



StatSoft

Porovnání nezávislých výběrů – neparametrická ANOVA

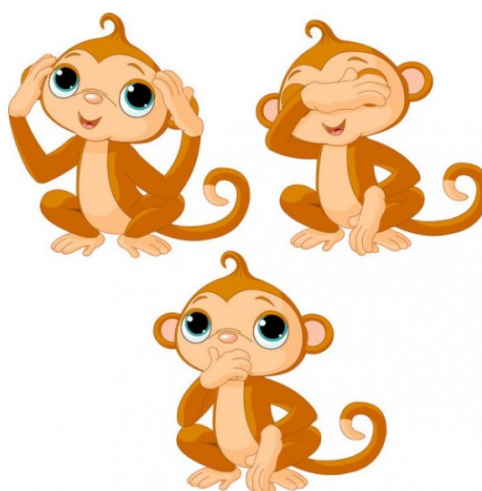
V tomto příspěvku si ukážeme práci v softwaru *STATISTICA*, konkrétně si vyzkoušíme aplikaci metody Kruskal-Wallisova ANOVA.

Data

V souboru [Kruskal_CZ.sta](#) jsou data týkající se skupiny opic stejného druhu, které byly náhodně rozděleny do tří experimentálních skupin.

Každému exempláři byla ukázána série předmětů. Úkolem bylo vybrat určitý předmět. Za správnou volbu byla opice odměněna. Pro korektní rozhodnutí bylo v první skupině důležité dobře určit tvar, ve druhé skupině barvu a ve třetí skupině velikost.

Na základě počtu pokusů potřebných k úspěšnému výběru chceme rozhodnout, zda jsou úlohy založené na rozpoznání tvaru, barvy a velikosti pro konkrétní druh opic stejně obtížné.



	1 skupina	2 počet pokusů
1	tvar	6
2	tvar	11
3	tvar	12
4	tvar	20
5	tvar	24
6	tvar	21
7	tvar	18
8	tvar	15
9	tvar	14
10	tvar	10
11	tvar	8
12	tvar	14

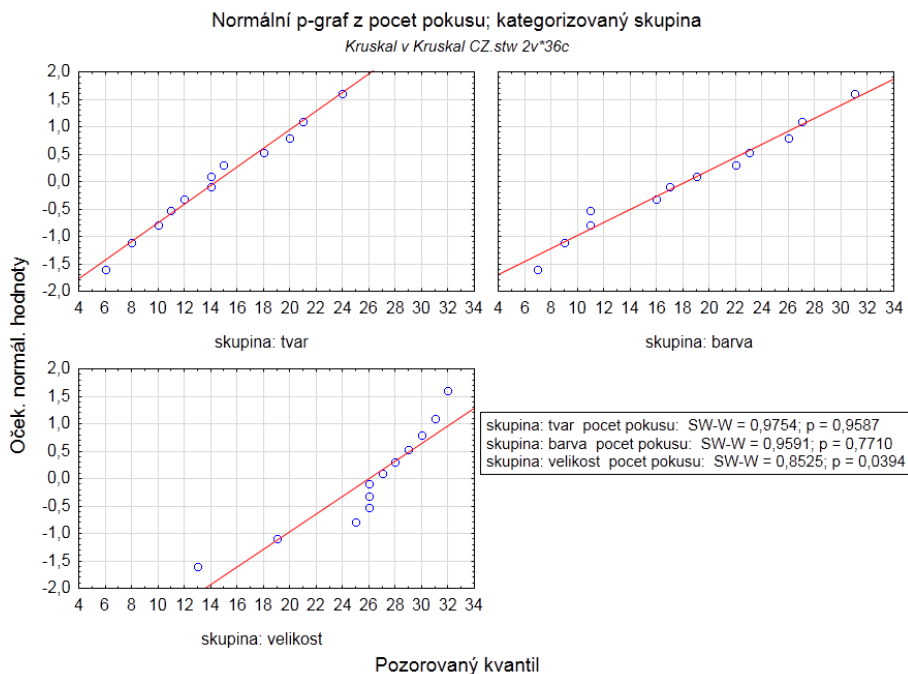
Řešení

Zdá se, že vhodnou metodou, jak tuto úlohu řešit by mohla být jednofaktorová ANOVA (analýza jednoduchého třídění). Než však začneme analýzu provádět, podívejme se, zda jsou splněny její předpoklady. Tyto by bylo vhodnější vyzkoušet např. na pilotním vzorku (na kterém bychom si opravdu pouze ověřili předpoklady). Nicméně obvykle však jiná data k dispozici nejsou, a proto ověříme předpoklady na datech, která máme k dispozici.

