

Univerzita Pardubice
Fakulta Ekonomicko- správní

Testy hypotéz s využitím programu MS EXCEL

Tomáš Borůvka

Bakalářská práce

2010

Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav matematiky
Akademický rok: 2009/2010

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Tomáš BORŮVKA**
Studijní program: **B6208 Ekonomika a management**
Studijní obor: **Management podniku - Management malých a středních podniků**
Název tématu: **Testy hypotéz s využitím programu MS Excel**

Z á s a d y p r o v y p r a c o v á n í :

Práce se zaměří na popis metod používaných při testování hypotéz s využitím MS EXCEL. S pomocí těchto metod budou v práci řešeny statistické úlohy. práce bude obsahovat:

- popis systému MS EXCEL
- popis a vysvětlení jednotlivých statistických funkcí v MS EXCEL
- testování hypotéz s využitím MS EXCEL.

Rozsah grafických prací: —
Rozsah pracovní zprávy: cca 30 stran
Forma zpracování bakalářské práce: tištěná/elektronická

Seznam odborné literatury:

- [1] GERBHARDT, Reiner. Excel 97 kompletní kapesní průvodce. Ing. Radka Halodová. 1. vyd. Praha : Grada Publishing, 1998. 448 s. ISBN 80-7169-592-0.
- [2] KUBANOVÁ, Jana. Statistické metody pro ekonomickou a statistickou praxi. 2. vyd. Bratislava : Statis, 2004. 254 s. ISBN 80-85659-37-9.
- [3] KUBANOVÁ, Jana, LINDA, Bohdan. Kritické hodnoty a kvantily vybraných rozdělení pravděpodobností. 1. dotisk vyd. Pardubice : Tiskařské středisko University Pardubice, 2007. 53 s. ISBN 80-7194-852-755-7.
- [4] ŠŤASTNÝ, Zdeněk. Matematické a statistické výpočty v Microsoft EXCELu. 1. vyd. Brno : Computer press, 1999. 254 s. ISBN 80-7226-141-X.

Vedoucí bakalářské práce: **Mgr. Ondřej Slavíček**
Ústav matematiky

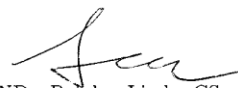
Datum zadání bakalářské práce: **30. června 2009**

Termín odevzdání bakalářské práce: **30. dubna 2010**



doc. Ing. Renáta Myšková, Ph.D.
děkanka

L.S.



doc. RNDr. Bohdan Linda, CSc.
vedoucí ústavu

V Pardubicích dne 13. července 2009

Prohlašuji:

Tuto práci jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č.121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 15.4.2010

Tomáš Borůvka

ANOTACE

Práce je věnována programu MS EXCEL a jeho využití při řešení statistických hypotéz. Práce se zabývá nejen řešením hypotéz ale i popisem používaných funkcí v MS EXCELU. V práci jsou uvedeny i příklady a postupy řešení některých výpočtů. Na závěr práce je uveden i praktický příklad využití hypotéz.

KLÍČOVÁ SLOVA

Testování hypotéz, MS EXCEL, Statistické funkce, Výpočty pomocí MS EXCEL.

TITLE

Experiments with hypothesis in computer program MS EXCEL

ANNOTATION

The work deals with a program MS EXCEL and his application on statistical hypothesis solution. It focuses on hypothesis solution and direction of used functions in the MS EXCEL. It contains examples and decision procedure of some calculations. It demonstrates use of hypothesis at the end of the work.

KEYWORDS

Experiments with hypothesis, MS EXCEL, Statistical functions, MS EXCEL assisted calculation.

OBSAH:

1. TABULKOVÝ PROCESOR MS EXCEL	7
2. FUNKCE V MS EXCEL	8
2.1 MATEMATICKÉ FUNKCE	8
2.1.1 Absolutní hodnota.....	8
2.1.2 Pi	8
2.1.3 Odmocnina	9
2.1.4 Sinus	9
2.1.5 Suma	9
2.2 STATISTICKÉ FUNKCE – ZÁKLADNÍ.....	10
2.2.1 Maximum.....	10
2.2.2 Minimum.....	11
2.2.3 Počet.....	11
2.2.4 Počet2.....	12
2.3 STATISTICKÉ FUNKCE	13
2.4 POPISNÁ STATISTIKA	14
2.4.1 Výběrové charakteristiky polohy	14
2.4.1.1 Aritmetický průměr	14
2.4.1.2 Geometrický průměr.....	15
2.4.1.3 Harmonický průměr.....	16
2.4.1.4 Výběrový medián	17
2.4.1.5 Výběrový modus	17
2.4.2 Výběrové charakteristiky variability.....	18
2.4.2.1 Populační disperze.....	18
2.4.2.2 Populační směrodatná odchylka	19
2.4.2.3 Výběrová disperze	19
2.4.2.4 Výběrová směrodatná odchylka.....	20
2.4.2.5 Výběrový variační koeficient.....	20
2.4.2.6 Výběrové variační rozpětí.....	21
2.4.3 Koeficienty asymetrie a excessu.....	21
2.4.3.1 Výběrový koeficient asymetrie.....	21
2.4.3.2 Výběrový koeficient excessu	22
3. TESTY HYPOTÉZ.....	23
3.1 TESTY VÝZNAMNOSTI	24
3.2 JEDNOVÝBĚROVÝ TEST VÝZNAMNOSTI PRO STŘEDNÍ HODNOTU NORMÁLNÍHO ROZDĚLENÍ PRAVDĚPODOBNOSTI SE ZNÁMÝM Σ	25
3.2.1 Výpočet kritické hodnoty pro Studentovo <i>T</i> -rozdělení pravděpodobnosti	28
3.3 JEDNOVÝBĚROVÝ TEST VÝZNAMNOSTI PRO STŘEDNÍ HODNOTU NORMÁLNÍHO ROZDĚLENÍ PRAVDĚPODOBNOSTI S NEZNÁMÝM Σ	29
3.3.1 Výpočet kritické hodnoty pro χ^2 rozdělení pravděpodobnosti.....	31
3.4 JEDNOVÝBĚROVÝ TEST VÝZNAMNOSTI PRO ROZPTYL	32
3.4.1 Výpočet kritické hodnoty pro <i>F</i> rozdělení pravděpodobnosti	33
3.5 DVOUVÝBĚROVÝ TEST VÝZNAMNOSTI PRO ROZPTYL F-TEST.....	34
3.6 DVOUVÝBĚROVÝ TEST VÝZNAMNOSTI PRO STŘEDNÍ HODNOTY SE ZNÁMÝM Σ Z-TEST	37
3.7 DVOUVÝBĚROVÝ T- TEST S ROVNOSTÍ ROZPTYLŮ.....	39
3.8 DVOUVÝBĚROVÝ T-TEST S NEROVNOSTÍ ROZPTYLŮ.....	42
3.9 DVOUVÝBĚROVÝ PÁROVÝ T-TEST NA STŘEDNÍ HODNOTU	44
4. VYUŽITÍ TESTŮ HYPOTÉZ.....	47
5. ZÁVĚR.....	53
6. POUŽITÉ ZDROJE	54

1. Tabulkový procesor MS EXCEL

Tato bakalářská práce se věnuje využití statistických funkcí v programu MS EXCEL 2003. Statistické funkce resp. statistika se využívá v mnoha odvětvích lidského života. Statistiku využívají manažeři při plánování výroby a chodu podniku. Statistiku využívá také stát při výpočtech různých ekonomických ukazatelů, používají ji pojišťovny, banky, burzovní makléři a mnoho dalších profesí a institucí. V dnešní době existuje mnoho statistických programů, které jsou k výpočtům mnohem vhodnější a lepší než je program MS EXCEL, jsou to například programy STATISTICA, SPSS, UNISTAT a mnoho dalších. Hlavní výhodou a předností tohoto programu je jeho velká dostupnost, prakticky v každé domácnosti se vyskytuje sada MS OFFICE, jejíž je EXCEL součástí.

EXCEL je typickým představitelem tabulkových procesorů, některá z jeho verzí je dostupná prakticky na každém počítači. Standardní součástí EXCELU je několik desítek statistických funkcí, které mohou být užity při statistických výpočtech. Je vybaven i poměrně kvalitní grafikou, která dovoluje pohodlné kreslení statistických grafů.

Základní člen EXCELU je buňka. Buňkou rozumíme políčko tabulky, do kterého umístíme požadovaný text, čísla nebo vzorce. Šířka buňky lze upravit dle požadavků tak, aby se celé číslo nebo text zobrazily kompletně. Buňka je definovaná souřadnicí sloupce a řádku. Sloupce jsou označeny písmeny. Řádky jsou označeny čísly. Množina buněk všech řádků a sloupců tvoří list. Označení listů najdeme jako záložky listů pod tabulkou. Listy můžeme přidávat, kopírovat, pojmenovat. Prostorové sestavě říkáme sešit. V aplikaci EXCEL jsou čísla uložena s přesností 15 platných číslic a s touto přesností se také provádějí výpočty. EXCEL umí počítat „pouze“ s daty, která se nacházejí v intervalu $\langle -1 \cdot 10^{307} ; 1 \cdot 10^{307} \rangle$.

2. Funkce v MS EXCEL

Program MS EXCEL 2003 disponuje mnoha funkcemi, jsou to například funkce logické, textové, statistické a matematické. Ve své bakalářské práci se však budu věnovat funkcím matematickým a hlavně statistickým.

2.1 Matematické funkce

Matematické funkce jsou v EXCELu jedny z nejdůležitějších funkcí. Jejich využití je velmi široké a použití jednoduché. Používají se i při výpočtu statistických funkcí. V následujících kapitolách stručně popíšu jen některé matematické funkce.

2.1.1 Absolutní hodnota

Zápis funkce: =ABS(argument)

Funkce vypočítá absolutní hodnotu zadaného čísla. Argument je reálné číslo, jehož absolutní hodnotu chcete zjistit.

Příklady:

ABS(2) - rovná se 2

ABS(-2) - rovná se 2

2.1.2 Pi

Zápis funkce: =PI()

Funkce vrátí číslo 3,14159265358979, matematickou konstantu π , s přesností na 14 desetinných míst. Do této funkce nezadááme žádný argument.

Příklady:

PI() =3.14.....

2.1.3 Odmocnina

Zápis funkce: =ODMOCNINA(argument)

Touto funkcí vypočítáme druhou odmocninu zadaného argumentu. Argument je číslo, jehož odmocninu chcete znát. Pokud je číslo záporné, vrátí funkce ODMOCNINA chybovou hodnotu #NUM!. EXCEL bohužel neumí pracovat s komplexními čísly.

Příklady:

$$\text{ODMOCNINA}(16) = 4$$

$$\text{ODMOCNINA}(-16) = \#NUM!$$

$$\text{ODMOCNINA}(\text{ABS}(-16)) = 4$$

2.1.4 Sinus

Zápis funkce: =SIN(Číslo)

„Číslo“ - je úhel v radiánech, jehož sinus chcete zjistit. Pokud je dané číslo ve stupních, pak jeho vynásobením hodnotou $\text{PI}()/180$ dostanete velikost úhlu v radiánech.

Příklady:

$$\text{SIN}(\text{PI}()) = 1,22\text{E-}16, \text{ což je přibližně nula. Sinus čísla } \square \text{ je nula.}$$

$$\text{SIN}(\text{PI}()/2) = \text{SIN}(90) = 1$$

$$\text{SIN}(-30) = -0,5$$

2.1.5 Suma

Sečte všechna čísla ve vybrané oblasti buněk.

Zápis funkce: =SUMA(číslo1;číslo2;...)

Číslo1, číslo2,... je 1 až 30 čísel, z nichž chceme sumu vypočítat.

Příklady:

SUMA(3; 2) - rovná se 5

SUMA("3"; 2; PRAVDA) = 6 (*Textové hodnoty jsou převáděny na čísla a logická hodnota PRAVDA je považována za číslo 1*).

2.2 Statistické funkce – základní

Statistika a statistické výpočty jsou v EXCELU silně podporovány. V této kapitole popíšu jen nejdůležitější funkce, které lze používat bez hlubších znalostí statistických metod. Další funkce popíšu spolu s teorií.

2.2.1 Maximum

Funkce maximum vrátí maximální hodnotu z daného seznamu argumentů. Tuto funkci využijeme později ke statistickým výpočtům například k výpočtu variačního rozpětí viz. dále

Zápis funkce: =MAX(číslo1;číslo2;)

Číslo1, číslo2,... je 1 až 30 čísel, mezi nimiž chcete nalézt maximální hodnotu.

Příklady: Jestliže oblast A1:A5 obsahuje čísla 10, 17, 19, 57, 2, pak:

MAX(A1:A5) rovná se 57

MAX(A1:A5;32) rovná se 57

MAX(A1:A5;132) rovná se 132

Poznámka: Funkce MAX je podobná funkci MIN. Jako argumenty můžete zadat čísla, prázdné buňky, logické hodnoty nebo čísla formátovaná jako text. Použijete-li jako argumenty chybové hodnoty nebo text, který nelze převést na číslo, bude výsledkem funkce chybová hodnota. Jestliže argumenty neobsahují žádná čísla, vrátí funkce MAX číslo 0

2.2.2 Minimum

Funkce minimum vypočítá minimální hodnotu z daného seznamu argumentů. Tuto funkci využijeme později ke statistickým výpočtům například k výpočtu variačního rozpětí viz. dále.

