

Statistické funkce

Microsoft Excel 2007 nabízí celou řadu nástrojů, jejichž prostřednictvím lze analyzovat statistická data. V programu je vestavěno mnoho funkcí, jež pomáhají při jednoduchých analytických úkolech, například PRŮMĚR, MEDIAN nebo MODE. Nestačí-li vám vestavěné statistické funkce, je k dispozici doplněk Analytické nástroje, jež poskytuje sadu nástrojů rozšiřujících vestavěné analytické schopnosti Office Excelu 2007. Pomocí těchto nástrojů lze vytvářet histogramy, pořadové statistiky a percentily, extrahovat vzorky z datových sad, provádět analýzy regrese, generovat speciální sady náhodných čísel, uplatňovat na svá data Fourierovu transformaci a další. V této kapitole se budeme věnovat většině důležitých statistických funkcí, jež jsou k dispozici v Excelu i v doplňku Analytické nástroje.

Analýza rozložení dat

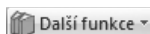
Ve statistice se kolekce hodnot označuje jako *rozložení*. Excel nabízí několik metod pro analýzu rozložení: vestavěné statistické funkce nebo funkce pořadových statistik a percentil spolu s nástrojem Pořadová statistika a percentily.

POZNÁMKA

Rozložení lze analyzovat také pomocí nástrojů Popisná statistika a Histogram, jež jsou součástí doplňku Analytické nástroje. Více informací naleznete v části „Práce s doplňkem Analytické nástroje“ na straně 551.

Práce s vestavěnými statistickými funkcemi

Vestavěné statistické funkce slouží k analýze skupiny (či populace) hodnot. V následujících částech se zaměříme na nejčastěji používané statistické funkce. Rychlý přístup k těmto funkcím získáte klepnutím na tlačítko Další funkce na kartě Vzorce a následným výběrem položky Statistická. Zobrazí se nabídka statistických funkcí.



Témata kapitoly:

- Analýza rozložení dat
- Lineární a exponenciální regrese
- Práce s doplňkem Analytické nástroje

POZNÁMKA

Excel nabízí také pokročilé statistické funkce LINREGRESE, LOGLINREGRESE, LINTREND a LOGLINTREND, jež pracují s poli. Více informací naleznete v části „Lineární a exponenciální regrese“ na straně 545.

Funkce PRŮMĚR

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; ...*), přičemž může obsahovat až 255 argumentů a ignoruje prázdné buňky a buňky obsahující logické či textové hodnoty. Chcete-li například vypočítat průměr hodnot v buňkách B4 až B15, mohli byste použít vzorec $=\text{(B4+B5+B6+B7+B8+B9+B10+B11+B12+B13+B14+B15)}/12$, podstatně jednodušší je však zadat vzorec $=\text{PRŮMĚR(B4:B15)}$.

Funkce MEDIAN, MODE, MAX, MIN a POČET

Do všech těchto funkcí zadáváte stejné argumenty, v podstatě pouze oblast nebo seznam čísel oddělených středníky, například (*číslo1; číslo2; ...*). Zadát lze až 255 argumentů, přičemž funkce ignorují text, chybové hodnoty a logické hodnoty. Následuje stručný popis jednotlivých funkcí:

- **MEDIAN:** Vypočítá medián zadaných čísel. Medián je číslo, které leží uprostřed množiny čísel. Je-li počet zadaných hodnot lichý, je vrácená hodnota průměrem dvou čísel, jež leží uprostřed množiny.
- **MODE:** Určí, jaká hodnota se nejčastěji vyskytuje v množině čísel. Pokud se žádná z čísel nevyskytuje více než jednou, vrátí funkce MODE chybovou hodnotu #N/A.
- **MAX:** Vrátí maximální hodnotu v oblasti.
- **MIN:** Vrátí minimální hodnotu v oblasti.
- **POČET:** Vrátí počet buněk obsahujících čísla v určené oblasti, včetně dat a vzorců, jejichž výsledkem jsou čísla.

POZNÁMKA

Chcete-li spočítat všechny neprázdné buňky bez ohledu na jejich obsah, můžete použít funkci POČET2. Další informace o této funkci naleznete v části „Funkce A a 2“ na straně 542.

Funkce SUMIF, SUMIFS a COUNTIF

Funkce SUMIF se podobá funkci SUMA, ale před sečtením buněk testuje každou buňku pomocí určené podmínky. Do této funkce se zadávají argumenty (*oblast; kritéria; oblast_součtu*). Argument *oblast* určuje oblast, kterou chcete testovat, argument *kritéria* určuje podmínku a argument *oblast_součtu* definuje buňky, jež mají být sečteny. Pracujete-li například s listem, jež obsahuje sloupec s názvy měsíců a názvem oblasti Měsíce, a sousední sloupec s čísly má název Tržby, vrátí vzorec $=\text{SUMIF(Měsíce; „Červen“; Tržby)}$ hodnotu v buňce Tržby, jež sousedí s popiskem Červen. Pomocí vzorce s podmínkou, například $=\text{SUMIF(Tržby; „>=999“; Tržby)}$, lze získat součet všech tržeb, jež jsou vyšší než 999 Kč.

Funkce SUMIFS pracuje podobně jako funkce SUMIF, avšak lze určit až 127 různých oblastí, jež mají být sečteny, z nichž každá může mít vlastní kritéria. Argument *oblast_součtu* se v této

funkci zadává na prvním místě: (*oblast_součtu; oblast_kritéria1; kritérium1; oblast_kritéria2; kritérium2;...*). Oblast součtu a každá oblast kritéria musí mít stejnou velikost a tvar. Použijeme podobný příklad jako u funkce SUMIF a budeme navíc předpokládat, že jsme také vytvořili definované názvy oblastí Měsíce, Celkem a Produkt1, Produkt2 atd. Vzorec =SUMIFS(**Celkem; Produkt3; „<=2700“; Měsíce; „Červen“**) vrátí celkové tržby za měsíc červen, pokud tržby za Produkt2 byly nižší nebo rovny 2 700 Kč.

ODKAZ

Vytváření vzorců si můžete usnadnit pomocí Průvodce podmíněným součtem. Více informací naleznete v části „Průvodce podmíněným součtem a Průvodce vyhledáváním“ na straně 474.

Podobně spočítá funkce COUNTIF buňky, jež odpovídají určenému kritériu. Do této funkce zadáváte argumenty (*oblast; kritérium*). S využitím stejného příkladu lze pomocí podmínky zjistit počet měsíců, v nichž se celkové tržby propadly pod 1 200 Kč, jako například ve vzorci =COUNTIF(**Tržby; „<1200“**).

ODKAZ

Další informace o podmínkách naleznete v části „Podmíněné testování“ na straně 473. Více informací o práci s názvy oblastí naleznete v části „Názvy buněk a oblastí buněk“ na straně 440.

Práce s funkcemi pro analýzu pořadových statistik a percentil

Excel nabízí několik funkcí, jež extrahují informace o pořadových statistikách a percentilech z množiny vstupních hodnot: PERCENTRANK, PERCENTIL, QUARTIL, SMALL, LARGE a RANK.

