by wartości prawdziwe parametrów rozkładu były wyznaczone tym bardziej dokładnie, im większa jest liczność próby losowej. Do tych formuł podstawione są wyniki obserwacji pozyskane w ramach próby losowej. Na tej podstawie otrzymujemy oszacowanie nieznanych parametrów rozkładu. Podkreślmy, że estymator jest tylko formułą wykorzystywaną do przybliżenia parametru rozkładu. Wartość liczbową otrzymywaną na podstawie tego oszacowania nazywa się mianem statystyki z próby. Zatem, wyniki obliczeń wartości średniej, wariancji, odchylenia standardowego, czy innych funkcji obliczanych na podstawie wyniku próby losowej są **statystykami z próby**. Dodajmy, że dany parametr populacji może być szacowany na podstawie wielu różnych formuł, różniących się od siebie precyzją (i właściwościami) oszacowania. Temat estymacji parametrów rozkładu prawdopodobieństwa jest bardzo obszerny (patrz: wykaz literatury umieszczony na stronie kursu), dlatego dla potrzeb tego ćwiczenia wystarczy, że rozważone zostaną tylko dwa estymatory parametrów  $\mu$  i  $\sigma$  odnoszone do rozkładu normalnego.

### Przykład

Załóżmy, że obowiązują warunki, o których mowa w poprzednim przykładzie. Inżynier zamierza oszacować dwa parametry rozkładu obserwacji masowego wskaźnika płynięcia MFR:  $\mu$  i  $\sigma$ , dlatego wykonuje próbę losową o liczności  $n = 16$ . Na tą chwilę liczność  $n$  jest wybrana arbitralnie. W wyniku przeprowadzonych badań, uzyskano następujące wyniki:

$X = \{ 7.8, 7.2, 7.4, 5.3, 6.9, 7.8, 8.5, 6.5, 8.0, 7.7, 5.7, 4.6, 7.7, 7.2, 5.5, 6.6 \}$  [g/10min]

Stosując dwie formuły: (1) i (2) otrzymujemy następujące oszacowanie wartości średniej rozkładu i odchylenia standardowego:

$\bar{x} = 6.9$ , oraz  $s = 1.1$  [g/10min]

Przypomnijmy, są to tylko przybliżenia dwóch stałych wartości  $\mu$  i  $\sigma$ . Jak wiemy z poprzedniej sekcji, że dokładne wartości  $\mu$  i  $\sigma$  to odpowiednio: 7 oraz 1 [g/10min]. O takim, czy innym wyniku pomiaru decyduje tu wyłącznie przypadek. Jak widzimy, oszacowanie wartości parametrów populacji są bazują na konkretnej realizacji próby losowej  $\{X_1, X_2, \dots, X_n\}$ . Gdybyśmy przeprowadzili kolejną próbę o liczności  $n = 16$ , uzyskalibyśmy zupełnie inny ciąg obserwacji, choć wartości oszacowań  $\mu$  i  $\sigma$  byłby zapewne zbliżone.

Z wyniku ostatniego przykładu widzimy, że nie dysponujemy tylko jednym oszacowaniem parametrów rozkładu  $\mu$  i  $\sigma$ , lecz raczej zbiorem ich wartości. Zbiór ten jest generowany na podstawie konkretnych realizacji prób doświadczalnych. Jeżeli warunki prowadzenia badań są ustalone, każde z uzyskanych przybliżeń poszukiwanych parametrów rozkładu  $\mu$  i  $\sigma$  jest równie wiarygodne. Interpretując pojedynczy wynik oszacowania parametru na podstawie próby jesteśmy skłonni przyznać, że jest to tylko jedna z możliwych wartości oszacowań (jedna z możliwych reprezentacji wartości prawdziwej). Zatem, na ile dokładnie przybliżamy wartości parametrów rozkładu na podstawie próby? Inaczej, na ile wąski/szeroki jest przedział zawierający wartości prawie wszystkich prawidłowych oszacowań parametru rozkładu? Odpowiedź na to pytanie prowadzi do tzw. przedziałów ufności.

## Przedział ufności wartości średniej populacji

W analogi do rozkładu obserwacji indywidualnych (populacji), gdybyśmy zadali sobie trud i wykonali wiele prób o licznosci  $n$  każda i za każdym razem obliczali wartość statystyki z próby, a następnie, wyniki tychże statystyk przedstawili na wykresie histogramu, otrzymalibyśmy tzw. **empiryczny rozkład statystyki z próby**. Typowe rozkłady statystyk z prób np. wartości średniej, wariancji, ilorazu wariancji są dobrze poznane od strony teoretycznej, a ich postacie zależą od licznosci próby. Są nimi odpowiednio: rozkład normalny oraz rozkład Chi kwadrat (przypadek obserwacji podlegających rozkładowi normalnemu).

Nawiązując do treści ćwiczenia, druga metoda estymacji parametrów populacji tzw. **estymacja przedziałowa**, wykorzystuje rozkłady statystyki z próby, a w tym konkretnym zadaniu – rozkład wartości średniej z próby. Wiedząc, że rozkład wartości średniej z próby podlega rozkładowi normalnemu (patrz ćwiczenie: *Rozkłady statystyk z próby. Centralne Twierdzenie Graniczne*) dowolny wynik wartości średniej z rozkładu o parametrach  $\mu$  i  $\sigma$  możemy przedstawić wykorzystując równanie (3):

$$P\left(\mu - z_{\alpha} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + z_{\alpha} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha, \quad (3)$$

gdzie:  $z$  jest zmienną standaryzowaną rozkładu normalnego. Równanie (3) wyraża prawdopodobieństwo zdarzenia polegającego na tym, że wartość średnia (arytmetyczna) z próby, znajdzie się w przedziale (2):

$$\mu - z_{\alpha} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + z_{\alpha} \frac{\sigma}{\sqrt{n}} \quad (4)$$

z prawdopodobieństwem równym  $1 - \alpha$ . Zauważmy, że zmniejszając wartość prawdopodobieństwa  $\alpha$ , zwiększamy szerokość przedziału, a więc i szansę p znalezienia w nim wartości średniej. Wartość  $\alpha$  wraz z parametrami  $\mu$  i  $\sigma$  oraz licznoscią  $n$  decyduje o położeniu i szerokości przedziału, w którym znajduje się potencjalny wynik z próby losowej.

Zakładając, że wartość drugiego parametru – odchylenia standardowego  $\sigma$  jest znana (dokładnie wyznaczona lub zakładana – przypadek występujący w rozważanym ćwiczeniu) oraz przekształcając nierówności (2), otrzymujemy formułę (3):

*Przedział ufności wartości średniej  $\mu$  populacji. Założenie:  $\sigma$  jest znane, a  $\bar{x}$  jest wyznaczone na podstawie próby losowej. Obserwacje mogą pochodzić z dowolnego rozkładu prawdopodobieństwa*

$$\bar{x} - z_{\alpha} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}} \quad (5)$$

Przekształcona nierówność przedstawia tzw. **przedział ufności wartości średniej  $\mu$  populacji**, a wartość  $1 - \alpha$  nazywa się mianem tzw. **poziomu ufności**. W naukach technicznych za  $\alpha$  przyjmuje się najczęściej wartość 0.05.

## Interpretacja przedziału ufności parametru wartości średniej populacji

Pomimo podobieństwa formuł (4) i (5), ich interpretacja jest zgoła odmienna:

- w formule (3) znane lub zakładane są:  $\mu$  i  $\sigma$  oraz  $\alpha$ , co daje nam pełną wiedzę o możliwych wynikach z próby  $\bar{x}$  i ich prawdopodobieństwie;
- nierówność (5) wyraża przedział, w którym, jak mamy nadzieję, znajduje się wartość  $\mu$  i mamy co do tego ufność rzędu  $1 - \alpha$ . **Co jest istotne**, nasz pogląd co do tego formułujemy wyłącznie na podstawie **pojedynczej próby doświadczalnej**. Jest to korzystne punktu widzenia ekonomiki prowadzenia badań doświadczalnych;
- interpretacja przedziału ufności wymaga przeprowadzenia pewnego eksperymentu myślowego lub symulacji komputerowej, która będzie rozważana w tym ćwiczeniu. Załóżmy, że mamy możliwość wielokrotnego ( $k$ -krotnego) przeprowadzenia próby losowej (o licznosci  $n$ ). Po przeprowadzeniu każdej z prób wyznaczamy przedział ufności wg formuły (5). Wówczas, na mocy przyjętych założeń poszukiwana wartość parametru  $\mu$  znajdowałaby się w średnio w  $k(1 - \alpha)$  wyznaczonych przedziałach ufności. Dla przykładu, gdy przyjmiemy za  $k = 100$  i za  $\alpha = 0.05$ , wtedy, przeciętnie rzecz biorąc, 95 tak wyznaczonych przedziałów ufności będzie zawierać wartość poszukiwanego parametru. Niestety, nie jesteśmy w stanie jednoznacznie wskazać, który z przedziałów ufności zawiera wartość prawdziwą, a który jej nie zawiera;
- kontynuując tą myśl, nawet jeżeli w myśl tej interpretacji, prognozowany odsetek 5% przedziałów ufności nie zawiera wartości prawdziwej parametru, konstrukcja przedziału ufności uprawnia nas do tego by sądzić, że jeden z jego końców znajduje się blisko wartości prawdziwej;
- przedział ufności może być zawężony przez zwiększenie licznosci próby (przez co błąd w oszacowaniu parametru może być zredukowany), ale przytoczona wyżej interpretacja przedziału ufności nie ulega zmianie;
- przedział ufności może być zawężony przez zwiększenie wartości prawdopodobieństwa  $\alpha$ , ale musimy się liczyć z tym, że wzrośnie także i odsetek przedziałów nie zawierających poszukiwanej wartości prawdziwej. Wprawdzie zmniejszenie wartości prawdopodobieństwa  $\alpha$  zwiększa szerokość przedziału ufności, ale zmniejsza się za to błąd oszacowania wartości prawdziwej;
- wyraz  $\frac{\sigma}{\sqrt{n}}$  w formule (5) to tzw. błąd standardowy - se (ang. standard error) wartości średniej;

**UWAGA:** Większość z tych uwag jest stosowna w odniesieniu do przedziałów ufności innych parametrów modeli probabilistycznych opisujących zjawiska losowe (np. wariancji, wskaźnika struktury, ilorazu wariancji). **Ważne jest to, by zapamiętać jaki jest sens pojęcia przedziału ufności i jak należy to pojęcie interpretować!**

## Przykład

Wykorzystajmy poprzednie przykłady do konstrukcji przedziału ufności wartości średniej populacji. Dysponując wcześniejszymi wynikami z próby:

$$X = \{ 7.8, 7.2, 7.4, 5.3, 6.9, 7.8, 8.5, 6.5, 8.0, 7.7, 5.7, 4.6, 7.7, 7.2, 5.5, 6.6 \}, \text{ g/10min}$$

wraz z oszacowaniami:

$$\bar{x} = 6.9, \text{ g/10min. oraz } s = 1.1, \text{ g/10min,}$$

wyznamy przedział ufności wartości średniej przy założeniu, że  $s$  jest dokładnym oszacowaniem rozrzutu obserwacji ( $s = \sigma$ ). Licznosc próby wynosi  $n = 16$ , a za wartość  $\alpha = 0.05$  ( $z_{\alpha/2} = 1.96$  – patrz: kalkulator prawdopodobieństwa lub tablice całki rozkładu  $Z$ ).

