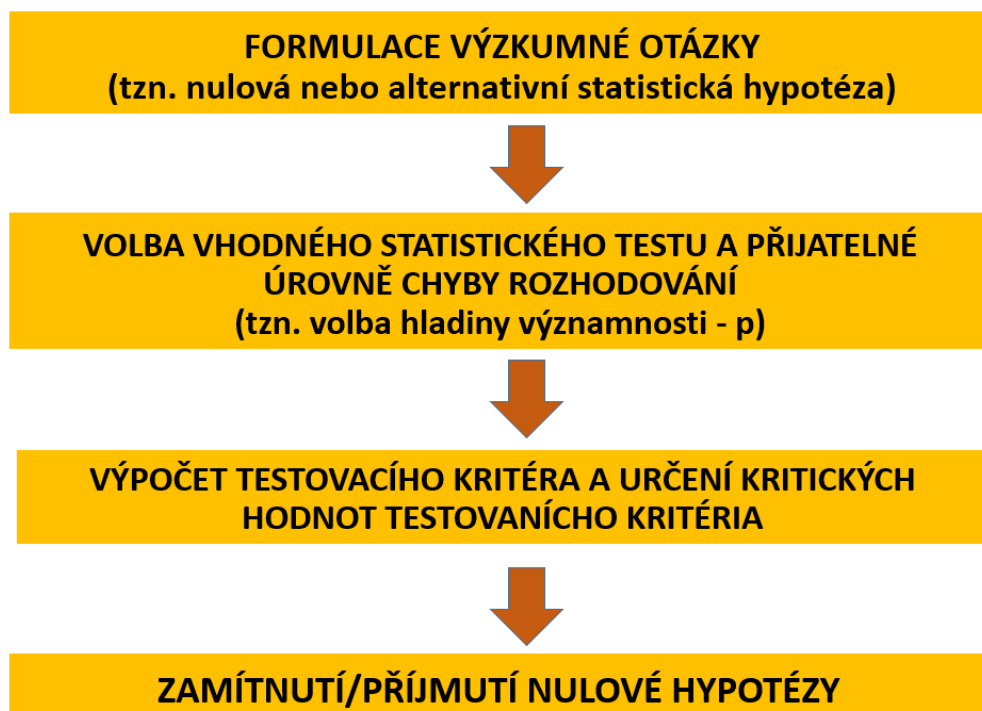


Testování hypotéz o rozdílu průměrů

1. Testování statistických hypotéz

Statistické metody využíváme k získání objektivních závěrů námi realizovaných experimentů. Tímto způsobem můžeme posoudit, jestli výsledky našich studií vyhovují či nevyhovují vznesené hypotéze, kterou jsme před započítím testování učinili. Nejčastěji porovnáváme vybrané parametry (např. průměry, rozptyly, mediány) mezi dvěma a více soubory dat, případně se provádí hodnocení vzájemného vztahu mezi sledovanými proměnnými. Pro získání relevantních závěrů nestačí pouze výpočet základních popisných charakteristik, ale je třeba provádět i testování relevantních statistických hypotéz (obrázek 1).

Obrázek 1: Postup při testování statistických hypotéz



Prvním krokem statistického testování je tedy formulace konkrétní výzkumné otázky. Je třeba provést nastavení tzv. nulové a alternativní statistické hypotézy. Nulová hypotéza (H_0) většinou předpokládá, že mezi sledovanými parametry je rozdíl statisticky nevýznamný, tudíž můžeme usuzovat, že dané parametry se sobě rovnají a testovaný efekt nebyl účinný. Naopak alternativní hypotéza je v rozporu s nulovou hypotézou. Příkladem může být testování potenciálně negativních

účinků herbicidu atrazinu na imunitní systém ryb. Nulová hypotéza v tomto případě může být definována jako tvrzení, že chronická expozice tímto herbicidem neovlivňuje počet leukocytů v krvi exponovaných jedinců. Alternativní hypotéza (H_1 nebo H_A) naopak tvrdí, že v důsledku působení uvedeného herbicidu dojde ke změně (zvýšení/snížení) počtu bílých krvinek. Pro rozhodnutí o tom, zda je pravdivá nulová nebo alternativní hypotéza, rozdujeme na základě výsledků statistického testu.

