PEiAR Laboratorium 9 **29.04.2024**

TEST LOSOWOŚCI PRÓBY, TESTY NIEZALEŻNOŚCI CECH

Zad.1 (W jaki sposób działa test serii Walda-Wolfowitza? Kiedy go wykorzystujemy?)
Z urny zawierającej 100 kul ponumerowanych wybieramy 10 kul kolejno bez zwracania. Które z następujących prób można uznać za losowe? Skorzystaj z paczki *'randtests'* i użyj testu serii.
[a] 5, 17, 21, 29, 33, 45, 56, 66, 72, 88
[b] 45, 12, 77, 64, 4, 93, 21, 37, 90, 95
[c] 63, 90, 47, 16, 86, 74, 97, 13, 26, 3
[d] 1, 81, 11, 21, 91, 71, 31, 61, 41, 51

Zad.2
Zmierzono wzrost studentów I roku Informatyki:
176, 182.5, 166, 175, 175.5, 161.5, 173, 165, 186, 170.5, 158, 163.5
oraz wzrost studentów II roku Informatyki:
168, 172, 163, 171.5, 177, 190, 172.5, 164, 183.5, 171, 157.5, 166.
Czy można twierdzić że wzrost studentów I roku ma ten sam rozkład co wzrost studentów II roku? Przeprowadź test Kołmogorowa-Smirnowa oraz test serii. W przypadku testu serii, najpierw utwórz ramkę frame_ z kolumną 'Height' przechowującą skonkatenowane oba wektory firstYear_ oraz secondYear_. Następnie dołącz kolumnę 'Year' z pierwszymi dwunastoma wartościami: 1 oraz dwunastoma kolejnymi: 2. Dalej, uporządkuj ramkę rosnąco według wzrostu. Można to zrobić używając funkcji *order()* w następujący sposób:
> *frame_ = frame_[order(frame_$Height)*, *]*
Uwaga: ważne jest by nie pominąć przecinka przed kwadratowym nawiasem zamykającym. Puste miejsce oznacza, że domyślnie, po uporządkowaniu ramki, chcemy widzieć wszystkie jej kolumny (np. argument *c(1, 3)*, spowodowałby widok jedynie pierwszej i trzeciej kolumny w uporz. ramce). Na końcu użyj testu serii na wektorze wartości z kolumny 'Year'.

Zad.3 (Testy niezależności: χ2 i Fishera)
W pewnej grupie społecznej zbadano wykształcenie i częstość chodzenia do teatru w ciągu jednego roku. Okazało się, że wśród osób z niższym wykształceniem, 23 osoby poszły do teatru 0 razy, 18 osób 1 raz natomiast 28 osób 2 razy.
Wśród osób ze średnim wykształceniem: 15 nie było w teatrze w danym roku ani raz, 30 osób było 1 raz i 11 osób było 2 razy. Natomiast wśród osób z wyższym wykształceniem 8 osób było 0 razy, 22 osoby 1 raz i 31 osób 2 razy.
Przy użyciu testu χ2 Pearsona (tego zwykłego oraz tego z paczki *'TeachingDemos'*) zweryfikuj hipotezę o niezależności częstotliwości uczęszczania do teatru od wykształcenia respondentów.

Zad.4
Zbadano, że wśród studentów pewnej grupy, cztery kobiety myją zęby 1 raz w tygodniu, dwie 3 razy i pięć 7 razy w tygodniu. Podobnie, 6 mężczyzn myje zęby 1 raz w tygodniu, czterech 3 razy i jeden pan myje zęby codziennie. Zweryfikuj hipotezę o niezależności płci w tej grupie od dbania o higienę jamy ustnej.