Stosując nierówność (5), po podstawieniu wyników do formuły, otrzymujemy:

$$P\left(6.9 - 1.96 \frac{1.1}{\sqrt{16}} \leq \mu \leq 6.9 + 1.96 \frac{1.1}{\sqrt{16}}\right) = 1 - 0.05 = 0.95$$

Przedział ufności (PU) wartości średniej rozkładu populacji wynosi:  $6.36 \leq \mu \leq 7.44$ . Na podstawie założeń konstrukcji przedziału ufności, mamy nadzieję, że w przedziale (6.36, 7.44) znajduje się poszukiwana wartość parametru  $\mu$ . Czy  $\mu$  faktycznie należy do tego przedziału? Na to pytanie z absolutną pewnością nie odpowiemy, ale metoda konstrukcji PU daje nam ufność co do tego twierdzenia rzędu 0.95. Widzimy, że przedział ufności daje nam dodatkowe informacje o precyzji oszacowania wartości  $\mu$ . Nadal nie wiemy ile dokładnie wynosi  $\mu$ , ale z dużą ufnością możemy przyjąć, że oszacowanie jest nie mniejsze niż 6.36 i nie większe niż 7.44. Przedział ufności zawiera wartość prawdziwą  $\mu$ , która z założenia wynosi 7. Wartość statystyki z próby: 6.9 jest bardzo zbliżona do wartości prawdziwej. Teoria podpowiada nam, że wynik leżący poza przedziałem ufności powinien trafić się raz na 20 tak wyznaczonych przedziałów ufności (powtórzonych prób).

## Przedział ufności wariancji $\sigma^2$

Nadal przyjmijmy, że próba losowa  $\{X_1, X_2, \dots, X_n\}$  pochodzi z populacji o rozkładzie normalnym. Wówczas, zmienność losowa drugiego z przybliżanych parametrów populacji - wariancji  $\sigma^2$  jest opisywana rozkładem Chi kwadrat ( $\chi^2$ ). Wprowadźmy zmienną Chi kwadrat  $\chi^2$ :

$$X^2 = \frac{(n-1)s^2}{\sigma^2}$$

Gdzie wariancja  $\sigma^2$  jest znana, natomiast  $S^2$  to wariancja z próby, która jest obliczana na podstawie próby o liczności  $n$ .  $df = n-1$  to tzw. liczba stopni swobody, parametr rozkładu  $\chi^2$ .

Przedział ufności wariancji jest zapisywany następująco:

$$P\left(X_{1-\frac{\alpha}{2}, n-1}^2 \leq X^2 \leq X_{\frac{\alpha}{2}, n-1}^2\right) = 1 - \alpha \quad (6)$$

$$\left(X_{1-\frac{\alpha}{2}, n-1}^2 \leq \frac{(n-1)s^2}{\sigma^2} \leq X_{\frac{\alpha}{2}, n-1}^2\right) = 1 - \alpha$$

gdzie:  $\chi^2$  jest statystyką Chi kwadrat, a  $(1-\alpha)$  to zakładany poziom ufności. Po podstawieniu do równania formuły wyrażającej statystykę  $\chi^2$  i przekształceniu otrzymujemy zapis przedziału ufności dla wariancji  $\sigma^2$ :

Przedział ufności wariancji  $\sigma^2$  populacji. Obserwacje z próby pochodzą z rozkładu normalnego o liczności  $n$ . Liczba stopni swobody:  $df = n-1$  określa jedną z krzywych rodziny rozkładów  $\chi^2$ . UWAGA: Ten przedział nie jest symetryczny

$$P\left(\frac{(n-1)s^2}{X_{\frac{\alpha}{2}, n-1}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{X_{1-\frac{\alpha}{2}, n-1}^2}\right) = 1 - \alpha \quad (7)$$

$$\frac{(n-1)s^2}{X_{\frac{\alpha}{2}, n-1}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{X_{1-\frac{\alpha}{2}, n-1}^2}$$

Przedział ufności dla wariancji (7) umożliwia oszacowanie zmienności losowej wariancji  $\sigma^2$  i/lub odchylenia standardowego  $\sigma$  na podstawie wyniku wariancji z próby  $s^2$ . Dokładność oszacowania wartości prawdziwej wariancji  $\sigma^2$  jest uzależniona od liczności próby oraz przyjętego poziomu ufności  $(1-\alpha)$ , który zazwyczaj wynosi 0.95.

## Przykład

Kontynuując przykład związany z badaniem MFR, mając dane oszacowanie odchylenia standardowego (lub wariancji) z próby  $s = 1.1$  % o licznosci  $n = 16$  i  $(1-\alpha) = 0.95$  przedział ufności wariancji MFR przedstawia się następująco:

$$\frac{(n-1)s^2}{X_{\frac{\alpha}{2}, n-1}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{X_{1-\frac{\alpha}{2}, n-1}^2}$$

Po podstawieniu wartości:  $s^2 = 1.21$ , oraz wartości statystyk Chi kwadrat dla liczby stopni swobody równej  $n - 1 = 15$ , odpowiednio:  $X_{\frac{\alpha}{2}, n-1}^2 = X_{0.025, 15}^2 = 27.488$  oraz  $X_{1-\frac{\alpha}{2}, n-1}^2 = X_{0.975, 15}^2 = 6.262$ ,

otrzymujemy:  $\frac{15 \cdot 1.21}{27.488} \leq \sigma^2 \leq \frac{15 \cdot 1.21}{6.262}$ ;

Ostatecznie, przedział wariancji wynosi:  $0.66 \leq \sigma^2 \leq 2.89$ , a przedział 95% losowej zmienności odchylenia standardowego wynosi:  $0.8 \leq \sigma \leq 1.7$ ;

Wartość drugiego parametru rozkładu normalnego  $\sigma$  - w tym zadaniu kojarzonego z rozrzutem pomiaru MFR - nie powinna być większa niż 1.7, i nie mniejsza niż 0.8 g/10min. Zauważmy, że podobnie jak w przypadku wartości średniej, przedział ufności odchylenia standardowego zawiera wartość prawdziwą parametru  $\sigma$ , która dla potrzeb tego przykładu wynosi 1 g/10 min.

Podsumowując, konstrukcja przedziałów ufności wyraża precyzję oszacowania wartości poszukiwanego parametru rozkładu obserwacji. Przedziały ufności mają bardzo ważne znaczenie z punktu widzenia opracowania, analizy i interpretacji wyników badań doświadczalnych.

## Ćwiczenie 2.1 Interpretacja przedziału ufności wartości średniej

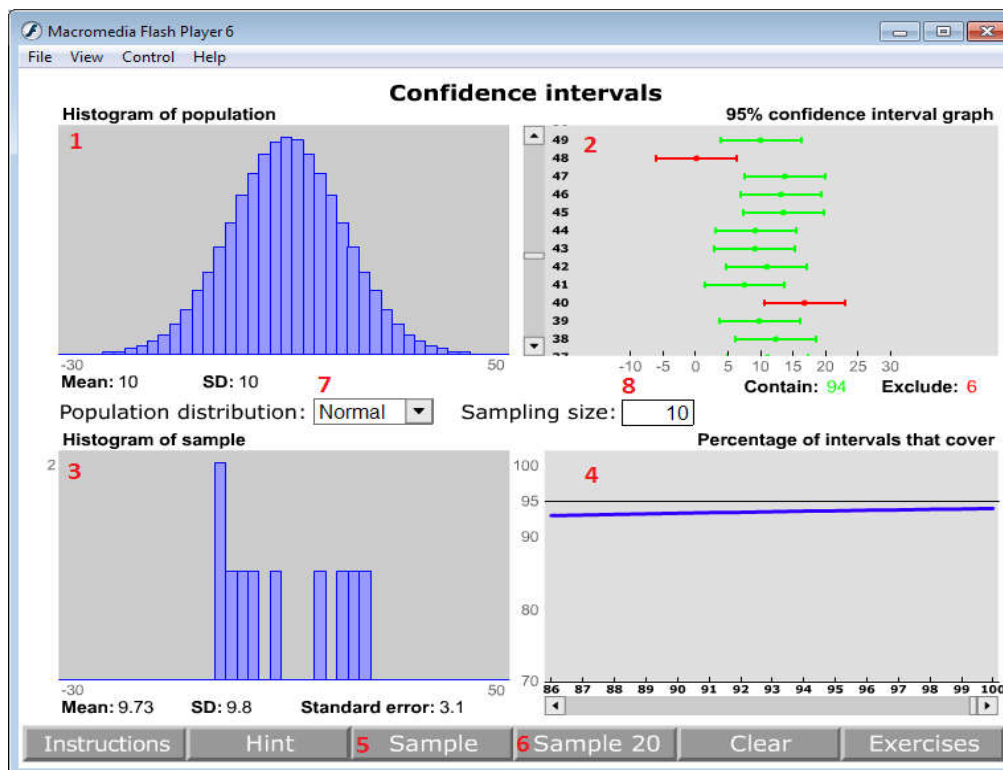
Celem ćwiczenia jest zapoznanie studenta z konstrukcją oraz interpretacją przedziału ufności wartości średniej populacji z próby. Dodatkowym celem jest przypomnienie i utrwalenie podstawowych pojęć z zakresu metod statystycznych: estymacja punktowa i przedziałowa, statystyka z próby, rozkład statystyki z próby.

Po uruchomieniu programu dołączonego do skryptu („przedzał\_ufnosc\_w.sredniej.exe”), panel użytkownika przedstawia cztery sekcje zaznaczone i opisane w legendzie do rys. 1.

Histogram (1) przedstawiony na rys. 1 obrazuje rozkład prawdopodobieństwa obserwacji indywidualnych (inaczej: rozkład populacji). W ćwiczeniu student ma do wyboru trzy funkcje opisujące rozkłady prawdopodobieństwa populacji (patrz: lista wyboru (7)) wraz z góry określonymi wartościami dwóch parametrów - wartości średniej  $\mu$  oraz odchylenia standardowego  $\sigma$ :

- Rozkład normalny - wartościami parametrów są:  $\mu = 10$ ,  $\sigma = 10$ ;
- Rozkład Gamma:  $\mu = 1.25$ ,  $\sigma = 0.65$ ;
- Rozkład dwumianowy:  $\mu = 0.4$ ,  $\sigma = 0.49$ ;

W rozważanym zadaniu, wartości prawdziwe  $\mu$  i  $\sigma$  wymienionych rozkładów są wprawdzie znane, ale tylko na potrzeby analizy oszacowań parametrów rozkładu oraz weryfikacji wyników teoretycznych przedstawionych we wprowadzeniu do ćwiczenia. Jak wcześniej zaznaczono, w warunkach rzeczywistych, wartości tych parametrów podlegają ocenie na podstawie wyniku z próby losowej.