Prvním předpokladem, který je potřeba ověřit, je normalita závislé proměnné ve všech skupinách.

Možností v softwaru *STATISTICA* je více, tentokrát k tomu využijeme normální pravděpodobnostní graf. Ze základní nabídky vybereme *Grafy* -> *2D Grafy* -> *Normální pravděpodobnostní grafy...* Na záložce *Zákl. nastavení* klikneme na tlačítko *Proměnné*: a vybereme proměnnou *počet pokusů*.

V oddílu *Statistiky* zaškrtneme volbu *Shapiro-Wilkův test* pro testování normality. Na záložce *Kategorizovaný* zaškrtneme v oddílu *Kategorie X* volbu *Zapnuto*, čímž se zpřístupní nabídka, kde klikneme na tlačítko *Změnit proměnnou*. Otevře se okno *Výběr kategorizační proměnné*, kde vybereme *skupina* a potvrdíme. Ještě jednou klikneme na *OK* a dostaneme následující normální pravděpodobnostní grafy pro jednotlivé skupiny.

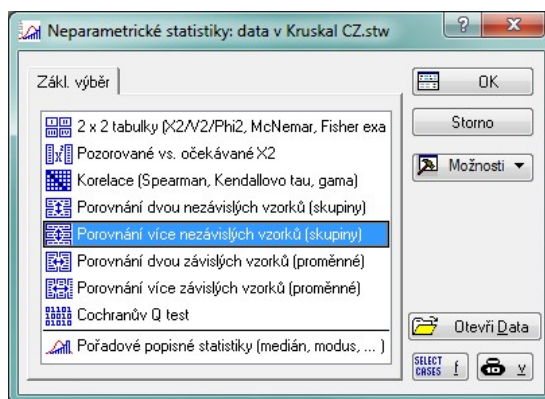
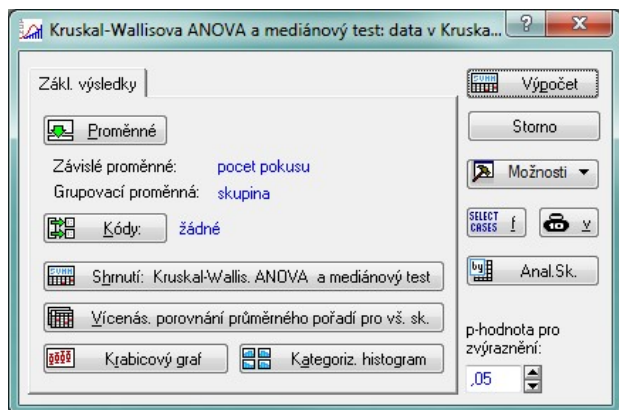


Z grafu i z hodnot Shapiro-Wilkova testu je patrné, že ve skupině *velikost* byl porušen předpoklad normality. Navíc počet pokusu v jednotlivých skupinách je velmi malý na to, abychom mohli počítat s výsledky centrální limitní věty. Místo ANOVY tedy použijeme její neparametrickou obdobu Kruskal-Wallisův test, jenž je založený na pořadí a nepředpokládá se, že data pocházejí z normálního rozdělení, nicméně se předpokládá, že pozorování jsou nezávislá (což v našem případě víme z povahy pokusu), ze spojitého rozdělení a že v každé skupině pocházejí pozorování z populací se stejným tvarem distribuce.

K dispozici je kromě Kruskal-Wallisovy ANOVY také mediánový test, který porovnává počet pozorování nad a pod mediánem v jednotlivých skupinách. Můžeme se tedy rozhodnout, jestli použijeme Kruskal-Wallisovu nebo mediánový test. Není ale korektní používat oba testy najednou.

Oba testy najdeme ve stejné nabídce následovně: ze základní nabídky vybereme *Statistiky* -> *Neparametrická statistika* -> *Porovnání více nezávislých vzorků (skupiny)*.