Funkce PERCENTRANK

Funkce PERCENTRANK vrátí pořadí hodnoty v množině dat vyjádřené procentuální částí množiny dat. Pomocí této funkce lze vytvářet procentuální tabulky propojené se vstupní oblastí, takže procenta jsou aktualizována v případě, že se změní vstupní hodnoty. Tuto funkci jsme využili k vytvoření pořadí ve sloupci E na obrázku 17.1.

ODKAZ

Soubor Výsledky SAT.xlsx naleznete v sekci Zkušební soubory na příloženém CD.

Do funkce PERCENTRANK zadáváte argumenty (*pole; x; desetiny*). Argument *pole* určuje vstupní oblast (v našem příkladu D2:D1001) a *x* určuje hodnotu, jejíž pořadí chcete získat. Argument *desetiny* je volitelný a určuje požadovanou přesnost čísla; pokud tento argument vynecháte, výsledek je zaokrouhlen na tři číslice (0,xxx nebo xx,x%).

V Excelu je k dispozici sada funkcí, díky nimž máte větší flexibilitu při provádění výpočtů v množinách dat obsahujících textové či logické hodnoty. K těmto funkcím patří PRŮMĚR2, COUNTA, MAXA, MINA, STDEVA, STDEVPA, VARA a VARPA, do nichž lze zadat až 255 argumentů (*hodnota1; hodnota2; ...*).

E2		fx =PERCENTRANK(D2:D1001;D2;4)								
	A	B	C	D	E	F	G	H	I	
1	ID studenta	Ústní	Matematika	Celkem	Percentil					
2	722-4499	418	518	936	3%					
3	605-3475	465	557	1022	34%					
4	546-3500	463	549	1012	27%					
5	655-8550	466	587	1053	55%					
6	812-6448	520	544	1064	62%					
7	814-6503	470	537	1007	24%					
8	332-3453	533	549	1082	73%					
9	55-0476	476	570	1046	49%					
10	745-0539	468	548	1016	30%					
11	675-1544	441	562	1003	21%					
12	836-6513	570	560	1130	94%					
13	146-4509	503	554	1057	57%					
14	45-9450	551	556	1107	86%					
15	924-3538	498	562	1060	59%					
16	664-9559	466	525	991	17%					
17	596-5425	414	549	963	8%					
18	914-1483	497	545	1042	46%					

Obrázek 17.1: Funkce PERCENTRANK propojí procentuální hodnoty se vstupními hodnotami.

FUNKCE A A 2

Verze těchto funkcí bez A nebo 2 ignorují buňky obsahující textové hodnoty. Obsahuje-li například oblast deseti buněk jednu textovou hodnotu, funkce PRŮMĚR tuto buňku ignoruje a vydělí součet 9 pro získání průměru, zatímco funkce PRŮMĚR2 považuje textovou hodnotu za součást oblasti a vydělí součet 10. To je užitečné v případech, že do svých výpočtů chcete zahrnout vždy všechny odkazované buňky, zvláště pracujete-li se vzorci, jež vrací po splnění určité podmínky textové příznaky. Více informací o funkcích STDEVA, STDEVPA, VARA a VARPA naleznete v části „Statistické funkce pracující se vzorky a populací“ na straně 543.

Funkce PERCENTIL A QUARTIL

Pomocí funkce PERCENTIL lze určit prahovou hodnotu. Do této funkce zadáváte argumenty (*pole*; *k*). Percentil *k* je nutné vyjádřit v rámci uzavřeného intervalu od 0 do 1. Chcete-li například zjistit, jaký výsledek na obrázku 17.1 představuje 87. percentil, můžete zadat vzorec =PERCENTIL(D2:D1001; 0,87).

Funkce QUARTIL, do níž zadáváte argumenty (*pole*; *kvartil*), funguje podobně jako funkce PERCENTIL, avšak vrací hodnotu, jež představuje nejnižší percentil nebo jakýkoli čtvrtinový kvartil ve vstupní množině. Argument *pole* určuje vstupní oblast. Argument *kvartil* určuje hodnotu, jež má být vrácena (viz tabulka 17.1).

Tabulka 17.1: Argument kvartil

Kvartil	Vrátí
0	Nejnižší hodnotu
1	25 procent
2	Hodnota medián (50 procent)
3	75 procent
4	Nejvyšší hodnota

DO DETAILU**VYUŽITÍ FUNKCÍ MIN, MEDIAN A MAX**

QUARTIL je výkonná funkce, pokud však nepotřebujete vrácení hodnot 25 nebo 75 procent, získáte rychlejší výsledky pomocí jiných funkcí, zvláště při práci s velkými množinami dat. Funkci MIN použijte místo QUARTIL(*pole*; 0), funkci MEDIAN místo QUARTIL(*pole*; 2) a funkci MAX místo QUARTIL(*pole*; 4).

Funkce SMALL a LARGE

Funkce SMALL a LARGE vrací k-tou nejmenší a k-tou největší hodnotu v množině dat; do obou funkcí se zadávají argumenty (*pole*; *k*), kde *k* je pořadí od nejmenšího nebo největšího čísla. Pokud byste například chtěli najít 15. nejvyšší výsledek na obrázku 17.1, můžete zadat vzorec =LARGE(D2:D1001; 15).

Funkce RANK

Funkce RANK vrátí pořadí argumentu (podle velikosti) v množině čísel. Do této funkce zadáváte argumenty (*číslo*; *odkaz*; *pořadí*). Argument *číslo* představuje číslo, jehož pořadí chcete nalézt, *odkaz* je oblast obsahující množinu dat a *pořadí* je číslo určující, zda se budou hodnoty třídit vzestupně či sestupně (výchozí). Chcete-li například zjistit pořadí výsledku ---1200 v datové sadě na obrázku 17.1, zadáte vzorec =RANK(1200; D2:D1001).

Při výchozím nastavení má nejvyšší hodnota pořadí 1, druhá nejvyšší 2 atd. Nemůže-li funkce RANK nalézt přesnou shodu mezi prvním argumentem a vstupní hodnotou, vrátí chybovou hodnotu #N/A.

Statistické funkce pracující se vzorky a populací

Rozptyl a směrodatná odchylka jsou statistická měřítka rozptylu skupiny či populace čísel. Směrodatná odchylka je druhá odmocnina rozptylu. Zpravidla spadá přibližně 68 procent normálně rozptýlené populace do jedné směrodatné odchylky od střední hodnoty a zhruba 95 procent spadá do dvou směrodatných odchylek. Velká směrodatná odchylka určuje, že populace je široce rozptýlená od střední hodnoty; malá směrodatná odchylka určuje, že populace je těsně shluknutá kolem střední hodnoty.

Čtyři statistické funkce – VAR, VARP, STDEV a STDEVP – počítají rozptyl a směrodatnou odchylku čísel v oblasti buněk. Než spočítáte rozptyl a směrodatnou odchylku skupiny hodnot, je nutné určit, zda tyto hodnoty představují celkovou populaci nebo pouze reprezentativní

vzorek dané populace. Funkce VAR a STDEV předpokládají, že hodnoty představují pouze vzorek celkové populace; funkce VARP a STDEVP předpokládají, že hodnoty představují celkovou populaci.