Rysunek 1 Panel programu „Confidence Interval” udostępnionego wraz z podręcznikiem D. Montgomery: „Applied Statistics and Probability for Engineers” wydanie V. (1) Histogram obserwacji indywidualnych (tu: rozkład normalny  $N(\mu=10, \sigma=10)$ ); (2) Przedziały ufności w. średniej wykreślone na podstawie wyników losowania  $n$  obserwacji (tu:  $n = 10$ ) z wybranej populacji (patrz: sekcja(1)) Kolor zielony przedziału oznacza, że przedział ufności zawiera wartość prawdziwą szacowanego parametru; Kolor czerwony, odpowiednio – nie zawiera; (3) Histogram z  $n$  obserwacji, a poniżej wyniki statystyk z próby: wartość średnia (Mean) i odch. stand. dla  $n=10$  (SD); (4) Wykres skumulowanego odsetka prób zakończonych sukcesem (wyrażonych w %); (5) Wygenerowanie/wybór jednej próby; (6) Wygenerowanie/wybór 20 kolejnych prób; (7) Wybór rozkładu prawdopodobieństwa obserwacji indywidualnych ; (8) Wybór liczności próby;

Bezpośrednio po uruchomieniu programu, domyślnym rozkładem populacji jest rozkład normalny (lista wyboru – (7)), a liczność próby  $n$  wynosi 10 (liczność próby zadawana w polu (8) może się zmieniać w zakresie od 10 – 100 obserwacji).

Po wyborze funkcji „Sample” (5), program generuje próbę losową z wybranego rozkładu obserwacji, a następnie oblicza wartości statystyk z próby: wartość średnią, odchylenie standardowe oraz błąd standardowy. Wyniki tych obliczeń są przedstawione pod wykresem histogramu obserwacji z próby (patrz: panel(3)). Na podstawie tych oszacowań, wyznaczany jest także przedział ufności wartości średniej populacji (formuła (5) przy zakładanym prawdopodobieństwie  $\alpha = 0.05$ ). Jego reprezentacja w postaci odcinka (wraz z wartością średnią z próby) jest kreślona na wykresie przedstawionym na panelu (2). Każda próba oraz każdy przedział ufności ma nadany indeks.

Podobnie, po wyborze funkcji „Sample 20”, generowanych (losowanych) jest 20 prób losowych o liczności  $n$ , na podstawie których są obliczane i kreślone przedziały ufności wartości średniej populacji.

Wykres umieszczony na panelu (4) to skumulowany odsetek prób zakończonych sukcesem polegającym na tym, że wyznaczony przedział ufności zawiera wartość prawdziwą szacowanego parametru rozkładu.



## Etap 1. Analiza oszacowania parametru populacji na podstawie pojedynczej próby losowej

Wybermy z listy rozkładów prawdopodobieństwa rozkład normalny, za licznosc próby przyjmijmy  $n = 16$  i wylosujmy pierwszą próbę (funkcja „Sample” (5)). Choć parametry populacji rozkładu normalnego ( $\mu = 10$ ,  $\sigma = 10$ ) są odmienne od tych rozważanych w przykładzie z wprowadzenia teoretycznego (patrz: pomiar MFR tworzywa), problem jest ten sam. Tak jak inżynier próbował ocenić rzeczywistą wartość wskaźnika MFR na podstawie pojedynczej próby doświadczalnej o licznosci  $n = 16$  ( $\alpha = 0.05$ ), tak teraz student posługując się programem symulacyjnym usiłuje oszacować wartość prawdziwą populacji.

Zgodnie z naszą wiedzą na temat konstrukcji przedziału ufności, po wylosowaniu pojedynczej próby i konstrukcji przedziału ufności (formuła (5)) mamy szansę 0.95 na to, że tak utworzony przedział będzie zawierać wartość prawdziwą (patrz sekcja: Interpretacja przedziału ufności parametru). Jest to dość duże prawdopodobieństwo, więc student powinien zauważyć odcinek o kolorze zielonym wykreślony na panelu (2) symulatora. Jeżeli odcinek został wykreślony kolorem czerwonym, oznacza to, że przedział ufności nie zawiera wartości prawdziwej. Przeciętnie, jeden na 20 studentów wykonujących to ćwiczenie uzyska taki wynik ( $1/\alpha = 20$ ). Czy ten wynik należy traktować jako nieprawidłowy? Czy mamy jakkolwiek podstawę by tak sądzić?

Nawet jeżeli otrzymamy taki wynik, proszę zwrócić uwagę na to, że jeden z końców przedziału ufności znajduje się blisko szacowanego parametru ( $\mu = 10$ ). **Pomimo, że według reguł ćwiczenia próba zakończyła się niepowodzeniem, honorujemy jej wynik.** Uznajemy, że wartość średnia z tej próby jest jednym z wielu możliwych do uzyskania wyników oszacowań. **W rzeczywistych warunkach (np. pomiar MFR) nie znamy wartości prawdziwej parametru, więc nie możemy ustalić czy próba zakończyła się powodzeniem, czy porażką.**

Teraz, zwiększymy licznosc próby. Niech  $n$  wyniesie 64 (patrz: pole (8)). Proszę wprowadzić tą wartość i jeszcze raz wybrać przycisk „Sample”. Z uwagi na zwiększenie licznosci próby, przedział ufności dla licznosci  $n = 64$  (formuła (5)) powinien być teraz wyraźnie węższy, niż dla  $n = 16$ . Jak widzimy na wykresie (2) precyzja oszacowania wartości parametru wzrosła, lecz interpretacja wyniku tej próby jest nadal taka sama. W przeważającej liczbie przypadków uzyskamy wynik zawierający wartość szacowanego parametru.

Gdy licznosc próby wzrośnie do  $n = 100$ , precyzja oszacowania parametru populacji  $\mu$  będzie jeszcze większa. Na podstawie formuły (5) możliwe jest obliczenie licznosci próby losowej jaką należy przyjąć, by błąd oszacowania nieznaney wartości parametru nie przekraczał arbitralnie wybranej dokładności. Jako ćwiczenie, zachęcam do wyznaczenia tej formuły.

## Etap 2. Analiza oszacowania parametru populacji na podstawie wielokrotnej próby losowej

Ten etap doświadczenia jest podobny do poprzedniego. Zadaniem studenta jest ponowny wybór rozkładu normalnego za rozkład populacji (obserwacji indywidualnych). Tym razem należy wybrać arbitralną licznosc próby  $n$ , a następnie wygenerować dużą liczbę prób (wielokrotny wybór funkcji (5) „Sample20”). Zgodnie z oczekiwaniami, odsetek przedziałów ufności zawierających wartość prawdziwą powinien wynieść średnio ok. 0.95 ogólnej liczby. Porównaj ten wynik do treści sekcji: *Interpretacja przedziału ufności parametru wartości średniej populacji.*

## Etap 3. Analiza oszacowania parametru populacji na podstawie pojedynczej próby losowej pochodzącej z rozkładu prawdopodobieństwa gamma

Zadaniem studenta jest wybór z listy rozkładów populacji – rozkładu gamma i odtworzenie etapów: 1 i 2 doświadczenia. *Komentarz: Sposób oszacowania i interpretacja przedziału ufności wartości średniej (formuła (5)) nie ulega zmianie.*

## Pytania kontrolne

1. Co to jest przedział ufności szacowanego parametru rozkładu prawdopodobieństwa?
2. Wyprowadź formuły obliczeniowe dla przedziałów ufności wartości średniej i wariancji?
3. Jaka jest interpretacja przedziału ufności i jaki jest związek przedziału ufności z rozkładem statystyki z próby?
4. Czy wartość oszacowanego parametru zawsze należy do przedziału ufności? Uzasadnij odpowiedź.
5. Co to jest poziom ufności? Wyjaśnij to pojęcie odnosząc do oszacowania przedziału ufności wartości średniej.
6. Co oznacza zapis:  $\mu \pm \sigma/\sqrt{n}$ ?
7. Co oznacza zapis:  $\bar{x} \pm s/\sqrt{n}$ ?

### 3. Weryfikacja hipotez statystycznych

#### Zagadnienia:

- Procedura weryfikacji hipotez statystycznych, parametrycznych;
- Błąd I i II rodzaju, wartość prawdopodobieństwa  $p$  wynikająca z próby;
- Moc  $P$  testu statystycznego;
- Statystyczna a praktyczna istotność różnicy średnich;
- Dobór licznosci próby z wykorzystaniem krzywych operacyjnych OC;

#### Wprowadzenie

Procedura weryfikacji hipotez statystycznych (inaczej: test statystyczny) dotyczy stwierdzeń odnoszonych do wartości jednego z parametrów rozważanego rozkładu prawdopodobieństwa. Z tego względu, procedura ta jest nazywana mianem parametrycznej. Dla przykładu, w przypadku rozkładu normalnego, gdzie rozważa się jedynie dwa parametry:  $\mu$  i  $\sigma$ , twierdzenie może dotyczyć położenia lub rozrzutu obserwacji.

Weryfikując hipotezę statystyczną formuluje się dwa wzajemnie wykluczające się twierdzenia dotyczące parametru rozkładu obserwacji. W nomenklaturze doświadczalnictwa nazywa się je mianem: hipotezy zerowej – oznaczoną symbolem:  $H_0$  i hipotezy alternatywnej:  $H_1$ . Procedura weryfikacji hipotezy statystycznej prowadzi do rozstrzygnięcia, które z tych dwóch twierdzeń jest prawdziwe, a które należy odrzucić. Z uwagi na nieuniknioną zmienność losową, nigdy nie jesteśmy w stanie definitywnie stwierdzić, która z hipotez,  $H_0$  czy  $H_1$ , jest prawdziwa. Możemy to jedynie orzec z arbitralnie przyjętym progiem prawdopodobieństwa. Ważną uwagą jest również to, że procedura weryfikacji hipotezy statystycznej bazuje na wyniku próby (bądź prób) doświadczalnych. To właśnie wynik próby i obliczona na jego podstawie tzw. wartość statystyki testowej stanowi przesłankę do przyjęcia bądź odrzucenia jednego z postulowanych twierdzeń.

W metodach statystycznych rozważa się również tzw. hipotezy nieparametryczne, które dotyczą twierdzeń odnoszonych do postaci rozkładu danych doświadczalnych (obserwacji). W takich sytuacjach jesteśmy zainteresowani zweryfikowaniem hipotezy o zgodności doświadczalnego rozkładu obserwacji z postulowanym rozkładem prawdopodobieństwa. Dla przykładu, w wielu sytuacjach zastosowanie wybranej metody analizy danych jest uzależnione od uprzedniego wykazania/sprawdzenia, że zbiór obserwacji lub pochodnych im statystyk podlega rozkładowi normalnemu (przykład.: analiza wariancji, czy analiza regresji).

Rozważmy wykorzystanie procedury weryfikacji hipotezy statystycznej w doskonaleniu jakości procesu wtryskiwania termoplastów. Z uwagi na to, że ten przykład będzie ilustracją wielu aspektów zastosowania weryfikacji hipotez statystycznych, scharakteryzujemy bliżej proces wtryskiwania.