Otevře se dialog *Kruskal-Wallisova ANOVA a mediánový test*. Klikneme na tlačítko *Proměnné*. Jako závislou proměnnou vybereme *počet pokusů*. Grupovací proměnná bude *skupina*. Klikneme na *OK*. Dialog by měl vypadat takto:



Klikneme na tlačítko **Shrnutí: Kruskal-Wallis. ANOVA a mediánový test**. Vytvoří se dvě tabulky.

První z nich obsahuje výsledky Kruskal-Wallisova testu. Na základě dosažené hladiny významnosti Kruskal-Wallisova testu ($p=0,0010$) můžeme říct, že jsme prokázali statisticky významný rozdíl v obtížnosti úloh založených na rozpoznání tvaru, barvy a velikosti.

Kruskal-Wallisova ANOVA založ. na poř.; počet pokusu (data v Kruskal CZ.stw)				
Nezávislá (grupovací) proměnná : skupina				
Kruskal-Wallisův test: $H(2, N=36) = 13,84438$ $p = ,0010$				
Závislá:	Kód	Počet platných	Součet pořadí	Prům. Pořadí
počet pokusu				
tvar	101	12	139,0000	11,58333
barva	102	12	200,0000	16,66667
velikost	103	12	327,0000	27,25000

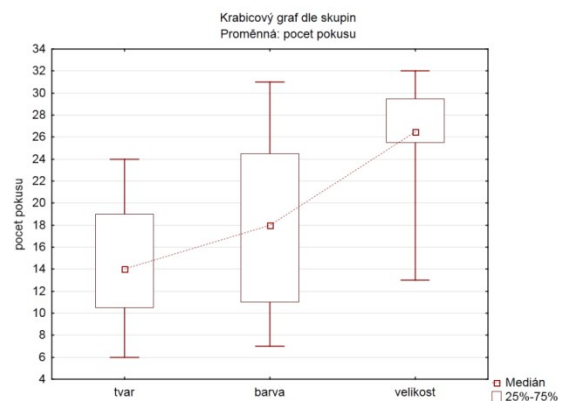
Poznamenejme, že Kruskal-Wallisův test nepracuje s původními hodnotami, ale s pořadovými čísly, která jim byla přiřazena. Z hodnot součtu pořadí pro jednotlivé skupiny vyplývá, že nejobtížnější bylo rozpoznat velikost, nejsnáze se identifikoval tvar.

Ke stejným závěrům dojdeme také pomocí mediánového testu, kdy zkoumáme počty případů nad a pod společným mediánem. Opět se potvrzuje, že nejlépe byly předměty rozpoznávány podle tvaru a nejhůře podle velikosti. Pokud bychom se pro mediánový test rozhodli, tak jeho výsledek ukazuje druhá tabulka:

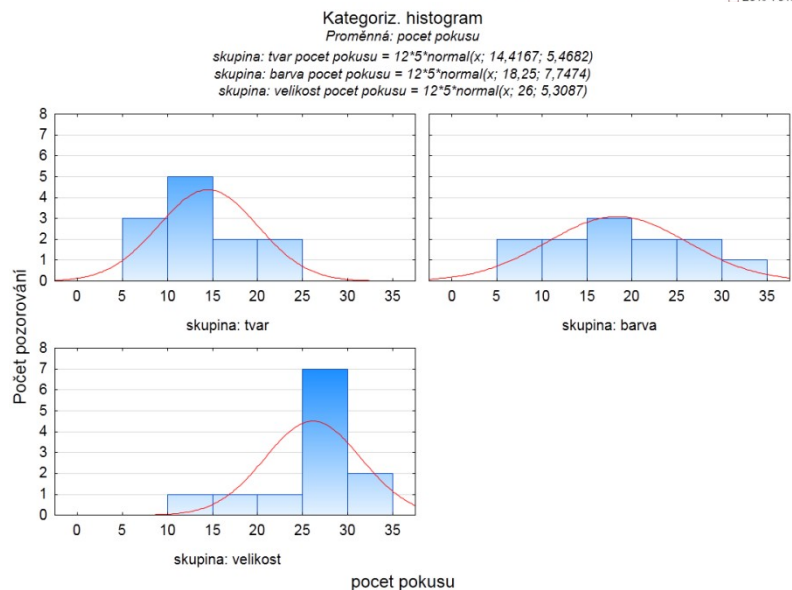
Mediánový test, celk. medián = 19,5000; počet pokusu (data v Kruskal CZ.stw)				
Nezávislá (grupovací) proměnná : skupina				
Chi-Kvadr. = 8,666667 $sv = 2$ $p = ,0131$				
Závislá:	tvar	barva	velikost	Celkem
počet pokusu				
<= Medián: pozorov.	9,00000	7,00000	2,00000	18,00000
očekáv.	6,00000	6,00000	6,00000	
poz.-oč.	3,00000	1,00000	-4,00000	
> Medián: pozorov.	3,00000	5,00000	10,00000	18,00000
očekáv.	6,00000	6,00000	6,00000	
poz.-oč.	-3,00000	-1,00000	4,00000	
Celkem: oček.	12,00000	12,00000	12,00000	36,00000