Jednoduché statistické výpočty: VAR a STDEV

Funkce VAR a STDEV počítají rozptyl a směrodatnou odchylku za předpokladu, že jejich argumenty představují pouze vzorek z celkové populace. Do těchto funkcí lze zadávat argumenty (*číslo1*; *číslo2*; ...) a je možné uvést až 255 argumentů. List na obrázku 17.2 ukazuje výsledky zkoušek pěti studentů a předpokládá, že výsledky v buňkách B4:E8 představují pouze část z celkové populace.

I4 f _x =VAR.VÝBĚR(B4:E8)										
	A	B	C	D	E	F	G	H	I	J
1	Výsledky zkoušek za 1. čtvrtletí									
2										
3	Student	Zkouška 1	Zkouška 2	Zkouška 3	Zkouška 4	Průměr		Celkový průměr	89,20	
4	Carothers, Andy	87	90	79	96	88,00		Rozdíl	30,80	
5	Groncki, Douglas	92	94	94	97	94,25		Standardní odchylka	5,55	
6	MacDonald, Scott	96	95	95	80	91,50				
7	Nusbaum, Tawana	85	87	87	88	86,75				
8	Rothenberg, Eric	81	88	88	85	85,50				
9										
10										

Obrázek 17.2: Funkce VAR a STDEV měří rozptyl vzorku výsledků zkoušek.

ODKAZ

Soubor VAR.xlsx naleznete v oblasti Zkušební soubory na přiloženém CD.

Buňka I4 používá vzorec =VAR(B4:E8) k výpočtu rozptylu pro tento vzorek výsledků zkoušek. Buňka I5 používá vzorec =STDEV(B4:E8) k výpočtu standardní odchylky.

Předpokládáme-li, že výsledky zkoušek v tomto příkladu jsou normálně rozptýlené, lze z toho vyvodit, že zhruba 68 procent studentů (obecné pravidlo) dosáhlo výsledků od 83,65 (průměr 89,20 minus směrodatná odchylka 5,55) do 94,75 (89,20 plus 5,55).

Statistické výpočty celkové populace: VARP a STDEVP

Pokud analyzovaná čísla představují celou populaci, a ne jen její vzorek, použijte k výpočtu rozptylu a směrodatné odchylky funkce VARP a STDEVP. Do těchto funkcí zadáváte argumenty (*číslo1*; *číslo2*; ...) a lze uvést až 255 argumentů.

Předpokládáme-li, že buňky B4:E8 v listu na obrázku 17.2 představují celkovou populaci, lze rozptyl a směrodatnou odchylku spočítat pomocí vzorců =VARP(B4:E8) a =STDEVP(B4:E8). Funkce VARP vrátí hodnotu 29,26 a funkce STDEVP vrátí hodnotu 5,41.

POZNÁMKA

Funkce STDEV, STDEVP, VAR a VARP nezahrnují do svých výpočtů textové hodnoty ani prázdné buňky. Chcete-li do výpočtů zahrnout i textové hodnoty a prázdné buňky, použijte verze s A na konci (STDEVA, STDEVPA, VARA a VARP). Více informací naleznete v části „Funkce A a 2“ na straně 542.

Lineární a exponenciální regrese

Excel nabízí několik maticových funkcí, jež provádějí lineární regresi (LINREGRESE, LINTREND, FORECAST, SLOPE a STEYX) a exponenciální regresi (LOGLINREGRESE a LOGLINTREND). Tyto funkce zadáváte jako maticové vzorce, jež mají maticové výsledky. V každé z těchto funkcí lze zadat jednu či více nezávislých proměnných. V následujícím přehledu popisujeme různé typy regrese:

- **Lineární regrese:** Vráti parametry lineárního trendu. Například na základě údajů o prodeji za jeden rok vám lineární regrese může pomoci sklonu přímky a průsečíku s osou y sdělit odhadované prodeje v březnu příštího roku. Sledováním přímky lze odhadnout budoucí prodeje, můžete-li bezpečně předpokládat, že růst zůstane lineární.
- **Exponenciální regrese:** Vypočítá exponenciální křivku, kterou prokládá zadaná data. Například řada měření růstu populace bude téměř vždy lépe zastoupena exponenciální křivkou než přímkou.
- **Vícenásobná regrese:** je analýza více než jedné množiny dat, jež často nabízí reálnější odhad. Provádět lze analýzy lineární i exponenciální vícenásobné regrese. Předpokládejme například, že chcete odhadnout přibližnou cenu za dům ve vašem regionu na základě čtverečních metrů, počtu koupelen, velikosti parcely a stáří. Pomocí vzorce vícenásobné regrese lze odhadnout cenu založenou na informacích získaných podle existujících domů.

REGRESE DO BUDOUCNOSTI?

Princip regrese může znít podivně, protože tento termín je obvykle spojován s minulostí, zatímco ve světě statistiky je regrese často používána k předpovídání budoucnosti. Řečeno jednoduše, **regrese** je statistická metoda, jež hledá matematický výraz popisující množinu hodnot.

Podniky se často snaží předpovídat budoucí prodeje a odhadovat procentuální výši prodejů na základě historie. Jednoduchá metoda procentuální výše prodejů určuje aktiva a pasiva, jež se mění spolu s prodeji, stanovuje vzájemný poměr a přiřazuje jim procentuální hodnoty. Ačkoli jsou takto založené předpovědi často dostačující pro pomalý a stálý krátkodobý růst, metoda ztrácí při rychlejším růstu přesnost.

Regresní analýzy používají složitější rovnice k analýze větších množin dat a převádějí je na souřadnice na přímce nebo křivce. Není tomu tak dávno, kdy se regresní analýza příliš nepoužívala z důvodu velkého počtu nutných kalkulací. S příchodem tabulkových procesorů, jako je Excel, jež nabízely vestavěné regresní funkce, se využívání regresních analýz mnohem více rozšířilo.

Výpočty lineární regrese

Rovnice $y = mx + b$ algebraicky popisuje přímku pro množinu dat s jednou nezávislou proměnnou, kde x je nezávislá proměnná, m představuje sklon přímky a b zastupuje průsečík s osou y . Pokud přímka představuje počet nezávislých proměnných v analýze vícenásobné lineární regrese s očekávaným výsledkem, rovnice regresní přímky bude mít následující formát:

$$y = m_1x_1 + m_2x_2 + \dots + m_nx_n + b$$

v níž y je závislá proměnná, x_1 až x_n je n nezávislých proměnných, m_1 až m_n představují koeficienty každé nezávislé proměnné a b je konstanta.

Funkce LINREGRESE

Funkce LINREGRESE používá tento obecnější vzorec k vrácení hodnot m_1 až m_n a hodnotu b , při známé množině hodnot pro y a známé množině hodnot pro každou nezávislou proměnnou. Tato funkce má formát =LINREGRESE(**pole_y**; **pole_x**; **b**; **stat**).

