

Ćwiczenie 2K-1

STATYSTYCZNA OCENA I WIZUALIZACJA WYNIKÓW OZNACZANIA Fe W ŻYWNOŚCI STAŁEJ Z UŻYCIEM PROGRAMU EXCEL

Wykonanie ćwiczenia

Wizualizacja wyników

Tworzenie czytelnych tabel i wykresów z prawidłowo opisanymi osiami, jednostkami, legendą i tytułem

Wpisz dane pomiarowe do tabeli

- nagłówki kolejnych kolumn: Nazwa próbki, Stężenie, SD stężenia, CV, Rozcieńczenie 1, Rozcieńczenie 2, Odważka, Zawartość w suchej masie, SD zawartości, Zawartość odniesienia, Odzysk;
- długie nazwy nagłówków podziel na dwa wiersze: zaznacz kursorem pozycję podziału →ALT + ENTER;
- policz SD ze średniej stężeń i CV; Użyj funkcji =MODUŁ.LICZBY aby uzyskać wartości dodatnie SD;
- policz zawartość i SD zawartości, odejmując od każdego zmierzonego stężenia dla próbek i CRM uśrednione stężenie próbki ślepej;
- ustaw formatowanie czcionki nagłówków na **bold**; wpisz odpowiednie jednostki w nawiasach kwadratowych;
- wstaw linie poziome na górze i na dole nagłówków i linię poziomą pod ostatnim wierszem tabeli;
- wyśrodkuj wszystkie komórki w tabeli;
- ustaw taką samą liczbę miejsc dziesiętnych w danej kolumnie: stężenia – 4, SD – 4, CV – 1;

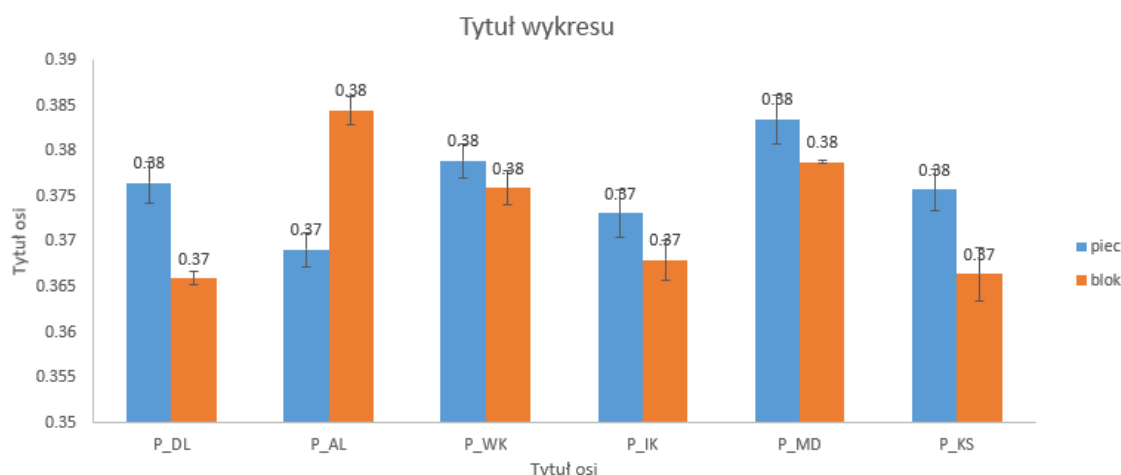
Wykres kalibracyjny

- wykres punktowy z 4 seriami pomiarów wzorców kalibracyjnych – porównanie czułości w kolejnych seriach pomiarowych; wskazanie wartości odstających
- do każdej linii trendu dodaj równanie krzywej na wykresie i wartości R-kwadrat
- Jaki wpływ na czułość i liniowość ma obecność wartości odstającej? Jak się zmieniają te parametry po usunięciu wartości odstającej?