#### Przykład

*Proces, w ramach którego formuje się tzw. wypraski wtryskowe składa się z trzech faz: uplastycznienia, wtrysku oraz ochładzania tworzywa w formie. O jakości tego procesu, który może być oceniany ze względu na wiele kryteriów (tzw. charakterystyk jakości), decyduje bardzo wiele czynników (tzw. zmiennych sterujących tj. zmiennych decydujących o przebiegu procesu). Pokróćce, czynniki te można sklasyfikować jako technologiczne (zmiennie nastawcze procesu), materiałowe (charakterystyki tworzywa), środowiskowe (oddziaływanie otoczenia procesu).*

*Na potrzeby aktualnej dyskusji umówmy się, że proces wtryskiwania przebiega stabilnie i podlega jedynie zakłóceniom losowym. Rozważaną charakterystyką jakości jest skurcz przetwórczy wyprasek  $s$  [%]. Rys. 1 przedstawia histogram obserwacji skurczu przetwórczego wraz z tzw. granicami specyfikacji:  $LSL = 4$ ,  $USL = 8$  oraz nominalnym  $T = 6$ . Obserwacje, które znajdują się poza przedziałem specyfikacji, reprezentują wady produktów. Na podstawie wyniku*

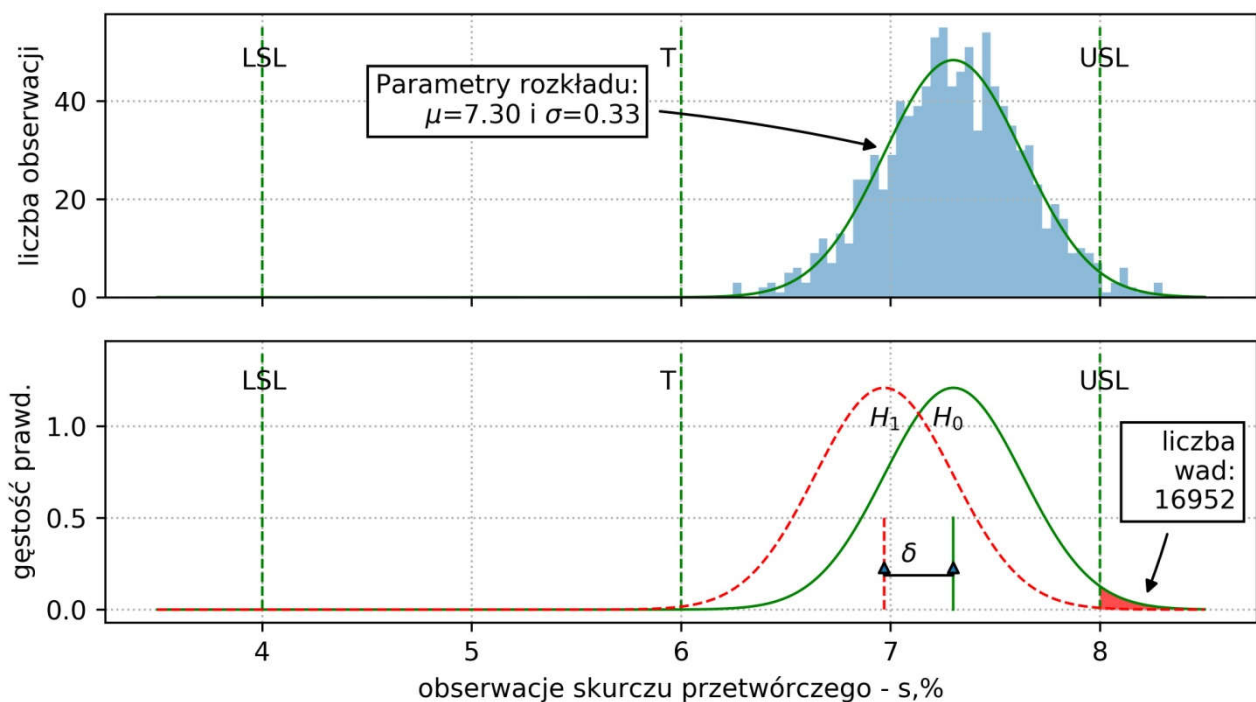
histogramu dopasowano i oszacowano parametry rozkładu prawdopodobieństwa. Kreśląc wykres normalności, zweryfikowano, że rozkład obserwacji może być przybliżony rozkładem normalny o parametrach  $\mu_0 = 7.3$  i  $\sigma = 0.33$ .

Dla uproszczenia przyjmijmy również, że operator procesu ma możliwość oddziaływania na przebieg procesu wtryskiwania tylko za pomocą jednej zmiennej sterującej: temperatury formy. Histogram obrazujący stan procesu odnosi się do warunków dla których  $T_f = 38^\circ\text{C}$ . Inżynier nadzorujący przebieg procesu wtryskiwania termoplastów może być zainteresowany wieloma aspektami doskonalenia jakości produkcji. Zgodnie z definicją operacyjną jakości: „**Jakość jest odwrotnie proporcjonalna do zmienności losowej procesu**”, inżynierowi zależy na tym, by rozrzut obserwacji reprezentowany przez krzywą rozkładu normalnego (rys. 1) był jak najwęższy (parametr  $\sigma$  jak najmniejszy) przy wartości średniej procesu zgodnej z wartością nominalną (parametr  $\mu = T$ ). Taki warunek minimalizuje liczbę wad produktów powstających w procesie. Stąd zasadne są dwa następujące pytania:

Jak dobrać wartości zmiennych sterujących procesem ( $T_f$ , ... i pozostałych) – tzw. warunków prowadzenia procesu, by wyniki pomiaru nadzorowanej charakterystyki jakości znajdowały w przedziale specyfikacji jak najbliżej nominalu  $T$ ? Aby udzielić racjonalnej odpowiedzi na to pytanie, wymagany jest model matematyczny badanego procesu (tu: tzw. model statyczny), na podstawie którego możliwe jest wyznaczenie wartości zmiennych sterujących (tzw. nastaw procesu) tak by był spełniony warunek:  $\mu = T$ , przy jednoczesnej minimalizacji rozrzutu obserwacji;

Jeżeli dogodne warunki prowadzenia procesu zostaną już ustalone, w jaki sposób utrzymywać ten stan, tak by liczba ew. wad produktów była akceptowalna tj. zgodna z przyjętym wskaźnikiem zdolności procesu;

W kontekście metod zapewniania jakości, dwa z wymienionych pytań są ściśle związane ze zagadnieniami Statystycznego Sterowania Procesem – SSP (ang. SPC Statistical Control Process). W pierwszym przypadku zadaniem operatora procesu jest optymalizacja procesu ze względu na przyjęte kryterium jakości (tu: skurcz przetwórczy wyprasek). Drugie pytanie odnosi się natomiast do technik nadzorowania procesu. **W praktyce obydwaj problemy są rozwiązywane z wykorzystaniem procedur weryfikacji hipotez statystycznych.**



Rysunek 1: Aktualny stan procesu wtryskiwania termoplastów. Histogram obserwacji skurczu przetwórczego wyprasek. Wartości LSL, USL i  $T$  oraz parametry rozkładu normalnego:  $\mu$  i  $\sigma$  odnoszą się do komentarza zawartego w przykładzie.. Opracowanie własne;

## Weryfikacja hipotezy statystycznej o równości dwóch średnich

Nawiążmy do przykładu doskonalenia procesu wtryskiwania termoplastów i przypuśćmy, że wartość średnia skurczu przetwórczego jest nieakceptowalna. Znacząca część produkcji jest niezgodna z przyjętą specyfikacją (patrz: rys. 1 – odsetek wad wynosi ok. 17000). Aby zapobiec problemom

występującym w procesie, inżynier zamierza zredukować aktualny poziom skurczu wyprasek, tak by jego wartość była zbliżona do wartości nominalnej  $T$  równej 6%. Inżynier nie dysponuje modelem matematycznym procesu i nie wie jak dobrać wartości zmiennych nastawczych procesu, by wprowadzić korektę do procesu. Niemniej, posiadając już pewne doświadczenie, zauważył, że zmiana jednej ze zmiennych procesu: temperatury formy  $T_f$ , zazwyczaj skutkuje zmniejszeniem skurczu przetwórczego termoplastów.

Inżynier zamierza sprawdzić, czy faktycznie zmiana nastaw procesu wtryskiwania (z wyższej na niższą temperaturę ochładzania wyprasek w formie wtryskowej  $T_f$ ) będzie przejawiać się w redukcji skurczu przetwórczego. W tym celu aranżuje następujące doświadczenie technologiczne. Ustala wszystkie pozostałe zmienne sterujące procesem, jawnie nieuwzględnione w eksperymencie na dotychczasowych poziomach nastaw procesu, i zmienia jedynie wartość temperatury formy  $T_f$  z 38°C na 32°C. Następnie, inżynier produkuje wypraski w liczbie  $n$  (liczność próby), mierzy ich skurcz przetwórczy i oblicza wartość średnią z próby  $\mu_1$ . Na podstawie tego pomiaru inżynier chce odpowiedzieć na pytanie: czy faktycznie zmiana temperatury formy pociąga za sobą korzystną zmianę skurczu przetwórczego wyprasek? Z uwagi na to, że w eksperymencie porównujemy ze sobą wyniki prób doświadczalnych, odnoszonych do różnych, ale ustalonych warunków prowadzenia badań (tzw. układów doświadczalnych), o takim eksperymencie mówimy jako o eksperymencie porównawczym. Temperatura formy –  $T_f$  - jedna ze zmiennych sterujących procesem, która jest poddana intencjonalnym zmianom, w nomenklaturze doświadczalnictwa nazywa się czynnikiem badanym. Zmienna poddana obserwacjom, w tym wypadku skurcz przetwórczy, jest traktowana jako czynnik wynikowy eksperymentu.

Zanim przejdziemy do opracowania wyniku tego doświadczenia, zwróćmy uwagę kilka istotnych kwestii. Przed wykonaniem eksperymentu inżynier wie, jaka jest aktualna wartość skurczu  $\mu$  przetwórczego oraz jego rozrzut  $\sigma$ . Zna te dwa parametry rozkładu obserwacji na podstawie wielu pomiarów pochodzących z nadzorowania procesu (rys. 1 :  $\mu_0 = 7.3$  i  $\sigma = 0.33$ ). Może zatem w przybliżeniu odtworzyć rozkład obserwacji indywidualnych (populacji). Podobnie, na mocy Centralnego Twierdzenia Granicznego - inżynier wie, że jeżeli będzie wykonywać (niezależne) próby doświadczalne o licznosci  $n$  każda, rozkład wartości średniej z próby będzie podlegać rozkładowi normalnemu o wartości średniej  $\mu_0$  i rozrzucie (odchyleniu standardowemu średniej)  $\sigma/\sqrt{n}$ . Zmieniając wartości temperatury formy  $T_f$ , inżynier przypuszcza, że zmianie ulegnie aktualna wartość średnia rozkładu populacji  $\mu_0$  (dla uproszczenia zakładamy, że  $\sigma$  nie ulega zmianie). To oznaczałoby, że intencjonalna zmiana czynnika badanego pociąga za sobą zmianę czynnika wynikowego. Potocznie powiedzielibyśmy, co często jest używane w żargonie naukowym, że zmiana temperatury formy ma wpływ na zmianę skurczu przetwórczego. Zauważmy, że interesuje nas tu tylko sam fakt, że między czynnikiem badanym a wynikowym występuje relacja przyczynowo-skutkowa i to właśnie chcielibyśmy wykazać w tym eksperymencie. Wynik eksperymentu nie wskazuje jaką wartość temperatury należałoby przyjąć, by skurcz przetwórczy mógł osiągnąć wartość nominalną  $T$ . Tego rodzaju problem jest rozwiązywany z wykorzystaniem planów doświadczalnych wieloczynnikowych oraz analizy regresji. Wynik eksperymentu ma na celu uzasadnienie zmiany temperatury  $T_f$  z 38 na 32 °C, o ile będzie to skutkowało redukcją skurczu przetwórczego  $s$  tworzywa.