Zápis funkce: =MIN(číslo1;číslo2;)

Číslo1, číslo2,... je 1 až 30 čísel, mezi kterými se má najít minimální hodnota.

Příklady: Jestliže buňky A1:A4 obsahují čísla 11, 17, 5, 37

MIN(A1:A5) rovná se 5

MIN(A1:A4; 0) rovná se 0

2.2.3 Počet

Funkce POČET sečte počet buněk, které obsahují argumenty. Počítají se argumenty, které jsou čísla, datum nebo textové reprezentace čísel. Argumenty představující chybové hodnoty nebo text, který nelze převést na čísla, se ignorují.

Zápis funkce: =POČET(hodnota1;hodnota2;...)

Hodnota1, hodnota2,... je 1 až 30 argumentů, které mohou obsahovat různé datové typy nebo na ně odkazovat. Počítají se však pouze čísla.

Poznámka: Prázdné buňky, logické hodnoty, text nebo chybové hodnoty se ignorují. Jestliže potřebujeme počítat logické hodnoty, texty nebo chybové hodnoty, použijeme funkci POČET2.

Příklady: Jestliže buňky A1:A7 obsahují: 11, neděle, 1.1.2009, " ", PRAVDA, "prázdná buňka", 23

POČET(A1:A7) rovná se 3

POČET(A4:A7) rovná se 1

POČET(A2:A7; 2) rovná se 3

2.2.4 Počet2

Funkce POČET2 určí počet neprázdných buněk. Pomocí funkce POČET2 můžete zjistit, kolik buněk v oblasti nebo v matici obsahuje data.

Zápis funkce: =POČET2(hodnota1;hodnota2;...)

Hodnota1, hodnota2,... je 1 až 30 argumentů reprezentujících hodnoty, které chcete spočítat. V tomto případě je hodnota libovolný typ informace včetně prázdného textu (""), avšak s výjimkou prázdných buněk. Pokud je argument matice nebo odkaz, prázdné buňky uvnitř této matice či odkazu se ignorují.

Poznámka: Jestliže nepotřebujete počítat logické hodnoty, texty nebo chybové hodnoty, použijte funkci POČET.

Příklady: Jestliže buňky A1:A7 obsahují: 11, neděle, 1.1.2009, " ", PRAVDA, "prázdná buňka", 23

POČET2(A1:A7) = 6

POČET2(A4:A7) = 3

POČET2(A1:A7; 2) = 7

POČET2(A1:A7; "dvě") = 7

2.3 Statistické funkce

Statistiku bychom neměli ve svých úvahách o využití EXCELU v práci nebo ve studiu a priori zavrhnout. Statistické funkce nám mohou podstatně pomoci získat cenné informace.

Statistický soubor je množina prvků, které jsou předmětem statistického zkoumání. Prvky mají určité společné vlastnosti, kterým se ve statistice říká znaky. Znak lze ztotožnit s vlastnostmi prvku. Vzhledem k tomu, že vlastnosti (některé měřitelné) lze popsat proměnnými, lze znak ztotožnit i s proměnnou. Nezabýváme se všemi znaky, ale jen vybranými, sledovanými znaky. Ty se potom zpracovávají statistickými metodami. Znak se rozděluje na kvantitativní (číselné) a kvalitativní (popisné).

Ve statistice je velice frekventovaným pojmem náhodná veličina. To je taková veličina, jejíž hodnoty jsou jednoznačně určeny výsledkem určitého náhodného pokusu, či výběru. Hodnoty náhodné veličiny se při opakování pokusu vlivem náhodných jevů mění. Např.: tělesná výška, teplota okolí, množství srážek, počet vadných výrobků. Náhodné veličiny lze rozdělit na nespojitě (diskrétní) a spojitě.

Diskrétní veličiny mohou nabývat pouze početného počtu hodnot, zatímco spojitě veličiny nabývají hodnoty z nějakého intervalu. Obor všech hodnot náhodné veličiny nazýváme definičním oborem. Jedním ze základních problémů při studiu náhodných veličin je určení pravděpodobnosti, s jakou daná náhodná veličina nabude určité konkrétní hodnoty nebo hodnoty z určitého intervalu. Pravidlo, kterým se tato pravděpodobnost řídí, se nazývá rozdělení pravděpodobnosti náhodné veličiny. Rozdělení pravděpodobnosti náhodné veličiny se udává distribuční funkcí $F(x)$. Ta přiřazuje ke každému jevu popisovanému touto veličinou určitou pravděpodobnost. Pravděpodobnost, že diskrétní náhodná veličina X bude mít po provedení náhodného pokusu hodnotu x značíme $P(X = x)$. Distribuční funkce je v intervalu $\langle 0; 1 \rangle$.

Rozdělení pravděpodobnosti spojitě náhodné veličiny se určuje prostřednictvím funkce, kterou označujeme jako hustota pravděpodobnosti $f(x)$.

Mezi základní popisy náhodné veličiny patří teoretická rozdělení náhodné veličiny. Podle charakteru náhodné veličiny se rozeznávají dvě základní skupiny teoretických rozdělení: rozdělení spojitě náhodné veličiny a rozdělení diskrétní náhodné veličiny.

2.4 Popisná statistika

Statistiku, zabývající se sběrem údajů o všech prvcích nějaké přesně vymezené skupiny a jejich zpracováním, nazýváme popisná. Základním úkolem této statistiky je poskytnout věcně správné informace o průběhu jevů a procesů. Statistický popis se zaměřuje především na dvě hlavní vlastnosti každého rozdělení a to na polohu, čili velikost hodnot a na variabilitu, neboli měnivost hodnot sledovaného znaku. Poloha a variabilita se vyjadřuje pomocí číselných veličin, tzv. charakteristik polohy a charakteristik variability.

2.4.1 Výběrové charakteristiky polohy

Charakteristiky polohy určují střed celého rozdělení, kolem kterého budou pozorované hodnoty náhodné veličiny při opakování pokusu náhodně kolísat. Základní charakteristikou polohy je aritmetický průměr.

2.4.1.1 Aritmetický průměr

Aritmetický průměr je velice důležitá funkce, která patří k základním charakteristikám náhodné veličiny. V EXCELu je používání této funkce velice jednoduché.

Zápis funkce: =PRŮMĚR(číslo1; číslo2.)

Vypočítá se vztahem:

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

Funkce PRŮMĚR vypočítává aritmetický průměr čísel v oblasti tak, že sečte řady číselných hodnot a výsledek pak vydělí počtem hodnot. Do této funkce lze zadávat argumenty (číslo1; číslo2; ...), při čemž může obsahovat až 255 argumentů a ignoruje prázdné buňky a buňky obsahující logické či textové hodnoty.

2.4.1.2 Geometrický průměr

Dalším průměrem je geometrický průměr. Funkce geometrický průměr vypočítá geometrický průměr ze zadaných argumentů.

Zápis funkce: =GEOMEAN(číslo1;číslo2; .)

Geometrický průměr se počítá vztahem:

$$G = \sqrt[n]{\prod X_i} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

Poznámky: Pokud je některý z argumentů ≤ 0 , vrátí funkce chybovou hodnotu #NUM!. Jestliže matice nebo odkaz obsahuje text, logické hodnoty nebo prázdné buňky, jsou tyto hodnoty ignorovány. Funkce geometrický průměr umí vypočítat průměr až ze 170 argumentů. Funkci GEOMEAN lze například použít k výpočtu průměrné míry růstu daného složeného úrokování s proměnným úrokem.

Příklad 1: Nově založený podnik vykázal v letech 2005 až 2009 čistý zisk viz. tabulka 1. Určete průměrné tempo růstu podniku.

Rok	2005	2006	2007	2008	2009
Zisk (v mil. Kč)	1,1	2,5	4,4	9,2	18,3

Tabulka č.1: Hodnoty zisku v letech 2005-2009.

	A	B	C	D	E	F
1						
2	rok	2005	2006	2007	2008	2009
3	zisk	1,1	2,5	4,4	9,2	18,3
4	průměrný růst(%)	458,9917267				
5	Zápis funkce:	GEOMEAN(B3:F3)*100				
6						
7						

Obr.1: Výpočet geometrického průměru.

2.4.1.3 Harmonický průměr

Dalším typem průměru, kterému se ještě budu věnovat je harmonický průměr. Je to převrácená hodnota aritmetického průměru převrácených hodnot

Zápis funkce: = HARMEAN(číslo 1 ;číslo2; ..)

Harmonický průměr se počítá vztahem:

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}}$$

Poznámky: Pokud bude některý argument ≤ 0 , vrátí funkce chybovou hodnotu #NUM!. Jestliže matice nebo odkaz obsahuje text, logické hodnoty nebo prázdné buňky, jsou tyto hodnoty ignorovány. Harmonický průměr je vždy menší než geometrický průměr, který je vždy menší než aritmetický průměr. Funkce harmonický průměr umí vypočítat průměr až z 255 argumentů. Používá se, jsou-li hodnoty znaku nerovnoměrně rozloženy kolem aritmetického průměru, nebo když jsou hodnoty extrémně nízké či vysoké.

Příklad 2: Máme stroj, který vyrobí jeden výrobek za 20 sekund, druhý, který vyrobí jeden výrobek za 30 sekund a třetí stroj, který vyrobí jeden výrobek za 6 sekund. Vypočtete průměrnou rychlost výroby jednoho výroku.

číslo stroje	potřebný čas (s)
1. stroj	20
2. stroj	30
3. stroj	6
H= HARMEAN(C4:C6)	12

Tab.2: Výpočet harmonického průměru.

2.4.1.4 Výběrový medián

Uspořádáme-li všechny hodnoty z náhodného výběru do neklesající posloupnosti, pak výběrový medián (\tilde{X}) je hodnota, která leží uprostřed této posloupnosti. Polovina čísel má tedy hodnotu, která je větší nebo rovna mediánu a polovina čísel má hodnotu, která je menší nebo rovna mediánu[3].

Zápis funkce: =MEDIAN(číslo1;číslo2;...)

Medián se vypočítá vztahem:

$$\tilde{X} = X_{\left(\frac{n+1}{2}\right)} \quad \text{- je-li } n \text{ liché číslo}$$

$$\tilde{X} = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} \quad \text{- je-li } n \text{ sudé číslo}$$

Poznámka: Pokud je v souboru sudý počet hodnot, vypočítá funkce MEDIAN průměr ze dvou prostředních hodnot. *Číslo1, číslo2,...* je 1 až 30 čísel, z nichž má být vypočten medián.

Příklady:

MEDIAN(1; 2; 3; 4; 5; 6; 7) rovná se 4

MEDIAN(2; 3; 4; 5) rovná se 3,5 - neboli průměr z prostředních čísel 3 a 4

2.4.1.5 Výběrový modus

Modus (\hat{X}) je hodnota, která se v daném statistickém souboru vyskytuje nejčastěji, tzn. že je to hodnota znaku s největší relativní četností[3].

Zápis funkce: =MODE(číslo1;číslo2)

Číslo1, číslo2,... je 1 až 30 čísel, z nichž má být vypočten modus.

Poznámky: Modus je též hodnota o maximální pravděpodobnosti, resp. Hustotě pravděpodobnosti.

Pokud soubor neobsahuje žádné duplicitní údaje, vrátí funkce chybovou hodnotu #N/A.

Jestliže matice nebo odkaz obsahuje text, logické hodnoty nebo prázdné buňky, jsou tyto hodnoty ignorovány, buňky s nulovou hodnotou jsou však započítávány.

V případě stejných relativních četností různých znaků určí EXCEL jako modus nejmenší číslo!

2.4.2 Výběrové charakteristiky variability

Charakteristiky variability udávají míru rozptýlení hodnot statistického znaku na číselné ose. Ze zkušeností plyne, že čím je větší variabilita sledovaného znaku, tím méně reprezentativní jsou charakteristiky polohy.

2.4.2.1 Populační disperze

Populační disperze se vypočítá jako součet čtverců odchylek všech hodnot náhodného výběru od aritmetického průměru dělený rozsahem výběru. Tato disperze je efektivní a vychýlená[3].

Funkce je definována vztahem:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}; \text{ kde } n \text{ je počet hodnot náhodného výběru, } X_i \text{ je hodnota náhodné}$$

veličiny, S^2 je populační disperze a \bar{X} je aritmetický průměr.

Zápis funkce: VAR(číslo1;číslo2))

Číslo1, číslo2, ... je 1 až 30 argumentů, vztahujících se ke vzorku základního souboru.

Poznámky: Předpokládá se, že jsou zadány hodnoty pro všechny jednotky základního souboru. Logické hodnoty, jako například PRAVDA a NEPRAVDA, a textové řetězce jsou ignorovány.

2.4.2.2 Populační směrodatná odchylka

Funkce charakterizuje variabilitu náhodné veličiny ve stejných jednotkách v jakých jsou zadány její hodnoty. Populační odchylka vyjadřuje, jak se hodnoty liší od průměrné hodnoty (střední hodnoty).

Funkce je definována vztahem:

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}; \text{ kde } n \text{ je počet hodnot náhodného výběru, } X_i \text{ je hodnota náhodné}$$

veličiny, S je populační směrodatná odchylka a \bar{X} je aritmetický průměr.

Zápis funkce: =SMODCH(číslo 1 ;číslo2)

Číslo1, číslo2 je 1 až 30 argumentů, vztahujících se k základnímu souboru.