Wykres kolumnowy

- zawartości Fe w próbkach z pieca mikrofalowego wstaw w osobnych seriach na wykresie kolumnowym;
- na osi x: nazwy próbek (nazwa produktu i numer lub inicjały lub nr laboranta); dodaj tytuł osi;
- na osi y: stężenie próbki; dodaj tytuł osi i jednostkę;
- usuń poziome i pionowe linie siatki w tle wykresu;
- dodaj etykiety danych do obu serii;
- dodaj legendę;
- słupki błędów: policz SD ze średniej stężeń i CV; Użyj funkcji =MODUŁ.LICZBY aby uzyskać wartości bezwzględne SD; Nanieś słupki błędów na bazie SD dla każdej serii pomiarowej;
- tytuł wykresu: „Stężenie Fe w mleku po mineralizacji”; ustaw 2 miejsca po przecinku w etykietach;

- ustaw początek osi y na 0 lub na wyższej wartości; Jak się zmienia czytelność wykresu?



Wykres skrzynka-wąsy (box-plot)

- zaznacz wartości średnie stężenia Fe dla próbek mleka po mineralizacji w piecu;

- wstaw wykres skrzynka-wąsy;

- zmień nazwę serii 1: „Piec”;

- zmień kolor wypełnienia skrzynki na jaśniejszy; zmień grubość wąsów na większą;

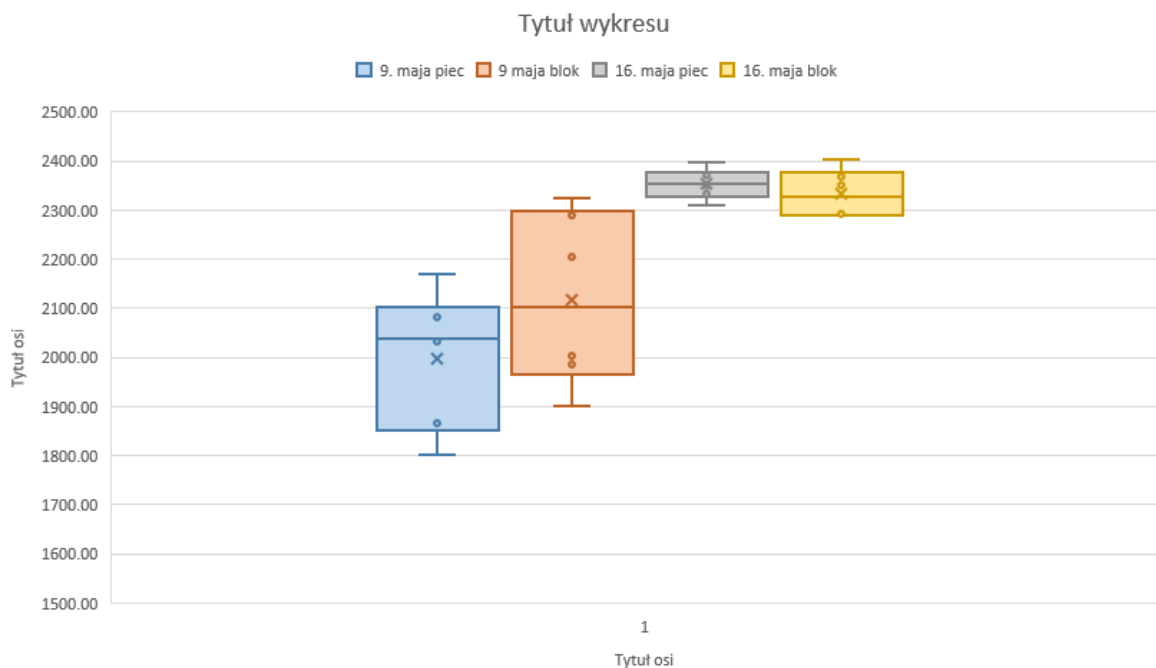
- formatowanie serii danych: pokaż punkty wewnętrzne;

