Wrocław, 21.12.2022 r.

**PODSTAWY ANALIZY DANYCH – PRAKTYCZNE WARSZTATY**

**ZADANIA DO SAMODZIELNEGO PRZEĆWICZENIA, LISTA NR 6**

W folderze „ONE WAY ANOVA” zamieszczonym w zakładce „Pliki z danymi” znajdą Państwo przykładów na jednoczynnikową analizę wariancji.

ANOVA wymaga, aby liczebność każdej z prób była odpowiednio duża (przyjmuje się 10 lub więcej obserwacji w każdej próbie). Dla naszych przykładów to założenie nie zawsze jest spełnione, trzeba o tym wiedzieć, ale na razie proszę się nie przejmować😊.

Pamiętajmy, że przed przeprowadzeniem testu statystycznego musimy przyjąć pewien poziom istotności, którego nie wolno zmieniać w trakcie wnioskowania. Przyjmijmy roboczo, że dla przykładów z tej listy przyjmowane poziomy istotności nie będą większe niż 0,05.

1. Dla podanych przykładów proszę przećwiczyć ANOVĘ jednoczynnikową i postarać się rozstrzygnąć, czy na przyjętym poziomie istotności należy odrzucić hipotezę zerową mówiącą, że wartości średnie badanej cechy w porównywanych populacjach są takie same.

Proszę posłużyć się funkcją aov. Aby wszystko działało poprawnie, należy przygotować dane tak, aby w jednej kolumnie (domyślnie „values”) znalazły się wszystkie wartości zmiennej objaśnianej, a w drugiej kolumnie (domyślnie „ind”) - wartości zmiennej grupującej (jakościowej), wskazującej, z której populacji pochodzi obserwacja.

Jeżeli mamy ramkę danych, w której wartości zmiennej objaśnianej są ułożone w kolumnach o nazwach odpowiadających wartościom zmiennej grupującej, do wymaganej postaci przejdziemy przy pomocy fukkcji „stack”:

ramka\_danych\_zmodyfikowana=stack(ramka\_danych)

Następnie tworzymy model One-Way-ANOVA:

model=aov(values~ind, data=ramka\_danych\_zmodyfikowana).

Materiał potrzebny do przećwiczenia kolejnych zadań omówimy już w Nowym Roku.

2. Czy dla rozważanych przykładów spełnione są założenia ANOVY o normalności rozkładów badanej cechy w każdej z populacji (można ocenić na podstawie wykresy normalności reszt)? Czy wszystkie rozkłady mają takie same wariancje?

Równość wariancji możemy zbadać testem Bartletta znajdującym się w pakiecie „base”: bartlett.test(values~ind, data=ramka\_danych\_zmodyfikowana) lub testem Levene’a z pakietu „car”: leveneTest(values~ind, data=ramka\_danych\_zmodyfikowana).

Test Levene’a działa naprawdę dobrze, gdy układ jest zrównoważony, tzn. we wszystkich próbach mamy tyle samo obserwacji oraz gdy obserwacji w próbie jest odpowiednio dużo.

Dla których z przykładów układy są zrównoważone?

3. Jeżeli dla któregoś z przykładów test F wskaże, że na przyjętym poziomie istotności należy odrzucić H0, proszę sprawdzić, które średnie (w populacjach) różnią się od siebie przy pomocy testu HSD Tukeya, będącego tzw. testem post-hoc:

TukeyHSD(model). Proszę pamiętać, że test Tukeya jest „gorszy” niż sama ANOVA, więc, co do zasady, jego wyniki interpretuje się jeśli test ANOVY sugeruje odrzucenie hipotezy o równości średnich we wszystkich populacjach.

4. Proszę wybrać dowolny przykład i spróbować ręcznie wyliczyć tabelę ANOVY (zawierającą sumy kwadratów, liczby stopni swobody i wartości MS) oraz obliczyć wartość statystyki F. Czy wyniki zgadzają się w tymi, które wylicza R? To ćwiczenie wydaje się żmudne, ale raz w życiu warto je wykonać.

5. Jeżeli dla któregoś z przykładów nie są spełnione założenia o normalności rozkładów lub równości wariancji, proszę zamiast ANOVY przeprowadzić test Kruskala-Wallisa, tzw nieparametryczny odpowiednik ANOVY (kruskal.test(values~ind, data=ramka\_danych\_zmodyfikowana). Proszę zwrócić uwagę, jak należy sformułować hipotezę zerową i alternatywną dla testu Kruskala-Wallisa.

Jeżeli dla wszystkich przykładów założenia są spełnione, proszę wybrać dowolny przykład i przećwiczyć na nim test Kruskala-Wallisa.

Proszę pamiętać, że jeśli można przeprowadzić ANOVĘ, to dodatkowe przeprowadzenie jej nieparametrycznego odpowiednika nie ma sensu. Można takie postępowanie porównać do poprawiania linijką pomiaru dokonanego prawidłowo przy pomocy suwmiarki. O powodach, dla których tak jest porozmawiamy niedługo😊

6. Proszę obejrzeć film szkoleniowy zamieszczony na YouTube pod adresem:

<https://www.youtube.com/watch?v=ZxS17lfPF94>

Link do filmu umieściłam też w naszym zespole na MS Teams. Na filmie omówiono interpretację wykresów interakcji dla ANOVY dwuczynnikowej (2×2).