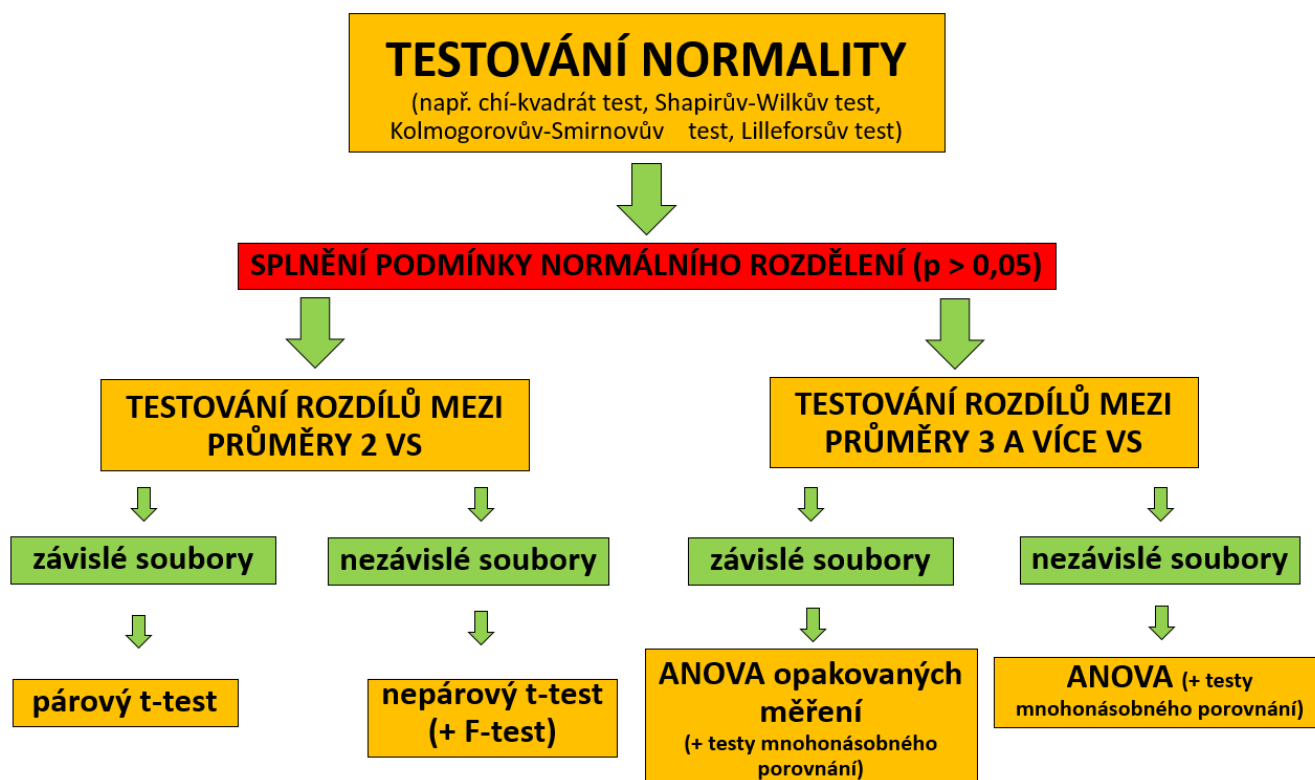
V dalším kroku statistického testování si na základě designu našeho experimentu zvolíme vhodnou statistickou metodu, která povede k rozhodnutí ve prospěch buď nulové, nebo alternativní hypotézy. Vlastní výpočet statistického testu se provádí ve většině případů s využitím vhodného statistického softwaru. Testování hypotéz sebou přináší riziko dvou chybných rozhodnutí – chyba I. a II. druhu. Chyby prvního druhu se dopustíme tehdy, zamítneme-li testovanou hypotézu, přestože platí. Pravděpodobnost vzniku chyby I. typu chceme zpravidla omezit. Požadujeme, aby tato pravděpodobnost nepřesáhla definovanou hodnotu, kterou nazýváme hladinu významnosti (označujeme p nebo α). Nejčastěji se hladina významnosti, tzn. adekvátní přijatelná úroveň rozhodování, volí 5 %, případně 1 % (tzn. 0,05, resp. 0,01). Chyba druhého druhu nastane, když je zamítnuta platná alternativní hypotéza.

2. Postup statistického testování

Při statistické analýze můžeme využívat řadu statistických testů. Podle rozdělení sledovaného statistického znaku rozlišujeme dvě základní skupiny statistických testů – parametrické a neparametrické testy. Parametrické testy využíváme pro hodnocení statistického znaku, u kterého předpokládáme normální rozdělení. Neparametrické testy jsou takové statistické testy, které nevyžadují znalost rozdělení a většinou je tedy využíváme pro data, u kterých není splněna podmínka normality. Obecně platí, že parametrické testy jsou početně náročnější, ale zároveň i silnější testy. Neparametrické testy mají tedy obecně menší sílu, ale jejich výhodou je především univerzálnost a možnost využití pro data kvalitativního i kvantitativního charakteru. Zároveň i jejich výpočet je mnohem jednodušší.

Postup statistického testování v případě splnění podmínky normálního rozdělení je uveden na obrázku 2. Zvolíme si vhodný statistický test dle designu našeho experimentu. Nejčastěji provádíme testování rozdílu mezi středními hodnotami. Máme-li v našem experimentu pouze dva soubory, zvolíme si příslušnou variantu Studentova t-testu. Pokud provádíme porovnání středních hodnot u třech a více souborů, je třeba zvolit složitější statistický test, kterým je například analýza rozptylu. Při výběru adekvátního statistického testu je třeba také brát v potaz informace o tom, zda testované soubory jsou nebo nejsou na sobě závislé.

Obrázek 2: Výběr statistického testu pro spojité proměnné při splnění podmínky normality



