

## ZAGADNIENIA NA EGZAMIN

Poniżej zebrane zostały wybrane pojęcia i zagadnienia omawiane w trakcie kolejnych wykładów. Ze względu na zdalną formę egzaminu znajomość tych pojęć i zagadnień będzie w większości sprawdzana w sposób praktyczny przy pomocy zadań dołączonych poniżej do każdego wykładu.

Jeżeli z treści zadania wynika np., że należy:

- zapisać wyrażenie opisujące wartość statystyki testowej np. w teście o równości średnich

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

to zakładając np. że odchylenia standardowe w obydwu próbach są takie same  $\sigma_1 = \sigma_2 = 1$  oraz, że rozmiary prób są równe  $n_1 = n_2 = 8$  a średnie otrzymane w próbach wynoszą  $\bar{x}_1 = 1,5$  i  $\bar{x}_2 = 1$ , wyrażenie powinno być zapisane np. w postaci:

$$Z = (1,5 - 1) / \text{pierwiastek}(1^2/8 + 1^2/8),$$

o ile wynika to z treści zadania może być również konieczne obliczenie wartości takiego wyrażenia, w analizowanym przykładzie statystyka testowa ma wartość  $Z = 1$ ,

- narysować wykres to:

należy się spodziewać, że trzeba będzie wskazać jeden z kilku wykresów, które zostaną dołączone jako odpowiedzi do pytania lub

naszkić wykres wyłącznie dla własnych potrzeb w celu wyciągnięcia wniosków.

## WYKŁAD 1

### Pojęcia:

doświadczenie, eksperyment, plan eksperymentu, macierz planu, wielkości wejściowe, wielkości wyjściowe,

### Pojęcia statystyczne

cechy mierzalne i niemierzalne,

szereg prosty, szereg rozdzielczy,

rozkład prawdopodobieństwa, dystrybuanta,

średnia arytmetyczna, kwanty rzędu  $p$ , mediana, odchylenie standardowe, rozstęp,

### Zagadnienia

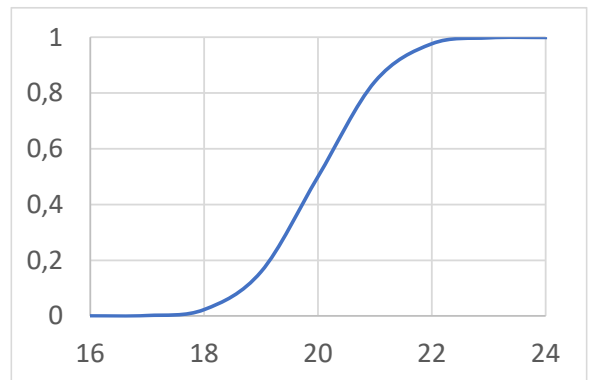
- wykres pudełkowy,
- histogram,
- wykres kwantyl-kwantyl,

### Zadania

#### Zad. 1.

Oszacuj na podstawie powyższego wykresu przedstawiającego dystrybuantę pewnego rozkładu: prawdopodobieństwa:

a)  $P(x \leq \dots)$ , b)  $P(x > \dots)$ , c) kwantyl rzędu ... .



--	--	--	--

#### Zad. 2.

W tabeli powyżej zapisane zostały wyniki pomiarów pewnego parametru 4 losowo wybranych detali. Wyznacz niezbędne miary i narysuj wykres pudełkowy z linią środkową ustaloną na podstawie średniej.

--	--	--	--	--	--	--	--	--	--

#### Zad. 3.

W tabeli powyżej zapisane zostały wyniki pomiarów pewnego parametru 10 losowo wybranych detali. Zakładając, że granice szeregów klasowych wynoszą ... skonstruuj szereg rozdzielczy i wykreśl histogram liczebności.

--	--	--	--	--

**Zad. 4.**

W tabeli powyżej zapisane zostały wyniki pomiarów pewnego parametru 5 losowo wybranych detali. Zakładając, że rzędy kwantyli z próby wynoszą: 0,1, 0,3, 0,5, 0,7 i 0,9 narysuj wykres kwanty – kwantyl porównujący rozkład z próby z pewnym rozkładem teoretycznym, którego kwantyle rzędów: 0,1, 0,3, 0,5, 0,7 i 0,9 wynoszą odpowiednio: ... . Czy wykres świadczy o zgodności obydwu rozkładów?

