Iwona Mróz,

Instytut Fizyki Doświadczalnej,

Uniwersytet Wrocławski

**„Podstawy statystyki i analizy danych” – materiały do wykładu nr 12 z dnia 10.01.2020 r**

Niniejsze materiały mają charakter roboczy. Bardzo proszę o zgłaszanie zauważonych błędów, braków, niedociągnięć i niejasności. Prośba dotyczy też przypisów. Z góry dziękuję za pomoc☺.

**Porównanie więcej niż dwóch grup niezależnych**

**Metoda parametryczna - jednoczynnikowa analiza wariancji (ang. one-way ANOVA, nazwa pochodzi od ANalysis Of Variance)**

***Opracowano na podstawie podręcznika Andrzeja Stanisza „Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny”, StatSoft Polska Sp. z o.o., Kraków 2006, tom 1, str. 260-262.***

Jednoczynnikowa analiza wariancji jest metodą parametryczną, która pozwala na porównanie więcej niż dwóch grup niezależnych. Jest to podstawowa metoda doświadczalnictwa, wymagająca starannego zaplanowania eksperymentu. Nieparametrycznym odpowiednikiem ANOVY jest test Kruskala Wallisa, omówiony poniżej.

Dla jednoczynnikowej analizy wariancji przyjmujemy, ze tylko jeden czynnik ma wpływ na wynik przeprowadzanych badań. Opracowano ANOVĘ dwu – lub wieloczynnikową (two-way ANOVA, multi-way ANOVA, odpowiednik metody ANOVA dla grup zależnych oraz inne metody oparte o analizę wariancji, jednak ich omówienie wymaga więcej czasu niż mamy do dyspozycji w tym kursie.

Rozpatrywany czynnik Może mieć kilka poziomów, nazywanych też zabiegami.

ANOVA porównuje średnie w populacjach, ale robi to przy pomocy oceny zmienności w sposób opisany poniżej.

Aby móc poprawnie stosować analizę wariancji należy przestrzegać zasad randomizacji (będziemy je omawiać na jednych z kolejnych zajęć). Ponadto, muszą być spełnione podstawowe założenia:

1. Badana cecha statystyczna jest mierzalna.

2. Próby pobieramy z *k* niezależnych populacji, dla każdej z tych populacji badana cecha musi mieć rozkład normalny. Źródła podają, ze ANOVA jest stosunkowo odporna na niewielkie odstępstwa od normalności, ale na razie zostawiamy ten problem doświadczonym statystykom.

3. Wszystkie rozkłady (dla wszystkich rozważanych populacji) mają jednakowe wariancje (założenie o jednorodności wariancji).

Z każdej z *k* populacji losujemy próbę o liczebności , tak, aby zachodziło , gdzie *n* to liczba wszystkich (niezależnych) obserwacji.

Hipotezy: zerowa i alternatywna mają postać:

H0:  (średnie w grupach (populacjach) są jednakowe).

H1: co najmniej dwie średnie różnią się między sobą.

Procedura wymaga wyliczenia średniej arytmetycznej ze wszystkich pomiarów (ang. grand mean) oraz średnich arytmetycznych dla każdej próby.

ANOVA wykorzystuje fakt, że zmienność opisywaną przez sumę kwadratów odchyleń pojedynczych pomiarów od średniej ze wszystkich pomiarów możemy rozbić na sumę kwadratów odchyleń pomiarów z danej grupy od średnich w tej grupie (bierzemy sumę kwadratów odchyleń dla wszystkich grup) oraz sumę kwadratów odchyleń średnich w grupach od „grand mean”.

Mianowicie, niech oznacza *j-ty* pomiar (*j = 1,2,…,ni*), w *i-tej* próbie (*i = 1,2,…,k*). Niech oznacza średnią w *i-tej* próbie, a to średnia ze wszystkich pomiarów (grand mean).

Oczywiście zachodzi .