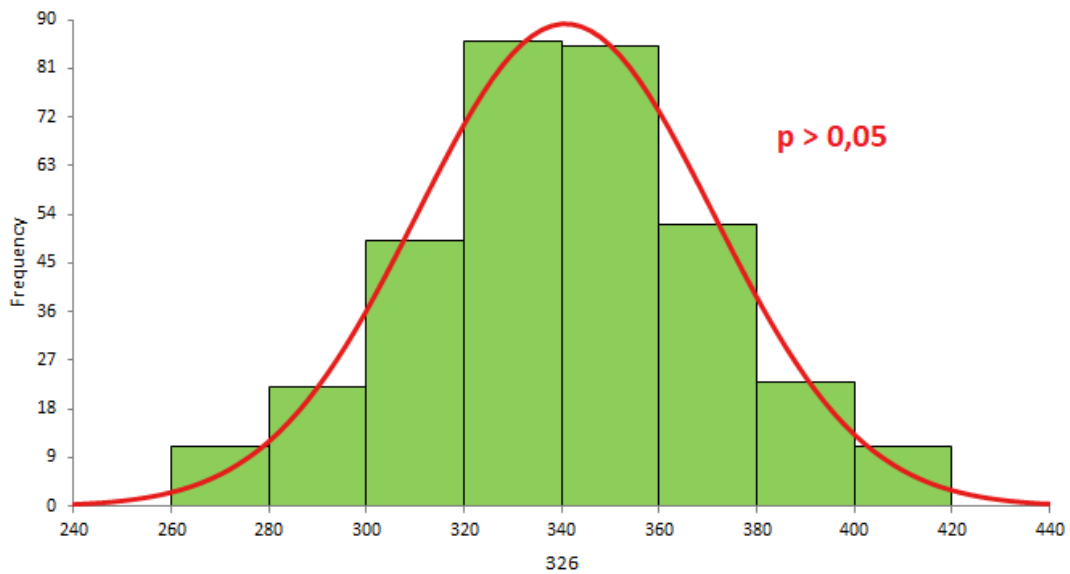
Poznámka: VS – výběrový soubor, ANOVA – analýza rozptylu

2. Testování normality

Podmínkou použití tzv. parametrických testů je normální rozdělení dat. Pro prvotní odhad, jestli data splňují podmínku normálního rozdělení, se využívá grafické sestavení histogramu (obrázek 3 a 4), který nám znázorňuje distribuci testovaných dat s využitím sloupcového grafu. Histogram sestavíme tak, že na ose x máme sledovanou veličinu a na ose y je četnost této veličiny v daném intervalu (třídě).

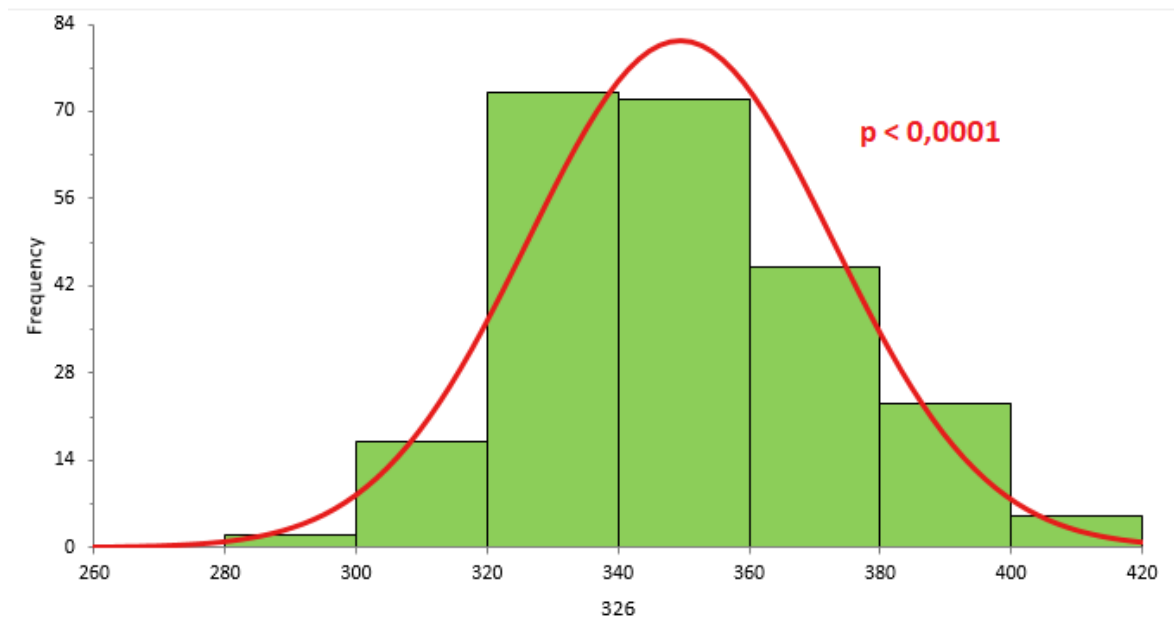
Přesnějším nástrojem je provedení příslušného testu normality. Testování normality se běžně provádí ve speciálním statistickém softwaru, kde je k dispozici řada testů normality lišících se svojí náročností a samozřejmě silou (např. Pearsonův χ^2 test, Kolmogorovův-Smirnovův test, Cramerův-von Misesův test, Andersonův-Darlingův test, Shapirův-Wilkův test). Nejčastěji se při testování normality budeme ve cvičeních setkávat se Shapirův-Wilkovým testem, který funguje efektivně především u souborů s malým rozsahem dat, kde dokáže odhalit případné porušení normality. Naopak se jeví jako málo spolehlivý při testování výběrových souborů, které obsahují velký počet identických hodnot. Postup použití tohoto testu v programu Unistat for Excel 6.5 je uveden na obrázku 5. Vyjde-li nám v statistickém testu hodnota p větší než 0,05, tak normalitu nezamítáme, tudíž přijímáme nulovou hypotézu.