## Zastosowanie procedury weryfikacji hipotez statystycznych

Aby stwierdzić, czy zmiana temperatury formy jest skutecznym sposobem redukcji skurczu przetwórczego, inżynier wykorzystuje procedurę weryfikacji hipotez statystycznych. Formuluje dwa przeciwstawne twierdzenia (lub inaczej: dwie wzajemnie wykluczające się hipotezy) dotyczące możliwego wyniku eksperymentu, które wyraża przez wartości średnie (parametry) dwóch populacji  $\mu_0$  i  $\mu_1$ :

Hipoteza zerowa	$H_0: \mu_0 = \mu_1$	Nie dochodzi do zmiany wartości średniej populacji. Zmiana czynnika badanego (temperatury formy) nie pociąga za sobą zmianę czynnika wynikowego (skurczu wyprasek); Następuje zmiana wartości średniej populacji.
Hipoteza alternatywna	$H_1: \mu_0 > \mu_1$	Zmiana czynnika badanego (temperatury formy) skutkuje zmianą czynnika wynikowego (skurczu wyprasek);

Te dwie sytuacje zilustrowano na rysunku 1 (jako  $H_0$  i  $H_1$ ). Aby rozstrzygnąć słusność jednej z hipotez, inżynier posługuje się rozkładem statystyki z próby (tzw. rozkładem referencyjnym), który w tym przypadku tego problemu ma postać rozkładu normalnego  $N(\mu_0, \sigma/\sqrt{n})$ . Wykorzystując ten fakt, inżynier dzieli możliwe do uzyskania wyniki prób doświadczalnych na dwa rozłączne obszary zdarzeń (patrz: rys. 2(a)):

- **typowe**, całkiem prawdopodobne z punktu widzenia hipotezy zerowej  $H_0$  (zachodzące często, gdy faktycznie obowiązuje hipoteza zerowa  $H_0$ );
- **nieprawdopodobne** z punktu widzenia aktualnie obowiązującej hipotezy zerowej  $H_0$ , stąd przemawiające za słusnością hipotezy alternatywnej  $H_1$ ;

Pierwszy obszar jest reprezentowany przez tzw. **przedział akceptacji** testu i jest to zbiór potencjalnych wyników doświadczenia zgodnych z hipotezą zerową  $H_0$ . Jeżeli wynik wartości średniej z próby będzie należeć do tego przedziału, wówczas uznamy, że nie doszło do zmiany wartości skurczu przetwórczego, gdyż takie zdarzenie jest całkiem typowe z punktu widzenia hipotezy  $H_0$ . I przeciwnie, jeżeli wynik wartości średniej będzie zaliczać się do tzw. **obszaru krytycznego** grupującego zdarzenia nieprawdopodobne przy założeniu słusności hipotezy  $H_0$ , wówczas zgodzimy się, że doszło do zmiany skurczu wtórnego pod wpływem zmiany temperatury formy. W konsekwencji odrzucimy hipotezę  $H_0$  i zaakceptujemy hipotezę alternatywną  $H_1$ . W starszych podręcznikach z rachunku prawdopodobieństwa i statystyki matematycznej tok rozumowania prowadzący do wyboru jednej z wzajemnie wykluczających się hipotez nazywano **zasadą praktycznej pewności**. W aktualnych pracach i podręcznikach akademickich przedstawioną procedurę określa się mianem **testowania hipotez statystycznych** lub **wnioskowaniem statystycznym**.

**Przed wykonaniem** jakichkolwiek prac doświadczalnych, znając postać rozkładu prawdopodobieństwa wartości średniej oraz licznosc próby, inżynier wyznacza dwa przedziały: akceptacji i krytyczny. Na mocy przyjętego rozkładu prawdopodobieństwa, przedział akceptacji wyznaczamy jako:

Jeżeli hipoteza  $H_0$  jest prawdziwa, formuła ta wyznacza przedział, w którym wynik wartości średniej z próby wystąpi z prawdopodobieństwem  $1-\alpha$ ;  $\mu_0$  i  $\sigma$  to parametry rozkładu obserwacji indywidualnych, natomiast  $z_{\alpha/2}$  jest zmienną standaryzowaną rozkładu normalnego);

$$P\left(\mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \quad (1)$$

Wartość prawdopodobieństwa  $\alpha$ , która jest przyjmowana w sposób arbitralny, traktuje się jako tzw. **poziom istotności testu statystycznego  $\alpha$**  (patrz dalej: błąd I rodzaju). W naukach technicznych i przyrodniczych za  $\alpha$  przyjmuje się zazwyczaj wartość  $\alpha = 0.05$ . Stąd, prawdopodobieństwo zaobserwowania wyniku wartości średniej w przedziale akceptacji:

$$\mu_0 - z_{0.025} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + z_{0.975} \frac{\sigma}{\sqrt{n}} \quad (2)$$

$$\mu_0 - 1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + 1.96 \frac{\sigma}{\sqrt{n}} \quad (3)$$

wynosi 0.95. Po podstawieniu stosownych inżynier oblicza końce przedziału akceptacji i jednocześnie wyznacza jego dopełnienie – przedział krytyczny.

### Przykład

W przypadku eksperymentu technologicznego związanego z wtryskiwaniem termoplastów, inżynier zna aktualne parametry rozkładu  $\mu$  i  $\sigma$ . Inżynier wybrał licznosc próby  $n = 5$  i na podstawie tych informacji wyznacza przedział akceptacji i przedział krytyczny testu. Po podstawieniu wartości średniej i rozrzutu populacji oraz licznosci próby do formuły (3) otrzymujemy przedział akceptacji testu statystycznego równości średnich :

$$7.3 - 1.96 \frac{0.33}{\sqrt{5}} \leq \bar{x} \leq 7.3 + 1.96 \frac{0.33}{\sqrt{5}}$$

$$7.01 \leq \bar{x} \leq 7.59$$

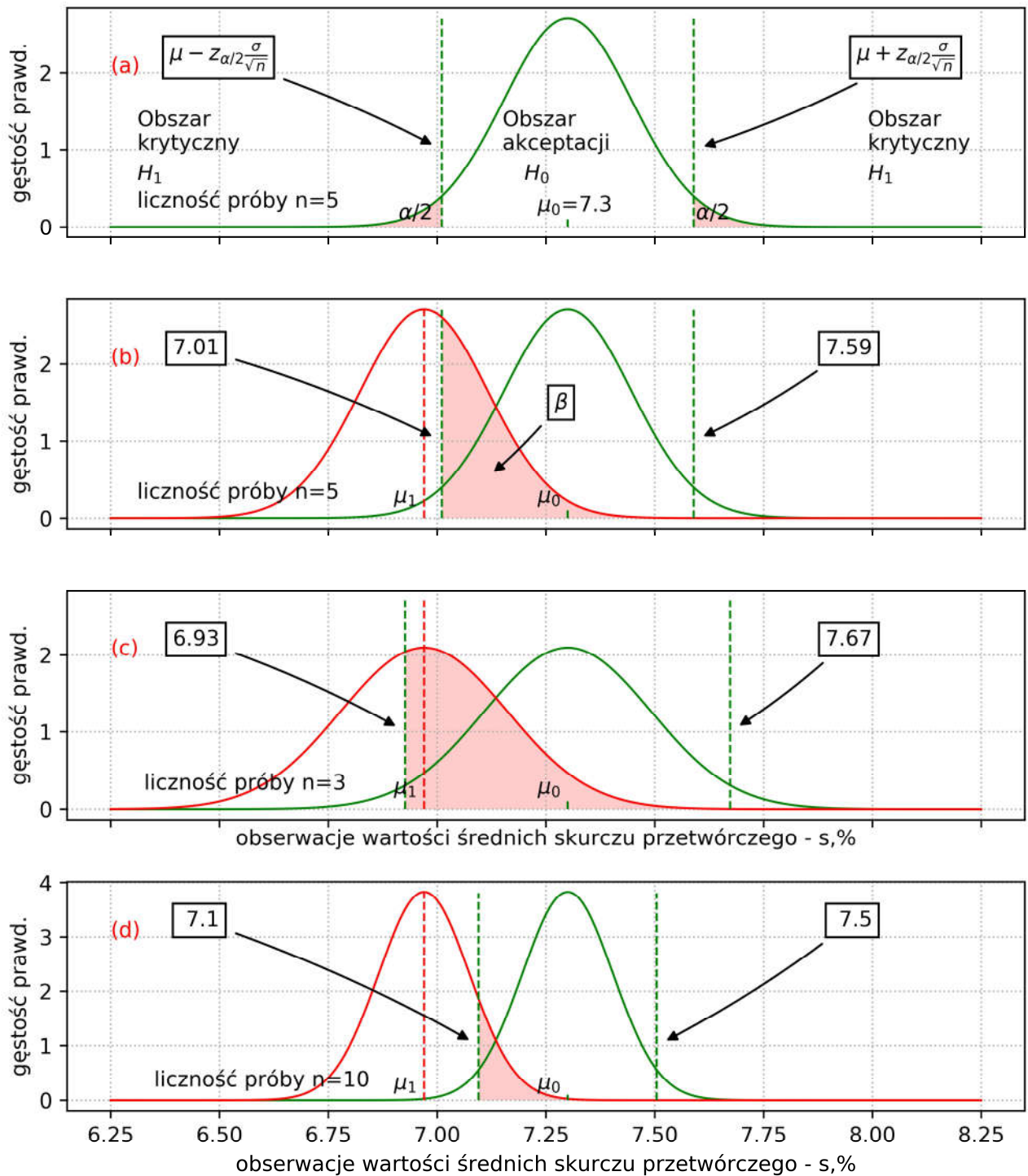
Wynik ten jest poglądowo przedstawiony na rys. 2(b).

### Dyskusja prawdopodobieństwa popełnienia błędu I i II rodzaju oraz mocy testu statystycznego

Korzystając z jakiegokolwiek formy procedury weryfikacji hipotez statystycznych, należy być świadomym, że **każda decyzja** o odrzuceniu lub akceptacji hipotezy zerowej  $H_0$  **jest**, w mniejszym lub większym stopniu **niejednoznaczna**. Niejednoznaczność w podejmowaniu decyzji wynika ze zmienności losowej, której nigdy nie jesteśmy w stanie wyeliminować. Naszą wątpliwość staramy się wyrazić przez przyjęcie pewnego progu prawdopodobieństwa popełnienia pomyłki polegającej na odrzuceniu hipotezy zerowej  $H_0$  przy założeniu, że jest ona prawdziwa. Z tą pomyłką wiążemy tzw. **błąd I rodzaju**, a odpowiadające jej, zakładane prawdopodobieństwo to wspomniany już wcześniej **poziom istotności testu statystycznego  $\alpha$** . Innymi słowy, przeprowadzając weryfikację hipotezy statystycznej, godzimy się na popełnienie błędu I rodzaju. Błąd polega on na przypadkowym odrzuceniu hipotezy zerowej  $H_0$ . Zauważmy, że przyjęta wartość prawdopodobieństwa  $\alpha$  wraz z licznoscią próby  $n$  jest wykorzystana do wyznaczenia przedziału akceptacji i przedziału krytycznego (równanie 2). Przeprowadzając procedurę weryfikacji hipotezy statystycznej na poziomie  $\alpha = 0.05$ , liczymy się z przypadkowym odrzuceniem hipotezy  $H_0$  średnio raz na 20 przeprowadzonych prób doświadczalnych, o ile rzeczywiście, hipoteza  $H_0$  jest prawdziwa.