Poznámky: Textové a logické hodnoty, jako například PRAVDA a NEPRAVDA, jsou ignorovány.

2.4.2.3 Výběrová disperze

Výběrová disperze se vypočítá jako součet čtverců odchylek všech hodnot náhodného výběru od aritmetického průměru dělený rozsahem výběru mínus jedna. Tato disperze je konzistentní a nevychýlená.

Funkce je definována vztahem:

$$\overline{S^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)}; \text{ kde } \overline{S^2} \text{ je hodnota výběrové disperze, } \bar{X} \text{ je aritmetický průměr, } n \text{ je}$$

počet hodnot náhodného výběru, X_i je hodnota náhodné veličiny.

Zápis funkce: VAR.VÝBĚR(číslo1;číslo2;...)

Číslo1, číslo2, ... je 1 až 30 argumentů, vztahujících se ke vzorku základního souboru.

Poznámky: Předpokládá se, že nejsou zadány hodnoty pro všechny jednotky základního souboru. Logické hodnoty, jako například PRAVDA a NEPRAVDA, a textové řetězce jsou ignorovány.

2.4.2.4 Výběrová směrodatná odchylka

Tato funkce se používá pro odhad směrodatné odchylky základního souboru určený z náhodného výběru. Směrodatná odchylka vyjadřuje, jak se hodnoty liší od průměrné hodnoty (střední hodnoty).

Funkce je definována vztahem:

$$\bar{S} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)}} ; \text{ kde } \bar{S} \text{ je výběrová směrodatná odchylka, } n \text{ je počet hodnot náhodného}$$

výběru, X_i je hodnota náhodné veličiny a \bar{X} je aritmetický průměr.

Zápis funkce: =SMODCH.VÝBĚR(číslo 1 ; číslo2)

Číslo1, číslo2 je 1 až 30 argumentů, vztahujících se k základnímu souboru.

Poznámky: Textové a logické hodnoty, jako například PRAVDA a NEPRAVDA, jsou ignorovány.

2.4.2.5 Výběrový variační koeficient

Výběrový variační koeficient je definován jako podíl výběrové směrodatné odchylky (\bar{S}) a aritmetického průměru (\bar{X}).

Funkce je definována vztahem:

$$V = \frac{\bar{S}}{\bar{X}}$$

Zápis funkce: = SMODCH.VÝBĚR(číslo 1 ; číslo2)/ PRŮMĚR(číslo1; číslo2)

Číslo1, číslo2 je 1 až 30 argumentů, vztahujících se k základnímu souboru.

Poznámky: EXCEL bohužel neumí vypočítat výběrový variační koeficient přímo, můžeme však použít složenou funkci.

2.4.2.6 Výběrové variační rozpětí

Jde o rozdíl největší a nejmenší hodnoty statistického znaku náhodného výběru.

Funkce je definována vztahem:

$$R_{\text{var}} = X_{\text{max}} - X_{\text{min}}$$

Zápis funkce: = MAX(číslo1;číslo2;) - MIN(číslo1;číslo2;)

Číslo1, číslo2 je 1 až 30 argumentů, vztahujících se k základnímu souboru.

2.4.3 Koeficienty asymetrie a excesu

Koeficienty se používají pro porovnání průběhu zkoumaného rozdělení pravděpodobností s průběhem normálního rozdělení pravděpodobností.

2.4.3.1 Výběrový koeficient asymetrie

Výběrový koeficient asymetrie (S_k) udává míru nesymetrie náhodného výběru kolem aritmetického průměru. Šikmost označuje stupeň asymetričnosti rozdělení veličiny kolem střední hodnoty (\bar{X}). Kladné zešikmení označuje rozdělení s asymetrickou stranou, která se vychyluje směrem k více kladným hodnotám. Záporné zešikmení označuje rozdělení s asymetrickou stranou, která se vychyluje směrem k více záporným hodnotám.

Funkce je definována vztahem:

$$S_k = \frac{n}{(n-1)(n-2)} \sum \left(\frac{X_i - \bar{X}}{S} \right)^3 ; \text{ kde } S_k \text{ je výběrový koeficient asymetrie, } n \text{ je počet hodnot}$$

náhodného výběru, X_i je hodnota náhodné veličiny, \bar{X} je aritmetický průměr a S je směrodatná odchylka.

Zápis funkce: = SKEW(číslo1;číslo2;)

Číslo1, číslo2 je 1 až 30 argumentů, jejichž šikmost chceme spočítat.

Poznámky: Argumenty musí být čísla nebo názvy, matice nebo odkazy obsahující čísla. Jestliže matice nebo odkaz obsahuje text, logické hodnoty nebo prázdné buňky, jsou tyto hodnoty ignorovány, buňky s nulovou hodnotou jsou však započítávány.

2.4.3.2 Výběrový koeficient excessu

Výběrový koeficient špičatosti (excesu) je charakteristika rozdělení náhodné veličiny, která porovnává dané rozdělení s normálním rozdělením pravděpodobnosti. Výběrový koeficient excessu udává koncentraci hodnot náhodného výběru kolem aritmetického průměru.

Funkce je definována vztahem:

$$E_k = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)} ; \text{ kde } E_k \text{ je výběrový koeficient}$$

excesu, n je počet hodnot náhodného výběru, X_i je hodnota náhodné veličiny a \bar{X} je aritmetický průměr a S je směrodatná odchylka.

Zápis funkce: = KURT(číslo1;číslo2;)

Číslo1, číslo2 je 1 až 30 argumentů, jejichž špičatost chceme spočítat.

Poznámky: Argumenty musí být čísla nebo názvy, matice nebo odkazy obsahující čísla. Jestliže matice nebo odkaz obsahuje text, logické hodnoty nebo prázdné buňky, jsou tyto hodnoty ignorovány, buňky s nulovou hodnotou jsou však započítávány.

Pokud existují méně než čtyři datové body nebo pokud se směrodatná odchylka vzorku rovná nule, vrátí funkce KURT chybovou hodnotu #DIV/0!.

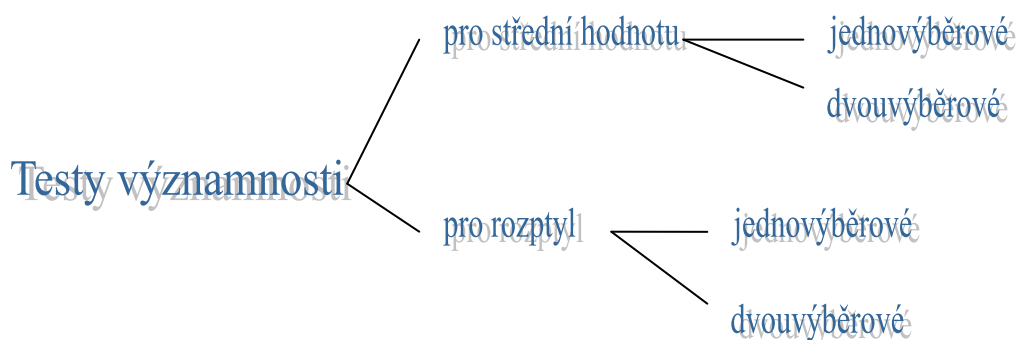
Aplikace EXCEL zobrazí chybu #DIV/0! v případě, že je nějaké číslo děleno nulou (0) nebo buňkou, která neobsahuje žádnou hodnotu.

3. Testy hypotéz

Testování statistických hypotéz patří mezi základní metody statistické indukce a nachází široké uplatnění v mnoha vědních oborech, např.: v ekonomii, technice, medicíně a dalších. Statistickou hypotézou se rozumí předpoklad o parametrech či tvaru rozdělení zkoumaného znaku. Hypotézu, jejíž platnost ověřujeme, nazveme nulovou hypotézou a označujeme ji zpravidla symbolem H_0 . Při testování hypotézy je třeba uvést, co bude platit, nebude-li platná nulová hypotéza. Toto druhé tvrzení nazýváme alternativní hypotéza, značíme ji H_1 a říkáme, že testujeme H_0 proti H_1 . Postup, kterým zjišťujeme platnost hypotézy, se nazývá testování hypotézy a provádí se pomocí statistického testu. Statistický test je jednoznačně dané pravidlo, které na základě realizace náhodného výběru určuje podmínky, za kterých hypotézu H_0 zamítneme nebo nezamítneme. Abychom mohli testovat hypotézu H_0 , musíme vhodně zvolit funkci náhodného výběru, pomocí které rozhodujeme o její platnosti. Tato funkce se nazývá testovací kritérium. Testovací kritérium je funkce náhodného výběru, jejíž tvar je závislý na testované hypotéze a rozdělení pravděpodobností základního souboru. Abychom mohli provést statistický test, je nutné, aby byla určena tzv. kritická oblast. Kritická oblast je podmnožina množiny hodnot testovacího kritéria, jejíž pravděpodobnost α je za předpokladu platnosti hypotézy tak malá, že náhodný jev „hodnota testovacího kritéria padne do kritické oblasti“ pokládáme za jev nemožný. Oblast přípustných hodnot je množina testovacího kritéria, které nepatří do kritické oblasti. Hladina významnosti testu je pravděpodobnost kritické oblasti α . Ve skutečnosti náhodný jev, že hodnota testovacího kritéria padne do kritické oblasti, není nemožný ale i v případě platnosti hypotézy může hodnota testovacího kritéria padnout do kritické oblasti s pravděpodobností α . Podle předem provedené dohody v tomto případě hypotézu zamítáme. Dopouštíme se tím chyby, kterou nazýváme chyba I. druhu (zamítnutí platné hypotézy). Pravděpodobnost chyby prvního druhu nazýváme též hladinou významnosti a značíme ji α . Zmenšit pravděpodobnost I. druhu můžeme zmenšením kritické oblasti. Současně s tím se ale zvětší oblast přípustných hodnot. Pak se může stát, že hodnota testovacího kritéria padne do oblasti přípustných hodnot, čímž se zvětší možnost přijetí této hypotézy. Dopustíme se tedy další chyby, kterou nazýváme chyba II. druhu (přijmeme neplatnou hypotézu). Její pravděpodobnost se značí β . Pravděpodobnost $1-\beta$ nazýváme síla testu a vyjadřuje, že hodnota testovacího kritéria padne správně“ do kritické oblasti [3]. Pro testování hypotéz v MS EXCEL musejí data v náhodných výběrech pocházet z normálního rozdělení pravděpodobností.

3.1 Testy významnosti

Při použití těchto testů musí rozdělení pravděpodobností základního souboru pocházet z normálního rozdělení pravděpodobností a testované hypotézy se týkají pouze parametrů základního souboru. U testů významnosti rozlišujeme mimo jiné testy jednovýběrové a testy dvouvýběrové. Testy zde popsané se používají u jednoho či dvou výběrů a jsou si dosti podobné. Je proto nutné mezi nimi velmi pečlivě volit. Záleží na tom, zda se testuje jednostranná hypotéza, např. že střední hodnota jednoho výběru je větší či menší než střední hodnota druhého výběru, nebo dvoustranná hypotéza, kdy se zjišťuje pravděpodobnost, že oba výběry mají střední hodnoty stejné anebo že se liší o určitou předem stanovenou hodnotu proti hypotéze, že tomu tak není. Záleží také na tom, zda se jejich rozptyly statisticky liší nebo nikoliv. Data v souborech se stejným rozptylem jsou homoskedastická, jinak jsou heteroskedastická.



Obr. 2: Rozdělení testů významnosti[5]

3.2 Jednovýběrový test významnosti pro střední hodnotu normálního rozdělení pravděpodobnosti se známým σ

Pro tento test uvažujeme náhodný výběr (X_1, X_2, \dots, X_n) o rozsahu n ze základního souboru $\mathbf{X} \sim N(\mu_0, \sigma)$ rozdělením pravděpodobností. Testovaná hypotéza H_0 má tvar:

H_0 : $EX = k$ a alternativní **H_1 :** $EX \neq k$ [3], kde EX je střední hodnota náhodného výběru (X_1, X_2, \dots, X_n) .

Zápis funkce: =ZTEST(pole, μ_0)

Pole je matice nebo oblast dat, proti které je testována hodnota μ_0 .

μ_0 je testovaná hodnota.

Funkce je definována vztahem:

$$Z = \frac{\bar{X} - k}{\sigma} \times \sqrt{n}$$

Náhodná veličina má $Z \sim N(0,1)$ rozdělení pravděpodobností. Kritická oblast W je definována jako množina těch hodnot testovacího kritéria Z , pro které platí: $W = \{Z: |Z| > z_\alpha\}$.

Poznámky: Pokud je pole prázdné, vrátí funkce ZTEST chybovou hodnotu #N/A.