Można pokazać, że zachodzi:

Pierwszy składnik po prawej stronie opisuje tzw. zmienność wewnątrzgrupową, a drugi składnik - zmienność między grupami. W drugim składniku po prawej stronie, sumowanie po *j* oznacza, że kwadrat różnicy wyznaczony dla *i*-tej próby mnożymy przez liczbę obserwacji w tej próbie. Z lewej strony mamy zmienność całkowitą.

Często używa się oznaczeń:

SS całkowita (całkowita suma kwadratów, sum of squares) = SS reszta (SS error) + SS pomiędzy grupami (SS effect lub SS Treatment)

Dla sum kwadratów możemy określić liczby stopni swobody. Mianowicie:

*df całkowite = n-1* , *df reszt = n-k*, *df grup = k-1.*

Zachodzi:

*df całkowite = df reszt + df grup*

Następnie wyznaczamy MS (mean squares) reszt (w grupach) i pomiędzy grupami:

 - średni kwadratowy błąd

 - średni kwadratowy efekt zabiegu

Ostatnim krokiem jest wyliczenie statystyki *F* (Fishera-Snedecora), o  *k-1* i *n-k* stopniach swobody.



Przy założeniu prawdziwości hipotezy zerowej statystyka ma rozkład Fishera-Snedecora), o  *k-1* i *n-k* stopniach swobody. Zbiór krytyczny konstruujemy dla danego poziomu istotności. Mówiąc obrazowo, statystyka F ocenia stosunek zmienności pomiędzy grupami do zmienności wewnątrzgrupowej. Jeżeli wartość statystyki dużo większa od 1, przemawia to za odrzuceniem hipotezy zerowej, bo „rozrzut” średnich jest większy od rozrzutu pomiarów w grupie.

UWAGI PRAKTYCZNE:

1. Przy sprawdzaniu założeń ANOVY, oceny jednorodności wariancji dokonujemy np. przy pomocy testów Levene’a lub Bartletta. Te testy sprawdzają hipotezy zerowe mówiące, z wszystkie wariancje w badanych populacjach są takie same.

2. Przy przeprowadzaniu analizy zawsze podaje się wartości wszystkich SS, df, MS oraz wartość testu F. Najczęściej wyniki przedstawia się w tabeli analizy wariancji.

**Testy post-hoc**

***Opracowano na podstawie podręcznika Andrzeja Stanisza „Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny”, StatSoft Polska Sp. z o.o., Kraków 2006, tom 2, str. 393, 414-417.***

W przypadku odrzucenia H0 w procedurze ANOVA chcemy się dowiedzieć, które pary średnich różnią się między sobą. Pozwalają na to tzw. testy post-hoc. Testy te stosujemy WYŁĄCZNIE w przypadku odrzucenia H0 ANOVY. Jeżeli nie ma podstaw do odrzucenia H0, stosowanie testów post-hoc jest bez sensu. Testów post-hoc jest bardzo dużo i żaden z nich nie jest bezwzględnie lepszy od pozostałych. Aby wybrać test, który będzie najlepszy dla naszych konkretnych zastosowań, poniżej przypominamy lub omawiamy ważne cechy testów statystycznych.

Błędy, które popełniamy przy wnioskowaniu statystycznym polegają na:

- Odrzuceniu hipotezy zerowej gdy jest ona prawdziwa (błąd pierwszego rodzaju, prawdopodobieństwo jego popełnienia to poziom istotności α.

- Przyjęciu hipotezy zerowej, gdy w rzeczywistości jest ona fałszywa (błąd drugiego rodzaju. prawdopodobieństwo jego popełnienia oznaczamy β).

Błędy pierwszego i drugiego rodzaju nie są niezależne, zależność ilustruje dołączony slajd. Wynika z niego, że nie można bezkarnie zmniejszać α, gdyż rośnie wtedy β. W różnych dziedzinach nauki akceptuje się różne poziomy β, podobnie jak to się dzieje z α.