Obrázek 3: Histogram výběrového souboru s normálním rozdělením (testováno Shapiro-Wilkovým testem)



Poznámka: Histogram byl zpracován v programu Analyse-It. Testování normality bylo provedeno v programu Unistat for Excel 6.5

Obrázek 4: Histogram výběrového souboru bez normálního rozdělení (testováno Shapiro-Wilkovým testem)



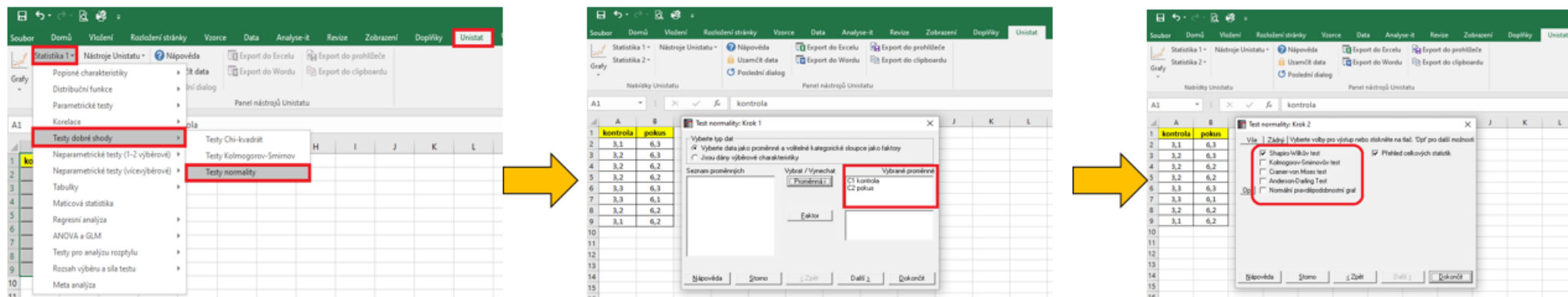
Poznámka: Histogram byl zpracován v programu Analyse-It. Testování normality bylo provedeno v programu Unistat for Excel 6.5

Obrázek 5: Postup testování normality s využitím programu Unistat for Excel 6.5

A. Označíme si zdrojová data a v hlavním menu si v nabídce Unistat zvolíme: **Statistika 1 → Testy dobré shody → Testy normality**.

B. Do kolonky „Vybrané proměnné“ si vybereme testované výběrové soubory a potvrdíme tlačítko „Další“.

C. Zvolíme si příslušný test normality a potvrdíme tlačítko „Dokončit“.



VYHODNOCENÍ

	Platná pozorování	Chybějící	Průměr	Směrodatná odchylka	Shapiro-Wilkův test	Pravděp.
kontrola	8	0	3,2000	0,0756	0,8489	0,9929
pokus	8	0	6,2250	0,0707	0,8272	0,9555

D. Dojde k vytvoření nového listu. Pokud hodnota pravděpodobnosti p je u všech souborů větší než 0,05, lze pro další testování použít parametrické testy (potvrzení normality).

3. Studentův t-test

Studentův t-test patří mezi velmi často využívané parametrické testy a používá se ke zjišťování rozdílů mezi středními hodnotami dvou souborů. Tento statistický test využíváme především v situacích, kdy potřebujeme mezi sebou porovnat kontrolní a pokusnou skupinu, tzn. potřebujeme například potvrdit účinnost testovaného preparátu či jiného pokusného zásahu. Porovnávat můžeme ovšem i dvě pokusné skupiny, ale při vyslovení závěru v tomto případě můžeme pouze tvrdit, že testované preparáty či pokusné zásahy se ve svých účincích liší nebo jsou shodné. Protože nemáme v tomto případě k dispozici kontrolní skupinu, nelze tvrdit, že testovaný preparát či pokusných zásah byl účinný a vedl ke změně sledovaného ukazatele. Podle designu použitého experimentu rozlišujeme několik základních variant tohoto testu:

- *jednovýběrový t-test*
- *dvouvýběrový párový t-test*
- *dvouvýběrový nepárový test.*

Zpracování t-testu lze provést v tabulkovém procesoru Excel, případně lze opět využít vhodný statistický software. Výsledkem těchto testů ve většině případů bývá hodnota pravděpodobnosti p . Pokud hodnota p je větší jak 0,05, tak můžeme potvrdit nulovou hypotézu, tudíž střední hodnoty testovaných souborů se neliší. Pokud ovšem hodnota p je menší jak 0,05, tak střední hodnoty testovaných souborů se signifikantně liší. Na obrázku 6 a 7 je uveden stručný postup výpočtu tohoto statistického testu v různých programech.

Jednovýběrový t-test slouží k porovnání střední hodnoty základního souboru (μ) a střední hodnoty výběrového souboru (\bar{x}). Zjišťujeme, zda výběrový soubor pochází z dané populace, která je charakterizována střední hodnotou μ , kterou považujeme za konstantu. Tuto konstantu většinou získáme z literatury. Často se jedná o fyziologickou hodnotu pro daný parametr.

Dvouvýběrový párový t-test se využívá pro zjištění statisticky významných rozdílů mezi středními hodnotami dvou výběrových souborů, které jsou na sobě závislé. Jedná se tedy o výběrové soubory, které vzniknou například dvěma měřeními provedenými opakovaně u stejné skupiny jedinců (např. před testovaným zásahem = kontrola, po testovaném zásahu = pokus). Tyto hodnoty představují páry, které jsou na sobě závislé.