Argument *pole_y* je množina hodnot y , které již znáte. Tímto argumentem může být jeden sloupec, jeden řádek nebo obdélníková oblast buněk. Je-li argumentem *pole_y* jeden sloupec, je každý sloupec v argumentu *pole_x* považován za nezávislou proměnnou. Podobně, je-li argumentem *pole_y* jeden řádek, je každý řádek v argumentu *pole_x* považován za nezávislou proměnnou. Pokud je argumentem *pole_y* obdélníková oblast buněk, lze použít pouze jednu nezávislou proměnnou; argument *pole_x* by v tomto případě měl představovat obdélník se stejnou velikostí a tvarem, jako argument *pole_y*. Vynecháte-li argument *pole_x*, použije Excel řadu 1, 2, 3, 4 atd.

Argumenty b a *stat* jsou volitelné. Uvedete-li jeden z nich, musí se jednat o logickou konstantu – buď PRAVDA, nebo NEPRAVDA. (Místo hodnoty PRAVDA lze uvést 1 a místo hodnoty NEPRAVDA je možné zadat 0.) Výchozí nastavení pro b a *stat* je PRAVDA a NEPRAVDA. Nastavíte-li argument b na NEPRAVDA, Excel zajistí, aby b (jako poslední část ve vzorci) byla 0. Pokud nastavíte argument *stat* na PRAVDA, obsahuje matice vrácená funkcí LINREGRESE následující validační statistiky:

se_1 až se_n	Standardní chybové hodnoty pro každý koeficient
se_b	Standardní chybová hodnota pro konstantu b
r^2	Koeficient determinace
se_y	Standardní chybová hodnota pro y
F	F-statistika
D_f	Stupně volnosti
ss_{reg}	Regresní součet čtverců
ss_{resid}	Reziduální součet čtverců

Než vytvoříte vzorec pomocí funkce LINREGRESE, je nutné vybrat dostatečně velkou oblast pro výslednou matici. Vynecháte-li argument *stat* (nebo jej výlučně nastavíte na hodnotu NEPRAVDA), bude výsledná matice zahrnovat jednu buňku pro každou z vašich nezávislých proměnných a jednu buňku pro b . Pokud zahrnete validační statistiky, výsledná matice bude vypadat jako následující příklad. Po výběru oblasti, jež má obsahovat výslednou matici, zadejte funkci a následně stiskněte kombinaci kláves Ctrl+Shift+Enter pro zadání funkce do každé buňky výsledné matice.

m_n	m_{n-1}	...	m_2	m_1	b
se_n	se_{n-1}	...	se_2	se_1	se_b
r^2	se_y				
F	D_f				
ss_{reg}	ss_{resid}				

Ať již použijete validační statistiky nebo ne, koeficienty a standardní chybové hodnoty pro nezávislé proměnné jsou vráceny v opačném pořadí oproti vašim vstupním datům. Jsou-li například čtyři nezávislé proměnné uspořádány do čtyř sloupců, vyhodnotí funkce LINREGRESE sloupec zcela vlevo jako x_1 , ale vrátí m_1 ve čtvrtém sloupci výsledné matice.

Obrázek 17.3 ukazuje jednoduchý příklad využití funkce LINREGRESE s jednou nezávislou proměnnou. Položky ve sloupci B v tomto listu představují měsíční odběr produktů pro malý podnik. Čísla ve sloupci A představují měsíce v období. Předpokládejme, že chcete vypočítat sklon a průsečík osy y regresní přímky nejlépe popisující relaci mezi odběrem produktů a měsíci. Jinými slovy, chcete popsat vývojovou tendenci dat. Vyberte oblast F6:G6, zadejte vzorec =LINREGRESE(B2:B19; A2:A19) a stiskněte kombinaci kláves Ctrl+Shift+Enter. Výsledné číslo v buňce F6 je 20,613, což je sklon regresní přímky; číslo v buňce G6 je 4002,065, což je průsečík osy y.

F6		fx {=LINREGRESE(B2:B19;A2:A19)}									
	A	B	C	D	E	F	G	H	I	J	K
1	Měsíc	Poptávka, tisíce	Trend								
2	1	4039	4022,678								
3	2	4057	4043,291								
4	3	4052	4063,904								
5	4	4094	4084,517								
6	5	4104	4105,13			Lineární odhad					
7	6	4110	4125,743			20,613003	4002,065				
8	7	4154	4146,356								
9	8	4161	4166,969								
10	9	4186	4187,582								
11	10	4195	4208,195								
12	11	4229	4228,808								
13	12	4244	4249,421								
14	13	4242	4270,034								
15	14	4283	4290,647								
16	15	4322	4311,26								
17	16	4333	4331,873								
18	17	4368	4352,486								
19	18	4389	4373,099								
20											

Obrázek 17.3: Funkce LINREGRESE vypočítá sklon a průsečík osy y regresní přímky.

ODKAZ

Soubor Analýza.xlsx naleznete v oblasti Zkušební soubory na příloženém CD.

Funkce LINREGRESE a LOGLINREGRESE vrací pouze souřadnice osy y použité pro výpočet přímek a křivek. Rozdíl mezi nimi spočívá v tom, že funkce LINREGRESE ukazuje přímku a funkce LOGLINREGRESE ukazuje exponenciální křivku. Proto je nutné pozorně zvolit správnou funkci pro požadovanou analýzu. Funkce LINREGRESE může být vhodnější k promítání prodejů a funkci LOGLINREGRESE spíše použijete pro statistické analýzy či populační trendy. Více informací naleznete v části „Funkce LOGLINREGRESE“ na straně 551.

DO DETAILU

VYUŽITÍ REGRESE PŘI VÝPOČTU CENY NEMOVITOSTÍ

Jedním z často používaných regresních modelů je Analýza konkurenčního trhu. Obchodníci s nemovitostmi využívají tento model k zjištění odhadované prodejní ceny za nemovitost, jež je založena na archivních údajích o prodeji porovnatelných domů v dané oblasti. Na následujícím obrázku vidíte příklad takového nástroje vytvořeného v Excelu:

Home Price Estimator For sample use only - results not guaranteed.

What You Want

Square Footage	2400
Number of Bathrooms	3
Number of Bedrooms	2
Age of Home	15
Estimated Price	\$230 209

To use this worksheet, first enter data for the house you want to find (in the blue cells), then enter data for selected homes in the same neighborhood that have recently been listed or sold.

Input Data

	Home 1	Home 2	Home 3	Home 4	Home 5
Square Footage	2500	3200	2200	3450	2675
Number of Bathrooms	2,5	3	2,75	3,75	2,5
Number of Bedrooms	4	3	4	6	4
Age of Home	25	7	40	22	68
Price	\$285 000	\$325 500	\$249 500	\$379 500	\$305 000

Tato aplikace používá funkci LINREGRESE k analýze statistik v oblasti Vstupní data a generuje matici výsledků založenou na podobných statistikách v oblasti Co požadujete. Matice LINREGRESE je ve skutečnosti umístěna ve skrytých řádkách pod viditelnou oblastí listu, jak vidíte na následujícím obrázku. První řádek hodnot v datové matici LINREGRESE je používán vzorcem Odhadovaná cena k odhadu hodnoty nemovitosti.

