SdI30 W12: *ANOVA*

1. **Wprowadzenie**

   Przykład 1

2. **Podstawowe pojęcia planowania eksperymentu**
3. **Planowanie eksperymentu**
4. **ANOVA dla układu całkowicie zrandomizowanego**
5. **Założenia ANOVA dla układu całkowicie zrandomizowanego**

   Przykład 1

6. **Przedziały ufności dla wartości oczekiwanych w eksperymencie całkowicie zrandomizowanym**

   Przykład 2

7. **Procedura wielokrotnych porównań Tukeya**

   Przykład 3

8. Test rangowy Kruskala-Wallisa

Przykład 4

Przykład 5

9. ANOVA dla zrandomizowanego układu bloków

Przykład 6

10. Podsumowanie

11. Zestaw zadań

1. Wprowadzenie

Analiza wariancji (ANOVA) jest zespołem metod statystyki matematycznej opartych na porównywaniu wariancji.

ANOVA jest podstawowym narzędziem statystyki eksperymentalnej, tj. szeroko rozbudowanej dla potrzeb doświadczalnictwa statystycznego metod planowania i oceny wyników eksperymentów.

W najprostszym przypadku ANOVA jest rozszerzeniem testu do weryfikacji hipotezy o równości wartości oczekiwanych w dwóch populacjach na większą liczbę populacji.

Typowym przykładem zagadnienia rozwiązywanego za pomocą ANOVA jest analiza wpływu czynników zewnętrznych na wynik przeprowadzanego doświadczenia.

Statystyka eksperymentalna wypracowała wiele metod planowania doświadczeń, jak np. bloki losowe, kwadraty łacińskie, analizę czynnikową i in. Każdy rodzaj doświadczenia ma odrębny schemat ANOVA.

W prezentowanym wykładzie rozważania są ograniczone do badania wpływu czynników zewnętrznych na wynik eksperymentu przy zmianie jednego z czynników (tzw. klasyfikacja jednoczynnikowa) lub wielu czynników równocześnie (klasyfikacja wieloczynnikowa).

2. Podstawowe pojęcia planowania eksperymentu

Zmienne potencjalnie związane ze zmienną objaśnianą nazywamy *czynnikami*. Regulowane w eksperymencie wartości przyjmowane przez czynnik nazywamy *poziomami* czynnika.

Przypuśćmy, że eksperyment jest przeprowadzany w celu zbadania twardości Y nowego typu plastiku jako funkcji dwóch czynników, tj. ciśnienia i temperatury w czasie formowania.

Jeżeli twardość plastiku jest mierzona dla ciśnień 100, 200, 300 kG/cm^2 i temperatur 200 i 300 stopni Fahrenheita ($°F$), to czynnik ciśnienie ma trzy ustalone poziomy, a czynnik temperatura dwa poziomy.

Kombinacje poziomów czynników, dla których jest obserwowana zmienna zależna nazywamy *zabiegami* eksperymentu.

Na przykład, jeżeli twardość nowego plastiku jest mierzona dla każdej z sześciu kombinacji ciśnienie-temperatura $(100, 200)$, $(200, 200)$, $(300, 200)$, $(100, 300)$, $(200, 300)$, $(300, 300)$, to eksperyment pociąga za sobą sześć zabiegów.

Zabiegi są podstawową częścią planowania eksperymentu, polegającą na działaniu zmierzającym do zmiany warunków eksperymentalnych.

Reasumując wprowadziliśmy dotąd trzy pojęcia związane z planowaniem eksperymentu: **czynnik**, **poziom** i **zabieg**.

3. Planowanie eksperymentu

Planowanie eksperymentu obejmuje cztery etapy:

1. Wybór czynników uwzględnionych w eksperymencie i identyfikacja parametrów, które są celem badań. Zwykle parametry informują o własnościach populacji związanej z wybranymi zabiegami.
2. Przyjęcie wielkości błędu standardowego, tj. decyzja o wielkości informacji jaką chcemy zdobyć.
3. Wybór zabiegów uwzględnionych w eksperymencie.
4. Decyzja w jaki sposób zabiegi są związane z jednostkami eksperymentalnymi.

Projektując eksperyment, którego celem jest porównanie kilku wartości oczekiwanych, wyróżniamy czynniki, które obok czynników losowych mogą naszym zdaniem również wpływać na wyniki obserwacji. Jest to równoznaczne z klasyfikacją badanej populacji według czynników, które być może są powodem większej różnicy między wartościami oczekiwanymi, niż to wynika z przyczyn losowych.

Klasyfikujemy więc populację według czynników systematycznych. Jeśli klasyfikacja dokonana jest według jednego czynnika, to mówimy o klasyfikacji pojedynczej lub jednoczynnikowej. Jeśli do klasyfikacji wykorzystanych jest kilka czynników, to mówi się o klasyfikacji wieloczynnikowej.

Wynik każdej obserwacji traktujemy jako rezultat nakładania się na siebie efektów dwóch rodzajów – przypadkowego i systematycznego, jeśli ten istnieje.

Istotą ANOVA jest rozbicie na addytywne składniki sumy kwadratów wariancji empirycznej wyników obserwacji.

Porównanie poszczególnych wariancji wynikających z działania danego czynnika z wariancją resztkową mierzącą losowy błąd poprzez zastosowanie testu F daje odpowiedź, czy dany czynnik odgrywa istotną rolę w kształtowaniu się wyników eksperymentu.

Sterowanie ilością informacji uzyskanej z eksperymentu jest podstawowym celem jego planowania. Sterowanie to odbywa się poprzez przyjęcie wielkości błędu standardowego.

Czynniki zaburzające informacje w eksperymencie są związane z błędami standardowymi wyznaczanych estymatorów parametrów i są proporcjonalne do miary szumu σ i odwrotnie proporcjonalne do natężenia sygnału (rozmiaru próbki).

Wybór zabiegów eksperymentu i ustalenie wielkości próbki decyduje o natężeniu informacji o populacji.