- na osi y: stężenie próbki; dodaj tytuł osi i jednostkę;

- legenda: seria 1 „Piec”, seria 2 „Blok”;

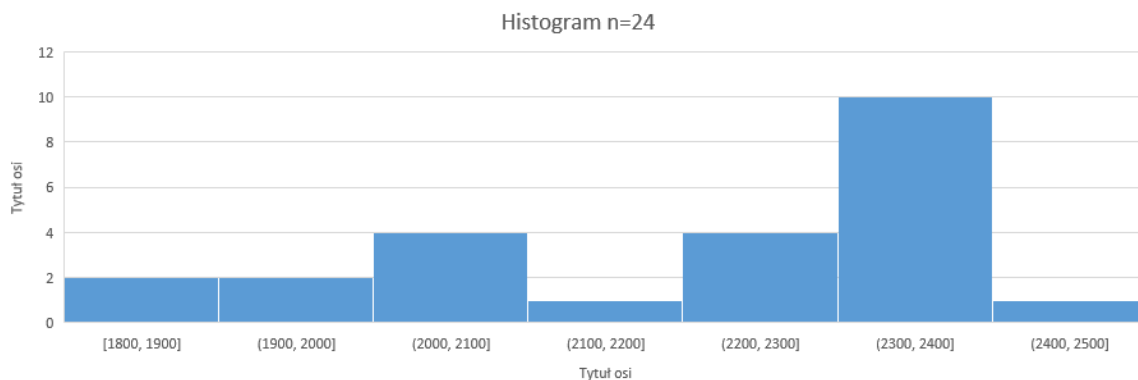
- ustaw minimum osi y na 0 lub na 1500; Jak się zmienia czytelność wykresu?

- Czy dodatkowe rozcieńczenie dla próbek, które przekroczyły zakres liniowości krzywej kalibracyjnej miało wpływ na wynik?



Histogram

- zaznacz wartości średnie stężenia Fe dla wszystkich próbek;
- wstaw wykres-histogram
- ustaw szerokość przedział; formatowanie liczb osi x: 0 miejsc dziesiętnych
- Jakim rozkładem charakteryzują się dane?



Testy statystyczne

Testy istotności stanowią ważny element kontroli jakości pomiarów chemicznych. Pozwalają one na podstawie wyników próby losowej podjąć decyzję o odrzuceniu hipotezy, którą się sprawdza, bądź stwierdzić brak podstaw do odrzucenia tej hipotezy.

Nadawanie nazw komórkom

Przed wykonaniem testów statystycznych nadaj nazwy seriom danych i ich wartościom średnim i SD, aby ułatwić pracę z formułami w Excelu. Dla każdej serii danych zawartości w suchej masie próbki żywności oblicz średnią =**ŚREDNIA** i odchylenie standardowe =**ODCH.STANDARD.PRÓBK**I lub =**ODCH.STANDARDOWE**, w seriach:

Seria	Nazwa serii (6 komórek)	Nazwa komórki ze średnią serii	Nazwa komórki z odchyleniem stand. serii
Mleko1	wynikiMleko1	średniaMleko1	sdMleko1
Mleko2	wynikiMleko2	średniaMleko2	sdMleko2
Mleko3	wynikiMleko3	średniaMleko3	sdMleko3

Można również nadać nazwę komórce z liczbą powtórzeń pojedynczej serii a także komórkom z wartościami odniesienia dla próbki żywności i CRM.