## WYKŁAD 2

### Pojęcia

prosty eksperyment porównawczy

### Pojęcia statystyczne

hipotezy statystyczne, poziom istotności, obszar krytyczny, rodzaje obszarów krytycznych,  $p - value$ ,

hipotezy parametryczne

### Zagadnienia

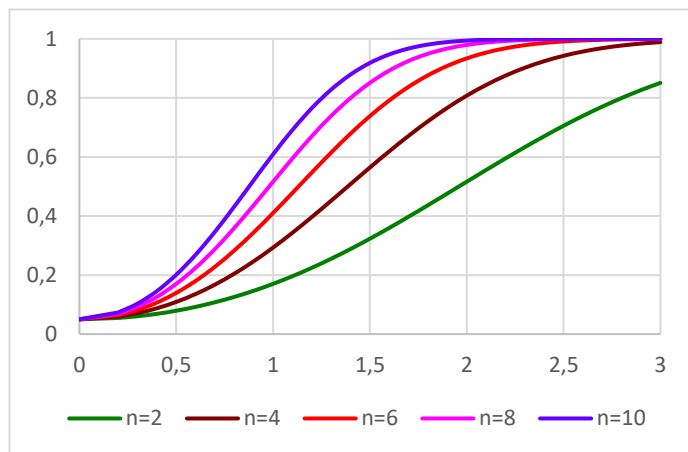
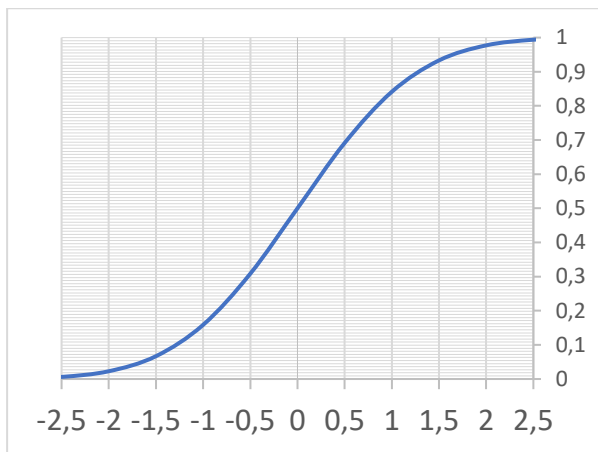
- warunki odrzucenia hipotezy zerowej w oparciu o obszar krytyczny i  $p - value$ ,
- weryfikacja hipotez o równości średnich dwóch populacji o znanych odchyleniach standardowych

### Zadania

**Zad. 1.** Badano wpływ dwóch poziomów pewnego czynnika na wartość parametru detalu. Wykonano  $n = \dots$  pomiarów dla obydwu poziomów badanego czynnika. Zakładając, że rozkład analizowanego parametru jest rozkładem normalnym i znane jest jego odchylenie standardowe, które niezależnie od poziomu czynnika wynosi  $\sigma = \dots$  – zweryfikować na poziomie istotności  $\alpha = \dots$  hipotezę o braku istotności wpływu poziomu czynnika na parametr detalu (przyjąć, że hipoteza  $H_1$  ma postać  $\mu_1 \neq \mu_2$ ).

Przyjmując, że obliczona na podstawie przeprowadzonego eksperymentu średnia wartość pomiarów dla pierwszego poziomu czynnika wynosi  $\bar{x}_1 = \dots$  a dla drugiego poziomu czynnika  $\bar{x}_2 = \dots$  należy:

- a) wyznaczyć wartość statystyki testowej w tym teście (zapisać wyrażenie i obliczyć jego wartość),
- b) wyznaczyć wartości graniczne obszaru krytycznego lub graniczny poziom istotności  $p - value$  (wartości należy oszacować w przybliżeniu na podstawie dołączonego wykresu dystrybuanty rozkładu normalnego standaryzowanego),
- c) wyciągnąć i uzasadnić wniosek odnośnie istotności wpływu poziomów czynnika na parametr detalu,
- d) określić moc testu (wykorzystując podany wykres mocy względem efektu tzn. różnicy pomiędzy średnimi uzyskanymi dla badanych poziomów czynnika wejściowego)
- e) ustalić liczebność próby  $n$  tak aby test dla otrzymanej różnicy średnich miał moc około  $\dots$  (na dołączonym wykresie mocy pokazane będą krzywe mocy dla różnych rozmiarów próby).