Można postawić pytanie: ***Czy możliwe jest obserwować zmienną Y i nie otrzymać żadnej informacji o interesującym nas parametrze***? Odpowiedź brzmi tak.

Dla zobrazowania tego faktu, przypuśćmy, że próbujemy dopasować model probabilistyczny pierwszego stopnia $Y = \beta_0 + \beta_1 x + \varepsilon$, do zbioru dziesięciu danych, z których wszystkie zostały zaobserwowane dla jednej wartości zmiennej x, powiedzmy, że $x = 5$. Wówczas nie mamy żadnych możliwości wyestymowania parametrów tego modelu.

Jedynym sposobem otrzymania informacji o parametrach modelu jest obserwacja zmiennej Y dla różnych wartości zmiennej x. Etap 4 planowania eksperymentu pozwala zmniejszać błąd eksperymentu.

Po zebraniu danych z zaplanowanego eksperymentu stosując metody ANOVA dokonujemy wnioskowana o wartościach oczekiwanych populacji związanych z p zabiegami.

4. ANOVA dla układu całkowicie zrandomizowanego

Najprostszym układem eksperymentalnym jest układ całkowicie zrandomizowany. Eksperyment przeprowadzony według układu całkowicie zrandomizowanego pozwala porównywać wartości oczekiwane $\mu_1, \mu_2, \dots, \mu_p$ dla p populacji związanych z zabiegami, na podstawie niezależnych prób losowych o liczebnościach $n_1, n_2, \dots, n_p$, otrzymanych dla kolejnych zabiegów.

Przyjmujemy oznaczenia:

p – liczba populacji,

σ_i^2– wariancja i-tej populacji, $i = 1, 2, \dots, p$,

μ_i– wartość oczekiwana i- tej populacji, $i = 1, 2, \dots, p$,

n_i– liczebność próby pobranej z i-tej populacji,

$n = \sum n_i$ – liczebność ogólna,

x_{ij} – i-ta obserwacja pochodząca z j-tej populacji,

$\bar{x}_j = \frac{1}{n_j}\sum_i x_{ij}$ – średnia arytmetyczna dla j-tej próby,

$\bar{x} = \frac{1}{n}\sum_j \sum_i x_{ij}$ – średnia arytmetyczna ogólna.

Tablica. Schemat układu całkowicie zrandomizowanego

Populacja 1 wartość oczekiwana μ_1, wariancja σ_1^2,	Populacja 2 wartość oczekiwana μ_2, wariancja σ_1^2,	... ...	Populacja p wartość oczekiwana μ_p, wariancja σ_p^2,
Niezależne pobieranie prób losowych			
Próba losowa 1 liczebność n_1, średnia arytmetyczna $\bar{x}_1$, wariancja z próby s_1^2,	Próba losowa 2 liczebność n_2, średnia arytmetyczna $\bar{x}_2$, wariancja z próby s_2^2,		Próba losowa p liczebność n_p, średnia arytmetyczna $\bar{x}_p$, wariancja z próby s_p^2.

Źródła zmienności poszczególnych obserwacji x_{ij} mogą wynikać z przyczyn głównych, w tym przypadku ze zróżnicowania na p populacji oraz z działania przyczyn losowych.

Zatem modelem obserwacji x_{ij} jest:

$$x_{ij} = \mu + \eta_j + \varepsilon_{ij} \tag{1}$$

gdzie

- μ jest oczekiwaną wartością badanej cechy bez uwzględnienia zróżnicowania na populacje,
- η_j przedstawia wpływ j-tej populacji na wartość badanej cechy,
- $\mu_j = \mu + \eta_j$ jest oczekiwaną wartością badanej cechy w j-tej populacji,

ε_{ij} jest błędem losowym modelu.

Składniki modelu (1) są oceniane na podstawie próby losowej. Stosując metodę najmniejszych kwadratów, tj. minimalizując sumę

$$S = \sum_{j=1}^{p} \sum_{i=1}^{n_j} \left(x_{ij} - \mu - \eta_j\right)^2$$

kolejno ze względu na μ i η_j, można wykazać, że estymatorami parametrów μ i η_j są odpowiednio średnia ogólna $\bar{x}$ oraz różnica średnich $\bar{x}_j - \bar{x}$.

Mamy więc

$$\hat{\mu} = \bar{x}, \qquad \hat{\eta}_j = \bar{x}_j - \bar{x}, \qquad \hat{\mu}_j = \bar{x}_j$$

Ponieważ

$$x_{ij} = \bar{x} + x_{ij} - \bar{x}_j + \bar{x}_j - \bar{x}$$

więc

$$x_{ij} - \bar{x} = \left(x_{ij} - \bar{x}_j\right) + \left(\bar{x}_j - \bar{x}\right)$$

Po podniesieniu obu stron do kwadratu i zsumowaniu według indeksów j oraz i otrzymamy

$$\sum_{j=1}^{p}\sum_{i=1}^{n_j}\left(x_{ij}-\bar{x}\right)^2$$

$$=\sum_{j=1}^{p}\sum_{i=1}^{n_j}\left(x_{ij}-\bar{x}_j\right)^2+2\sum_{i=1}^{n_j}\left(x_{ij}-\bar{x}_j\right)\sum_{j=1}^{p}\left(\bar{x}_j-\bar{x}\right)$$

$$+\sum_{j=1}^{p}n_j\left(\bar{x}_j-\bar{x}\right)^2$$

ponieważ

$$\sum_{i=1}^{n_j}\left(x_{ij}-\bar{x}_j\right)=0$$

więc

$$\sum_{j=1}^{p}\sum_{i=1}^{n_j}\left(x_{ij}-\bar{x}\right)^2=\sum_{j=1}^{p}\sum_{i=1}^{n_j}\left(x_{ij}-\bar{x}_j\right)^2+\sum_{j=1}^{p}n_j\left(\bar{x}_j-\bar{x}\right)^2$$

Jest to podstawowa tożsamość analizy wariancji. Niezależne próby pochodzące z p populacji nazywamy grupami.