Aby nadać nazwę pojedynczej komórce albo serii komórek, zaznacz je, i wybierz jedną z opcji z rysunków poniżej: wpisz nazwę w okienku po lewej stronie okna programu lub w zakładce Formuły wybierz Definiuj nazwę.

wyniki9majaPIEC	
H	I
1	
odważka [g]	zawartość [mg/kg]
2	
3	
4	
5	0.2 2080.42
6	0.2 2167.08
7	0.2 1865.42
8	0.2 2039.17
9	0.2 2031.67
10	0.2 1799.58
11	0.2 985.83
12	0.2 686.67
13	średnia: 1997.22
14	SD: 137.9519508

H	I
odważka [g]	zawartość [mg/kg]
0.2	2080.42
0.2	2167.08
0.2	1865.42
0.2	2039.17
0.2	2031.67
0.2	1799.58
0.2	985.83
0.2	686.67
średnia:	1997.22
SD:	137.9519508

Nazwa nie może zaczynać się od liczby, nazwa powinna mieć ponad 4 znaki długości, niedozwolone są spacje lub znaki poza podkreśleniem „_”

Test Q-Dixona – identyfikacja błędu grubego w serii danych

- Zastosuj test Q-Dixona, osobno do każdej z 3 serii danych – zawartości w suchej masie próbek mleka w proszku.

Test Q-Dixona wymaga, aby dane były posortowane rosnąco, jednakże Excel umożliwia wskazanie najniższych i najwyższych wartości w serii danych bez potrzeby ich sortowania. Zastosuj do tego celu funkcję =MIN.K i =MAX.K w następujący sposób, np. dla serii 9 maja piec:

=MIN.K(wynikiMleko1 ; 1) - najmniejsza wartość w serii – x_1

=MIN.K(wynikiMleko1 ; 2) - druga najmniejsza wartość w serii – x_2

=MAX.K(wynikiMleko1 ; 1) - największa wartość w serii – x_n

=MAX.K(wynikiMleko1 ; 2) - druga największa wartość w serii – x_{n-1}

Wyniki powyższych funkcji można podstawić do wzoru na test Q-Dixona:

☛ obliczyć wartość rozstępu R zgodnie ze wzorem

$$R = x_n - x_1 \quad (1)$$

☛ obliczyć wartość parametrów Q_1 i Q_n wg wzorów:

$$Q_1 = \frac{x_2 - x_1}{R} \quad (2)$$

$$Q_n = \frac{x_n - x_{n-1}}{R} \quad (3)$$

Uzyskane parametry Q_1 i Q_n porównać z wartościami krytycznymi w tabelach statystycznych.

Zastosuj formatowanie warunkowe → Reguły wyróżniania komórek → Większe niż... aby podświetlić na czerwono parametry przekraczające wartość krytyczną.

Test F-Snedecora – porównanie rozrzutu wyników dwóch serii

Test F-Snedecora jest stosowany w celu porównania wariacji dla dwóch zbiorów wyników. Zastosuj test F dla par danych wybranych próbek mleka w proszku.

W przypadku gdy zbiory wyników są równoliczne, wzór upraszcza się do postaci:

$$F = \frac{\frac{n_1}{n_1 - 1} SD_1^2}{\frac{n_2}{n_2 - 1} SD_2^2} \quad (4) \qquad F = \frac{SD_1^2}{SD_2^2} \quad (5)$$

F – parametr testu; n_1, n_2 – liczność zbiorów;

SD_1, SD_2 – odchylenia standardowe dwóch zbiorów;

$$f_1 = n_1 - 1, f_2 = n_2 - 1 \quad (6)$$

f_1, f_2 – liczba stopni swobody dla dwóch zbiorów

Wartości SD należy dobrać tak, aby w liczniku znalazła się wartość większa.

- oblicz wartość p

Wartość p, p-wartość, prawdopodobieństwo testowe (ang. p-value, probability value) – prawdopodobieństwo, że zależność jaką zaobserwowano w losowej próbie z populacji mogła wystąpić przypadkowo, wskutek losowej zmienności prób, choć w populacji wcale nie występuje. Jest to narzędzie służące jedynie do podstawowej kontroli błędów, i świadczy o wartości dowodowej danych jedynie pośrednio.