Moc testu wyliczamy jako *1*- β. Określa ona prawdopodobieństwo odrzucenia hipotezy zerowej gdy jest ona fałszywa. Moc testu zależy m.in. od wielkości prób i poziomu istotności. Różne testy sprawdzające podobne hipotezy statystyczne mają różną moc przy tym samym α. Testy parametryczne są z reguły mocniejsze od ich nieparametrycznych odpowiedników, dlatego powinny być testami pierwszego wyboru.

Moc testu statystycznego wiąże się z innymi cechami, takimi jak czułość czy konserwatywność.

Czułość testu odpowiada za jego zdolność do wykrywania statystycznie istotnych różnic. Zaletą takich testów jest większa moc, wadą – możliwość wzbudzenia „fałszywego alarmu” i zwiększeniem prawdopodobieństwa popełnienia błędu pierwszego rodzaju.

Konserwatywność testu to cecha, która zmniejsza zdolność testu do wykrywania istotnych różnic. Za to, zwiększa się prawdopodobieństwo, że wykryte różnice rzeczywiście istnieją w populacji. Wraz ze wzrostem konserwatywności testu maleją jego czułość i moc.

***Cechy testów statystycznych takie jak czułość i konserwatywność maja szczególne znaczenie przy wyborze odpowiedniego testu post-hoc.***

Przykładem testu konserwatywnego jest test Scheffego. W literaturze można znaleźć informację, że ten test jest wręcz przesadnie konserwatywny. Z kolei testem czułym jest test Duncana. Spotkałam się z opinią, że dla początkujących dobry jest dość konserwatywny test Tukeya. Oczywiście wybór testu post-hoc zależy od rozwiązywanego problemu badawczego.

Istnieją też testy post-hoc zostały do wykonania po odrzuceniu H0 testem Kruskala Wallisa (nieparametrycznym odpowiedniku ANOVY).

**Nieparametryczny odpowiednik metody ANOVA – test Kruskala-Wallisa**

***Opracowano na podstawie podręcznika Andrzeja Stanisza „Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny”, StatSoft Polska Sp. z o.o., Kraków 2006, tom 1, str. 371-373.***

Test Kruskala-Wallisa służy do porównania rozkładów cechy dla więcej niż dwóch grup niezależnych. Wykonujemy go wtedy, gdy nie są spełnione założenia metody ANOVA. Przeprowadzenie testu wymaga nadania obserwacjom rang (zob materiały do wykładu nr 12, test U Manna-Whitneya).

Założenia:

- Danych jest k populacji, badana cecha statystyczna ma w populacji rozkład typu ciągłego,

- dane można rozpatrywać w skali porządkowej (można je uporządkować),

- wylosowano k próbek o liczebnościach n1, n2,..., nk, przy czym ni  5 dla i = 1,...,k.

- Oznaczamy dystrybuanty w populacjach: F1(x), ... , Fk(x).

Hipotezy:

Weryfikacja hipotezy zerowej:

H0: F1(x) = F2(x) = ... = Fk(x) (wszystkie dystrybuanty są równe)

Wobec hipotezy alternatywnej:

H1: Istnieje przynajmniej jedna para dystrybuant Fi(x) , Fj(x), ij, dla których Fi(x)  Fj(x).

Używana statystyka H ma postać:

H = (12/(n(n+1)))(Ti2/ni) – 3(n+1)

Sumowanie przebiega po wszystkich próbach, i = 1,...,k.

n oznacza całkowitą liczbę obserwacji, n = n1+n2+...+nk,

Ti jest sumą rang nadanych wartościom w i-tej próbie.

Jeżeli hipoteza zerowa jest prawdziwa to statystyka H ma asymptotyczny rozkład 2 o k-1 stopniach swobody (pojęcia będą wyjaśnione).

Statystyka H w powyższej postaci nie uwzględnia rang wiązanych. Jeżeli takie rangi pojawiają się w rozpatrywanych próbach, stosujemy poprawkę:

P = 1-((l3-l))/(n3-n)

gdzie sumowanie przebiega po wszystkich próbach, a l jest liczbą pomiarów mających tę samą rangę wiązaną.