Pierwszy składnik jest sumą kwadratów odchyleń między poszczególnymi obserwacjami a ich średnimi grupowymi i nazywany jest *sumą kwadratów odchyleń wewnątrz grup* albo *resztkową sumą kwadratów*.

Drugi jest sumą kwadratów odchyleń średnich grupowych od średniej ogólnej i nazywany jest *sumą kwadratów pomiędzy grupami*.

Wprowadzamy oznaczenia

$SS = \sum_{j=1}^{p} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}\right)^2$ – ogólna suma kwadratów (total sum of squares),

$SST = \sum_{j=1}^{p} n_j \left(\bar{x}_j - \bar{x}\right)^2$ – suma kwadratów odchyleń między grupami (sum of squares for treatments),

$SSE = \sum_{j=1}^{p} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right)^2$ – suma kwadratów wewnątrz grup (sum of squares for error),

Przy tych oznaczeniach podstawową tożsamość analizy wariancji możemy krótko zapisać

$$SS = SST + SSE$$

Zauważmy, że jeżeli średnie grup są „blisko” siebie, to są również bliskie średniej ogólnej i w konsekwencji suma SST jest „mała”. W szczególności, jeżeli $\bar{x}_1 = \bar{x}_2 = \cdots = \bar{x}_p$, to $SST = 0$.

Jeżeli istnieją duże różnice pomiędzy średnimi dla grup, to pewne średnie dla grup różnią się znacznie od średniej ogólnej, co objawia się „dużą” sumą SST.

Oznaczać to może powód do odrzucenia hipotezy zerowej o równości wartości oczekiwanych dla wszystkich grup, tj. hipotezy

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_p$$

na rzecz hipotezy alternatywnej, że nie wszystkie wartości oczekiwane są sobie równe.

Sumę SS obliczamy na podstawie odchyleń n obserwacji od średniej ogólnej, a więc ma ona $n - 1$ stopni swobody.

Suma SST jest obliczana z p odchyleń niezależnych średnich grupowych od średniej ogólnej, więc ma $p - 1$ stopni swobody.

Suma SSE jest obliczana z odchyleń n obserwacji od p średnich grupowych, więc ma $n - p$ stopni swobody.

Sumy SST i SSE podzielone przez odpowiadające im ilości stopni swobody nazywamy średnimi kwadratami:

$MST = \frac{SST}{p-1}$ jest średnim kwadratem dla grup (means square)

$MSE = \frac{SSE}{n-p}$ jest średnim kwadratem dla błędu (mean square for error)

MSE jest stosowany jako estymator wariancji błędu modelu σ^2.

W obliczeniach analizy wariancji przyjmujemy, że dysponujemy danymi w postaci jak w tabeli

Grupa 1	Grupa 2	...	Grupa p
x_{11}	x_{12}	...	x_{1p}
x_{21}	x_{22}	...	x_{2p}
...	...	...	...
$x_{n_1 1}$	$x_{n_2 2}$	...	$x_{n_p p}$

Na podstawie danych empirycznych chcemy odpowiedzieć na pytanie:

Czy różnice między średnimi grupowymi wynikają tylko z przyczyn losowych, czy też mają one charakter systematyczny?

Jeśli miałby miejsce pierwszy przypadek, to średni kwadrat dla grup MST byłby estymatorem jednakowej wariancji dla wszystkich p populacji.

Jeżeli natomiast różnice między średnimi byłyby systematyczne, to należy uwzględnić wkład tego systematycznego zróżnicowania do zmienności międzygrupowej.

Ponieważ statystyka

$$F = \frac{MST}{MSE}$$

ma rozkład F Snedecora o $(p-1, n-p)$ stopniami swobody, więc możemy ją wykorzystać do weryfikacji hipotezy zerowej.

Hipotezę zerową odrzucimy, jeśli obliczona statystyka F_0 należy do prawostronnego zbioru krytycznego, tj.

$$F_0 > F_{1-\alpha;p-1;n-p} \text{ lub } \alpha_0 = F_{F(p-1;n-p)}(F_0) < \alpha.$$

gdzie $F_{1-\alpha;p-1;n-p}$ jest kwantylem rzędu $1-\alpha$ rozkładu F Snedecora o $(p-1, n-p)$ stopniach swobody.

5. Założenia ANOVA dla układu całkowicie zrandomizowanego

Przedstawiony został model teoretyczny postaci:

$$x_{ij} = \mu + \eta_j + \varepsilon_{ij}$$

Aby model ten można było zastosować w konkretnym eksperymencie muszą być spełnione odpowiednie założenia:

1. Składniki losowe modelu są niezależnymi zmiennymi losowymi o rozkładzie $\mathcal{N}(0, \sigma)$.

2. Wariancje dla grup są równe, tj. $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_p^2$.

3. Zmienne losowe η_j, $j = 1,2, \dots, p$ są niezależne.

Przy badaniu równości wielu wariancji weryfikujemy hipotezę zerową

$$H_0\colon \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_p^2$$

przeciw hipotezie alternatywnej, że nie wszystkie wariancje są sobie równe.

Zwykle stosujemy jeden z następujących testów:

1. Test Bartletta. Stosowany przy założeniu, że

a) badana cecha X w p populacjach ma rozkłady

$$\mathcal{N}(\mu_j, \sigma_j),\ j = 1, 2, \dots, p,$$

b) z j-tej populacji pobrano losowo próbę o liczebności n_j, $j = 1, 2, \dots, p$.

2. Test Hartleya. Stosowany w przypadku, gdy

$$n_1 = n_2 = \cdots = n_p \text{ i } n_p \geq 5.$$

3. Test Cochrana. Stosowany przy tym samym założeniu co test Hartleya.

Wyniki analizy wariancji są zestawiane w formie zbiorczej zwanej tablicą ANOVA.