D26		={LINREGRESE(prices;inputs;PRAVDA;PRAVDA)}								
A	B	C	D	E	F	G	H	I	J	K
25										
26		Výstupní pole	58,05735057	15739,49919	-25459,2084	100,0201939	34188,10582			
27			0	0	0	0	0			
28			1	0	#N/A	#N/A	#N/A			
29			#NUM!	0	#N/A	#N/A	#N/A			
30										
31										
32		Převedení vstupů	2500	2,5	4	25	285000			
33			3200	3	3	7	325500			
34			2200	2,75	4	40	249500			
35			3450	3,75	6	22	379500			
36			2675	2,5	4	68	305000			
37										
38										
39										
40										

V tomto sešitu je skryto 25 řádků a 37 sloupců, je zapnuta ochrana listu a buňky jsou uzamčeny s položkami, jež jsou povoleny pouze ve vyznačených vstupních oblastech. Tip: jak uvádí poznámka na listu, lze k dosažení odhadované ceny zadávat ceny domů, skutečné prodejní ceny jsou však reálnější, pokud je nelze zjistit.

ODKAZ

Soubor Ceny domů.xlsx naleznete v oblasti Zkušební soubory na příloženém CD.

Funkce LINTREND

Funkce LINREGRESE vrátí matematický popis přímky, jež nejlépe odpovídá známým datům. Funkce LINTREND hledá body ležící na této přímce, jež spadají do kategorie neznámých dat. Čísla vrácená funkcí LINTREND lze využít k načrtnutí vývojové tendence – přímky, jež pomáhá dávat smysl datům. Pomocí funkce LINTREND lze *předpovědět* či provádět uvážené odhady o budoucích datech na základě tendencí vyvozených ze známých dat. (Buďte opatrní. Ačkoli lze pomocí funkce LINTREND načrtnout přímku, jež nejlépe odpovídá známým datům, nemůže vám tato funkce sdělit, zda tato přímka dobře předpovídá budoucnost. S odhadem vám mohou pomoci validační statistiky vrácené funkcí LINREGRESE.) Funkce LINTREND má formát **=LINTREND(pole_y; pole_x; nová_x; b)**.

První dva argumenty představují známé hodnoty vašich závislých a nezávislých proměnných. Stejně jako ve funkci LINREGRESE je argument *pole_y* jeden sloupec, jeden řádek nebo obdélníková oblast. Argument *pole_x* je také stejný jako ve funkci LINREGRESE. Třetí a čtvrtý argument je volitelný. Vynecháte-li argument *nová_x*, považuje jej funkce LINTREND za shodný s argumentem *pole_x*. Pokud uvedete argument *b*, musí být hodnota tohoto argumentu PRAVDA nebo NEPRAVDA (či 1 nebo 0). Má-li argument *b* hodnotu PRAVDA, počítá se konstanta *b* běžným způsobem.

Chcete-li vypočítat datové body tendenční přímky, jež nejlépe odpovídají známým datům, jednoduše z této funkce vynechejte třetí a čtvrtý argument. Výsledná matice bude mít stejnou velikost jako oblast *pole_x*. Na obrázku 17.4 jsme funkci LINTREND použili k vyhledání hodnoty bodů na regresní přímce, jež popisují množinu dat z příkladu na obrázku 17.3. K vytvoření těchto hodnot jsme vybrali oblast C2:C19 a zadali vzorec **=LINTREND(B2:B19; A2:A19)** jako maticový stisknutím kombinace kláves Ctrl+Shift+Enter.

K vytvoření předpovědi z existujících dat je nutné doplnit oblast pro *nová_x*. Zadat lze libovolný počet buněk. Výsledná matice bude mít stejnou velikost jako oblast *pole_x*. Na obrázku 17.5 jsme pomocí funkce LINTREND vypočítali odhadované prodeje pro 19., 20. a 21. měsíc. Abychom získali tyto hodnoty, zadali jsme **19** až **21** do oblasti A21:A23, vybrali oblast C21:C23 a poté zadali vzorec **=LINTREND(B2:B19; A2:A19; A21:A23)** jako maticový stisknutím kombinace kláves Ctrl+Shift+Enter.

C2		fx [=LINTREND(B2:B19;A2:A19)]										
	A	B	C	D	E	F	G	H	I	J	K	
1	Měsíc	Poptávka, tisíce	Trend									
2	1	4039	4022,678									
3	2	4057	4043,291									
4	3	4052	4063,904									
5	4	4094	4084,517			Lineární odhad						
6	5	4104	4105,13			20,613003	4002,065					
7	6	4110	4125,743									
8	7	4154	4146,356									
9	8	4161	4166,969									
10	9	4186	4187,582									
11	10	4195	4208,195									
12	11	4229	4228,808									
13	12	4244	4249,421									
14	13	4242	4270,034									
15	14	4283	4290,647									
16	15	4322	4311,26									
17	16	4333	4331,873									
18	17	4368	4352,486									
19	18	4389	4373,099									

Obrázek 17.4: Funkce LINTREND vytvoří datové řady, jež lze načrtnout jako přímku v grafu.

Funkce FORECAST

Funkce FORECAST je podobná funkci LINTREND s tím rozdílem, že vrací jen jednu předpovědanou hodnotu. Tato funkce má formát **=FORECAST(x; pole_y; pole_x)**.

Argument x je datový bod, u něhož chcete předpovědět hodnotu. Například místo funkce LINTREND bychom mohli použít funkci FORECAST k předpovědi hodnoty v buňce C23 na obrázku 17.5 zadáním vzorce **=FORECAST(21; B2:B19; A2:A19)**, kde argument x odkazuje na 21. Datový bod na regresní přímce. Pomocí této funkce lze vypočítat jakýkoli bod v budoucnosti.

C21		fx {=LINTREND(B2:B19;A2:A19;A21:A23)}										
	A	B	C	D	E	F	G	H	I	J	K	
1	Měsíc	Poptávka, tisíce	Trend									
2	1	4039	4022,678									
3	2	4057	4043,291									
4	3	4052	4063,904									
5	4	4094	4084,517									
6	5	4104	4105,13									
7	6	4110	4125,743									
18	17	4368	4352,486									
19	18	4389	4373,099									
20												
21	19		4393,712									
22	20		4414,325									
23	21		4434,938									
24												
25												

Obrázek 17.5: Pomocí funkce LINTREND lze předpovědět prodeje pro měsíce 19, 20 a 21.

Funkce SLOPE

Funkce SLOPE vrátí směrnici regresní přímky proložené zadanými body v oblastech pole_y a pole_x. Směrnice je vertikální vzdálenost dělená horizontální vzdáleností mezi dvěma body na přímce a vyjadřuje rychlost změny podél regresní přímky. Její hodnota je stejná jako první číslo v matici vrácené funkcí LINREGRESE. Jinými slovy, funkce SLOPE vypočítá trajektorii přímky použité funkcemi FORECAST a LINTREND k výpočtu hodnot datových bodů. Funkce SLOPE má formát **=SLOPE(pole_y; pole_x)**.