Funkcja **=F.TEST(seria1 ; seria2)** oblicza wartość prawdopodobieństwa dla założenia, że wariancje obu porównywanych serii pomiarowych nie różnią się w sposób statystycznie istotny (gdy $p > 0,05$).

- wyszukaj parametr F krytyczny z tablic statystycznych

- oblicz parametr F krytyczny z funkcji **=ROZKŁ.F.ODWR.PS(0,05 ; f1 ; f2)**

Zastosuj formatowanie warunkowe → Reguły wyróżniania komórek → Większe niż... aby podświetlić na czerwono parametry przekraczające wartość krytyczną.

- zastosuj dodatek Analiza danych → Test F: z dwiema próbami dla wariancji

Test t-Studenta – porównanie średnich dwóch serii pomiarowych

Podstawowym zadaniem tego testu jest porównanie wartości średniej arytmetycznej wyników pomiaru ze znaną wartością (odniesienia/rzeczywistą/oczekiwaną/certyfikowaną). W analizie chemicznej test t-studenta stosuje się, m.in. żeby ocenić, czy wyniki uzyskane przez danego analityka, laboratorium czy urządzenie pomiarowe obarczone są błędem systematycznym. Innym zastosowaniem testu t-studenta jest porównanie serii wyników pomiaru dla próbki uzyskanych przez różnych laborantów, na różnych instrumentach lub przy zmiennych warunkach procedury analitycznej.

Hipoteza zerowa (H_0) w tym teście statystycznym zakłada, że *istniejąca różnica między porównywalnymi wartościami (wartość rzeczywista vs wartość otrzymana) jest tylko wynikiem występowania błędów losowych.*

Aby stwierdzić, czy różnica między wartością średnią wielkości mierzonej (x_{sr}) a wartością odniesienia (x_{cert}) jest statystycznie istotna, korzystamy z wyrażenia (gdy test F-Snedecora nie wykazał istotnych różnic wariancji między seriami):

$$t_{eksp} = \frac{|x_{sr1} - x_{sr2}|}{\sqrt{SD_1^2 \cdot SD_2^2}} \sqrt{n} \quad (7)$$

Gdzie x_{sr} – wartość średnia serii pomiarowej, x_{cert} – wartość odniesienia, n – liczba powtórzeń w serii pomiarowej ($n=n_1=n_2$), SD – odchylenie standardowe

Gdy serie pomiarowe nie są równoliczne $n_1 \neq n_2$ to należy zastosować poniższy wzór:

$$t_{eksp} = \frac{|x_{\dot{s}r1} - x_{\dot{s}r2}|}{\sqrt{S_{zbior}^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad (8)$$

$$S_{zbior}^2 = \frac{SD_1^2 + SD_2^2}{2} \quad (9)$$

Gdy test F-Snedecora wskazał na istotne różnice pomiędzy wariancjami dwóch serii pomiarowych, należy zastosować wzór:

$$t_{eksp} = \frac{|x_{\dot{s}r1} - x_{\dot{s}r2}|}{\sqrt{S^2}} \quad (10) \quad s^2 = \frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2} \quad (11)$$

obliczamy liczbę stopni swobody (f):

$$f = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)^2} \quad (12)$$

Funkcja **=T.TEST(seria1 ; seria2 ; strony ; typ)** oblicza wartość prawdopodobieństwa dla założenia, że wariancje obu porównywanych serii pomiarowych nie różnią się w sposób statystycznie istotny (gdy $p > 0,05$). Zastosuj tę funkcję dla porównywanych serii pomiarowych. Parametr testu – strony – należy wpisać 2 dla testu dwustronnego; parametr testu – typ – należy wpisać 2, gdy wariancje nie różnią się istotnie lub 3, gdy wariancje różnią się istotnie.