Zad.5 (Wykresy - ciąg dalszy…)
Narysuj wykres funkcji *y = x·sin(x)* na przedziale [-30, 30].
I sposób:
wykonamy zadanie używając funkcji *curve(...)*. Zdefiniuj powyższą funkcję pod zmienną *fun_* i wywołaj polecenie *curve()* z szeregiem doprecyzowujących argumentów:
type = ... // *'l', 's', 'o', 'h', 'c'*,..., domyślnie *'l'*//
from =..., to=... // równoważnie *xlim=c(..., ...)* //

n =... // *101* domyślnie //
ylim =... // ustaw *[-50, 50]*
cex.axis =...// *1* domyślnie //
lwd=...// *1* domyślnie (wartości naturalne) //
bty=... // *'o', 'n', 'u'*, domyślnie *'o'* //
lty=... // *1, 2, 3, 4, 5, 6* (domyślnie *1*) //
xlab=...,
main=...,
col=...// *1, 2, 3, 4, 5, ...* (domyślnie *1*) //
Poeksperymentuj z wartościami parametrów aby domyślić się za co odpowiadają.
Dorysuj na wykresie dwie przerywane proste *y = x* oraz *y = - x* stykające się z lokalnymi ekstremami badanej funkcji. Użyj w tym celu funkcji *abline(...)* // której kiedyś używaliśmy do rysowania poziomej linii na zadanej wysokości: *abline(h=...)* //przyjmującej tym razem – jako swoje dwa pierwsze argumenty – najpierw (!) współczynnik *b* oraz współczynnik *a* funkcji liniowej *y = ax + b*, której wykres ma zostać narysowany.
Chcemy też aby znaki podpisujące oś pionową ułożone były poziomo, w tym celu w definicji wykresu dodaj kolejny argument *las = 1* (możliwe wartości *0, 1, 2* lub *3*).
Na końcu, chcemy do rysunku dodać informację tekstową (ze strzałką) wskazującą na to że wypoziomowaliśmy znaki podpisujące pionową oś. W tym celu rysujemy strzałę:
>*arrows(..arg1.., ..arg2.., ..arg3.., ..arg4.., angle = 20, length = 0.15, code = 2, lwd = 2, col = 4)*
// najważniejsze są tu cztery pierwsze argumenty stanowiące współrzędne początku strzały (dwa pierwsze argumenty) oraz współrzędne końca strzały (dwa kolejne argumenty) // a potem umieszczamy tekst poleceniem:
>*text(..arg1.., ..arg2.., labels='...treść annotacji…')*
II sposób:
użyjemy paczki ***'ggplot2'***. W celu wygenerowania dowolnego wykresu trzeba najpierw zdefiniować jego bazę / konstruktor:
>*base_ = ggplot(data.frame(x = c(-30, 30)), aes(x))*
Następnie, zbudujmy na zdefiniowanym konstruktorze wykres funkcji fun_:
>*plot_ = base_ + stat_function(fun = fun_, geom = 'line', n = 100)*
Dalej, naniesiemy na istniejący już wykres, nowy zielony punktowy wykres funkcji wartość bezwzględna:
>*plot_1 = plot_ + stat_function(fun = abs, geom = 'point', n = 100, colour = 'green')*
i jeszcze czerwony schodkowy wykres funkcji *cosinus*
>*plot_2 = plot_1 + stat_function(fun = cos, geom = 'step', n = 60, colour = 'red')*

Narysuj poznanymi sposobami wykresy funkcji:

a) $y = \dfrac{\left|\left(\log_3(x-2)+\ln(x-4)\right)\right|}{2}$ na przedziale *[4, 30]*

b) $y = e^{|\sin(2x)\cdot\cos(2x)|}$ na przedziale *[-20, 20]* .

Zad.6 (Testy zgodności – dla więcej niż 2 prób…)
Pewna ciągła cecha X została zbadana na losowo wybranych osobnikach w czterech różnych populacjach. Otrzymano próbki:
x1: 11.2, 5, 45, 3.42, 24, 54, 33.5, 18
x2: 72.5, 6, 22.4, 70, 101, 23, 42, 33
x3: 32, 44.6, 44.9, 64, 2.3, 8, 93, 100, 22, 34
x4: 20, 19, 18, 2, 3.6, 4.6, 8
Zweryfikuj hipotezę, że rozkłady tej cechy we wszystkich czterech populacjach są identyczne. Użyj testu Kruskala – Wallisa o identyczności rozkładów.
Uwaga – jako argument należy podać listę zbudowaną ze wszystkich próbek

> *kruskal.test(list_)*
Zilustruj otrzymany wynik testu przy użyciu wykresu pudełkowego dla analizowanej listy, np.:
> boxplot(list_, boxwex=0.7, las=1, col=4)