Podpunkty a)-c) patrz przykład 3 (slajdy 19-21), podpunkty d)-e) posługiwanie się wykresem mocy omówione zostało na wykładzie dla testu dla jednej średniej w przykładzie 6 (slajd 37), różnica w interpretacji wykresu w przypadku testu dla dwóch średnich sprowadza się do osi poziomej wykresu, która w zadaniu będzie przedstawiała różnicę średnich dla analizowanych poziomów czynnika, np. z dołączonego rysunku wynika, że różnica o wielkości 1 będzie wykrywana przez test z mocą 0,4 dla liczebności próby  $n = 6$ .

## WYKŁAD 3

### Pojęcia:

eksperyment jednoczynnikowy, plan randomizowany kompletny

### Zagadnienia

dekompozycja zmienności zmiennej zależnej, całkowita zmienność zmiennej ( $SS_T$ ), zmienność wyjaśniona modelem ( $SS_\tau$ ), zmienność niewyjaśniona modelem ( $SS_e$ ),

analiza wariancji: model, reszty, założenia analizy wariancji

test Levene'a, testy post-hoc, test NIR Fishera (zastosowanie)

pojęcie efektu, miary efektu:  $f$  Cohena,  $RMSSE$ , *rozpiętość standaryzowana*,

typy zmienności wg Cohena

### Zadania

#### Zad. 1.

Badano wpływ  $a = \dots$  poziomów pewnego czynnika na wartość pewnego parametru detalu. Wykonano po  $r = \dots$  doświadczeń dla każdego poziomu czynnika wykonując eksperyment wg planu randomizowanego kompletnego. Z przeprowadzonej analizy wariancji wynika, że zmienność wyjaśniona modelem  $SS_\tau = \dots$  a zmienność niewyjaśniona  $SS_e = \dots$ . Należy:

- wyznaczyć wartość statystyki  $F$  (zapisać wyrażenie i obliczyć jego wartość),
- wyznaczyć wartości graniczne obszaru krytycznego lub graniczny poziom istotności  $p$  – *value* (wartości należy oszacować w przybliżeniu na podstawie dołączonego wykresu dystrybuanty rozkładu  $F$ , będzie dołączonych kilka wykresów należy wybrać odpowiedni biorąc pod uwagę liczbę stopni swobody statystyki, dołączone wykresy nie muszą odpowiadać właściwym – wartości należy odczytać z wykresów),
- wyciągnąć i uzasadnić wniosek odnośnie istotności wpływu poziomów czynnika na wartość parametru detalu.

(patrz przykład 1, slajd 19)

#### Zad. 2.

Wykonując analizę wariancji dla eksperymentu, w którym badano wpływ  $a = \dots$  poziomów pewnego czynnika na wartość pewnego parametru detalu dla  $r = \dots$  powtórzeń doświadczeń dla każdego poziomu czynnika w przeprowadzonym teście istotności otrzymano wartość statystyki  $F = \dots$  z odpowiadającą jej wartością granicznego poziomu istotności  $p$  – *value* =  $\dots$ . Należy:

- wyciągnąć i uzasadnić wniosek odnośnie istotności wpływu poziomów czynnika na parametr detalu,
- zweryfikować założenia analizy na podstawie dołączonych wykresów sporządzonych dla reszt (wykres normalności, wykres reszt w funkcji numeru doświadczenia, wykres reszt w

funkcji wartości przewidywanych z modelu)

c) ocenić poprawność wniosków analizy z punktu a).

(patrz przykład 1, slajd 22-23)

**Zad. 3.**

W tabeli zestawione zostały wyniki analizy wariancji i testów post-hoc. Czy wpływ poziomów czynnika może być uznany za istotny, dla których poziomów czynnika różnica pomiędzy otrzymanymi wartościami parametru jest statystycznie istotna. Odpowiedź uzasadnij.