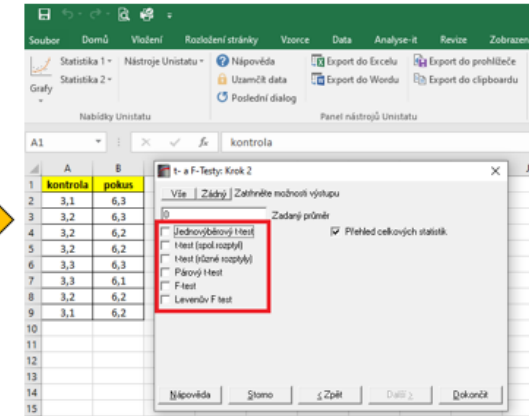
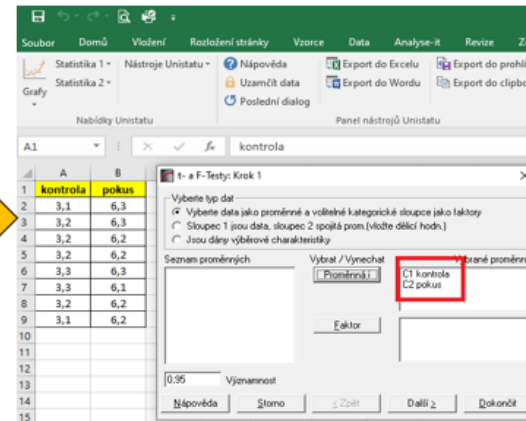
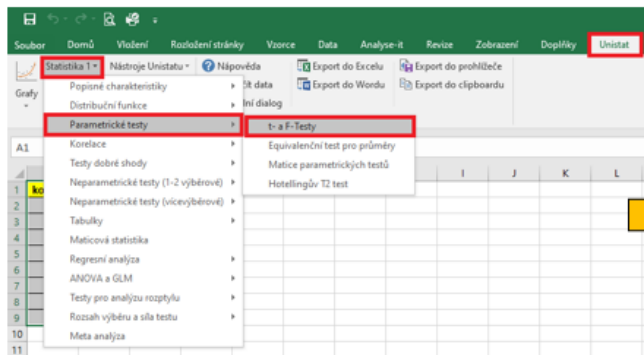
Dvouvýběrový nepárový t-test se využívá pro zjištění rozdílů mezi středními hodnotami dvou nezávislých výběrových souborů. V prvním kroku testování je ovšem třeba provést F-test, který nám zjistí, zda rozptyly těchto dvou souborů jsou shodné nebo různé. Tuto informaci potřebujeme znát pro další testování, kdy si vybíráme z varianty nepárového t-testu pro shodné rozptyly nebo nepárového t-testu pro různé rozptyly.

Obrázek 6: Postup provádění t-testu v programu Unistat for Excel 6.5

A. Označíme si zdrojová data a v hlavním menu si v nabídce Unistat zvolíme: **Statistika 1 → Parametrické testy → t- a F-testy.**

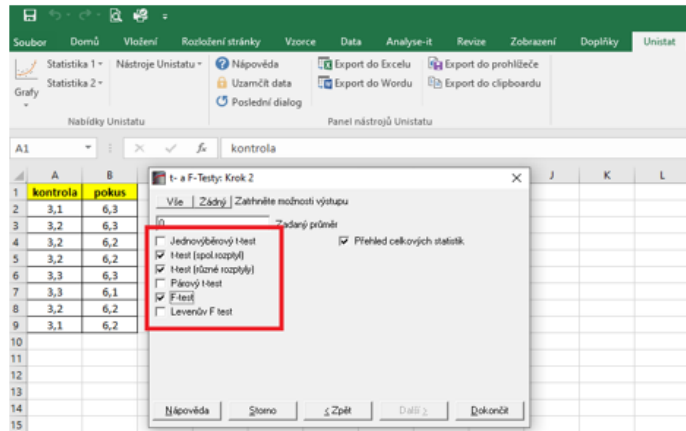
B. Do kolonky „Vybrané proměnné“ si vybereme testované výběrové soubory a potvrdíme tlačítko „Další“.

C. Zvolíme si příslušný test, který odpovídá designu našeho experimentu a zaškrtneme „Přehled celkových tabulek. Dále potvrdíme tlačítko „Dokončit“.



D. Budeme-li například vyhodnocovat data z našeho experimentu s využitím nepárového t-testu, zaškrtneme si v nabídce t-test pro společné a různé rozptyly a také F test.

E. Dojde k vytvoření nového listu, kde jsou výsledky pravděpodobností zvolených testů. Nejprve zkontrolujeme výsledek F-testu a podle tohoto si zvolíme příslušnou variantu t-testu (v našem případě t-test pro společné rozptyly). Vzhledem k tomu, že v t-testu nám hodnota p vyšla menší než 0,05, můžeme říct, že mezi průměry jsou signifikantní rozdíly.



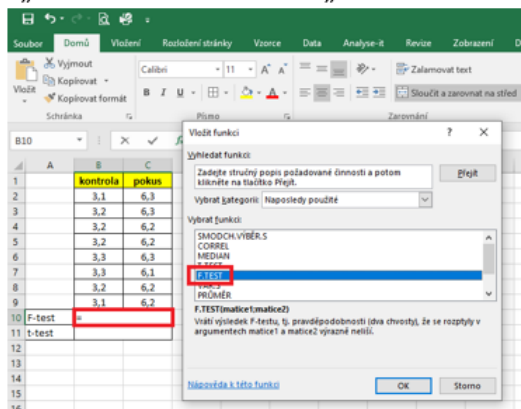
VYHODNOCENÍ

List1 - Microsoft Excel - Analyse-it®												
Nabídka Unistatu												
t- a F-Testy												
Pro kontrola a pokus												
	Platná pozorování	Chybějící	Průměr	Směrodatná odchylka	Rozdíly	Směrodatná chyba	t-statistika	Štupně volnosti	jednostr. pravděp.	dvostr. pravděp.	p < 0,05	
3				0,0732	-3,6250	0,0366	-82,6587	14,0000	0,0000	0,0000	-3,1035	-2,9485
4					-3,6250	0,0366	-82,6587	13,9381	0,0000	0,0000	-3,1035	-2,9485
6												
7		Rozptyly 1	Rozptyly 2	Statistika F	st. vol. čitatele	st. vol. jmenovatele	Pravostr. pravděp.	dvostr. pravděp.	Dolní 95%	Horní 95%		
8		F-test		0,0057	0,0050	1,1429	7	7	0,4323	0,8647	0,2288	5,7085
9												
10												