Źródło (Source)	Stopnie swobody (Degrees of freedom)	Sumy kwadratów (Sum of squares)	Średnie kwadraty (Mean squares)	Statystyka F (F ratio)
między grupami (treatments)	$p-1$	SST	MST	F
wewnątrz grup (Error)	$n-p$	SSE	MSE	
Ogółem (Total)	$n-1$	SS		

Przykład 1. Dział marketingu sieci sklepów planuje założenie nowego sklepu w jednym z czterech miast o porównywalnej liczbie mieszkańców. Za jeden z ważniejszych czynników w podjęciu decyzji o lokalizacji dział uznał miesięczne dochody rodzinne ich mieszkańców. W tym celu wylosowano po kilka rodzin z tych miast i odnotowano ich roczne dochody (w tys. zł). Dane zestawione są w tabeli.

Miasto 1	Miasto 2	Miasto 3	Miasto 4
25	32	27	18
27	35	32	23
31	30	48	29
17	46	25	26
29	32	20	42
30	22	12	
	19	18	
	51		
	27		

Na poziomie istotności $\alpha = 0{,}05$ zbadać, czy są wystarczające podstawy do uznania istotnych różnic w przeciętnych dochodach rodzinnych mieszkańców badanych miast.

Rozwiązanie. W eksperymencie uwzględniono jeden czynnik „miasto” dla czterech poziomów. Mamy więc do czynienia z całkowicie zrandomizowanym eksperymentem z $p = 4$ grupami (populacjami). Niech $\mu_1, \mu_2, \mu_3, \mu_4$ reprezentują oczekiwany dochód rodzinny dla badanych miast. Z materiału eksperymentalnego mamy:

$$n_1 = 6, n_2 = 9, n_3 = 7, n_4 = 5, n = 27$$

$$S_y = \sum_{i=1}^{n} y_i = 773, S_{y^2} = \sum_{i=1}^{n} y_i^2 = 24493,$$

$$CM = \frac{S_y^2}{n} = \frac{5997529}{27} = 22130{,}704$$

$$SS = S_{y^2} - CM = 2362,3$$

$$SST = 227,6,\ SSE = 2134,7,\ MST = 75,9,\ MSE = 92,8.$$

Sprawdzamy hipotezę zerową

$$H_0\colon\ \mu_1 = \mu_2 = \mu_3 = \mu_4$$

przeciw hipotezie alternatywnej, że co najmniej w dwóch miastach oczekiwane przeciętne dochody różnią się istotnie.

Obliczamy statystykę testową

$$F_0 = \frac{MST}{MSE} = 0,82$$

oraz istotność testu:

$$\alpha_0 = 1 - F_{F(3;23)}(F_0) = 0,4975$$

Tablica ANOVA

Źródło	stopnie swobody	Sumy kwadratów	Średnie kwadraty	Statystyka F
Miasto	3	227.6	75.9	0.82
Błąd	23	2134.7	92.8	
Ogółem	26	2362.3		

Obszar krytyczny jest postaci $F > F_{(1-\alpha;3;23)} \approx 3{,}028$

Ponieważ obliczona wartość statystyki F_0 nie należy do obszaru krytycznego, więc na poziomie istotności $\alpha = 0{,}05$ nie ma podstaw do odrzucenia hipotezy zerowej, że przeciętne dochody w badanych miastach są równe.

6. Przedziały ufności dla wartości oczekiwanych w eksperymencie całkowicie zrandomizowanym

Przedział ufności dla wartości oczekiwanej j-tej populacji $(j = 1, 2, \dots, p)$

$$\bar{T}_j \pm t_{\frac{\alpha}{2};n-p}\sqrt{\frac{MSE}{n_j}}$$

Przedział ufności dla różnicy wartości oczekiwanych i-tej oraz j-tej populacji

$$\left(\bar{T}_i - \bar{T}_j\right) \pm t_{\frac{\alpha}{2};n-p}\sqrt{MSE} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

gdzie $\bar{T}_j = \frac{1}{n_j}\sum_i x_{ij}$ jest średnią arytmetyczną dla j-tej grupy.

Przykład 2. Nawiązując do danych z przykładu 1 wyznaczyć 95-procentowe przedziały ufności dla

a) oczekiwanego dochodu rodzinnego w miastach 1 i 2,

b) różnicy wartości oczekiwanych dochodów rodzinnych dla miasta 1 i 2.

Rozwiązanie. Z tablicy ANOVA mamy $MSE = 92{,}8$.

Średnie dochody wyznaczone z prób wynoszą:

$$\bar{T}_1 = 26{,}5,\ \bar{T}_2 = 32{,}7,\ \bar{T}_3 = 26{,}0,\ \bar{T}_4 = 27{,}6,$$

a) 95-procentowe przedziały ufności dla μ_1 i μ_2 wynoszą:

$$\bar{T}_1 \pm t_{0,025;23}\sqrt{\frac{MSE}{n_1}} = 26{,}5 \mp 3{,}9$$

$$\bar{T}_2 \pm t_{0,025;23}\sqrt{\frac{MSE}{n_2}} = 32{,}7 \mp 3{,}2$$

b) 95-procentowy przedział ufności dla różnicy $\mu_2 - \mu_1$

$$(\bar{T}_2 - \bar{T}_1) \pm t_{0,025;23}\sqrt{MSE}\sqrt{\frac{1}{n_2} + \frac{1}{n_1}} = (-4,4; 16,6)$$

7. Procedura wielokrotnych porównań Tukeya

Procedura Tukeya oparta jest na studentyzowanym rozstępie (studentized range)

$$q = \frac{\bar{y}_{max} - \bar{y}_{min}}{\frac{s}{\sqrt{n_0}}}$$

gdzie

$\bar{y}_{max}$ i $\bar{y}_{min}$ są odpowiednio największą i najmniejszą średnią grupową,

n_0 jest liczbą obserwacji w każdej próbie,

$s^2 = MSE$ jest estymatorem wariancji błędu modelu.

Istota tej procedury polega na wyznaczeniu krytycznej wartości dla różnicy $\bar{y}_{max} - \bar{y}_{min}$ pomiędzy największą i najmniejszą średnią z próbek implikującej różnice odpowiadających im wartości oczekiwanych populacji.

Każda inna para średnich grupowych różniąca się co do bezwzględnej wartości co najmniej o wartość krytyczną również implikuje różnice odpowiadających im wartości oczekiwanych populacji z których pochodzą.