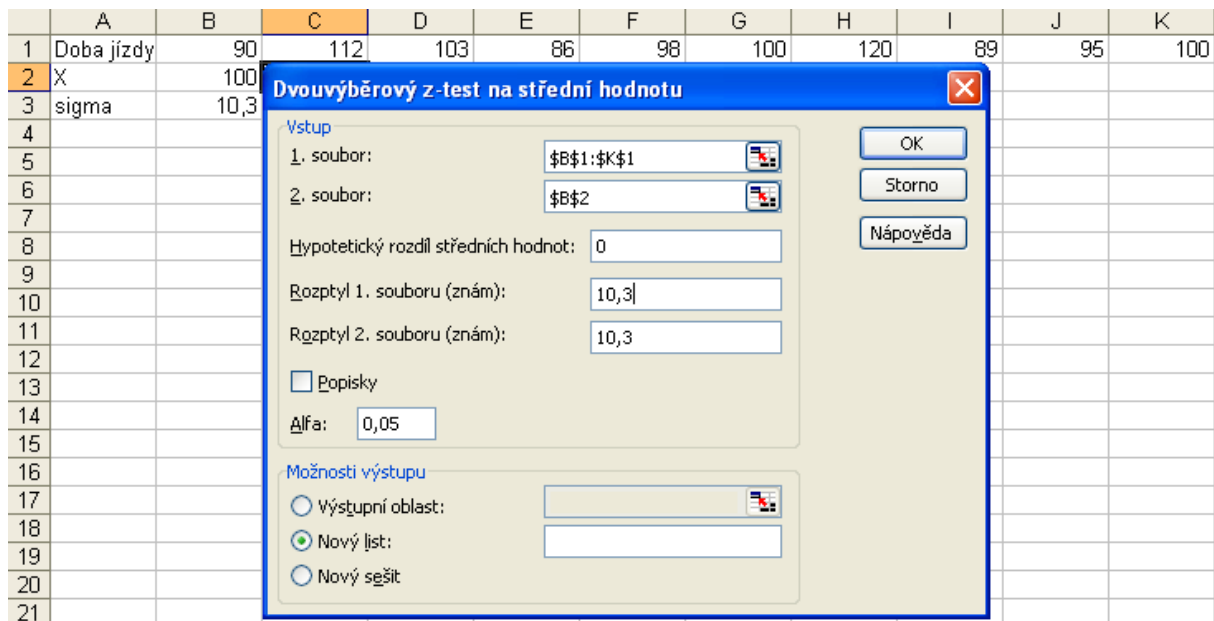
Příklad 3: Podle jízdního řádu je jízdní doba nedělního spoje mezi Prahou a Brnem 100 minut. Po deset neděl byl sledován příjezd tohoto spoje do Prahy a za předpokladu, že autobus vyjel z Brna včas, byly zaznamenány jízdní doby viz. tabulka. Na hladině významnosti $\alpha = 0,05$ testujte, zda doba uvedená v jízdním řádu odpovídá skutečnosti, jestliže víte, že hodnoty pocházejí ze základního souboru s normálním rozdělením pravděpodobností se směrodatnou odchylkou $\sigma = 10,3$.

jízda	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
doba jízdy	90	112	103	86	98	100	120	89	95	100

Tab.3: Doba jízdy spoje Brno- Praha

Řešení: Budeme testovat hypotézu, že doba jízdy autobusu odpovídá jízdnímu řádu. Testovaná hypotéza H_0 má tvar: **H_0 :** $EX = 100$ a alternativní **H_1 :** $EX \neq 100$.

Postup řešení 1: Nejjednodušší způsob výpočtu jednostranného z-testu spočívá v použití testu z nabídky Nástroje → Analýza dat → Dvouvýběrový Z-test. Tento test je sice určen k výpočtu testu dvouvýběrového, avšak když místo druhého souboru dat zadáme **k** vypočítá se nám test jednovýběrový. Výsledek naší hypotézy je pak v kolonce **z** a kritická hodnota je v kolonce **z krit (2)**.



Obr. 3: Výpočet jednovýběrového Z-testu na střední hodnotu.

	A	B	C	D
1	Dvouvýběrový z-test na střední hodnotu			
2				
3		<i>Soubor 1</i>	<i>Soubor 2</i>	
4	Stř. hodnota	99,3	100	
5	Znamý rozptyl	10,3	10,3	
6	Pozorování	10	1	
7	Hyp. rozdíl stř. hodnot	0		
8	z	-0,20796		
9	P(Z<=z) (1)	0,417629		
10	z krit (1)	1,644854		
11	P(Z<=z) (2)	0,835259		
12	z krit (2)	1,959964		

Tab. 4: Výsledek jednovýběrového Z-testu.

Postup řešení 2: Druhým způsobem výpočtu Z-testu je použití složeného vzorce. Tento způsob je však o něco složitější, avšak jeho výsledek je přesnější. Výpočet provádíme následujícím způsobem:

- 1) Vypočítáme si aritmetický průměr.
- 2) Spočítáme počet prvků.
- 3) Vypočítáme hodnotu Z-testu.

	A	B	C	D	E	F	G	H	I	J	K
1	Doba jízdy	90	112	103	86	98	100	120	89	95	100
2	k	100									
3	sigma	10,3									
4	průměr	99,3	PRŮMĚR(B1:K1)								
5	n	10	POČET(B1:K1)								
6	Z-test=	-0,21491	(B4-B2)/B3*ODMOCNINA(B5)								

Obr. 4: Výpočet jednovýběrového Z-testu vzorcem

Konečný tvar Z-testu:
$$\frac{(\text{PRŮMĚR}(\text{B1:K1}) - 100)}{10,3 * \text{ODMOCNINA}(\text{POČET}(\text{B1:K1}))}$$

Kritickou oblast **W** najdeme buďto v publikaci Kritické hodnoty a kvantily vybraných rozdělení pravděpodobností [4] nebo nám tuto hodnotu umí EXCEL vypočítat (viz. 3.2.1). Kritická hodnota je v tomto případě 1,959964.

Závěr: Z výpočtu vyplývá že **kritická hodnota** $> |z|$ tzn., že výsledek testu nespadá do oblasti kritických hodnot. Hypotézu H_0 nezamítáme. Doba uvedená v jízdním řádu odpovídá na hladině významnosti $\alpha = 0,05$ skutečnosti.

3.2.1 Výpočet kritické hodnoty pro Studentovo T-rozdělení pravděpodobnosti

Kritická hodnota nám odděluje oblast přípustných hodnot od hodnot kritických. Tuto mez můžeme najít v tabulkách [4] a nebo ji podle vzorce vypočítat. Pro výpočet použijeme funkci TINV, která vypočítá hodnotu distribuční funkce Studentova T-rozdělení pravděpodobnosti.

Zápis funkce: = TINV(pravděpodobnost;volnost)

Pravděpodobnost je pravděpodobnost daného dvojstranného t-rozdělení.

Volnost je počet stupňů volnosti charakterizující rozdělení.

Poznámky: Pokud některý z argumentů není číselného typu, vrátí funkce TINV chybovou hodnotu #HODNOTA!. Pokud pravděpodobnost < 0 nebo pokud pravděpodobnost > 1, vrátí funkce TINV chybovou hodnotu #NUM!. Není-li argument Volnost celé číslo, bude zkrácen. Pokud volnost < 1, vrátí funkce TINV chybovou hodnotu #NUM!.

3.3 Jednovýběrový test významnosti pro střední hodnotu normálního rozdělení pravděpodobnosti s neznámým σ

Uvažujeme náhodný výběr (X_1, X_2, \dots, X_n) o rozsahu n ze základního souboru X s $N(\mu_0, \sigma)$ rozdělením pravděpodobností. Testovaná hypotéza H_0 má tvar: **$H_0: EX=k$** a alternativní **$H_1: EX \neq k$** ^[3].

Funkce je definována vztahem:

$$T = \frac{\bar{X} - k}{S} \times \sqrt{n-1} \quad ; \text{ kde } S \text{ je populační směrodatná odchylka základního souboru, } k \text{ je}$$

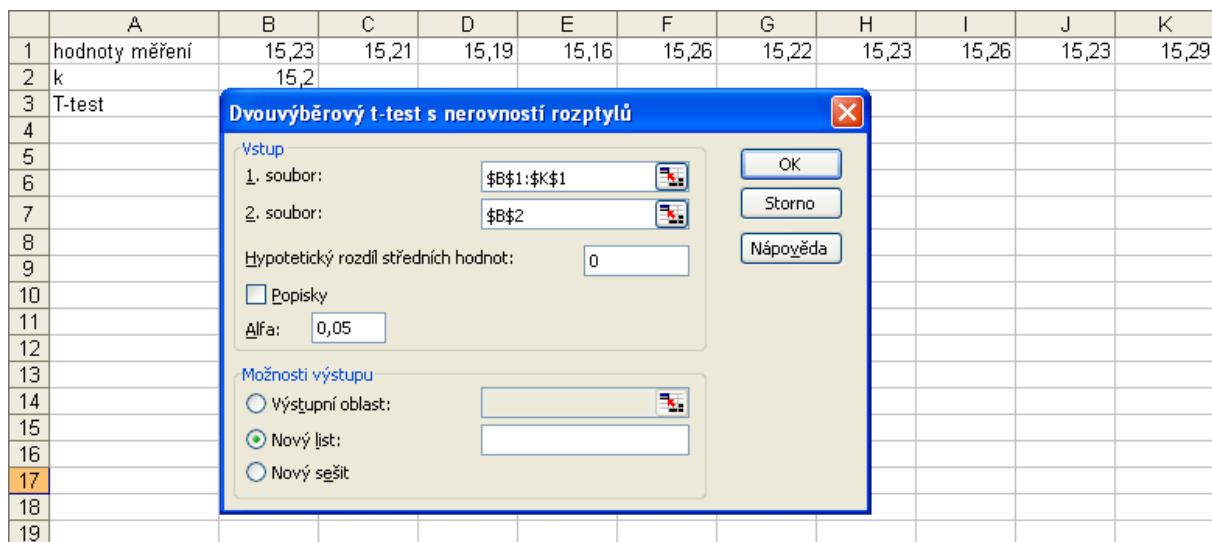
zadaná konstanta, \bar{X} je aritmetický průměr a n je počet hodnot.

Náhodná veličina má T Studentovo rozdělení pravděpodobnosti s $n-1$ stupni volnosti. Kritická oblast W je definována jako množina těch hodnot testovacího kritéria T , pro které platí: $W = \{T: |T| > t_{\alpha, n-1}\}$.

Příklad 4: Pro kontrolu správnosti nastavení měřicího přístroje bylo provedeno 10 zkušebních měření se správnou hodnotou $\mu_0=15,2$. Byly zjištěny tyto výsledky: 15,23; 15,21; 15,19; 15,16; 15,26; 15,22; 15,23; 15,26; 15,23; 15,29. Předpokládáme, že chyba měření má normální rozdělení pravděpodobnosti. Ověřte na hladině významnosti $\alpha=0,05$, zda je chyba měření zatížena systematickou chybou.

Řešení: Budeme testovat hypotézu, že střední hodnota hodnot měřených přístrojem je správná, tj. rovna 15,20. Použijeme jednovýběrový T-test o střední hodnotě s neznámým σ . Testovaná hypotéza H_0 má tvar: **$H_0: EX=15,2$** a alternativní **$H_1: EX \neq 15,2$** .

Postup řešení 1: Nejjednodušší způsob výpočtu jednostranného T-testu spočívá v použití testu z nabídky **Nástroje** → **Analýza dat** → **Dvouvýběrový T-test**. Tento test je sice určen k výpočtu testu dvouvýběrového, avšak když místo druhého souboru dat zadáme **k** vypočítá se nám test jednovýběrový. Výsledek naší hypotézy je pak v kolonce **t**. Nevýhodou tohoto testu je, že nám nespočítá kritické hodnoty.



Obr. 5: Výpočet jednovýběrového T- testu pomocí nástroje Analýza dat.

	A	B	C	D	E
1	Dvouvýběrový t-test s nerovností rozptylů				
2					
3		<i>Soubor 1</i>	<i>Soubor 2</i>		
4	Stř. hodnota	15,228	15,2		
5	Rozptyl	0,001373333	#DIV/0!		
6	Pozorování	10	1		
7	Hyp. rozdíl stř. hodnot	0			
8	Rozdíl	0			
9	t stat	2,389296521			
10	P(T<=t) (1)	#NUM!			
11	t krit (1)	#NUM!			
12	P(T<=t) (2)	#NUM!			
13	t krit (2)	#NUM!			

Tab. 5: Výsledek jednovýběrového T- testu.

Postup řešení 2: Druhou možností řešení jednovýběrového T- testu je použití složeného vzorce. Tento způsob je však o něco složitější, avšak jeho výsledek je přesnější. Výpočet provádíme následujícím způsobem: 1) Spočítáme počet prvků.

- 2) Vypočítáme si aritmetický průměr.
- 3) Vypočítáme směrodatnou odchylku.
- 4) Vypočítáme hodnotu jednovýběrového T-testu.

	A	B	C	D	E	F	G	H	I	J	K
1	hodnoty měření	15,23	15,21	15,19	15,16	15,26	15,22	15,23	15,26	15,23	15,29
2	k	15,2									
3	n	10	POČET(B1:K1)								
4	průměr	15,228	PRŮMĚR(B1:K1)								
5	směrodatná odchylka	0,035157	SMODCH(B1:K1)								
6	T- test	2,389297	(B4-B2)/B5*ODMOCNINA(B3-1)								

Obr. 6: Výpočet jednovýběrového T-testu na střední hodnotu.

Konečný tvar T-testu:

$$=(\text{PRŮMĚR}(B1:K1)-15,2)/\text{SMODCH}(B1:K1)*\text{ODMOCNINA}(\text{POČET}(B1:K1)-1)$$

Kritickou oblast **W** najdeme buďto v publikaci Kritické hodnoty a kvantily vybraných rozdělení pravděpodobností [4] nebo nám tuto hodnotu umí EXCEL vypočítat (viz. 3.2.1). Kritická hodnota je v tomto případě 2,2622.