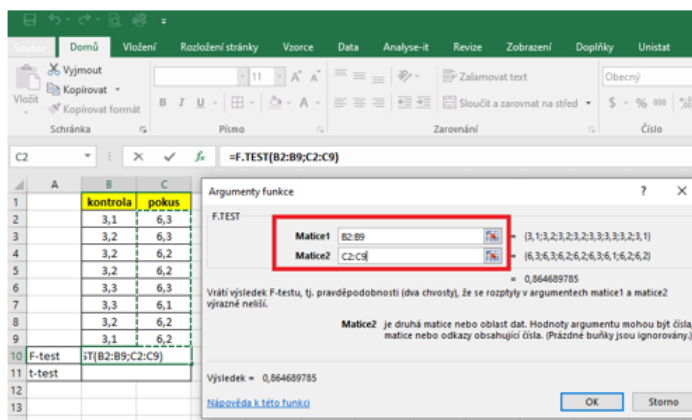
p > 0,05

Obrázek 7: Postup provádění t-testu v tabulkovém procesoru Excel

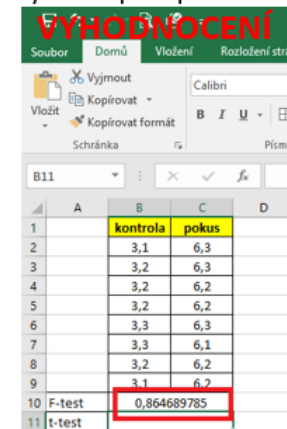
A. Máme-li používat nepárový t-test, musíme nejprve udělat F-test. U párového t-testu už F-test neděláme. V nabídce „Vzorce“ si zvolíme funkci „F-test“.



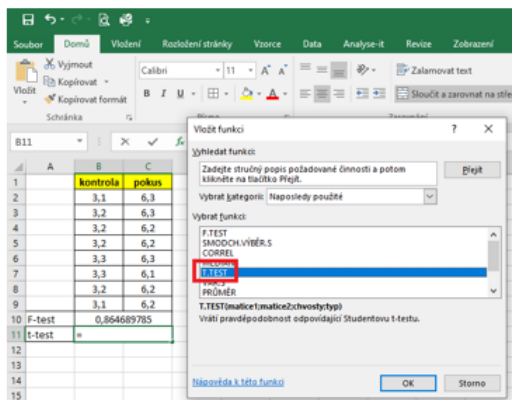
B. Do kolonky Matice 1 a Matice 2 označíme data porovnávaných výběrových souborů.



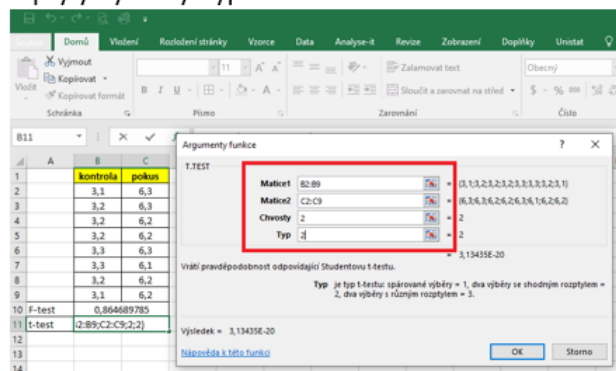
C. Protože hodnota pravděpodobnosti F-testu vyšla 0,865, budeme při dalším testování volit nepárový t-test pro společné rozptyly.



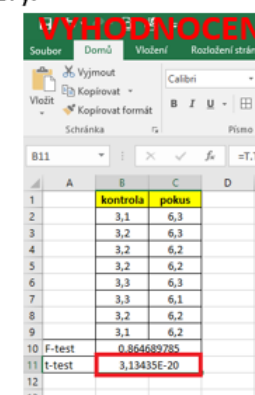
D. Pro výpočet t-testu si v nabídce „Vzorce“ zvolíme funkci t-test.



E. Do kolonky Matice 1 a Matice 2 si označíme data porovnávaných souborů. V kolonce „Chvosty“ si zvolíme číslo 2 a v kolonce typ si zvolíme v tomto případě nabídku 2 (společné rozptyly). Pokud bychom měli párový t-test, volíme typ 1. Pro nepárový t-test s různými rozptyly by to byl typ 3.



F. Protože hodnota pravděpodobnosti t-testu je menší než 0,000, můžeme říct, že mezi soubory jsou vysoce významné rozdíly mezi průměry.