Założenia: niezależne próby losowe o tej samej liczebności n_0 pobrane z populacji $\mathcal{N}(\mu_j, \sigma)$, $j = 1, 2, \ldots, p$.

Czynności:

1. Wyliczenie odległości $\omega = q_\alpha(p;\ \nu)\dfrac{s}{\sqrt{n_0}}$,

gdzie

p liczba grup,

$s = \sqrt{MSE}$,
ν – liczba stopni swobody związana z MSE,
n_0 – liczba obserwacji w każdej grupie,
$q_\alpha(p;\ \nu)$ – wartość krytyczna studentyzowanego rozstępu.

2. Klasyfikowanie p wartości oczekiwanych. Jeżeli dowolna para średnich próbkowych różni się więcej niż ω, to istotnie różnią się wartości oczekiwane odpowiadających im populacji.

Jeżeli w rezultacie przeprowadzonej analizy wariancji okaże się, że istnieje podstawa do odrzucenia hipotezy zerowej, to w takim przypadku możemy zbadać, które z porównywanych wartości oczekiwanych istotnie różnią się, a między którymi nie ma istotnych różnic.

Jeśli byłoby to uzasadnione może powstać potrzeba połączenia populacji w jednorodne grupy (homogeneity group).

Sposób grupowania metodą Tukeya dostępny jest w pakietach statystycznych. Omówimy go w kontekście przykładu.

Przykład 3. Jednym z aspektów jakości samochodów osobowych jest koszt naprawy uszkodzeń spowodowanych drobnymi ulicznymi stłuczkami. Decydujące znaczenie mają tu zderzaki. Producent rozważa wprowadzenie nowego typu zderzaków spośród czterech zaprojektowanych typów. Zainstalowano po siedem zderzaków każdego typu na pojazdach popularnej klasy i poddano je próbom zderzania ze ścianą z prędkością 30 km/h. Następnie oszacowano koszty napraw powstałych uszkodzeń (w j.m.). Wyniki są przedstawione w tablicy.

Typ zderzaka			
1	2	3	4
315	285	269	255
288	292	277	287
293	263	273	265
306	249	252	279
299	275	263	241
310	266	251	312
282	252	272	310

a) Przyjmując 5-procentowy poziom istotności zbadać, czy są istotne różnice w kosztach usuwania uszkodzeń dla badanych czterech typów zderzaków.

b) W przypadku występowania różnic ustalić typy zderzaków różniących się ze względu na koszty usuwania awarii.

Rozwiązanie. a) Próby są niezależne. Porównywane są wartości oczekiwane dla czterech populacji kosztów usuwania awarii. Weryfikowana jest hipoteza zerowa

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

przeciw hipotezie alternatywnej

H_1: Co najmniej dwie wartości oczekiwane istotnie różnią się.

Do rozstrzygnięcia hipotezy zastosowana jest statystyka

$$F = \frac{MST}{MSE}$$

Obszar krytyczny wyznacza nierówność z użyciem kwantyla

$$F > F_{1-\alpha;k-1;n-k}$$

Do obliczeń wykorzystana jest procedura ONEWAY.

Wprowadzamy dane i ustalamy Range Test TUKEY.

Obliczona z próby wartość statystyki testowej wynosi $F = 5{,}20$. Ponieważ kwantyl $F_{0,95;3;24} = 3{,}01$, więc są podstawy do odrzucenia hipotezy zerowej. Stwierdza się istotne zróżnicowanie oczekiwanych kosztów usuwania awarii samochodów dla badanych typów zderzaków.

b) Do ustalenia, które typy zdarzaków istotnie różnią się ze względu na koszty napraw powypadkowych zastosujemy metodę Tukeya wielokrotnego porównania.

Średnie arytmetyczne z prób wynoszą:

$$\bar{T}_1 = 299{,}0;\ \ \bar{T}_2 = 268{,}9;\ \ \bar{T}_3 = 265{,}3;\ \ \bar{T}_4 = 278{,}4$$

Kwantyl $q_\alpha(p;\ \nu)$ odczytany z tablicy kwantyli (R. Zieliński) dla $\alpha = 0{,}05, p = 4, \nu = 24$ wynosi $q_{0,05}(4; 24) = 3{,}90$.

Średni kwadrat dla błędu $MSE = 308{,}2$, a wspólne liczebności prób $n_0 = 7$. Stąd krytyczna wartość odległości

$$\omega = q_\alpha(p;\ \nu)\sqrt{\frac{MSE}{n_0}} = 25{,}9$$

Obliczone średnie arytmetyczne ustawiamy w porządku niemalejącym, tj.

$\bar{T}_3$	$\bar{T}_2$	$\bar{T}_4$	$\bar{T}_1$
265,3	268,9	278,4	299,0

Zauważmy, że

$$\bar{T}_1 - \bar{T}_3 = 33{,}7 > 25{,}9$$

oraz

$$\bar{T}_1 - \bar{T}_2 = 30{,}1 > 25{,}9$$

Żadne inne różnice nie przewyższają krytycznej wartości ω. Stąd stwierdzamy, że μ_2 i μ_3 istotnie różnią się od μ_1.

Natomiast nie ma podstaw do uznania istotnych różnic między μ_2, μ_3 i μ_4.

Wniosek. Producent samochodów nie powinien montować zderzaków pierwszego typu. Wybór spośród pozostałych typów zderzaków winien się dokonać na podstawie innego kryterium, np. kosztu produkcji zderzaka.

8. Test rangowy Kruskala-Wallisa

Testy parametryczne wymagają spełniania dość silnych założeń. W przypadku, gdy założenia te nie są spełnione korzystamy z testów nieparametrycznych, które wymagają słabszych założeń niż testy parametryczne.

Testy nieparametryczne są szczególnie zalecane, gdy próba jest mała i założenie normalności nie może być przyjęte lub badane zmienne nie są mierzalne.

Uwaga. Jeżeli można przyjąć założenie normalności mając małą próbę lub dysponujemy bardzo dużą próbą, to wnioskowanie powinno być oparte na testach parametrycznych jako mocniejszych.