Závěr: Z výpočtu vyplývá **kritická hodnota** < |T|, tzn., že výsledek testu spadá do oblasti kritických hodnot. Hypotézu H_0 zamítáme. Chyba měření je na hladině významnosti $\alpha=0,05$ zatížena statistickou chybou.

3.3.1 Výpočet kritické hodnoty pro χ^2 rozdělení pravděpodobnosti

Kritická hodnota nám odděluje oblast přípustných hodnot od hodnot kritických. Tuto mez můžeme najít v tabulkách [4] a nebo ji podle vzorce vypočítat. K výpočtu kritické hodnoty χ^2 rozdělení pravděpodobnosti nám slouží funkce CHIINV.

Zápis funkce: = CHIINV(pravděpodobnost;volnost)

Pravděpodobnost je pravděpodobnost χ^2 rozdělení pravděpodobnosti.

Volnost je počet stupňů volnosti.

Poznámky: Pokud některý z argumentů není číselného typu, vrátí funkce CHIINV chybovou hodnotu #HODNOTA!. Jestliže je argument Pravděpodobnost < 0 nebo pravděpodobnost > 1, vrátí funkce CHIINV chybovou hodnotu #NUM!. Není-li argument Volnost celé číslo, bude zaokrouhlen směrem dolů. Pokud je argument Volnost < 1 nebo Volnost ≥ 1010 , vrátí funkce CHIINV chybovou hodnotu #NUM!.

3.4 Jednovýběrový test významnosti pro rozptyl

Uvažujeme náhodný výběr (X_1, X_2, \dots, X_n) o rozsahu n ze základního souboru X s $N(\mu_0, \sigma)$ rozdělením pravděpodobností. Testujeme nulovou hypotézu, že rozptyl základního souboru je roven konstantě k . Testovaná hypotéza H_0 má tvar: $H_0: DX=k$ a alternativní $H_1: DX \neq k$ [3].

Funkce je definována vztahem:

$\chi = \frac{nS^2}{k}$; kde χ je rozptyl, S^2 je populační rozptyl základního souboru, n je počet hodnot základního souboru a k je testovaná konstanta.

Příklad 5: Při uvedení do provozu byl dávkovač seřízen tak, aby směrodatná odchylka velikostí dávek byla přesně rovna hodnotě 0,4. Po čase byla provedena kontrola, zda se provozní parametry dávkovače nezměnily. Změřením 11 vzorků (dávek) se zjistily tyto jejich hodnoty: 50,12; 49,65; 48,85; 50,56; 50,23; 49,13; 49,10; 50,77; 49,34; 49,86; 50,45. Vhodným testem rozhodněte, zda lze spolehlivě tvrdit ($\alpha=0,05$), že se přesnost dávkovače zhoršila (tj. zmenšila), předpokládáme-li normalitu rozdělení velikosti dávek.

Řešení: Budeme testovat hypotézu, že rozptyl velikostí dávek není větší než jeho dřívější hodnota $0,4^2$ proti alternativě, že je tento rozptyl větší. Použijeme tedy jednostrannou variantu jednovýběrového testu o rozptylu: $H_0: \chi \geq 0,16$ proti $H_1: \chi < 0,16$.

Postup řešení: EXCEL bohužel neumí vypočítat jednovýběrový test pro rozptyl přímo, musíme tedy použít složenou funkci.

Výpočet provádíme následujícím způsobem: 1) Spočítáme počet prvků.

2) Vypočítáme si rozptyl.

3) Vypočítáme hodnotu jednovýběrového χ -testu.

4) Vypočítáme kritickou hodnotu pro χ^2 rozdělení

	A	B	C	D	E	F	G	H	I	J	K	L
1	hodnoty dávek	50,12	49,65	48,85	50,56	50,23	49,13	49,1	50,77	49,34	49,86	50,45
2	k	0,16										
3	n	11	POČET(B1:L1)									
4	rozptyl	0,430125	VAR.VÝBĚR(B1:L1)									
5	χ - test	29,5711	POČET(B1:L1)*VAR.VÝBĚR(B1:L1)/0,16									
6	χ krit	18,30704	CHIINV(0,05;10)									

Obr. 7: Výpočet jednovýběrového jednostranného χ -testu.

Konečný tvar χ -testu je: =POČET(B1:L1)x VAR.VÝBĚR(B1:L1)/0,16

Kritickou hodnotu najdeme buďto v publikaci Kritické hodnoty a kvantily vybraných rozdělení pravděpodobností [4], nebo nám tuto hodnotu umí EXCEL vypočítat (viz. kap. 3. 3. 1). Kritická hodnota je v tomto případě 18,30704.

Závěr: Z výpočtu vyplývá že **kritická hodnota** $< \chi$, tzn., že výsledek testu spadá do oblasti kritických hodnot. Hypotézu H_0 zamítáme. Provozní přesnost dávkovače se na hladině významnosti $\alpha=0,05$ nezhoršila.

3.4.1 Výpočet kritické hodnoty pro F rozdělení pravděpodobnosti

Kritická hodnota nám odděluje oblast přípustných hodnot od hodnot kritických. Tuto mez můžeme najít v tabulkách[4] a nebo ji podle vzorce vypočítat. Pomocí funkce FINV vypočítáme kritickou hodnotu pro F rozdělení pravděpodobnosti.

Zápis funkce:= FINV(prst;volnost1;volnost2)

Prst je pravděpodobnost α rozdělení F. Do tabulky zadáváme hodnotu α ze zadání dělenou dvěma!

Volnost1 je počet stupňů volnosti v čitateli.

Volnost2 je počet stupňů volnosti ve jmenovateli.

Poznámky:

Pokud některý z argumentů není číselného typu, vrátí funkce FINV chybovou hodnotu #HODNOTA!.

Jestliže je argument Prst < 0 nebo Prst > 1 , vrátí funkce FINV chybovou hodnotu #NUM!.

Není-li jeden z argumentů Volnost1 nebo Volnost2 celé číslo, bude zkrácen.

Pokud je argument Volnost1 < 1 nebo Volnost1 ≥ 1010 , vrátí funkce FINV chybovou hodnotu #NUM!.

Je-li argument Volnost2 < 1 nebo Volnost2 ≥ 1010 , vrátí funkce FINV chybovou hodnotu #NUM!.

3.5 Dvouvýběrový test významnosti pro rozptyl **F-Test**

Tento test je mimo jiné nezbytný pro výběr správného testu o rovnosti středních hodnot dvou výběrů. Před jeho provedením je nutné se přesvědčit, že oba výběry nejsou závislé. Testuje se shodnost variancí dvou výběrů podle Fischer-Snedecorova rozdělení pravděpodobnosti. Nulová hypotéza oboustranného testu předpokládá rovnost rozptylů, alternativní hypotéza jejich nerovnost. Testuje se poměr rozptylů obou výběrů, přičemž větší rozptyl musí být v čitateli (poměr větší nebo rovný 1). Hypotéza o rovnosti rozptylů se zamítá, jestliže je tento poměr větší než kritická hodnota F-rozdělení. Pokud by hodnota F vyšla menší než 1, je třeba oba výběry v tabulce prohodit a test opakovat.

Ve vstupním dialogu je hladina významnosti označena jako „alfa“, tato alfa odpovídá jednostrannému testu, a proto musíte pro dvojstranný test zadat hladinu významnosti dělenou dvěma, výstup obsahuje rozptyly a střední hodnoty, počty stupňů volnosti (zde označené jako rozdíly), ale hlavně hodnotu F a kritickou $F_{\text{krit}}(1)$ (tj. pro jednostranný test). Pokud však zadáte $\alpha = \alpha/2$, $F_{\text{krit}}(1)$ bude odpovídat dvojstrannému testu.

Data v souborech se stejným rozptylem jsou homoskedastická, jinak jsou heteroskedastická. Na této vlastnosti závisí testy středních hodnot, jako je dvouvýběrový T-test s rovností rozptylů nebo dvouvýběrový T-test s nerovností rozptylů (viz dále).

Zápis funkce: =FTEST(pole 1;pole2)

Funkce je definována vztahem:

$$F = \frac{\max\{\overline{S_1^2}, \overline{S_2^2}\}}{\min\{\overline{S_1^2}, \overline{S_2^2}\}}; \text{ kde } \overline{S_1^2}, \overline{S_2^2} \text{ jsou výběrové rozptyly z náhodných výběrů X a Y.}$$

Náhodná veličina má Fischer-Snedecorovo rozdělení pravděpodobnosti s $n_1 - 1$ a $n_2 - 1$ stupni volnosti. Kritická oblast W je definována jako množina těch hodnot testovacího kritéria F , pro které platí: $W = \{F: F > F_2\}$ kde $F_2 = F_{F(n_1-1, n_2-1)}^{-1}(\frac{2-\alpha}{2})$. Hodnotu F_2 nám EXCEL vypočítá jako $F_{\text{krit}}(1)$ nebo ji nalezneme v tabulkách [4].

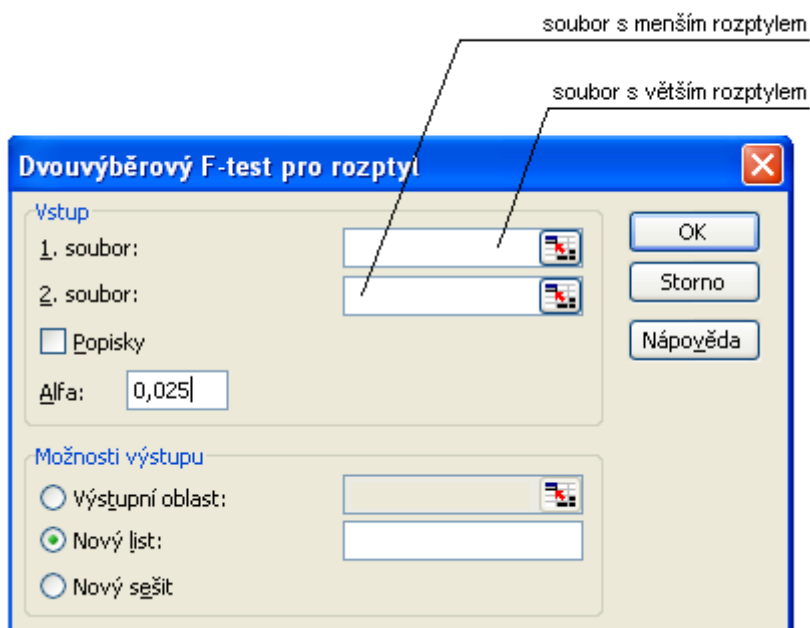
Příklad 6: Nově vzniklá linka pro stáčení nealkoholických nápojů vybírá dávkovač. Máme na výběr ze dvou nabídek. O výběru rozhodne přesnost dávkování. Rozhodněte na hladině významnosti $\alpha = 0,05$, zda je přesnost dávkovačů stejná za předpokladu, že náhodný výběr pochází z normálního rozdělení pravděpodobností. Bylo provedeno 10 měření viz. tabulka.

dávkovač X	1,47	1,51	1,49	1,48	1,52	1,46	1,47	1,48	1,49	1,50
dávkovač Y	1,49	1,50	1,48	1,47	1,49	1,51	1,46	1,47	1,49	1,50

Tab. 5: Výsledky dávkovačů

Řešení: Máme ověřit, že rozptyly obou dávkovačů jsou přibližně stejné, proto použijeme dvouvýběrový F-test. Testujeme hypotézu $H_0: DX=DY$ proti alternativní hypotéze $H_1: DX \neq DY$.

Postup řešení 1: K výpočtu dvouvýběrového F-testu použijeme test z nabídky Nástroje → Analýza dat → Dvouvýběrový F- test. Jedinou možnou chybou je zadání hodnot s větším rozptylem do 1. souboru dat, v tomto případě by nám F- test vyšel větší než 1 a to není možné.



Obr. 8: Postup při vyplňování hodnot.

	A	B	C	D
1	Dvouvýběrový F-test pro rozptyl			
2				
3		<i>Soubor 1</i>	<i>Soubor 2</i>	
4	Stř. hodnota	1,487	1,486	
5	Rozptyl	0,000357	0,000249	
6	Pozorování	10	10	
7	Rozdíl	9	9	
8	F	1,433036		
9	P(F<=f) (1)	0,300288		
10	F krit (1)	4,025994		
11				
12				

Tab. 6: Výpočet dvouvýběrového F-testu.

Postup řešení 2: Druhou možností řešení dvouvýběrového F- testu je použití vzorce. Výpočet provádíme následujícím způsobem:

- 1) Spočítáme rozptyl souboru X.
- 2) Spočítáme rozptyl souboru Y.
- 3) Vypočítáme hodnotu F- testu.
- 4) Vypočítáme kritickou hranici F rozdělení.

	A	B	C	D	E	F	G	H	I	J	K
1	dávkovač X	1,47	1,51	1,49	1,48	1,52	1,46	1,47	1,48	1,49	1,5
2	dávkovač Y	1,49	1,5	1,48	1,47	1,49	1,51	1,46	1,47	1,49	1,5
3	rozptyl X	0,000357	VAR.VÝBĚR(B1:K1)								
4	rozptyl Y	0,000249	VAR.VÝBĚR(B2:K2)								
5	F- test	1,433036	B3/B4								
6	F krit	4,025994	FINV(0,025;9;9)								

Obr.. 9: Výpočet dvouvýběrového F- testu pomocí vzorce

Konečný tvar F- testu: =VAR.VÝBĚR(B1:K1)/VAR.VÝBĚR(B2:K2)

Závěr: Z výpočtu vyplývá že kritická hodnota > F, tzn., že výsledek testu nespadá do oblasti kritických hodnot. Hypotézu H_0 nezamítáme. Rozptyly obou dávkovačů jsou na hladině významnosti $\alpha=0,05$ přibližně stejné.