(patrz przykład 1, slajdy 19, 32, 30, 34)

$SS_{\tau}$	$MS_{\tau}$	$SS_e$	$MS_e$	$F$	$p$

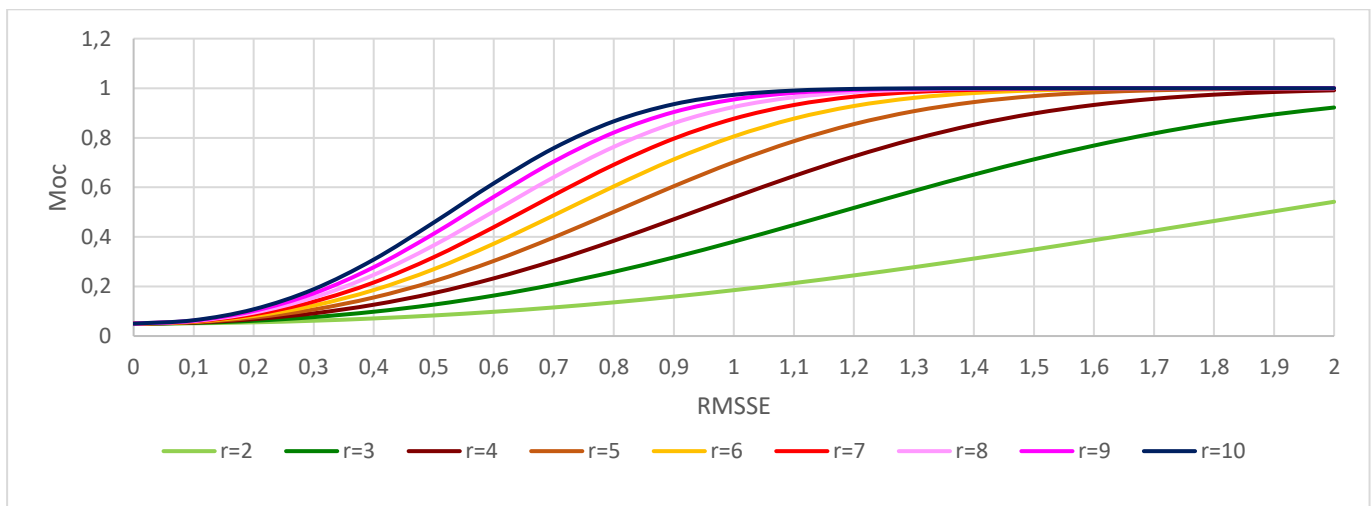
Poziom czynnika	1	...	$a$
1			
...			
$a$			

**Zad. 4.**

Zaplanowano eksperyment, którego celem było zbadanie wpływu  $a = \dots$  poziomów pewnego czynnika na wartość pewnego parametru detalu. Założono, że wykonanych zostanie  $r = \dots$  powtórzeń doświadczeń dla każdego poziomu czynnika. Czy jeżeli zgodnie z założeniami uzyskanie dla wybranych poziomów czynnika średnich wartości parametru równych odpowiednio  $\mu_1 = \dots, \mu_2 = \dots, \dots, \mu_a = \dots$  dla  $\sigma = \dots$  to:

- a) moc testu wynosi  $\dots$  i test można uznać za mocny?
- b) aby zapewnić moc testu równą 0,8 należy wykonać  $r = \dots$  powtórzeń doświadczeń na każdym poziomie.

Patrz przykład 1 (slajdy 43-44). Moc testu należy odczytać z dołączonego wykresu mocy (z wykresu poniżej wynika np., że gdyby planowana w eksperymencie wartość efektu RMSSE wynosiła 0,9 to dla  $r = 5$  moc wyniosłaby 0,6 i żeby zapewnić moc równą 0,8 należałoby wykonać  $r = 7$  powtórzeń.