Testy nieparametryczne stosowane w ANOVA:

1) Test Kołmogorowa-Smirnowa do badania zgodności rozkładów dla dwóch populacji.

2) Test Smirnowa zgodności trzech rozkładów, stosowany przy założeniu równoliczności wszystkich prób.

3) Test rangowy Kruskala-Wallisa.

Test Kruskala-Wallisa jest stosowany do porównywania p populacji o dowolnych rozkładach ciągłych z dystrybuantami $F_1(x), F_2(x), \dots, F_p(x)$.

Z populacji tych losujemy niezależnie n_i elementów do i-tej próby ($i = 1, 2, \dots, p$).

Weryfikujemy hipotezę zerową

$$H_0\colon F_1(x) = F_2(x) = \cdots = F_p(x)$$

że rozkład wszystkich p populacji jest taki sam, przeciw hipotezie alternatywnej

H_1: Co najmniej dwie populacje mają istotnie różniące się rozkłady.

Do sprawdzenia hipotezy zerowej stosujemy statystykę H określoną wzorem:

$$H = \left(\frac{12}{n(n+1)}\sum_{j=1}^{p}\frac{T_j^2}{n_j}\right) - 3(n+1)$$

p jest liczbą populacji,

$$n = n_1 + n_2 + \cdots + n_p$$

T_j jest sumą rang w j-tej próbie, przy czym 1 nadane jest najmniejszej a n największej obserwacji (rangowanie jest wspólne dla wszystkich obserwacji, a powtarzające się obserwacje mają uśrednione rangi).

Statystyka H ma w przybliżeniu rozkład chi-kwadrat z $(p - 1)$ stopniami swobody, więc obszar krytyczny wyznaczony za pomocą kwantyla jest postaci:

$$H > \chi^2_{1-\alpha;p-1}$$

Przykład 4. W celu porównania trzech metod nauczania matematyki przeprowadzono badania testowe. Pobrano czteroelementowe próby losowe spośród każdej z trzech grup studentów, nauczanych metodami A, B i C. Wiadomości studentów sprawdzono odpowiednim testem.

Wyniki tego testu są podane w tablicy.

Metoda	Student			
	1	2	3	4
A	75	80	79	82
B	74	78	68	72
C	93	83	89	87

Otrzymanym wynikom testów odpowiadają rangi zestawione w tablicy.

Metoda	Ranga studenta				Suma rang
	1	2	3	4	
A	4	7	6	8	25
B	3	5	1	2	11
C	9	10	11	12	42

Statystyka testowa obliczona z eksperymentu wynosi

$$H_0 = \left(\frac{12}{(12)(13)}\,\frac{625+121+1764}{4}\right) - 39 = 9{,}2692$$

Ponieważ $H_0 > \chi^2_{0,95;2} = 5{,}991$, więc hipotezę zerową należy odrzucić.

Przykład 5. Producent kalkulatorów ma podjąć decyzję, w który z trzech typów baterii je wyposażać. Dwa czynniki determinują decyzję: cena i czas eksploatacji baterii. Przeprowadzono eksperyment polegający na wyposażeniu w trzy typy baterii po pięć kalkulatorów. Te 15 kalkulatorów eksploatowano do rozładowania baterii. Czasy eksploatacji oraz ich rangi są zestawione w tabeli:

Typ I (ranga)	Typ II (ranga)	Typ III (ranga)
15.5 (5)	20.5 (13)	13.3 (1.5)
17.2 (8)	18.3 (9)	18.9 (10)
13.3 (1.5)	21.2 (15)	19.5 (12)
20.6 (14)	15.7 (6)	15.2 (4)
19.2 (11)	16.3 (7)	13.8 (3)
Suma (39.5)	(50)	(30.5)

Rozwiązanie. Weryfikowana jest hipoteza nieparametryczna o jednakowym rozkładzie czasów zdatności dla trzech typów baterii. Do sprawdzenia hipotezy wykorzystana jest procedura KRUSKAL

Wyniki obliczeń przedstawione są w tablicy

level	sample size	average rank
typ1	5	7.9
typ2	5	10.0
typ3	5	6.1 .

test statistic H = 1.90841 Significent level = 0.385119

Stąd na poziomie istotności $\alpha = 0{,}05$ brak podstaw do odrzucenia nieparametrycznej hipotezy zerowej.

9. ANOVA dla zrandomizowanego układu bloków

Przypuśćmy, że chcemy porównać czasy potrzebne do komputerowego przetworzenia dziennych operacji finansowych w banku dla systemów przetwarzania danych oferowanych przez trzy firmy softwarowe, powiedzmy A, B i C.

Eksperyment całkowicie zrandomizowany można osiągnąć np. poprzez wybór 15 dni i losowe przydzielenie po 5 dziennych operacji finansowych dla każdego z systemów przetwarzania danych.

Lepszym sposobem przeprowadzenia eksperymentu, zapewniającym zdobycie więcej informacji w przeliczeniu na przeciętny czas przetwarzania jest wykorzystanie dowodów operacji bankowych tylko z 5 wylosowanych dni i przetworzenie tych danych przez wszystkie trzy systemy.

Aby wyeliminować odchylenia powstałe z porównań z dnia na dzień losowane są kolejności systemów do przetwarzania dziennych danych. Każde zadanie może być przedstawione jako blok złożony z trzech jednostek eksperymentalnych.

Bloki są zrandomizowane, ponieważ systemy komputerowe są losowo przydzielane jednostkom eksperymentalnym w bloku.

Systemy testowane do przetwarzania dziennych danych finansowych są uruchamiane w losowej kolejności po to, aby uniknąć ewentualnych obciążeń spowodowanych przez inne nieznane lub niemierzalne czynniki (wielkości dziennych obrotów, złożoności transakcji, itp.), które mogą wpływać na czasy przetwarzania danych.

Układ zrandomizowanych bloków służący do porównania p zabiegów zawiera b bloków, a każdy blok zawiera po p losowo ustawionych jednostek eksperymentalnych.