4. Analýza rozptylu a testy mnohonásobného porovnání

Při statistickém hodnocení složitějších experimentů jsme často nuceni provádět porovnání středních hodnot třech a více souborů mezi sebou. Pro tyto účely už nelze použít výše zmíněný Studentův t-test, ale je třeba využít složitější statistické metody. Jednou z možností je aplikace analýzy rozptylu (ANOVA). Při testování jednoho faktoru se využívá tzv. jednofaktorová ANOVA (one-way ANOVA). Typickým příkladem může být například hodnocení vlivu dvou různých přípravků (A a B) na koncentraci vybraného biochemického ukazatele. Do těchto experimentů se běžně také zařazuje kontrolní skupina, proto celkový počet porovnávaných výběrových souborů je v tomto případě 3. Pokud bychom mimo typu přípravku prováděli ještě hodnocení účinků v různých časových intervalech, tak je třeba využít již dvoufaktorovou analýzu rozptylu, kdy hodnocenými faktory bude nejen typ testovaného přípravku, ale i čas odběru biologického vzorku. Podmínkou použití analýzy rozptylu je nejenom požadavek normálního rozdělení, ale zároveň musí být splněna podmínka homogenity rozptylu. Homogenitu rozptylu můžeme ověřit například pomocí Levenova testu, který ovšem musíme spočítat pomocí vhodného statistického programu. Další podmínkou je také nezávislost výběrových souborů. Nelze tedy tuto metodu využít pro hodnocení výběrových souborů, které vznikly opakovaným měřením. Pro tyto účely by bylo třeba zvolit variantu analýzy rozptylů pro opakovaná měření.

Metodou analýzy rozptylu zjistíme, zda mezi testovanými výběrovými soubory je, případně není signifikantní rozdíl mezi středními hodnotami. Pokud nám hodnota pravděpodobnosti vyjde menší než 0,05, zamítáme nulovou hypotézu o shodě středních hodnot a je třeba provést v dalším kroku testy mnohonásobného porovnání, které nám určí, kde konkrétně se rozdíly nacházejí. Existuje řada různých testů mnohonásobného porovnání, liší se především ve své síle. Pokud chceme zjistit rozdíly mezi všemi skupinami mezi sebou, tak se jako vhodný post hoc test často volí Tukey-HSD test. Jedná se o konzervativní test, který se vyznačuje dostatečnou silou. Pokud chceme provést porovnání pouze vůči kontrolní skupině, lze použít Dunnettův test.

Výpočet analýzy rozptylu i testů mnohonásobného porovnání nelze provést v tabulkovém procesoru Excel. Je třeba využít vhodný statistický program. Na obrázku 8 a 9 je uveden schématický návod použití těchto testů ve statistickém programu Unistat for Excel 6.5.

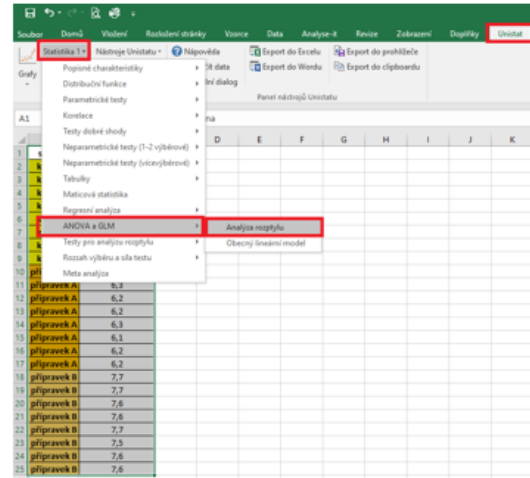
Obrázek 8: Postup analýzy rozptylu ve statistickém programu Unistat for Excel 6.5

A. Při použití metody ANOVA je třeba dodržet správné formátování zdrojových dat.

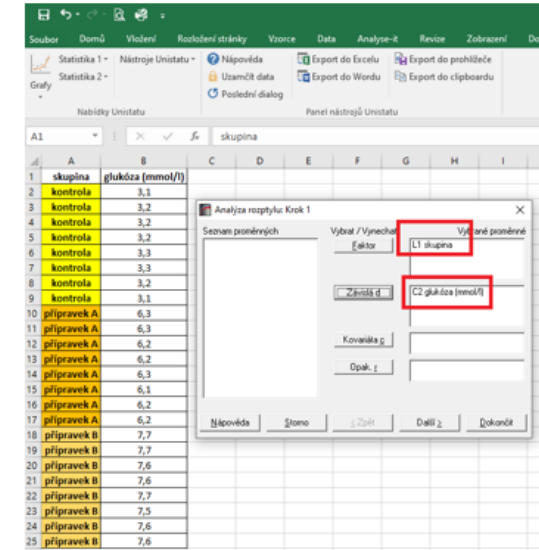
A	B
skupina	glukóza (mmol/l)
kontrola	3,1
kontrola	3,2
kontrola	3,2
kontrola	3,2
kontrola	3,3
kontrola	3,3
kontrola	3,2
kontrola	3,1
přípravek A	6,3
přípravek A	6,3
přípravek A	6,2
přípravek A	6,2
přípravek A	6,3
přípravek A	6,1
přípravek A	6,2
přípravek A	6,2
přípravek B	7,7
přípravek B	7,7
přípravek B	7,6
přípravek B	7,6
přípravek B	7,6
přípravek B	7,7
přípravek B	7,7
přípravek B	7,5
přípravek B	7,6
přípravek B	7,6



B. Označíme si zdrojová data a v hlavním menu si v nabídce Unistat zvolíme: **Statistika 1** → **ANOVA a GLM** → **Analýza rozptylu**. V další nabídce si zvolíme „NE“.



C. Zvolíme si příslušný „**Faktor**“ a „**Závislou**“. Použijeme tlačítko **další** a v následujícím dialogovém okně si vybereme nabídku „**Klasický experiment**“ a zvolíme „**Dokončit**“.