3.6 Dvouvýběrový test významnosti pro střední hodnoty se známým σ Z-test

Uvažujeme, že náhodný výběr $(X_1, X_2, \dots, X_{n_1})$ je vytvořen ze základního souboru X s $N(\mu_1, \sigma_1)$ rozdělením pravděpodobností, náhodný výběr $(Y_1, Y_2, \dots, Y_{n_2})$ je tvořen ze základního souboru Y s $N(\mu_2, \sigma_2)$ rozdělením pravděpodobností. Dále předpokládáme, že náhodné veličiny $X_1, X_2, \dots, X_{n_1}, Y_1, Y_2, \dots, Y_{n_2}$ jsou nezávislé a parametry σ_1 a σ_2 známe.

Testujeme nulovou hypotézu, že střední hodnota základního souboru X se rovná střední hodnotě základního souboru Y . **$H_0: EX = EY$** proti **$H_1: EX \neq EY$**

Funkce je definována vztahem:

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}; \text{ kde } \sigma_1 \text{ a } \sigma_2 \text{ jsou známé směrodatné odchylky náhodných výběrů } X \text{ a } Y,$$

\bar{X}, \bar{Y} jsou střední hodnoty náhodných výběrů a n_1, n_2 jsou rozsahy výběrů X a Y .

Náhodná veličina Z má $N(0,1)$ rozdělení pravděpodobností. Kritická oblast W je definována jako množina těch hodnot testovacího kritéria Z , pro které platí: $W = \{Z: |Z| > z_\alpha\}$.

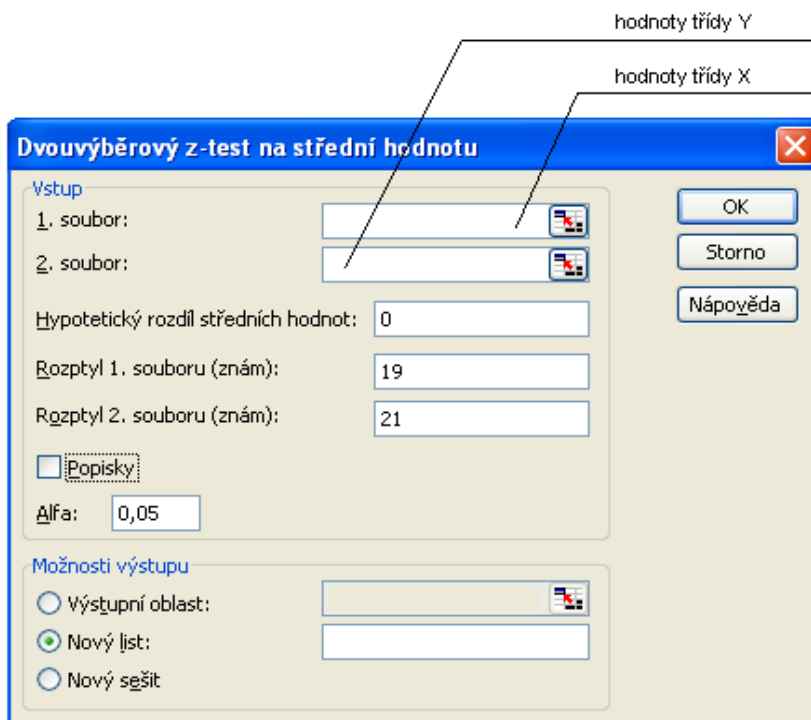
Příklad 7: Učitel matematiky chtěl zjistit, zda-li je úroveň vědomostí jeho žáků v jeho třídách přibližně stejná. Každý student dostal test se 100 otázkami. Počty správných odpovědí jsou uvedeny v tabulce. Na hladině významnosti $\alpha = 0,05$ zjistěte, zda jsou významné rozdíly ve vědomostech studentů mezi třídami. Uvažujme, že hodnoty v tabulce mají normální rozdělení pravděpodobnosti a je známo, že směrodatná odchylka počtu správných odpovědí ve třídě X je 19 a ve třídě Y je 21.

třída X	44	63	38	21	70	32	18	41	30	24	37	14	29	81	34
třída Y	33	45	71	18	25	63	38	22	15	48	36	84	21	65	20

Tab. 7: Tabulka počtu správných odpovědí.

Řešení: Naším úkolem je ověřit, zda-li jsou vědomosti studentů z různých tříd přibližně stejné. Směrodatné odchylky jsou zadány, a proto použijeme dvouvýběrový Z-test. Testujeme hypotézu **$H_0: EX = EY$** proti alternativní hypotéze **$H_1: EX \neq EY$** .

Postup řešení : K výpočtu dvouvýběrového Z-testu použijeme nástroj z nabídky Nástroje → Analýza dat → Dvouvýběrový Z-test. Postup vyplňování hodnot je jednoduchý a je uveden na následujícím obrázku.



Obr. 9: Postup při vyplňování hodnot Z-testu.

	A	B	C	D	E
1	Dvouvýběrový z-test na střední hodnotu				
2					
3		<i>Soubor 1</i>	<i>Soubor 2</i>		
4	Stř. hodnota	38,4	40,26667		
5	Známy rozptyl	19	21		
6	Pozorování	15	15		
7	Hyp. rozdíl stř. hodnot	0			
8	z	-1,1430952			
9	P(Z<=z) (1)	0,12649953			
10	z krit (1)	1,64485363			
11	P(Z<=z) (2)	0,25299906			
12	z krit (2)	1,95996398			

Tab. 8: Výpočet dvouvýběrového Z -testu.

Kritickou hodnotu nám v tomto případě vypočítá samotný T- test. Další možností jejího určení je její výpočet viz. kap. 3. 2. 1. nebo ji nalezneme v tabulkách [4].

Závěr: Z výpočtu vyplývá $z_{krit(2)} > |P(Z \leq z)|$ tzn., že výsledek testu nespadá do oblasti kritických hodnot. Hypotézu H_0 nezamítáme. Vědomosti studentů z různých tříd jsou na hladině významnosti $\alpha=0,05$ přibližně stejné.

3.7 Dvouvýběrový T- test s rovností rozptylů

Jsou-li data homoskedastická, tj. variance obou výběrů jsou stejné, dá se provést testování hypotézy o určitém rozdílu středních hodnot tímto nástrojem. Pokud jsou však heteroskedastická, použije se pro stejný účel další T-test (viz část 3.8). Rozhodující je pro tuto volbu F-test (viz 3.5).

Nulová hypotéza oboustranného testu má tvar $H_0: EX=EY$ a alternativní $H_1: EX \neq EY$. Nejčastěji se testuje nulový rozdíl, takže hypotézy testují shodnost či rozdílnost obou středních hodnot.

Vstupní dialog předpokládá dva výběry dat, nejlépe ve sloupcích (ve vstupním dialogu položky Soubor 1, Soubor 2). Dále se zadá výstupní oblast, hypotetický rozdíl středních hodnot (např. 0) a alfa (hladina významnosti). Kritické hodnoty a pravděpodobnosti se spočítají pro jednostranný i oboustranný test a označí symboly (1), resp. (2).

Výstup pro oboustranné testování obsahuje hodnotu t-kritéria (deklarovaného jako t stat (2)), která může být i záporná. V testu ji pak bereme v absolutní hodnotě. Pokud testujeme určitý nenulový rozdíl, je důležité jej správně polarizovat; musí být vypočten jako předpokládaný průměr levého výběru minus pravého výběru.

Zápis funkce: =TTEST(Soubor 1;Soubor 2;Strany;Typ)

Poznámky: Strany – určuje, zda se jedná o jednostranný (1) nebo oboustranný (2) test.

Typ – určuje, o jaký typ T-testu se jedná: 1- párové hodnoty; 2- dva výběry se shodným rozptylem, 3- dva výběry s různým rozptylem.

Funkce je definována vztahem:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{n_1 S_1^2 + n_2 S_2^2}} \times \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}; \text{ kde } \bar{S}_1^2 \text{ a } \bar{S}_2^2 \text{ jsou výběrové rozptyly náhodných}$$

výběrů X a Y, \bar{X}, \bar{Y} jsou střední hodnoty náhodných výběrů a n_1, n_2 jsou rozsahy výběrů X a Y.

Náhodná veličina má Studentovo rozdělení pravděpodobnosti s n_1+n_2-2 stupni volnosti. Kritická oblast W je definována jako množina těch hodnot testovacího kritéria T, pro které platí: $W = \{T: |T| > t_{\alpha, n_1+n_2-2}\}$.

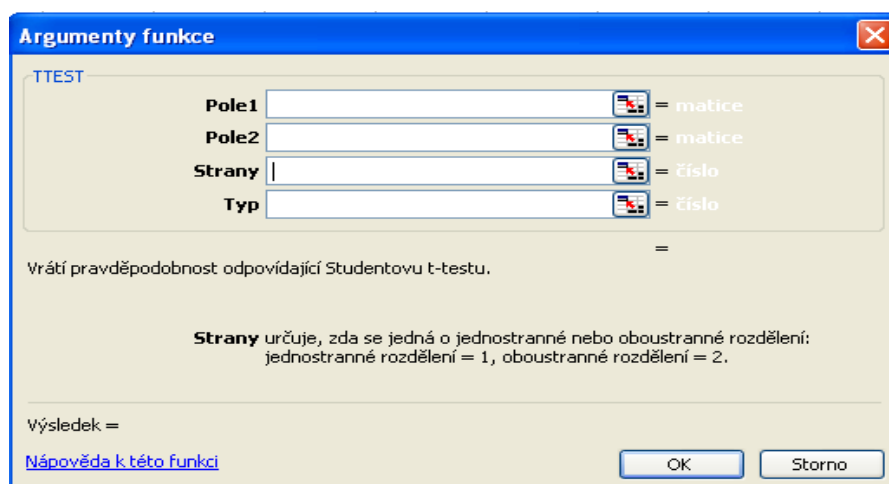
Příklad 8: Zjistěte, zda dva různé druhy nosníků mají stejnou nosnost a zda by bylo možné jeden nahradit druhým. U každého druhu bylo provedeno 8 zkoušek nosnosti v tunách na m². Výsledky jsou uvedeny v tabulce. Předpokládejte, že nosnost nosníku je náhodná veličina s normálním rozložením pravděpodobností. ($\alpha = 0,05$) a $\sigma_1 = \sigma_2$.

1.druh	3,4	3,5	3,2	3,6	3,8	3,9	3,0	3,1
2.druh	3,3	3,6	3,8	2,9	3,9	2,8	2,9	3,8

Tab. 9: Nosnosti nosníků.

Řešení: Naším úkolem je zjistit, zda-li je pevnost nosníků přibližně stejná (dají se zaměnit). V zadání máme dva výběry jejichž rozptyly jsou přibližně stejné a proto použijeme dvouvýběrový T-test s rovností rozptylů. Testovaná hypotéza má tvar: $H_0: EX=EY$ a alternativní $H_1: EX \neq EY$. Předpokládáme, že střední hodnoty obou výběru jsou stejné.

Postup řešení: K výpočtu dvouvýběrového T-testu s rovností rozptylů použijeme test z nabídky Nástroje → Analýza dat → Dvouvýběrový T- test s rovností rozptylů.



Obr. 9: Tabulka pro vyplňování hodnot dvouvýběrového T-testu.

	A	B	C	D	E
1	Dvouvýběrový t-test s rovností rozptylů				
2					
3		<i>Soubor 1</i>	<i>Soubor 2</i>		
4	Stř. hodnota	3,4375	3,375		
5	Rozptyl	0,105535714	0,210714		
6	Pozorování	8	8		
7	Společný rozptyl	0,158125			
8	Hyp. rozdíl stř. hodnot	0			
9	Rozdíl	14			
10	t stat	0,314347307			
11	P(T<=t) (1)	0,378946441			
12	t krit (1)	1,761310115			
13	P(T<=t) (2)	0,757892882			
14	t krit (2)	2,144786681			

Tab. 10: Výpočet dvouvýběrového T-testu.

Postup řešení 2: Druhou možností řešení tohoto příkladu je použití funkce TTEST. Zobrazí se nám stejná tabulka jako na obr. 9. do kolonky Typ zadáme hodnotu 2, která odpovídá dvouvýběrovému T- testu s rovností rozptylů.

Kritickou hodnotu nám v tomto případě vypočítá samotný T- test. Další možností jejího určení je její výpočet viz. kap. 3. 2. 1. nebo její hodnotu nalezneme v tabulkách [4].

Závěr: Z výpočtu vyplývá **t krit(2)>|t stat|** tzn., že výsledek testu nespadá do oblasti kritických hodnot. Hypotézu H_0 nezamítáme. Pevnost nosníků je na hladině významnosti $\alpha=0,05$ přibližně stejná, dají se zaměnit.

3.8 Dvouvýběrový T-test s nerovností rozptylů

Platí zde totéž jako u dvouvýběrového T-testu s rovností rozptylů, pouze by měl být tento test používán při výběrech s odlišnými rozptyly (heteroskedasticitě dat). Vhodným předběžným testem na shodnost rozptylů je F-test - hladinu významnosti je třeba zvolit stejnou pro všechny související testy.