W tablicy podane są oznaczenia stosowane w ANOVA tego układu

b – liczba obserwacji w zabiegu, p – liczba obserwacji w bloku

$n = bp$ – liczba wszystkich obserwacji w eksperymencie

	Zabieg 1	Zabieg 2	...	Zabieg p
Sumy dla zabiegów	T_1	T_2	...	T_p
Średnie dla zabiegów	$\bar{T}_1$	$\bar{T}_2$	...	$\bar{T}_p$

	Blok 1	Blok 2	...	Blok b
Sumy dla bloków	B_1	B_2	...	B_b
Średnie dla bloków	$\bar{B}_1$	$\bar{B}_2$	...	$\bar{B}_b$

Formuły obliczeniowe ANOVA dla układu bloków losowych podane są w tablicy.

$CM = \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$ – poprawka dla średniej (correction for mean)

$SS = \sum_{i=1}^{n} y_i^2 - CM$ – ogólna suma kwadratów

$SST = \frac{1}{b}\sum_{j=1}^{p} T_j^2 - CM$ – suma kwadratów dla zabiegów

$SSB = \frac{1}{p}\sum_{i=1}^{b} B_i^2 - CM$ – suma kwadratów dla bloków

$SSE = SS - (SST + SSB)$ – suma kwadratów dla błędu

$MST = \frac{SST}{p-1}$ – średni kwadrat dla zabiegów

$MSB = \frac{SSB}{b-1}$ – średni kwadrat dla bloków

$MSE = \frac{SSE}{n-p-b+1}$ – średni kwadrat dla błędu

W układzie bloków zrandomizowanych suma kwadratów SS jest rozłożona na trzy składowe:

$$SS = SSB + SST + SSE$$

Układ bloków losowych można wykorzystać do testowania tej samej hipotezy zerowej i alternatywnej co dla układu całkowicie zrandomizowanego, mianowicie:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_p$$

H_1: Co najmniej dwie wartości oczekiwane dla zabiegów są różne.

Statystyka testowa

$$F = \frac{MST}{MSE}$$

przy założeniu prawdziwości hipotezy zerowej ma rozkład

$$F(p-1; n-p-b+1)$$

Ponieważ suma kwadratów dla bloków SSB jest miarą rozrzutu wyjaśnioną przez różnice między średnimi bloków, możemy także testować

H_0: Wartości oczekiwane dla wszystkich bloków są sobie równe,

H_1: Co najmniej dwie wartości oczekiwane dla bloków są różne.

Statystyka testowa

$$F = \frac{MSB}{MSE}$$

przy założeniu prawdziwości hipotezy zerowej ma rozkład

$$F(b-1; n-p-b+1)$$

Głównym celem jest testowanie różnic między wartościami oczekiwanymi dla zabiegów, niemniej układ bloków zrandomizowanych pozwala również testować efektywność blokowania, tj. istnienia różnic między blokami.

Jeżeli będzie brak podstaw do stwierdzenia istotnych różnic między oczekiwanymi wartościami dla bloków, to blokowanie nie zmniejsza błędu eksperymentu, co oznacza nieefektywność blokowania.

W takiej sytuacji pomijana jest informacja wynikająca z blokowania, ponieważ blokowanie zmniejszyło jedynie liczbę stopni swobody związanej z estymatorem błędu modelu nie dając wiele w zamian.

Testy F dla układu losowego bloków podane są w tablicy.

Test porównania wartości oczekiwanych dla zabiegów

Weryfikowana jest hipoteza zerowa H_0: $\mu_1 = \mu_2 = \cdots = \mu_p$
przeciw H_1: Co najmniej dwie wartości oczekiwane dla zabiegów są różne.

Statystyka testowa: $F = \frac{MST}{MSE}$

Obszar krytyczny: $F > F_{1-\alpha;p-1;n-p-b+1}$

Test porównania wartości oczekiwanych dla bloków

Weryfikowana jest hipoteza zerowa

H_0: Wartości oczekiwane dla wszystkich bloków są sobie równe,
przeciw hipotezie H_1: Co najmniej dwie wartości oczekiwane dla bloków są różne.

Statystyka testowa: $F = \frac{MSB}{MSE}$

Obszar krytyczny: $F > F_{1-\alpha;b-1;n-p-b+1}$

Przedziały ufności dla różnic wartości oczekiwanych ze względu na zabiegi i bloki podane są w tablicy

Przedział ufności dla różnicy wartości oczekiwanych związanych z zabiegami

$$\left(\bar{T}_i - \bar{T}_j\right) \mp t_{1-\frac{\alpha}{2};n-p-b+1} s\sqrt{\frac{2}{b}}$$

Przedział ufności dla różnicy wartości oczekiwanych związanych z blokami

$$\left(\bar{B}_i - \bar{B}_j\right) \mp t_{1-\frac{\alpha}{2};n-p-b+1} s\sqrt{\frac{2}{p}}$$

Przykład 6. Wycena prywatyzowanego przedsiębiorstwa państwowego poprzedzona jest szczegółową analizą wartości majątku, potencjału produkcyjnego, możliwości przestawienia produkcji, sposobów zabezpieczenia socjalnego pracowników, itp. Szacowanie wartości majątku przeprowadzają specjalistyczne firmy zajmujące się wyceną. Przeszacowanie wartości przedsiębiorstwa zmniejsza szanse prywatyzacji firmy, natomiast zaniżenie wartości zmniejsza przychód z prywatyzacji.

W celu zmniejszenia ryzyka popełnienia błędu przedstawiciel odpowiedniego ministerstwa zamierza porównać średnie oszacowania wartości trzech niezależnych firm wyceniających majątek zanim zleci jednej z nich dokonanie oszacowania wartości rynkowej prywatyzowanego przedsiębiorstwa.

Przedstawiciel zebrał informacje o wycenie majątku tych samych czterech przedsiębiorstw przez każdą z rozważanych trzech firm wyceniających.

Uzyskane dane o wycenach (w mln zł) są podane w tabeli.