Budeme-li chtít najít směrnici regresní přímky, jež popisuje datové body z příkladu na obrázku 17.5, můžeme zadat vzorec **=SLOPE(B2:B19; A2:A19)** jako matici. Výsledná hodnota bude 20,613.

Funkce STEYX

Funkce STEYX vrátí standardní chybu při výpočtu lineární regrese. Standardní chyba je určena množstvím chyb při odhadu y pro jednotlivá x . Tato funkce má formát **=STEYX(pole_y; pole_x)**. Uplatníme-li tuto funkci na list znázorněný na obrázku 17.5, vrátí vzorec **=STEYX(B2:B19; A2:A19)** standardní chybovou hodnotu 12,96562.

Výpočty exponenciální regrese

Na rozdíl od lineární regrese, jež načrtává hodnoty na přímce, popisuje exponenciální regrese křivku výpočtem matice hodnot. Rovnice exponenciální regresní křivky vypadá následovně:

$$y = b * m_1^{x^1} * m_2^{x^2} * \dots * m_n^{x^n}$$

Máte-li pouze jednu nezávislou proměnnou, bude mít rovnice tuto podobu:

$$y = b * m^x$$

Funkce LOGLINREGRESE

Funkce LOGLINREGRESE funguje podobně jako funkce LINREGRESE s tím rozdílem, že slouží k analýze dat, jež nejsou lineární a vrací souřadnice exponenciální křivky místo přímky. Výsledkem funkce LOGLINREGRESE je matice hodnot, jež opisuje parametry křivky. Tato funkce má formát **=LOGLINREGRESE(pole_y; pole_x; b; stat)**.

Do funkce LOGLINREGRESE zadáváte stejné argumenty jako do funkce LINREGRESE a výsledkem je matice ve stejném formátu. Nastavíte-li volitelný argument *stat* na PRAVDA, vrátí funkce také validační statistiky. Více informací o základních rovnicích funkce LOGLINREGRESE naleznete v části „Funkce LINREGRESE“ na straně 546.

POZNÁMKA

Funkce LINREGRESE a LOGLINREGRESE vrátí pouze souřadnice osy y použité pro výpočty přímkou a křivkou. Rozdíl mezi nimi spočívá v tom, že funkce LINREGRESE určuje přímkou a funkce LOGLINREGRESE exponenciální křivku. Je tedy nutné věnovat pozornost volbě vhodné funkce. Funkce LINREGRESE může být vhodnější pro odhady prodeje a funkce LOGLINREGRESE se více hodí pro statistické analýzy či populační trendy.

Funkce LOGLINTREND

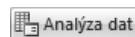
Zatímco funkce LOGLINREGRESE vrátí matematický popis exponenciální regresní křivky, jež nejlépe odpovídá množině známých dat, funkce LOGLINTREND vyhledá body ležící na této křivce. Funkce LOGLINTREND funguje jako její lineární protějšek, LINTREND a má formát **=LOGLINTREND(pole_y; pole_x; nová_x; b)**. Více informací o argumentech funkce LOGLINTREND naleznete v části „Funkce LINTREND“ na straně 549.

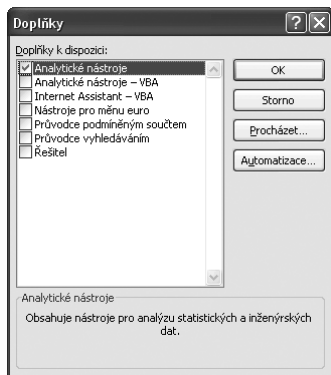
Práce s doplňkem Analytické nástroje

Doplňek Analytické nástroje je součástí Excelu 2007, ačkoli o tom možná ani nevíte. *Doplňky* jsou malé balíčky nástrojů, jež se více či méně snadno integrují do uživatelského prostředí Excelu. Nejprve je však nutné doplňky nainstalovat. V následujících částech si popíšeme instalaci doplňku Analytické nástroje a jeho využití.

Instalace doplňku Analytické nástroje

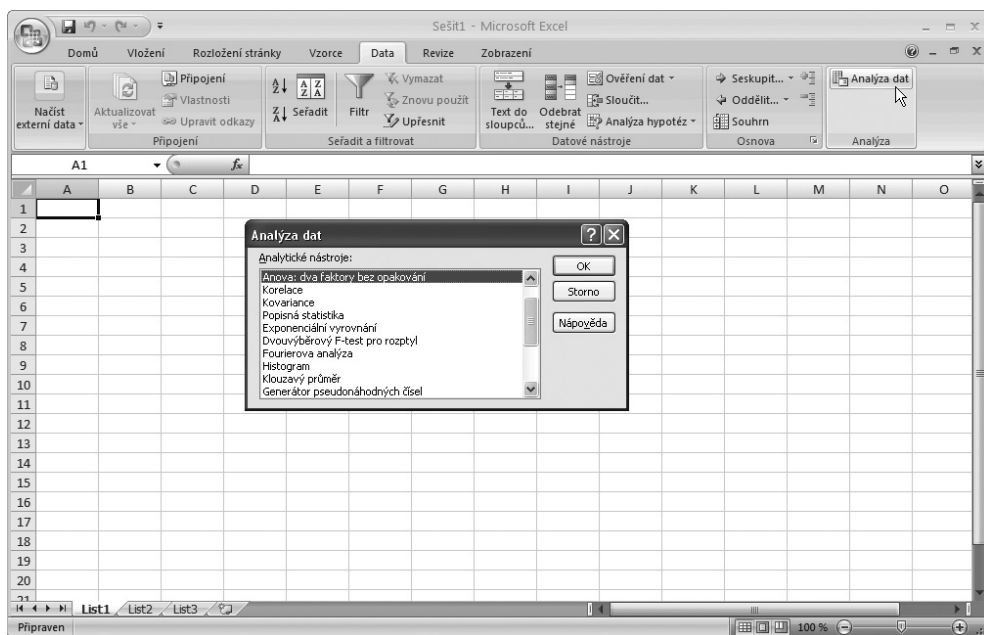
Chcete-li zjistit, zda máte nainstalovaný doplňek Analytické nástroje, klepněte na pásu karet na kartu Data. Pokud zde vidíte tlačítko analýza dat, můžete začít s doplňkem pracovat. Jestliže toto tlačítko nevidíte, klepněte na tlačítko Office, Možnosti aplikace Excel a poté vyberte kategorii Doplňky. V seznamu Spravovat v dolní části dialogového okna vyberte položku doplňky aplikace Excel a poté klepněte na tlačítko Přejít. Otevře se dialogové okno, jež vidíte na obrázku 17.6. V dialogovém okně Doplňky označte zaškrtačiví políčko Analytické nástroje a klepněte na tlačítko OK pro zahájení instalace. Excel vás požádá o potvrzení. Klepněte na tlačítko Ano.





Obrázek 17.6: Doplňek Analytické nástroje nainstalujete prostřednictvím dialogového okna Doplňky.