VYHODNOCENÍ

Zdroj variability	Součet čtverců	St. vol.	Průměrný čtverec	Stat F	Pravd
Hlavní efekty	82,835	2	41,418	6797,641	0,0000
skupina	82,835	2	41,418	6797,641	0,0000
Vysvětleno	82,835	2	41,418	6797,641	0,0000
Chyba	0,128	21	0,006		
Celkem	82,963	23	3,607		

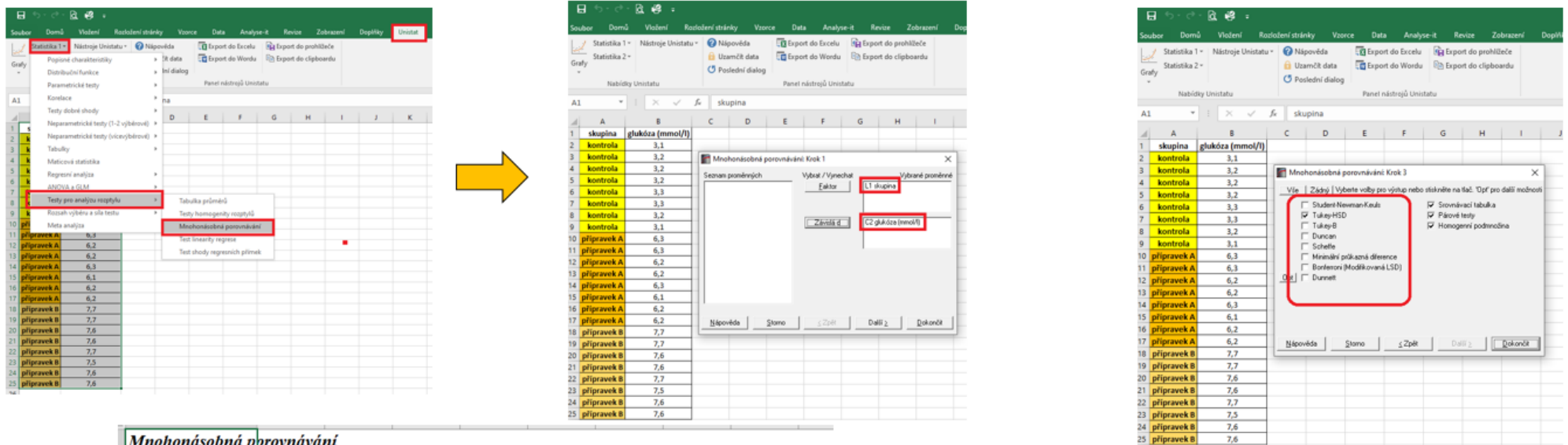
D. Dojde k vytvoření nového listu, kde je výsledek pravděpodobnosti testu (řádek „Skupina“). Vzhledem k tomu, že nám vyšla hodnota menší než 0,000, znamená to, že mezi skupinami existují signifikantní rozdíly mezi průměry. Je třeba tedy provést testy mnohonásobného porovnání pro zjištění, kde konkrétně se statisticky významné rozdíly nacházejí.

Obrázek 9: Postup testů mnohonásobného porovnání ve statistickém programu Unistat for Excel 6.5

A. Pro testy mnohonásobného porovnání využijme shodné formátování dat jako v případě testu ANOVA. Označíme si zdrojová data a v hlavním menu si v nabídce Unistat zvolíme: **Statistika 1 → Testy pro analýzu rozptylu → Mnohonásobné porovnání**. V další nabídce si zvolíme „NE“.

B. V dalším okně si zvolíme „Faktor“ a „Závislou“ (viz analýza rozptylu). V následujícím dialogovém okně necháme přednastavené hodnoty, které nám program nabízí a zvolíme nabídku „Další“.

C. Vybereme si vhodný typ testu mnohonásobného porovnání a zvolíme nabídku „Dokončit“. Pro porovnání všech skupin mezi sebou se nejčastěji volí Tukey-HSD. Pokud porovnáme pouze vůči kontrole, tak si vybíráme Dunnnettův test.



Mnohonásobná porovnání
Tukey-HSD
Pro glukóza (mmol/l), tříděno podle skupina
Střední kvadratická chyba: 6,09292857142733E-03, Stupně volnosti: 21
** označuje významně odlišné páry. Homogenní podskupiny jsou ve vertikálních sloupcích.
Párový test je významný, pokud q hodnota je větší než tabulková hodnota q

Skupina	Přip.	Průměr	kontrola	přípravek A	přípravek B
kontrola	8	3,2000		**	**
přípravek A	8	6,2250	**		**
přípravek B	8	7,6568	**	**	

Skupiny	Rozdíl	Smérodlatná chyba	q Stat	Tabulka q	Důležitá	Dolní 95%	Horní 95%	Výsledek
přípravek B - kontrola	4,4568	0,0390	161,4917	3,5646	0,0000	4,3584	4,5551	**
přípravek A - kontrola	3,0250	0,0390	109,6118	3,5646	0,0000	2,9266	3,1234	**
přípravek B - přípravek A	1,4318	0,0390	51,8799	3,5646	0,0000	1,3334	1,5301	**

VYHODNOCENÍ

D. Dojde k vytvoření nového listu, kde jsou uvedeny hodnoty pravděpodobností pro jednotlivé kombinace skupin.

Zdroje:

Hendl, J. Přehled statistických metod – analýza a metaanalýza dat. 2015. Vydavatelství Portál Praha, 736 s.

Lepš, J. Biostatistika. 1996. Jihočeská univerzita v Českých Budějovicích, 166 s.

Meloun, M., Militký, J. 2012. Kompendium statistického zpracování dat. Nakladatelství Karolinum, Univerzita Karlova v Praze. 982 s.