Výsledky můžeme zobrazit také graficky. V dialogu **Kruskal-Wallisova ANOVA a mediánový test** klikneme na tlačítko **Krabicový graf**. Ještě je třeba vybrat proměnnou, kterou chceme v grafu zobrazit, což je *počet pokusů* a zvolit typ krabicového grafu. Nás zajímá krabicový graf znázorňující medián, dolní a horní kvartil a rozpětí (tj. první volba v dialogu **Typ krabicového grafu**), protože medián odpovídá povaze Kruskal-Wallisova testu.

Také krabicový graf potvrzuje, že nejsnáze byl rozpoznán tvar a nejvíce pokusů opice potřebovaly ke správnému určení velikosti předmětu.



Dalším způsobem, jak si můžeme prohlédnout rozložení hodnot závislé proměnné v jednotlivých skupinách, je kategorizovaný histogram. V dialogu **Kruskal-Wallisova ANOVA a mediánový test** klikneme na tlačítko **Kategoriz. histogram**. Otevře se dialog, ve kterém vybereme proměnnou, kterou v grafu zobrazíme, tj. *počet pokusů*. Klikneme na **OK**. Výsledek je na obrázku.



Opět se potvrzuje, že opice lépe rozpoznávaly tvar (tj. distribuce je mírně zešikmená vlevo) než barvu a velikost.

Výsledky se zdají být zřetelně nejhorší v případě velikosti.

Již víme, že mezi skupinami je významný rozdíl. Ale mezi kterými z nich? Na tyto otázky dostaneme odpověď, pokud klikneme na tlačítko *Vícenás. porovnání průměrného pořadí pro vš. sk.* v dialogu *Kruskal-Wallisova ANOVA a mediánový test*.

Vícenásobné porovnání z' hodnot; počet pokusu (data v Kruskal CZ.stw)			
Nezávislá (grupovací) proměnná : skupina			
Kruskal-Wallisův test: H (2, N= 36) =13,84438 p =,0010			
Závislá: počet pokusu	tvar R:11,583	barva R:16,667	velikost R:27,250
tvar		1,181851	3,642426
barva	1,181851		2,460575
velikost	3,642426	2,460575	

Vícenásobné porovnání p hodnot (oboustr.); počet pokusu (data v Kruskal CZ.stw)			
Nezávislá (grupovací) proměnná : skupina			
Kruskal-Wallisův test: H (2, N= 36) =13,84438 p =,0010			
Závislá: počet pokusu	tvar R:11,583	barva R:16,667	velikost R:27,250
tvar		0,711794	0,000810
barva	0,711794		0,041614
velikost	0,000810	0,041614	

Na základě vícenásobného porovnání můžeme říct, že naše exempláře dosahovaly významně lepších výsledků při určování tvaru a barvy než velikosti. Rozdíl v úspěšnosti při určování tvaru a barvy není statisticky významný.

Více informací o pozadí metody se můžete dozvědět na našem odborném kurzu [Analýza rozptylu](#), případně v kurzu [Neparametrická statistika](#).

Závěr

Jak se zdá z výsledků analýzy, můžeme konstatovat, že mezi faktory existují rozdíly a pro opice jsou tyto úlohy odlišné svou obtížností. Nejjednodušší je rozpoznání tvaru, velikost způsobuje naopak největší problémy. Zvolený test prokázal signifikantní rozdíl mezi barvou a velikostí, tvarem a velikostí.