## WYKŁAD 4

### Pojęcia:

eksperyment dwuczynnikowy,  
efekt główny, efekt interakcji  
miary efektu:  $f$  Cohena,  $RMSSE$ , *stopień niecentralności*,

### Zadania

#### Zad.1.

Rozważmy eksperyment dwuczynnikowy, w którym obydwie zmienne niezależne  $A$  i  $B$  przyjmują wartości na dwóch poziomach: niskim i wysokim:  $A -$ ,  $A +$ ,  $B -$ ,  $B +$ . Oblicz średnią wartość efektów  $A$  i  $B$  oraz efekt ich interakcji. Narysuj wykres przedstawiający wpływ poziomu czynników na wartość zmiennej zależnej i odpowiedz na pytania:

- Czy efekt  $A$  zależy od poziomu czynnika  $B$ ?
- Czy efekt  $B$  zależy od poziomu  $A$ ?
- Czy interakcja czynników wpływa na wartość zmiennej zależnej.

(patrz przykład 1, slajdy 4-9)

#### Zad.2.

Rozważmy eksperyment dwuczynnikowy, w którym zmienne niezależne  $A$  i  $B$  przyjmują wartości na kilku poziomach. W tabeli obok podane zostały wyznaczone eksperymentalnie średnie wartości zmiennej zależnej na każdym z poziomów zmiennych  $A$  i  $B$ . Narysuj wykres przedstawiający wpływ poziomu czynników na wartość zmiennej zależnej i skomentuj otrzymane wyniki:

- czy i jak czynniki wpływają na zmienną zależną,
- czy efekt interakcji jest istotny,
- jak dobrać poziom czynnika  $A$  aby otrzymać określoną wartość lub zmianę zmiennej zależnej.

(patrz przykład 2, slajdy 20-21)

#### Zad. 3.

W tabeli zestawione zostały wyniki analizy wariancji – skomentuj otrzymane wyniki. Czy wpływ czynników i ich interakcji jest statystycznie istotny?

(patrz przykład 1, slajdy 32, 19)

<i>efekt</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i>
$A$				
$B$				
$A * B$				



**Zad. 4.**

Na wykresach przedstawione zostały wykresy reszt: w funkcji wartości przewidywanych z modelu, w funkcji poziomu czynników wejściowych oraz wykres normalności. Czy wszystkie założenia analizy wariancji są spełnione? Odpowiedź uzasadnij.

*(patrz przykład 2, slajdy 25-27)*

## WYKŁAD 5

### Zagadnienia

Analiza regresji, metoda najmniejszych kwadratów,  
dekompozycja zmienności zmiennej zależnej, całkowita zmienność zmiennej ( $SS_T$ ), zmienność wyjaśniona modelem ( $SS_r$ ), zmienność niewyjaśniona modelem ( $SS_e$ ), ocena jakości dopasowania funkcji regresji: współczynnik determinacji

### Zadania

#### Zad.1.

Celem badań jest określenie funkcji obiektu w postaci:  $y = \dots$  Podaj funkcje bazowe, zapisz macierz wejść zakładając, że zmienne wejściowe są ograniczone  $\dots \leq x_i \leq \dots$  a eksperyment ma być przeprowadzony na poziomach niskich i wysokich zmiennych.

(patrz przykład 2, slajdy 27, 30)

#### Zad.2.

W podanych tabelach zestawione zostały wyniki analizy regresji, celem której było określenie funkcji obiektu w postaci:  $y = \dots$  Tabela 1 zawiera wyniki testu istotności dla znalezionej funkcji regresji, Tabela 2. wyniki testów istotności dla współczynników funkcji.

Tabela 1					
$SS_e$	$MS_e$	$SS_r$	$MS_r$	$F$	$p$

Tabela 2				
	$b^*$	$b$	$t$	$p$
wyraz wolny				
$\vdots$				

Na podstawie wyników analizy:

- zapisz postać znalezionej funkcji regresji podając w wyrażeniu znalezione wartości współczynników
- ocień czy znaleziona funkcja w sposób istotny statystycznie opisuje związek pomiędzy zmiennymi wejściowymi a zmienną wyjściową,
- ocień, które współczynniki są istotne statystycznie,
- ocień, która zmienna wejściowa ma największy wpływ na wartość zmiennej wyjściowej.

(patrz slajdy 39-52)