Po klepnutí na tlačítko Analýza dat na kartě Data se otevře dialogové okno Analýza dat, jež vidíte na obrázku 17.7.



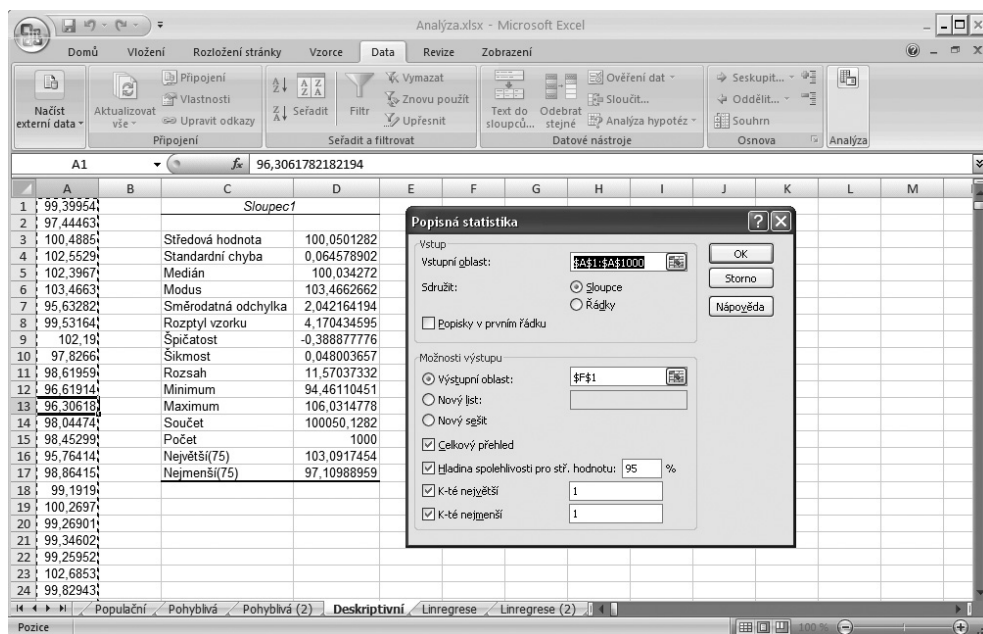
Obrázek 17.7: Dialogové okno Analýza dat otevřete klepnutím na stejnojmenné tlačítko na kartě Data.

Práce s nástrojem Popisná statistika

Nástroj Popisná statistika vytváří sestavu statistických kritérií s jednou variantou pro data ve vstupní oblasti. Pro každou proměnnou ve vstupní oblasti obsahuje výstupní oblast tohoto

nástroje podrobný popis statistik, jak vidíte na obrázku 17.8. Tento nástroj lze spustit klepnutím na tlačítko Analýza dat na kartě Data, výběrem položky Popisná statistika a následným klepnutím na tlačítko OK. Otevře se dialogové okno Popisná statistika (viz obrázek 17.8).

Nástroj Popisná statistika vyžaduje, aby se vstupní oblast skládala z jedné či více proměnných a z výstupní oblasti. Je také nutné určit, zda mají být proměnné uspořádány do sloupců nebo řádků. Zahrnete-li i řádek s popisky, je třeba označit zaškrtnuté políčko Popisky v prvním řádku. Excel pak použije popisky k rozpoznání proměnných ve vstupní tabulce. Možnost Celkový přehled vyberte pouze v případě, že požadujete podrobnou výstupní tabulku, kterou vidíte na obrázku 17.8; v opačném případě ponechejte toto zaškrtnuté políčko prázdné.



Obrázek 17.8: Tuto statistickou tabulku jsme vygenerovali popisem sloupce s 1 000 hodnotami obsaženými ve sloupci A pomocí nástroje Popisná statistika.

Stejně jako jiné nástroje v doplňku Analytické nástroje, vytváří nástroj Popisná statistika tabulku s konstantami. Pokud tabulka s konstantami nevyhovuje vašim potřebám, lze většinu stejných statistických dat získat prostřednictvím jiných nástrojů z doplňku Analytické nástroje nebo pomocí vzorců s funkcemi. V tabulce 17.2 naleznete přehled statistik a vzorců.

Tabulka 17.2: Vzorce popisných statistik

Statistika	Vzorec
Střední hodnota	=PRŮMĚR(<i>číslo1</i> ; <i>číslo2</i> ; ...)
Standardní chyba	Podobně jako =STEYX(<i>pole_y</i> ; <i>pole_x</i>), avšak používá rozložení ± místo standardního běžného rozložení.
Medián	=MEDIAN(<i>číslo1</i> ; <i>číslo2</i> ; ...)

Statistika	Vzorec
Modus	=MODE(číslo1; číslo2; ...)
Standardní odchylka	=STDEV(číslo1; číslo2; ...)
Rozptyl	=VAR(číslo1; číslo2; ...)
Špičatost	=KURT(číslo1; číslo2; ...)
Šikmost	=SKEW(číslo1; číslo2; ...)
Oblast	=MAX(číslo1; číslo2)-MIN(číslo1; číslo2; ...)
Minimum	=MIN(číslo1; číslo2; ...)
Maximum	=MAX(číslo1; číslo2; ...)
Suma	=SUMA(číslo1; číslo2; ...)
Počet	=COUNT(číslo1; číslo2; ...)
k-tá největší hodnota	=LARGE(pole; k)
k-tá nejmenší hodnota	=SMALL(pole; k)
Interval spolehlivosti	Podobně jako =CONFIDENCE(alfa; sm_odch; velikost), ale používá jiný algoritmus

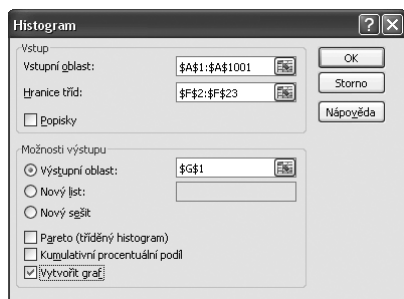
Vytváření histogramů

Histogram je graf (obvykle jednoduchý sloupcový graf), jenž použije skupinu dat k výpočtu individuální a kumulativní četnosti oblasti buněk dat a tříd dat, jež spadají do několika intervalů (označovaných jako hranice tříd).

Nástroj Histogram si představíme na příkladu tabulky s 1 000 výsledky testů. (Vstupní oblast musí obsahovat pouze numerická data.) Abychom viděli rozložení celkových výsledků v 50bodových intervalech, začneme nastavováním rozložení hranic tříd, jež je znázorněno ve sloupci F na obrázku 17.9.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID studenta	Ústní	Matematika	Celkem										
2	722-4499	418	518	936		600								
3	605-3475	465	557	1022		650								
4	546-3500	463	549	1012		700								
5	655-8550	466	587	1053		750								
6	812-6448	520	544	1064		800								
7	814-6503	470	537	1007		850								
8	332-3453	533	549	1082		900								
9	55-0476	476	570	1046		950								
10	745-0539	468	548	1016		1000								
11	675-1544	441	562	1003		1050								
12	836-6513	570	560	1130		1100								
13	146-4509	503	554	1057		1150								
14	45-9450	551	556	1107		1200								
15	924-3538	498	562	1060		1250								
16	664-9559	466	525	991		1300								
17	596-5425	414	549	963		1350								
18	914-1483	497	545	1042		1400								
19	266-3469	581	519	1100		1450								
20	116-2492	402	568	970		1500								
21	524-6464	488	528	1027		1550								

Obrázek 17.9: Sloupec F obsahuje přehled rozložení hranic tříd.