Firma wyceniająca	Wycena przedsiębiorstwa 1	2	3	4
A	4.6	6.2	5.0	6.6
B	4.9	6.3	5.4	6.8
C	4.4	5.9	5.4	6.3

a) Przeprowadzić analizę wariancji dla przeprowadzonych wycen. Na poziomie istotności $\alpha = 0{,}05$ sprawdzić, czy są istotne różnice między oczekiwanymi wycenami dla zabiegów i bloków.

b) Wyznaczyć 90-procentowy przedział ufności dla różnic między oczekiwanymi wycenami dla firm wyceniających A i B.

Rozwiązanie. a) W przeprowadzonym eksperymencie dane zostały zebrane zgodnie z planem zrandomizowanych bloków. Możemy spodziewać się bardziej zbliżonych wycen majątku tego samego zakładu przez firmy przeprowadzające wycenę niż między wycenami czterech zakładów.

Obliczenia elementów analizy wariancji dają:

$$\sum_{i=1}^{12} y_i = 67{,}8,\ \sum_{i=1}^{12} y_i^2 = 390{,}28,$$

$$CM = \frac{\left(\sum_{i=1}^{12} y_i\right)^2}{12} = 383{,}07,\ SS = \sum_{i=1}^{12} y_i^2 - CM = 7{,}21,$$

$$SST = 0{,}26,\ SSB = 6{,}763333,\ SSE = 0{,}186667$$

$$MST = 0{,}13,\ MSB = 2{,}254,\ MSE = 0{,}03111$$

Tablica ANOVA

Źródło	Stopnie swobody	*SS*	*MS*	*F*
Firma wyceniająca (*T*)	2	0.260	0.130	4.19
Przedsiębiorstwo (*B*)	3	6.763	2.254	72.71
Błąd	6	0.187	0.031	
Ogółem	11	7.210		

Statystyka F do testowania równości wartości oczekiwanych wycen sporządzanych przez trzy firmy ma rozkład $F(2;6)$. Ponieważ $F_0 = 4{,}19 < F_{0{,}95;2;6} = 5{,}14$, więc na poziomie istotności $\alpha = 0{,}05$ nie ma podstaw do odrzucenia hipotezy zerowej. Taką samą decyzję można podjąć na podstawie obliczonej istotności testu $\alpha_0 \approx 0{,}07$.

Statystyka F do testowania hipotezy

H_0: Nie ma różnic między przeciętnymi wycenami przedsiębiorstw

ma rozkład $F(3; 6)$. Ponieważ $F_0 = 72{,}71 > F_{0,95;3;6} = 4{,}76$, więc na poziomie istotności $\alpha = 0{,}05$ stwierdzamy istotne różnice między przeciętnymi wycenami przedsiębiorstw.

b) Podstawiając do wzoru na przedział ufności dla różnicy $\mu_A - \mu_B$ przeciętnych wycen przez firmy A i B otrzymujemy 95-procentową realizację przedziału ufności

$$(-0{,}491;\ -0{,}009)$$

Ponieważ jednostkami są miliony zł, więc różnica pomiędzy przeciętnymi wycenami dla firm wyceniających A i B zawiera się w przedziale od –491 tys. zł do –9 tys. zł.

10. Podsumowanie

Celem wykładu było zapoznanie studentów z kierunku matematyka w technice z podstawowymi elementami planowania eksperymentu.

Omówione były założenia oraz metody wnioskowania dla układów całkowicie zrandomizowanych.

W kontekście tych eksperymentów przedstawione zostały konstrukcje przedziałów ufności dla wartości oczekiwanych.

Pokazane zostały możliwości zastosowań procedury wielokrotnych porównań Tukeya oraz testu rangowego Kruskala-Wallisa.

Na koniec rozszerzono możliwości ANOVA dla zrandomizowanego układu bloków.

Teoria obrazowana była praktycznymi przykładami.

11. Zestaw zadań W12

1. (Krysicki 5.2). Zmierzono długości czasów świecenia trzech typów żarówek, otrzymując (w h):

dla typu 1: 1802, 1992, 1854, 1880, 1761, 1900;
dla typu 2: 1664, 1755, 1823, 1862;
dla typu 3: 1877, 1710, 1882, 1720, 1950.

Na poziomie istotności $\alpha = 0{,}05$ zweryfikować hipotezę, że wartości przeciętne czasów świecenia żarówek tych typów są jednakowe.

Odp.: Na założonym poziomie istotności nie ma podstaw do odrzucenia hipotezy o równości wariancji oraz nie ma podstaw do odrzucenia hipotezy o równości przeciętnych czasów świecenia żarówek.

2. (Krysicki 5.3). Spośród trzech odmian ziemniaków każdą uprawiano na 12 działkach tej samej wielkości i rodzaju. Działki te podzielono na 4 grupy po 3 działki i dla każdej grupy zastosowano różny rodzaj nawozu. Plony w q zestawione w tabeli:

Odmiana	Nawóz			
	1	2	3	4
1	5,6 6,1 5,9	6,6 6,7 6,6	7,7 7,3 7,4	6,3 6,4 6,3
2	5,7 4,9 5,1	6,5 6,7 6,6	6,9 7,1 6,5	6,6 6,7 6,7
3	6,3 6,1 6,3	6,5 6,4 6,2	6,6 6,6 6,8	6,3 6,1 6,0

Na poziomie istotności $\alpha = 0{,}05$ zweryfikować następujące hipotezy:

a) wartości przeciętne plonów dla różnych odmian nie różnią się istotnie niezależnie od stosowanego nawozu,

b) wartości przeciętne plonów dla różnych nawozów nie różnią się istotnie niezależnie od odmiany,
c) interakcja między odmianami i nawozami jest równa 0.

3. Rozwiązać zadanie 5.7 z Krysickiego.
4. Korzystając ze wspomagania komputerowego rozwiązać przykład 3 z wykładu.
5. Korzystając ze wspomagania komputerowego rozwiązać przykład 4 z wykładu.
6. Korzystając ze wspomagania komputerowego rozwiązać przykład 6 z wykładu.