Weryfikując hipotezę statystyczną, możliwe jest popełnienie jeszcze jednego błędu, który polega na akceptacji hipotezy zerowej  $H_0$  przy założeniu, że jest ona fałszywa. Tą hipotetyczną, zakładaną pomyłkę określa się **błędem II rodzaju**. Rolą tego pojęcia jest wyrażenie trafności podjętej decyzji. Błąd ten, a raczej związane z nim prawdopodobieństwo  $\beta$ , oblicza się na podstawie wyniku próby doświadczalnej. Im mniejsze jest prawdopodobieństwo  $\beta$ , tym konkluzja związana z odrzuceniem hipotezy zerowej  $H_0$  jest silniejsza. Z tego powodu, w praktyce wprowadza się jeszcze jedno prawdopodobieństwo nazywane mocą testu  $P$  (ang. Power). Prawdopodobieństwo  $P = 1 - \beta$  odnosi się do sytuacji, w której prawidłowo odrzucamy hipotezę zerową (tzn. przy założeniu, że hipoteza  $H_0$  jest fałszywa). W praktyce doświadczalnej, szacowanie prawdopodobieństw  $\beta$  lub  $P$  ma sens jedynie wtedy, gdy wynik próby doświadczalnej prowadzi do odrzucenia hipotezy zerowej. Drugim zastosowaniem prawdopodobieństw  $\beta$  lub  $P$  jest budowa krzywych operacyjnych OC (ang. Operating Curve) wykorzystywanych do doboru licznosci próby  $n$  (patrz: dalsza część skryptu). Tabela 1 grupuje możliwe do uzyskania wyniki testów statystycznych wraz z odpowiadającymi im błędami I i II rodzaju. Problem w weryfikacji hipotez statystycznych polega na tym, że nigdy z absolutną pewnością nie rozstrzygniemy, który z przedstawionych wariantów jest prawdziwy. Ostateczna

decyzja jest zawsze obciążona błędem. Rolą osoby prowadzącej badania doświadczalne jest taki dobór liczności próby, by błąd ten zminimalizować do akceptowalnych granic.



Rysunek 2 Wykresy ilustrują rozkłady wartości średniej z próby (statystyki testowej); W szczególności, na rysunku (a) przedstawiono rozkład wartości średniej skurczu przetwórczego dla  $n = 5$  wraz obszarem: akceptacji i krytycznym dla praw. błędu I rodzaju  $\alpha = 0.05$ ; Rysunki (b), (c) i (d) przedstawiają wyniki procedury weryfikacji hipotezy statystycznej o równości średnich  $H_0$  vs.  $H_1$  dla różnych licznosci prób  $n = 3, 5$  i  $10$ . Wartości krytyczne oddzielające obszary akceptacji od obszarów krytycznych zaznaczono zieloną linią przerywaną wraz z przytoczeniem ich wartości; Opracowanie własne;



Tabela 1 Wyniki procedury weryfikacji hipotezy statystycznej wraz z możliwymi do wystąpienia błędami I i II rodzaju

Hipoteza \ Wynik	Przyjęto $H_0$	Odrzucono $H_0$
$H_0$ jest prawdziwa	Właściwa podjęta decyzja	Błędna decyzja Błąd I rodzaju
$H_0$ jest fałszywa	Błędna decyzja Błąd II rodzaju	Właściwie podjęta decyzja

## Przykład

Wróćmy do problemu inżyniera jakości związanego z redukcją wad występujących w procesie wtryskiwania termoplastów i przeprowadźmy dalszą część procedury weryfikacji hipotezy statystycznej o równości średnich. Jak pamiętamy, inżynier zna aktualne warunki prowadzenia procesu, wybrał licznosc próby  $n = 5$ , i na ich podstawie tych informacji wyznaczył przedział akceptacji i przedział krytyczny testu – obszary obowiązywania hipotezy zerowej  $H_0$  i alternatywnej  $H_1$ . Kolejno, inżynier zmienił intencjonalnie poziom czynnika badanego – temperatury formy  $T_f$  (jednej ze zmiennych sterujących procesem wtryskiwania) z 38 na 32°C i po okresie ustabilizowania procesu wyprodukował wypraski w liczbie  $n$  szt., które następnie poddał badaniu. W wyniku oceny skurczu każdej z wyprasek, inżynier obliczył wartość średnią z próby (tzw. statystykę testową). Średni skurcz wyniósł  $\bar{x} = 6.97[\%]$ . Teraz wartość tej statystyki jest porównywana ze wcześniej wyznaczonymi przedziałami. Okazuje się, że wartość średnia należy do przedziału krytycznego testu, stąd inżynier podejmuje decyzję **o odrzuceniu hipotezy zerowej na poziomie istotności  $\alpha = 0.05$  i akceptacji hipotezy alternatywnej**. W nomenklaturze doświadczalnictwa, dla inżyniera oznacza to, że **zmiana czynnika badanego** (tu: temperatury formy) **wywołuje efekt zmiany czynnika wynikowego** (skurczu przetwórczego tworzywa). O takiej zmianie czynnika wynikowego mówimy jako o **znamiennej (istotnej) statystycznie**. Nie oznacza to jeszcze, że ta zmiana musi być **istotna praktycznie**, czyli, że wnosi zamierzony efekt, na którym zależy inżynierowi (patrz sekcja: „Różnica istotna statystycznie a różnica istotna praktycznie”).

Przyjmijmy na chwilę, że wynik testu wynika uzasadnia zmianę temperatury formy, a przez to i redukcję skurczu przetwórczego, a więc zbliżenie się do wartości nominalnej procesu  $T$ . Nie ukrywajmy, że działanie inżyniera jest nieco spontaniczne i przypomina często stosowaną „metodę prób i błędów”. Inżynier tak naprawdę nie wie, jak dużą zmianę temperatury należy wprowadzić do procesu, by osiągnąć zamierzony efekt skurczu tworzywa. Zmiana temperatury formy była przecież dowolna, stąd działanie inżyniera traktujemy raczej jako przypadkowe. Niemniej, po pierwsze, zastosowana technika wnioskowania jest solidną podstawą do rozwinięcia bardziej metodycznych działań rozwiązujących problem doboru nastaw procesu (planowanie doświadczeń czynnikowych). Po drugie, wykazano, że temperatura tworzywa jest faktycznie czynnikiem mającym wpływ na skurcz przetwórczy i jest to **cenna wiedza technologiczna**.

## Szacowanie prawdopodobieństwa błędu II rodzaju i mocy testu statystycznego

Na ile trafna jest decyzja podjęta przez inżyniera? Przecież o takim, a nie innym wyniku eksperymentu mógł równie dobrze zdecydować przypadek (patrz: błąd I rodzaju), a nie intencjonalna zmiana wprowadzona do procesu. Rzeczywiście, taki wynik jest całkiem możliwy do uzyskania, i jak zauważono wcześniej, szansa na odrzucenie hipotezy  $H_0$  wynosi  $\alpha = 0.05$ . W praktyce, aby ocenić na ile jest zasadna decyzja o odrzuceniu hipotezy zerowej, zazwyczaj szacuje się moc testu statystycznego lub wyznacza się tzw. wartość prawdopodobieństwa  $p$  odpowiadającą wynikowi statystyki testowej (ang. p-Value). Obydwa prawdopodobieństwa, charakteryzują trafność podjętej decyzji (**ze wskazaniem na moc testu** – zdaniem autora skryptu). Jeżeli moc testu przyjmuje bardzo duże wartości prawdopodobieństwa, jesteśmy skłonni uważać, że uzyskany wynik testu jest wiarygodny. Podobnie, jeżeli obliczona wartość prawdopodobieństwa  $p$  przyjmuje zdecydowanie niższe wartości niż zakładany poziom istotności testu  $\alpha$  - mamy zaufanie do wyniku testu.

## Przykład

Obliczmy moc testu statystycznego  $P$  oraz wartość prawdopodobieństwa  $p$  wcześniej przeprowadzonego testu statystycznego równości średnich. Posłużymy się pomocniczym rysunkiem 2(b) przedstawiającym wynik testu. Ilustracja zawiera już nie jedną, lecz dwie krzywe rozkładu normalnego. Jedna krzywa  $N(\mu_0, \sigma/\sqrt{n})$ , tak jak poprzednio reprezentuje hipotezę zerową testu  $H_0$ , druga  $N(\mu_1, \sigma/\sqrt{n})$  odpowiada warunkom wynikającym z przeprowadzonego testu i dotyczy hipotezy alternatywnej  $H_1$ . Zakładamy, że obydwa rozkłady charakteryzują się tym samym rozrzutem  $\sigma/\sqrt{n}$ , a wartość

średnia wyniku próby jest zarazem wartością średnią rozkładu prawdopodobieństwa  $\mu_1 = 6.97$ . Moc testu jest prawdopodobieństwem prawidłowego odrzucenia hipotezy zerowej, dlatego zakłada, że obowiązującym rozkładem jest ten związany z hipotezą alternatywną. W myśl definicji mocy testu:  $P = 1 - \beta$ , jednym ze sposobów wyznaczenia tego prawdopodobieństwa jest scalkowanie krzywej rozkładu normalnego  $N(\mu_1, \sigma/\sqrt{n})$  w przedziale, gdzie obowiązuje hipoteza zerowa  $H_0$  i odjęcie tego wyniku od jedności.

Moc testu statystycznego średnich jest liczona w przedziale akceptacji przy założeniu, że hipoteza alternatywna jest prawdziwa (wartość średnia wynosi 6.97)...

$$1 - \beta = 1 - P\left(\mu_0 - z_{\alpha} \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu_0 + z_{\alpha} \frac{\sigma}{\sqrt{n}} \middle|_{H_1}\right)$$

... stąd otrzymujemy ...

$$1 - P(7.01 \leq \bar{x} \leq 7.59 |_{H_1}) = 1 - 0.3912$$

... i ostatecznie moc testu wynosi:

$$P = 0.6088$$

W rozważanym przykładzie, interpretacja mocy testu jest następująca. Jeżeli wartość średnia wyniku z próby jest faktycznie parametrem rozkładu prawdopodobieństwa  $N(\mu_1, \sigma/\sqrt{n})$ , odpowiadającym nowym warunkom prowadzenia procesu  $H_1$ , to wielokrotne odtwarzanie eksperymentu w ok. 60 % będzie prowadziło do odrzucenia hipotezy zerowej. Wprawdzie wynik wartości średniej doprowadził do odrzucenia hipotezy zerowej, lecz nie jest przekonujący – średnio co trzecia próba może prowadzić do akceptacji hipotezy zerowej.