Tento test testuje opět nulovou hypotézu $H_0: EX=EY$ s alternativní hypotézou $H_1: EX \neq EY$, předpokládá se však statistická nerovnost směrodatných odchylek.

Zápis funkce: =TTEST(Soubor 1;Soubor 2;Strany;Typ)

Poznámky: Strany – určuje, zda se jedná o jednostrannou nebo oboustrannou hypotézu.

Typ – určuje, o jaký typ T-testu se jedná, 1- párové hodnoty, 2- dva výběry se shodným rozptylem, 3- dva výběry s různým rozptylem. V našem případě zvolíme hodnotu 3.

Funkce je definována vztahem:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}}}; \text{ kde } S_1^2 \text{ a } S_2^2 \text{ jsou výběrové rozptyly náhodných výběrů X a Y, } \bar{X}, \bar{Y}$$

jsou střední hodnoty náhodných výběrů a n_1, n_2 jsou rozsahy výběrů X a Y.

Náhodná veličina má Studentovo rozdělení pravděpodobnosti. Kritická oblast W je definována jako množina těch hodnot testovacího kritéria T, pro které platí: $W = \{T: |T| > KH\}$. KH je kritická hranice, která se vypočítá podle vztahu:

$$KH = \frac{\frac{S_1^2}{n_1 - 1} \times t_{\alpha, n_1 - 1} + \frac{S_2^2}{n_2 - 1} \times t_{\alpha, n_2 - 1}}{\frac{S_1^2}{n_1 - 1} + \frac{S_2^2}{n_2 - 1}}; \text{ kde } S_1^2 \text{ a } S_2^2 \text{ jsou výběrové rozptyly náhodných}$$

výběrů X a Y, \bar{X}, \bar{Y} jsou střední hodnoty náhodných výběrů a n_1, n_2 jsou rozsahy výběrů X a Y a t_α je kritická hodnota Studentova rozdělení pravděpodobností.

Příklad 9: Máme k dispozici údaje o hodinových výdělcích taxikářů ze dvou měst A a B. Předpokládejme, že hodnoty výdělků pocházejí ze Studentova rozdělení pravděpodobnosti. Hodnoty výdělků jsou uvedeny v tab. 11 a 12. Ověřte na hladině významnosti $\alpha = 5\%$ tvrzení, že dlouhodobý průměr výdělků taxikářů je v obou městech stejný. Dále víme, že $\sigma_1 \neq \sigma_2$.

68	133	144	106	154	175	141	148	75	50	130	151
199	134	183	137	127	101	157	119	112	115	88	168
142	103	135	115	195	133	105	82	78	143	85	85
179	124	113	97	80	84	135	99	116	133	118	200
145	165	123	155	131	98	148	44	125	82	110	111

Tab. 11: Hodinové výdělky taxikářů ve městě A.

148	127	174	132	125	139	158	140	108	146	125	154	167
132	128	127	111	134	111	118	105	150	109	112	114	133
115	81	132	112	148	162	124	159	198	134	134	157	108
158	73	137	154	138	168	151	136	117	104	141	171	148
145	151	140	113	147	146	105	141	128	167	152	131	

Tab. 12: Hodinové výdělky taxikářů ve městě B.

Řešení: Naším úkolem je ověřit, zda jsou výdělky taxikářů v obou městech přibližně stejné. Víme, že rozptyly obou výběrů se liší. Použijeme dvouvýběrový T- test s nerovností rozptylů. Testujeme hypotézu $H_0: EX=EY$ proti $H_1: EX \neq EY$.

Postup řešení 1: K výpočtu dvouvýběrového T- testu s nerovností rozptylů použijeme test z nabídky Nástroje → Analýza dat → Dvouvýběrový T- test s nerovností rozptylů.

Obr. 10: Tabulka pro vyplňování hodnot dvouvýběrového T-testu.

	A	B	C	D
1	Dvouvýběrový t-test s nerovností rozptylů			
2				
3		<i>Soubor 1</i>	<i>Soubor 2</i>	
4	Stř. hodnota	123,85	135,2031	
5	Rozptyl	1223,621	508,8946	
6	Pozorování	60	64	
7	Hyp. rozdíl stř. hodnot	0		
8	Rozdíl	100		
9	t stat	-2,13244		
10	P(T<=t) (1)	0,01771		
11	t krit (1)	1,660234		
12	P(T<=t) (2)	0,03542		
13	t krit (2)	1,983971		

Tab. 13: Výpočet dvouvýběrového T- testu s nerovností rozptylů.

Postup řešení 2: Druhou možností řešení tohoto příkladu je použití funkce TTEST. Zobrazí se nám stejná tabulka jako na obr. 10. Do kolonky Typ zadáme hodnotu 3, která odpovídá dvouvýběrovému T- testu s nerovností rozptylů.

Kritickou hodnotu nám v tomto případě vypočítá samotný T- test. Další možností jejího určení je její výpočet viz. kap. 3. 2. 1. nebo její hodnotu nalezneme v tabulkách [4].

Závěr: Z výpočtu vyplývá $t_{krit}(2) < |t_{stat}|$ tzn., že výsledek testu spadá do oblasti kritických hodnot. Hypotézu H_0 zamítáme. Průměrné mzdy taxikářů ve městech A a B se na hladině významnosti $\alpha = 0,05$ liší.

3.9 Dvouvýběrový párový t-test na střední hodnotu

Dvouvýběrový párový t-test na střední hodnotu je opět založen na Studentově rozdělení pravděpodobnosti, stejně jako testy předchozí. Tento test se používá, jestliže na každé z n vybraných statistických jednotek naměříme dva statistické znaky, přičemž měření jsou nezávislá. Neověřujeme tedy shodu spárovaných hodnot, ale spíše jejich rozdíl (např. klinické testy před a po léčbě pacienta by se měly lišit). To je v souladu s volbou hladiny významnosti (opět většinou 0,05), což má za následek velkou pravděpodobnost při potvrzení odlišnosti (95%).

Testované hypotézy zapíšeme ve tvaru $H_0: Ed=0$ s alternativní hypotézou $H_1: Ed \neq 0$, přičemž $D = X - Y$. Kritická hodnota se bere pro $n - 1$ stupňů volnosti, kde n je počet dvojic.

Zápis funkce: =TTEST(Soubor 1;Soubor 2;Strany;Typ)

Poznámky: Strany – určuje, zda se jedná o jednostranné (1) nebo oboustranné rozdělení (2).

Typ – určuje, o jaký typ T-testu se jedná, 1- párové hodnoty, 2- dva výběry se shodným rozptylem, 3- dva výběry s různým rozptylem. V našem případě zvolíme hodnotu 1. Pokud jsou oba výběry totožné nebo jsou odlišné, ale mají stejný rozptyl, vznikne chyba "dělení nulou".

Funkce je definována vztahem:

$$T = \frac{\bar{D} - k}{S_D} \times \sqrt{n - 1}; \text{ kde } D \text{ je rozdíl hodnot dvou souborů, } S_D \text{ je výběrové variační rozpětí}$$

dvojic a \bar{D} je aritmetický průměr rozdílů.

Příklad 10: Deset dobrovolníků se přihlásilo k testování nové diety. Hmotnosti pacientů před a po dietě jsou uvedeny v tabulce 14. Testujte na hladině významnosti $\alpha = 0,05$, zda má dieta vliv na snížení hmotnosti. Data v tabulce pocházejí z normálního rozdělení pravděpodobnosti.

číslo pacienta	1	2	3	4	5	6	7	8	9	10
váha před dietou	113,7	100	112,6	112,8	104,1	91,5	122,4	110,7	129,8	109,7
váha po dietě	101	89,5	105,5	95,1	106,4	85,3	101,5	111,1	115,8	107,1
úbytek váhy	12,7	10,5	7,1	17,7	-2,3	6,2	20,9	-0,4	14	2,6

Tab.14: Hmotnosti pacientů

Postup řešení 1: K výpočtu dvouvýběrového párového T- testu použijeme test z nabídky Nástroje → Analýza dat → Dvouvýběrový párový T- test.

	A	B	C	D
1	Dvouvýběrový párový t-test na střední hodnotu			
2				
3		<i>Soubor 1</i>	<i>Soubor 2</i>	
4	Stř. hodnota	110,73	101,83	
5	Rozptyl	116,2668	90,509	
6	Pozorování	10	10	
7	Pears. korelace	0,72266		
8	Hyp. rozdíl stř. hodnot	0		
9	Rozdíl	9		
10	t stat	3,679348		
11	P(T<=t) (1)	0,00254		
12	t krit (1)	1,833113		
13	P(T<=t) (2)	0,005081		
14	t krit (2)	2,262157		

Tab. 15: Výpočet dvouvýběrového párového T- testu.

Postup řešení 2: Druhou možností řešení tohoto příkladu je použití funkce TTEST. Zobrazí se nám stejná tabulka jako na obr. 10. Do kolonky Typ zadáme hodnotu 1, která odpovídá dvouvýběrovému párovému T- testu.

Kritickou hodnotu nám v tomto případě vypočítá samotný T- test. Další možnosti jejího určení je její výpočet viz. kap. 3. 2. 1. nebo její hodnotu nalezneme v tabulkách [4].

Závěr: Z výpočtu vyplývá **t krit(2) < |t stat|** tzn., že výsledek testu spadá do oblasti kritických hodnot. Hypotézu H_0 zamítáme. Dieta nemá na hladině významnosti $\alpha=0,05$ vliv na hmotnost pacientů.

4. Využití testů hypotéz

V předešlých kapitolách jsem popisoval parametrické statistické testy. Využití hypotéz je velice široké, neboť v dnešní době je statistika hojně využívána. Pro příklad praktického využití statistických hypotéz jsem si vybral hypotézu, zda má vliv místo bydliště na průměrnou mzdu. Nejprve budu porovnávat Královéhradecký a Pardubický kraj a poté Královéhradecký kraj a Prahu.

Příklad 11: Máme k dispozici údaje o průměrných mzdách obyvatel Královéhradeckého a Pardubického kraje z let 2005 - 2007. Předpokládejme, že průměrné mzdy pocházejí z normálního rozdělení pravděpodobnosti. Průměrné výdělky obyvatel Královéhradeckého kraje jsou uvedeny v tabulce 16 a Pardubického kraje v tabulce 17. Ověřte na hladině významnosti $\alpha=0,05$ tvrzení, že průměrné mzdy obyvatel obou krajů jsou přibližně stejné.

Královéhradecký kraj		
2005	2006	2007
14 249	14 962	16 197
19 375	18 102	18 579
16 294	17 399	18 822
21 808	23 610	24 103
21 246	22 639	23 188
16 005	17 106	18 525
13 657	14 508	15 940
12 990	13 744	14 812
12 238	13 356	14 948
12 346	12 274	14 332
16 205	15 679	18 038
15 540	17 171	17 230
15 742	16 365	17 776
19 377	20 252	21 641
16 677	18 812	19 948
17 859	19 255	20 482
17 015	18 228	19 044
19 922	20 204	22 611
13 583	13 828	14 564
22 141	24 572	26 371
16 054	16 744	18 094
14 107	14 987	15 371
9 540	9 332	10 583
16 826	17 631	18 259
27 058	29 758	25 962
15 493	16 612	18 145
21 480	22 423	23 050
16 256	17 307	18 164
13 018	14 355	15 552

Pardubický kraj		
2005	2006	2007
13 403	14 004	15 575
16 185	17 362	18 537
17 567	18 767	20 624
17 594	19 237	20 255
15 953	17 163	18 325
13 709	14 927	15 908
11 703	12 502	13 484
11 605	11 908	11 211
11 024	12 247	12 372
16 141	17 027	18 627
17 370	18 281	19 469
15 831	16 278	16 656
24 236	25 965	25 974
17 581	19 331	21 391
16 738	17 549	19 131
15 833	17 370	18 061
18 171	19 405	20 514
13 642	14 182	15 017
22 776	22 885	25 275
15 370	16 476	17 750
14 327	15 067	16 959
8 989	9 144	10 545
16 205	17 388	18 372
29 134	30 016	31 167
15 772	16 624	18 010
19 215	20 012	21 420
16 577	17 528	18 550
15 396	16 622	17 343
13 127	13 921	14 903

Tab.16: Průměrné mzdy v Královéhradeckém kraji.

Tab.17: Průměrné mzdy v Pardubickém kraji.