- oblicz liczbę stopni swobody f
- wyszukaj parametr t krytyczny z tablic statystycznych
- oblicz parametr t krytyczny z funkcji **=ROZKŁ.T.ODWR.DS(0,05 ; f)**
- Zastosuj formatowanie warunkowe → Reguły wyróżniania komórek → Większe niż... aby podświetlić na czerwono parametry przekraczające wartość krytyczną.
- zastosuj dodatek Analiza danych → Test t: z dwiema próbami zakładający równe wariancje
Lub Test t: z dwiema próbami zakładający nierówne wariancje

Test t-Studenta – porównanie średniej serii pomiarowej z wartością odniesienia

Zastosuj test t-Studenta, analogicznie jak powyżej, dla pomiarów żywności stałej – mleka w proszku – i CRMu. W przypadku mleka producent nie podaje niepewności zawartości żelaza, a więc należy zastosować odpowiedni wariant wzoru na test t-Studenta (patrz literatura).

- oblicz liczbę stopni swobody f
- wyszukaj parametr t krytyczny z tablic statystycznych
- oblicz parametr t krytyczny z funkcji **=ROZKŁ.T.ODWR.DS(0,05 ; f)**
- Zastosuj formatowanie warunkowe → Reguły wyróżniania komórek → Większe niż... aby podświetlić na czerwono parametry przekraczające wartość krytyczną.

ANOVA – analiza wariancji

Zastosuj analizę wariancji dla więcej niż 2 serii pomiarowych – ANOVA – za pomocą dodatku: Analiza danych → Analiza wariancji jednoczynnikowa. Dane wejściowe muszą być ułożone jako serie w sąsiadujących kolumnach.

- oceń na podstawie parametru Wartość-p czy występują istotne statystycznie różnice w zmienności między 4 seriami danych.

Statystyka opisowa

- Oblicz podstawowe parametry statystyki opisowej dla obu kolumn:

Parametr	Mleko	Jednostka
Średnia		
Odchylenie standardowe SD		
Względne odchylenie standardowe RSD		
Współczynnik zmienności CV		
Minimum		
Maksimum		
Zakres		
Mediana		
1. kwartyl		
3. kwartyl		
Rozstęp międzykwartkowy IQR		

W tabelce podane są różne miary centralne i dyspersji danych. Średnia i SD to wielkości parametryczne, mediana, kwartyle i IQR to wielkości nieparametryczne. Nieparametryczne metody nie zakładają żadnego rozkładu danych i są odporniejsze na wartości odstające niż metody parametryczne.

- zastosuj dodatek Analiza danych → Statystyka opisowa

Wymagania

- testy statystyczne (Q-Dixona, F-Snedecora, t-Studenta); cel, przebieg;
- Podstawowe parametry statystyczne, miary rozrzutu i położenia (średnia, odchylenie standardowe, względne odchylenie standardowe, mediana, kwartyl, percentyl, rozstęp międzykwartkowy, zakres);
- sposoby wizualizacji danych (histogram, wykres skrzynka-wąsy);
- testowanie hipotez statystycznych, hipoteza zerowa, hipoteza alternatywna, poziom istotności, wartość prawdopodobieństwa p, ANOVA;
- zaokrąglanie wyniku pomiaru wraz z niepewnością.

Literatura

1. „Ocena i kontrola jakości wyników analitycznych”, Wydawnictwo Naukowo-Techniczne Warszawa 2007, praca zbiorowa pod redakcją Piotra Konieczki i Jacka Namieśnika.
2. „Analiza statystyczna w laboratorium analitycznym”, Wojciech Hyk, Zbigniew Stojek;
3. „Podstawy chemii analitycznej tom 1”, D.A. Skoog, D.M. West, F.J. Holler, S.R. Crouch, Wydawnictwo Naukowe PWN, Warszawa 2007, rozdział 7;
4. <https://mirosławmamczur.pl/category/wykresy/page/2/>
5. https://brain.fuw.edu.pl/edu/index.php/Wnioskowanie_Statystyczne_-_wyk%C5%82ad