## Liczność próby a moc testu statystycznego

Bardzo ważnym aspektem weryfikacji hipotez statystycznych jest dobór licznosci próby. Uważnie przyjrzyjmy się formułom (1 ÷ 3). Zauważamy, że prawdopodobieństwo prawidłowego odrzucenia hipotezy zerowej tj. moc testu, jest tym większa, im większa jest licznosc próby (patrz: czynnik  $1/\sqrt{n}$ ). Jeżeli licznosc próby  $n$  próby będzie zwiększana systematycznie do bardzo dużych wartości, teoretycznie rzecz biorąc, jesteśmy w stanie wykazać każdą, nawet najmniejszą różnicę między wartościami średnimi  $\delta = \mu_1 - \mu_0$ . Nie jest to pomyślnie stwierdzenie, gdyż pamiętajmy, licznosc próby z założenia musi być wybrana przed rozpoczęciem badań doświadczalnych. Ta uwaga prowadzi do zaskakującej konkluzji: **wynik testu statystycznego jest w pewnych sytuacjach zależny od wyboru licznosci próby! Badając ten sam proces z wykorzystaniem dwóch różnych licznosci prób, możemy otrzymać dwa odmienne wyniki testu statystycznego.**

## Przykład

Spróbujmy jeszcze raz przeprowadzić procedurę weryfikacji hipotezy statystycznej z przykładu 2, wybierając inną niż wcześniej licznosc próby. Dla przykładu, niech licznosc wynosi  $n_2 = 3$ , przy zachowaniu poprzedniego wyniku wartości średniej  $\bar{x} = 6.97\%$ , i odchylenia standardowego  $\sigma = 0.33$  oraz  $\alpha = 0.05$ . Studentowi zaleca się wykonanie obliczeń związanych z przeprowadzeniem testu statystycznego oraz podjęciem decyzji o ostatecznym wyniku procedury.

Wynik testu może okazać się być zaskakujący. Próba o licznosci  $n_2 = 3$  prowadzi do akceptacji hipotezy zerowej, a więc nie daje podstaw do stwierdzenia, że średnie różnią się istotnie statystycznie na przyjętym poziomie prawdopodobieństwa  $\alpha = 0.05$ . Czyżby procedura weryfikacji hipotez statystycznych nie dawała jednoznacznych, wiarygodnych wyników?

## Różnica istotna statystycznie a różnica istotna praktycznie

Istota tego problemu tkwi w pojęciu praktycznej różnicy wartości średnich. Przeprowadzając próbę doświadczalną inżynier chciałby poznać, czy zmiana warunków prowadzenia procesu będzie skutkować zmianą widoczną w pomiarze wartości średniej. Wprawdzie przed wykonaniem eksperymentu inżynier **nie zna** prawdziwej różnicy w wartościach średnich  $\delta$ , odpowiadających hipotezom  $H_0$  i  $H_1$ , ale może za to **przyjąć** tzw. **różnicę praktyczną średnich  $\delta$** , którą chciałby wykazać za pomocą testu statystycznego. Ta różnica to tzw. **efekt**. Wiedząc jaki efekt  $\delta$  ma wykazać test statystyczny i jednocześnie wybierając (wysoką) moc testu  $P$ , inżynier może świadomie wybrać licznosc próby. Jeżeli rzeczywista różnica w wartościach średnich jest większa lub równa przyjętej

wartości  $\delta$ , to test statystyczny powinien wykazać ten efekt z mocą testu P. Dla przykładu, jeżeli za moc testu P przyjmiemy duże wartości prawdopodobieństwa np. rzędu 0.95, wówczas mamy dużą szansę na poprawne odrzucenie hipotezy zerowej  $H_0$ , o ile rzeczywiście  $H_0$  jest fałszywa. Takie podejście jest pomocne w rozwiązaniu problemu niejednoznaczności wyniku testu statystycznego zasygnalizowanego w poprzednim przykładzie. W praktyce doświadczalnej, aby zminimalizować szansę popełnienia błędu w testowaniu hipotez statystycznych, posługujemy się tzw. krzywymi operacyjnymi OC (ang. Operating Curve).

Krzywe operacyjne OC to wykresy przedstawiające zależność: mocy testu P w funkcji zmiennej opisującej (i) relatywną zmianę parametru poddanego testowi statystycznemu oraz (ii) liczności próby. W przypadku testu wartości średnich jest to różnica między średnimi  $\delta$ . Krzywe OC umożliwiają obiektywny, optymalny wybór liczności próby doświadczalnej.

Na rysunku 3 przedstawiono tzw. krzywe operacyjne OC wykorzystywane do przyjmowania liczności próby w teście wartości średnich. Oś rzędnych wykresu reprezentuje praktyczną różnicę średnich tzn. efekt, który zamierzamy wykazać. Aby krzywe OC mogły być wykorzystane do każdej sytuacji badawczej (tu: dotyczącej testu różnicy wartości średnich), efekt ten jest wyrażony w jednostkach odchylenia standardowego (4).

*Różnica praktyczna d wyrażona w jednostkach odchylenia standardowego, gdzie  $\delta$  to zakładany efekt jaki ma wykazać test statystyczny, a  $\sigma$  to znane/zakładane odchylenie standardowe;*

$$d = \frac{\mu_1 - \mu_0}{\sigma} = \frac{\delta}{\sigma} \quad (4)$$

Z kolejną osi odciętych wykresu 3 zaznaczono prawdopodobieństwo popełnienia błędu II rodzaju  $\beta$ , lub moc testu P, gdy prawdopodobieństwo jest odczytywane w stronę niższych wartości ( $P = 1 - \beta$ ). Przed przystąpieniem do realizacji badań doświadczalnych należy przyjąć efekt  $\delta$ , jaki ma wykryć test statystyczny, a następnie i wyrazić go w jednostkach odchylenia standardowego d (równanie 4). Dodatkowo, należy zdecydować się na wybór wartości mocy testu (lub wartość  $\beta$ ). Tu, sugeruje się wysoką wartości prawdopodobieństwa P (lub niską wartość błędu  $\beta$ ). Punkt o współrzędnych: (d, P) lub (d,  $\beta$ ) wykresu wskazuje najbliższą położoną krzywą opatrzoną indeksem liczności próby n.

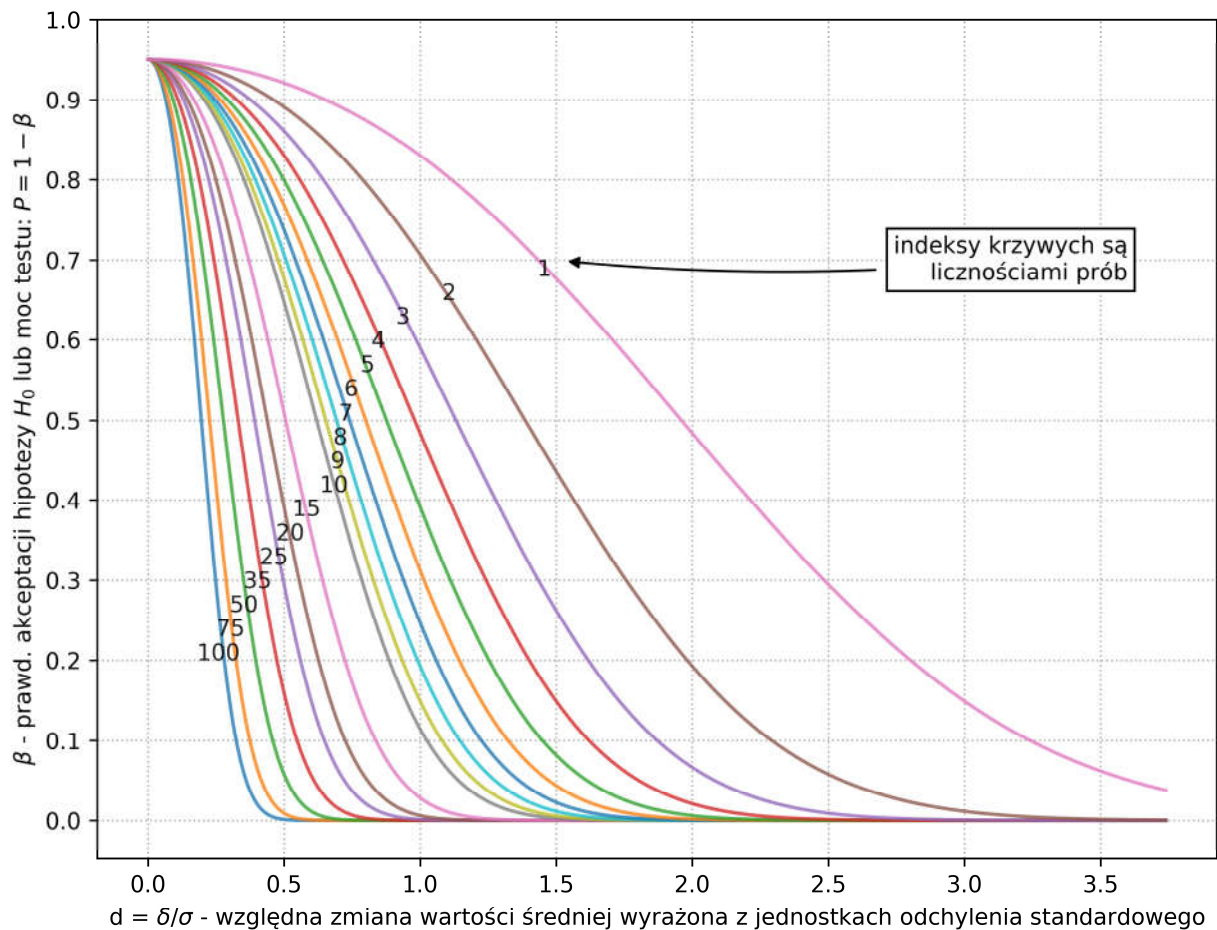
**Zaznaczmy jedną istotną uwagę.** Prowadząc badania doświadczalne bez wcześniejszego doboru (optymalnej) liczności próby decydujemy się przypadkowy wynik procedury weryfikacji hipotezy statystycznej. **W wyniku wyboru zbyt dużej liczności próby możemy wykazać statystyczną istotność różnic między średnimi, pomimo że wykazana różnica może wcale nie mieć żadnego praktycznego znaczenia.** Oczywiście, możliwa jest i odwrotna sytuacja. **Gdy liczność próby jest zbyt mała, test statystyczny może nie wykazać praktycznej różnicy średnich, pomimo, iż faktycznie taka różnica ma miejsce.**

Z tej dyskusji wynika następujący wniosek:

***Liczność próby doświadczalnej nie może być ani zbyt wysoka, ani zbyt niska, lecz optymalna i zgodna z przyjętymi założeniami testu (tu: przyjętą różnicą oraz mocą testu).***

## Wskazówki praktyczne dotyczące realizacji badań doświadczalnych

Z uwagi na to, że opracowanie zawiera wiele wątków teoretycznych i praktycznych, więc jako podsumowanie zbierzmy wszystkie uwagi zawarte w skrypcie i przedstawmy spójną procedurę doświadczalną mającą na celu opracowanie wyniku eksperymentu porównawczego wraz z dodatkowym przykładem.