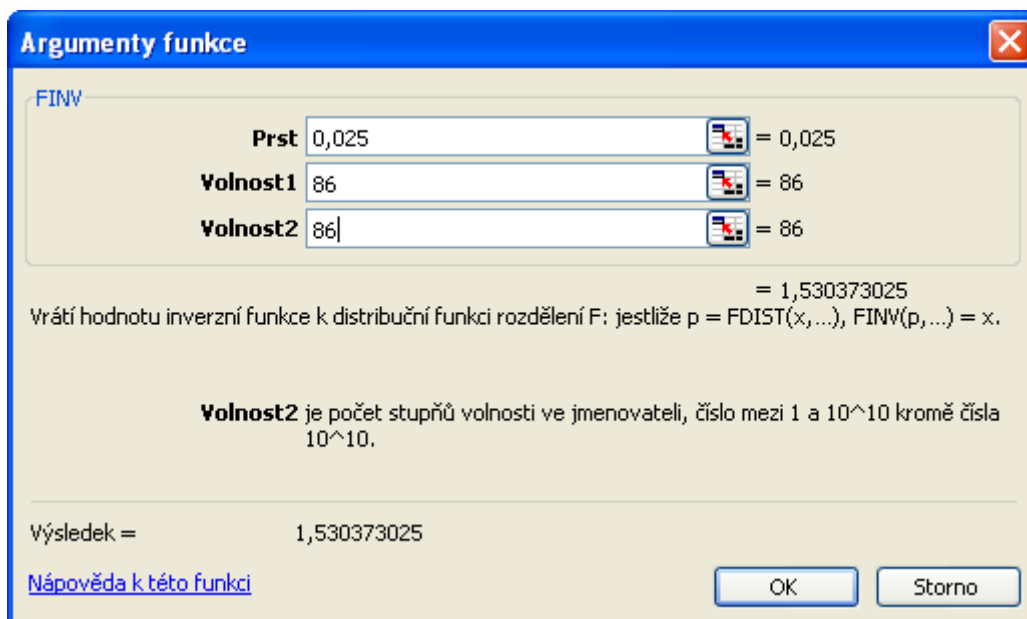
Řešení: Naším úkolem je ověřit, jestli jsou průměrné mzdy v obou krajích přibližně stejné. Budeme tedy testovat hypotézu o rovnosti středních hodnot. Nejprve však musíme provést F- test, abychom zjistili, zda jsou data homoskedatická či heteroskedastická. Testujeme hypotézu $H_0: EX=EY$ proti $H_1: EX \neq EY$.

Postup řešení: **1) Použijeme F- test.**

	A	B	C	D
1	Dvouvýběrový F-test pro rozptyl			
2				
3		<i>Soubor 1</i>	<i>Soubor 2</i>	
4	Stř. hodnota	17261,92	17651,11	
5	Rozptyl	17589883	15185420	
6	Pozorování	87	87	
7	Rozdíl	86	86	
8	F	1,15834		
9	P(F<=f) (1)	0,248482		
10	F krit (1)	1,530373		
11				

Tab.18: Dvouvýběrový F- test pro rozptyl.

2) Přepočítáme kritickou hodnotu F- testu.



Obr.11: Výpočet kritické hranice F- rozdělení.

3) Porovnáme výsledek F- testu s kritickou hranicí.

$$F = 0,861822, F_{krit} = 1,530373025$$

$F < F_{krit}$ data jsou homoskedastická → použijeme dvouvýběrový T- test s rovností rozptylů.

4) Vypočítáme dvouvýběrový T- test s rovností rozptylů.

	A	B	C	D
1	Dvouvýběrový t-test s rovností rozptylů			
2				
3		Soubor 1	Soubor 2	
4	Stř. hodnota	17614,77	17261,92	
5	Rozptyl	15159347	17589883	
6	Pozorování	87	87	
7	Společný rozptyl	16374615		
8	Hyp. rozdíl stř. hodnot	0		
9	Rozdíl	172		
10	t stat	0,575094		
11	P(T<=t) (1)	0,28299		
12	t krit (1)	1,653761		
13	P(T<=t) (2)	0,565979		
14	t krit (2)	1,973852		

Tab.19: Výpočet dvouvýběrového T- testu s rovností rozptylů.

5) Porovnáme výsledek T- testu s kritickou hranicí.

$$T \text{ stat} = 0,0575094, t \text{ krit} = 1,973852$$

Závěr: Z výpočtu vyplývá $|t \text{ stat}| < t_{krit}(2)$, výsledek testu nespadá do oblasti kritických hodnot → hypotézu H_0 nezamítáme.

Průměrné výdělky v Královéhradeckém a Pardubickém kraji jsou na hladině významnosti $\alpha = 0,05$ přibližně stejné.

Příklad 12: Máme k dispozici údaje o průměrných mzdách obyvatel Královéhradeckého kraje a Prahy z let 2005 - 2007. Předpokládejme, že průměrné mzdy pocházejí z normálního rozdělení pravděpodobnosti. Průměrné výdělky obyvatel Královéhradeckého kraje jsou uvedeny v tabulce 20 a Prahy v tabulce 21. Ověřte na hladině významnosti $\alpha = 0,05$ tvrzení, že průměrné mzdy obyvatel obou krajů jsou přibližně stejné.

Královéhradecký kraj		
2005	2006	2007
14 249	14 962	16 197
16 294	17 399	18 822
21 808	23 610	24 103
22 359	24 630	25 131
21 246	22 639	23 188
16 005	17 106	18 525
13 657	14 508	15 940
12 990	13 744	14 812
12 238	13 356	14 948
12 346	12 274	14 332
16 205	15 679	18 038
15 540	17 171	17 230
15 742	16 365	17 776
19 377	20 252	21 641
16 677	18 812	19 948
17 859	19 255	20 482
17 015	18 228	19 044
19 922	20 204	22 611
13 583	13 828	14 564
22 141	24 572	26 371
16 054	16 744	18 094
14 107	14 987	15 371
9 540	9 332	10 583
16 826	17 631	18 259
27 058	29 758	25 962
15 493	16 612	18 145
21 480	22 423	23 050
16 256	17 307	18 164
16 596	17 762	18 536
13 018	14 355	15 552

Tab. 20: Průměrné mzdy v Královéhradeckém kraji.

Hlavní město Praha		
2005	2006	2007
14 989	14 660	14 916
23 857	25 674	27 194
28 213	17 008	21 715
18 447	5 087	8 849
28 364	29 318	31 407
22 732	24 550	25 932
25 315	27 438	27 784
11 303	12 440	14 177
13 175	14 092	15 331
13 358	14 295	14 262
25 235	25 892	27 309
30 689	33 073	35 022
18 990	19 781	22 180
25 017	28 790	29 963
17 679	18 419	20 126
21 497	23 443	25 250
26 770	29 129	29 421
22 850	24 508	26 917
14 497	16 333	17 271
31 834	36 422	41 013
20 975	23 062	24 075
23 271	26 174	26 874
12 910	14 291	14 460
28 524	28 388	29 102
45 152	48 057	51 055
25 526	27 388	29 474
25 422	27 054	29 409
18 598	19 429	21 527
19 178	21 106	21 925
19 287	20 393	21 560

Tab. 21: Průměrné mzdy v Praze.

Řešení: Naším úkolem je ověřit, jestli jsou průměrné mzdy v obou krajích přibližně stejné. Budeme tedy testovat hypotézu o rovnosti středních hodnot. Nejprve však musíme provést F- test, abychom zjistili, zda jsou data homoskedastická či heteroskedastická. Testujeme hypotézu $H_0: EX=EY$ proti $H_1: EX \neq EY$.

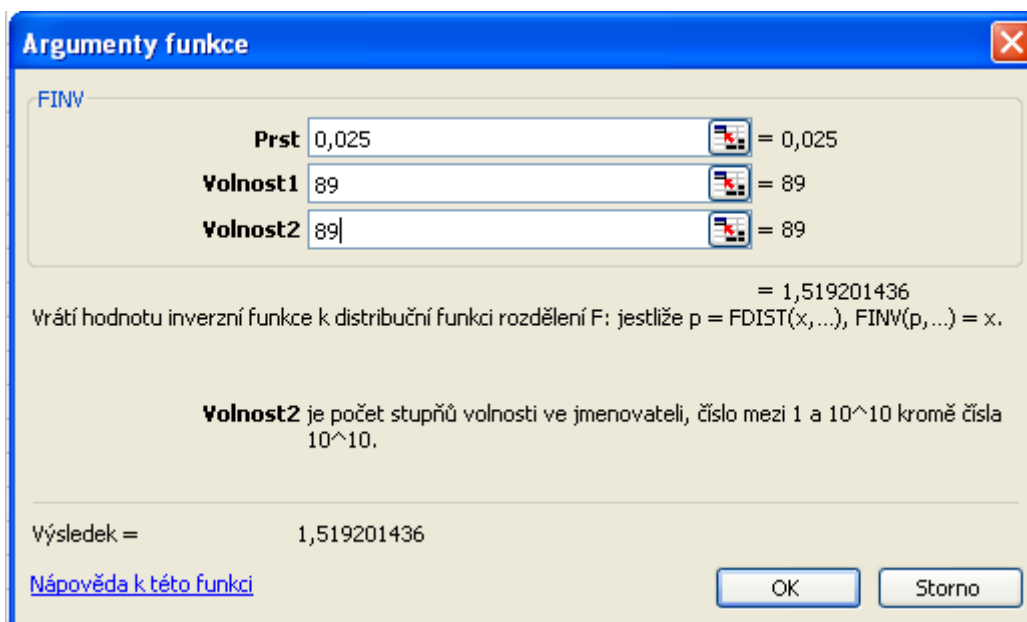
Postup řešení:

1) Použijeme F- test.

	A	B	C	D
1	Dvouvýběrový F-test pro rozptyl			
2				
3		<i>Soubor 1</i>	<i>Soubor 2</i>	
4	Stř. hodnota	23498,31	17828,94	
5	Rozptyl	63618164	16042591	
6	Pozorování	90	90	
7	Rozdíl	89	89	
8	F	3,965579		
9	P(F<=f) (1)	2,07E-10		
10	F krit (1)	1,519201		
11				
12				

Tab. 22: Výsledek dvouvýběrového F- testu pro rozptyl.

2) Přepočítáme kritickou hodnotu F- testu.



Obr. 12: Výpočet kritické hranice F- rozdělení.

3) Porovnáme výsledek F- testu s kritickou hranicí.

$$F = 3,965579, F_{\text{krit}} = 1,519201436$$

F > **F_{krit}** data jsou heteroskedastická → použijeme dvouvýběrový T- test s nerovností rozptylů.

4) Vypočítáme dvouvýběrový T- test s nerovností rozptylů.

	A	B	C	D
1	Dvouvýběrový t-test s nerovností rozptylů			
2				
3		<i>Soubor 1</i>	<i>Soubor 2</i>	
4	Stř. hodnota	23498,31	17828,94	
5	Rozptyl	63618164	16042591	
6	Pozorování	90	90	
7	Hyp. rozdíl stř. hodnot	0		
8	Rozdíl	131		
9	t stat	6,026069		
10	P(T<=t) (1)	7,98E-09		
11	t krit (1)	1,656569		
12	P(T<=t) (2)	1,6E-08		
13	t krit (2)	1,978239		

Tab. 23: Výpočet dvouvýběrového T- testu s nerovností rozptylů.

5) Porovnáme výsledek T- testu s kritickou hranicí.

$$T \text{ stat} = 6,026069, t \text{ krit} = 1,978239$$

Závěr: Z výpočtu vyplývá $|t \text{ stat}| > t \text{ krit}(2)$, výsledek testu spadá do oblasti kritických hodnot \rightarrow hypotézu H_0 zamítáme.

Průměrné výdělky v Královéhradeckém kraji a Hlavním městě Praze se na hladině významnosti $\alpha = 0,05$ liší.

5. Závěr

V bakalářské práci se věnuji programu MS EXCEL a jeho využití při řešení statistických hypotéz. Program MS EXCEL je pro svoji dostupnost velmi vhodný pro řešení statistických úloh. V práci se snažím detailně popsat postup pro řešení jednotlivých statistických parametrických testů. Na začátku práce popisuji nejdůležitější matematické funkce, které jsem uplatnil při řešení statistických hypotéz. V práci jsou dále definovány a vysvětleny nejdůležitější statistické pojmy, se kterými se v práci setkáme. V třetí části práce se věnuji popisu jednotlivých testů. U každého testu je uveden i příklad jeho využití. Postup při výpočtu příkladu je detailně vysvětlen a nechybí ani obrázky postupů. V závěru práce se věnuji využití statistických testů v praxi. Konkrétně srovnávám průměrné mzdy obyvatel České republiky a zjišťuji, zda jsou průměrné mzdy v jednotlivých krajích České republiky přibližně stejné.

6. Použité zdroje

- [1] BROŽ, Milan. Mistrovství v Microsoft EXCEL 2000 a 2002. 1. vyd. Praha : Computer Press, 2002. 648 s. ISBN 80-7226-809-0.
- [2] GERBHARDT, Reiner. EXCEL 97 kompletní kapesní průvodce. Ing. Radka Halodová. 1. vyd. Praha : Grada Publishing, 1998. 448 s. ISBN 80-7169-592-0.
- [3] KUBANOVÁ, Jana. Statistické metody pro ekonomickou a statistickou praxi. 2. vyd. Bratislava : Statis, 2004. 254 s. ISBN 80-85659-37-9.
- [4] KUBANOVÁ, Jana, LINDA, Bohdan. Kritické hodnoty a kvantily vybraných rozdělení pravděpodobností. 1. dotisk vyd. Pardubice : Tiskařské středisko University Pardubice, 2007. 53 s. ISBN 80-7194-852-755-7.
- [5] ŠŤASTNÝ, Zdeněk. Matematické a statistické výpočty v Microsoft EXCELu. 1. vyd. Brno : Computer press, 1999. 254 s. ISBN 80-7226-141-X.
- [6] *Český statistický úřad* [online]. 2010 [cit. 2010-04-13]. Dostupné z WWW: <<http://www.czso.cz>>.
- [7] *Microsoft Office online* [online]. 2010 [cit. 2010-04-13]. Dostupné z WWW: <<http://www.office.microsoft.com>>.