Rysunek 3 Weryfikacja hipotezy statystycznej o równości średnich – porównanie wartości średniej z wartością referencyjną - krzywe operacyjne OC wykorzystywane do doboru licznosci próby. Oś odciętych reprezentuje zmianę wartości średniej; Oś rzędnych przedstawia prawdopodobieństwo akceptacji hipotezy zerowej  $H_0$  (lub moc testu  $1 - \beta$ ); Indeksy krzywych odpowiadają licznosci próby; Opracowanie własne;

1. **Sformułowanie problemu.** Wróćmy do sedna problemu przedstawionego w pierwszym przykładzie. Przypomnijmy, że inżynierowi zależy na zmianie wartości skurczu przetwórczego ogółu wyprasek, tak by zredukować liczbę niezgodności/wad procesu wtryskiwania. Inżynier przypuszcza, że obniżenie temperatury formy wtryskowej przyniesie korzystną zmianę – redukcję skurczu, a więc zmianę w kierunku wartości nominalnej  $T$ . W tym celu zamierza zweryfikować swój pogląd. Celem poznawczym eksperymentu jest wykazanie związku przyczynowo-skutkowego między skurczem przetwórczym  $s$  a temperaturą formy  $T_f$ . Celem praktycznym eksperymentu jest redukcja wad występujących w procesie.

2. **Wybór licznosci próby.** Inżynier przygotowuje się do eksperymentu. Znając:

- aktualne parametry rozkładu normalnego obserwacji skurczu przetwórczego:  $\mu_0 = 7.3$  oraz  $\sigma = 0.33$ ;
- granice specyfikacji: ( $LSL = 4$ ,  $USL = 8$ ) wraz z nominalnym  $T = 6$ ;

inżynier ocenia aktualny stan procesu. Z całkowania krzywej rozkładu normalnego  $N(\mu_0, \sigma)$  wynika, że szacunkowa liczba wad wynosi ok. 1700 PPM (ang. Parts Per Million – liczba wad przypadająca na 1 milion potencjalnie wyprodukowanych produktów). Ta wartość jest niedopuszczalna. W pierwszym kroku doskonalenia procesu, inżynier zamierza zredukować tę liczbę do ok. 1000 PPM. Przyjmując założenie, że wartość  $\sigma$  nie zmieni się podczas prowadzenia prac doświadczalnych oraz

wykorzystując właściwości zmiennej standaryzowanej  $z$  - inżynier oblicza różnicę praktyczną  $\delta$  (różnica między aktualną wartością średnią  $\mu_0$  a wartością średnią  $\mu_1$  dla której wadliwość wynosi 1000 PPM). Ta różnica to efekt  $\delta$  jaki zamierza wykazać eksperymentalnie inżynier. Znając ten efekt i aktualną wartość rozrzutu skurczu przetwórczego, inżynier wyznacza  $d$  - względną zmianę wartości średniej w jednostkach odchylenia standardowego – formuła (8). Inżynierowi zależy na wysokiej wiarygodności wyniku eksperymentu, stąd za moc testu przyjmuje prawdopodobieństwo popełnienia pomyłki równe  $\beta = 0.1$  (lub moc testu  $P = 1 - \beta = 0.9$ ). Innymi słowy, jeżeli rzeczywista różnica w wartościach średnich wynosi co najmniej  $\delta$ , to test statystyczny powinien to wykazać z prawdopodobieństwem rzędu 0.9. Ta zakładana różnica wynosi  $\delta = \sigma = 0.3$  ( $d = 1$ ). Gdyby udało się zmienić aktualny poziom wartości średniej do  $\mu_1 = \mu_0 - \sigma$ , wówczas liczba produkowanych wad wyniesie ok. 900, co jest bliskie przyjętym założeniom. To jest cel minimum, jaki chciałby osiągnąć inżynier. Dysponując krzywymi OC (patrz: rys. 3), widzimy, że aby wykazać taki efekt w eksperymencie z błędem  $\beta < 0.1$  (lub mocą testu większą niż 0.9) należy wybrać licznosc próby większą lub równą  $n = 10$ .

**komentarz:** Nawiązując do licznosci  $n = 5$  rozważanej we wcześniejszym przykładzie, pamiętamy, że wynik pierwszego testu był wprawdzie pozytywny, wykazano różnicę w wartościach średnich, ale jednocześnie był mało wiarygodny – moc testu wyniosła tylko 0.6088 (lub ryzyko popełnienia błędu II rodzaju  $\beta$  aż: 0.3912). Z kolei, w kolejnym przykładzie, gdy za licznosc wybrano  $n = 3$  test nie wykazał zamierzonego efektu z uwagi na niską moc testu (patrz: krzywe OC – rys. 3). Jakkolwiek, z krzywych operacyjnych OC widzimy, że licznosc próby  $n = 3$  lub 5 wystarcza w zupełności do wykrycia dużych efektów zmian wartości średniej (np.  $\delta = 2\sigma$  z mocą wyższą niż 0.9). Problem w tym, że tak mała licznosc próby nie pozwala wiarygodnie wykryć mniejszych efektów rzędu np.  $\delta = \sigma$ , co utrudnia podjęcie wiarygodnych decyzji o ew. zmianach w procesie.

**3. Realizacja badań doświadczalnych i podjęcie decyzji o ew. zmianie w procesie.** Inżynier przeprowadza próbę doświadczalną. Zmienia wartość temperatury formy z aktualnej  $T_f = 38^\circ\text{C}$  na  $32^\circ\text{C}$  i produkuje wypraski w liczbie  $n = 10$ . Następnie, mierzy wartość skurczu przetwórczego wyprasek i oblicza wartość średnią, która wynosi  $\bar{x} = 6.97\%$ . Przeprowadza procedurę weryfikacji hipotezy statystycznej o równości średnich przedstawioną w przykładzie 2. Na podstawie jej wyniku podejmuje decyzję o zmianie temperatury formy do  $T_f = 32^\circ\text{C}$ . Rolą studenta jest przeprowadzenie procedury weryfikacji hipotezy o równości średnich wraz z wyznaczeniem mocy testu (lub odczytaniem z krzywych OC). Rozwiązanie problemu przedstawiono na rys. 2d.

### Ćwiczenie 3.1 Weryfikacja hipotezy statystycznej o równości średnich

Proces nanoszenia warstwy ochronnej materiału termoutwardzalnego na powierzchnię betonu jest prowadzony w warunkach (temperatura powierzchni)  $T_p = 12^\circ\text{C}$  oraz (temperatura materiału)  $T_m = 22^\circ\text{C}$ , co odpowiada średniej wydajności procesu  $w_k = 1.6 \text{ m}^2/\text{l}$  i średnim rozrzucie obserwacji  $\sigma_{wk} = 0.1 \text{ m}^2/\text{l}$  (rysunek 4). Operator procesu postanowił zmienić warunki prowadzenia procesu zwiększając temperaturę ogrzewania materiału termoutwardzalnego do wartości  $T_m = 34^\circ\text{C}$  - sądząc, że zastosowana zmiana wpłynie na zwiększenie wydajności procesu.

Pomiar wydajności procesu w nowych warunkach ( $T_p = 12^\circ\text{C}$ ,  $T_m = 24^\circ\text{C}$ ) wyniósł  $w_k = 1.68 \text{ m}^2/\text{l}$  przy tym samym rozrzucie obserwacji  $\sigma_{wk} = 0.1 \text{ m}^2/\text{l}$ . Licznosc próby na podstawie której oszacowano wydajność procesu nanoszenia warstw wierzchnich wynosiła  $n = 9$  obserwacji.

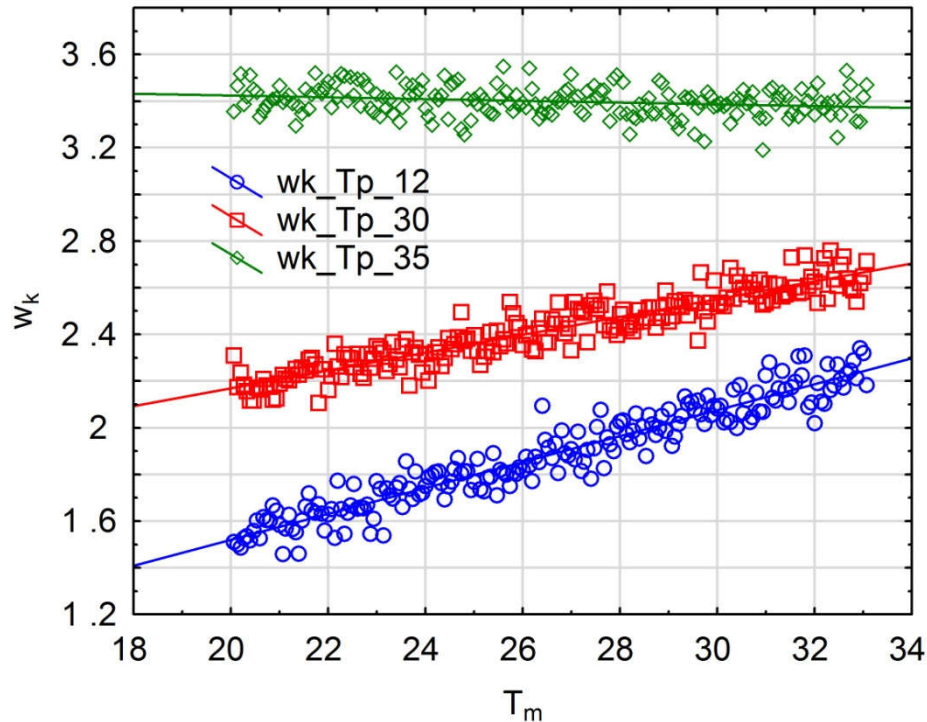
- (a) Przeprowadź test statystyczny weryfikujący osąd operatora procesu, dotyczący poprawy wydajności prowadzenia procesu.

$H_0: \mu = 1.6$  na rzecz hipotezy alternatywnej:

$H_1: \mu \neq 1.6$

wybierając poziom istotności testu równy  $\alpha = 0.05$ ;

- (b) Ile wynosi wartość prawdopodobieństwa  $P$  przeprowadzonego testu.  
 (c) Jaka jest wartość błędu drugiego rodzaju  $\beta$  ?  
 (d) Jaki będzie wynik testu statystycznego, jeżeli licznosc próby wynosiłaby  $n = 3$  ?  
 (e) Ile wyniosłaby wymagana licznosc próby  $n$ , gdyby bład  $\beta$  miał być ograniczony warunkiem:  $\beta < 0.1$ . Załóż, że bład I rodzaju wynosi  $\alpha=0.05$ , a zmiana wartości średniej wynosi  $3\sigma_{wk}$  (poprawa wydajności o ok. 20%);



Rysunek 4 Wydajność procesu nanoszenia warstw wierzchnich w funkcji czynników badanych: temperatury materiału  $T_m$  i temperatury powierzchni  $T_p$ .

## Pytania kontrolne

1. Na czym polega procedura weryfikacji hipotezy statystycznej?
2. Na czym polega popełnienie błędu I rodzaju? Czym jest poziom istotności testu statystycznego?
3. Jak oszacować bład II rodzaju. Co to jest moc testu statystycznego?
4. Jaką rolę pełnią statystyki testowe/referencyjne w procedurze weryfikacji hipotezy statystycznej?
5. Czym są krzywe operacyjne OC i jaka jest ich rola w testowaniu hipotez statystycznych?
6. Czym jest różnica praktyczna, a czym różnica statystycznie istotna/znamienna?
7. Jaką rolę pełni tzw. obszar krytyczny statystyki testowej w procedurze weryfikacji hipotez statystycznych?