Obrázek 17.10: Toto dialogové okno se otevře po výběru nástroje Histogram v dialogovém okně Analýza dat.

Rozložení hranic tříd nemusí být rovnoměrné jako na obrázku 17.9, musí však být ve vze-
stupném pořadí. Na kartě Data klepněte na tlačítko Analýza dat, vyberte nástroj Histogram
a klepněte na tlačítko OK. Na obrázku 17.10 vidíte dialogové okno Histogram.

Nástroj Histogram může pracovat se třemi položkami informací: s umístěním dat (v tomto
případě D2:D1001), umístěním hranic tříd (F2:F22) a buňkou v levém horním rohu oblasti,
v níž se má analýza zobrazit (G1). Po klepnutí na tlačítko OK vypíše Excel analýzu do sloupců
G a H, jak ukazuje obrázek 17.11.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID studenta	Ústní	Matematika	Celkem			Bin	Frequency	Cumulative %					
2	722-4499	418	518	936		600	600	0	0,00%					
3	605-3475	465	557	1022		650	650	0	0,00%					
4	546-3500	463	549	1012		700	700	0	0,00%					
5	655-8550	466	587	1053		750	750	0	0,00%					
6	812-6448	520	544	1064		800	800	0	0,00%					
7	814-6503	470	537	1007		850	850	0	0,00%					
8	332-3453	533	549	1082		900	900	3	0,30%					
9	55-0476	476	570	1046		950	950	48	5,10%					
10	745-0539	468	548	1016		1000	1000	155	20,60%					
11	675-1544	441	562	1003		1050	1050	329	53,50%					
12	836-6513	570	560	1130		1100	1100	291	82,60%					
13	146-4509	503	554	1057		1150	1150	150	97,60%					
14	45-9450	551	556	1107		1200	1200	23	99,90%					
15	924-3538	498	562	1060		1250	1250	1	100,00%					
16	664-9559	466	525	991		1300	1300	0	100,00%					
17	596-5425	414	549	963		1350	1350	0	100,00%					
18	914-1483	497	545	1042		1400	1400	0	100,00%					
19	266-3469	581	519	1100		1450	1450	0	100,00%					
20	116-2492	402	568	970		1500	1500	0	100,00%					
21	524-5404	488	520	1007		1550	1550	0	100,00%					

Obrázek 17.11: Podle této analýzy byly tři výsledky alespoň 900, ale méně než 950; 48 je alespoň 950, ale méně než 1 000; atd.

POZNÁMKA

V tomto modelu jsme vytvořili vlastní hranice tříd, může je však určit i nástroj Histogram. Ponecháte-li pole Hranice tříd prázdné, vytvoří se rovnoměrně rozložené intervaly hranic tříd, přičemž minimální a maximální hodnoty ve vstupní oblasti jsou využity jako počáteční a koncové body. Počet intervalů je roven druhé odmocnině počtu vstupních hodnot.

Při práci s nástrojem Histogram byste měli mít na paměti následující fakta:

- Ve sloupci Četnost uvádí histogram počet vstupních hodnot, jež se rovnají nebo jsou větší než hodnota hranice třídy, ale menší než následující hodnota hranice třídy.
- Poslední hodnota v tabulce uvádí počet vstupních hodnot, jež se rovnají nebo jsou větší než poslední hodnota hranice třídy.
- Chcete-li výstupní hodnoty seřadit sestupně, označte v dialogovém okně Histogram zaškrtnuté políčko Pareto (tříděný histogram).
- Pokud chcete vytvořit tabulku s kumulativními procentuálními podíly každé úrovně hranice tříd, vyberte možnost Kumulativní procentuální podíl.
- Vyberete-li v dialogovém okně Histogram možnost Vytvořit graf, vytvoří nástroj Histogram zároveň i graf a umístí jej vedle tabulky četnosti.

ODKAZ

Podrobné informace o grafech naleznete v části 6 této knihy.

DO DETAILU

PŘI POUŽITÍ VZORCŮ V HRANICÍCH TŘÍD

Nástroj Histogram duplikuje sloupce hodnot hranic tříd ve sloupci Hranice tříd, což je výhodné v případě, že výstup umístíte například do jiného listu v sešitu. Protože však nástroj Histogram kopíruje hodnoty hranic tříd, je dobré, pokud oblast hranic tříd obsahuje číselné konstanty, a ne vzorce. Pracujete-li se vzorci, měli byste zajistit, aby neobsahovaly relativní odkazy; v opačném případě by po zkopírování oblasti nástrojem Histogram mohly mít vzorce neočekávané výsledky.

Analýza rozložení pomocí funkce ČETNOSTI

Pomocí nástroje Histogram lze snadno vytvořit novou tabulku četnosti vždy, když změníte vstupní hodnoty, tento nástroj však generuje statická čísla (číselné konstanty). Pokud byste chtěli vytvořit raději vzorce propojené se vstupními hodnotami, lze použít vestavěnou maticovou funkci ČETNOSTI, jež vypočítá, kolikrát se určené hodnoty vyskytují v populaci, a přijímá argumenty (*data; hodnoty*). Na obrázku 17.12 vidíte funkci ČETNOSTI, jež pracuje s daty znázorněnými na obrázku 17.9.

Chcete-li využít funkci ČETNOSTI, nastavte sloupec hodnot hranic tříd a poté vyberte celou oblast, v níž se má zobrazit výstup, což je v našem případě oblast G2:G21 – buňky ve sloupci G, jež přímo sousedí s hodnotami hranic tříd ve sloupci F. (Musí se jednat o sloupec, protože funkce ČETNOSTI nemůže použít řádek nebo oblast s více řádky jako výstupní oblast.) Poté zadejte vzorec, přičemž určíte vstupní oblast jako první argument a oblast hranic tříd jako druhý argument. Stisknutím kombinace kláves Ctrl+Shift+Enter uzamknete maticový vzorec. Více informací o maticích naleznete v části „Práce s maticemi“ na straně 465.

Práce s nástrojem Pořadová statistika a percentily

Předpokládejme, že chcete seřadit výsledky znázorněné na obrázku 17.9. Data byste mohli seřadit sestupně, tedy s nejlepším výsledkem nahoře a nejhorším dole ve sloupci. K vyhledání

pořadí kteréhokoli výsledku lze vedle seřazených výsledků vytvořit vzestupné řady čísel, přičemž vedle nejlepšího výsledku bude hodnota 1 a vedle nejhoršího výsledku hodnota 1000.

G2		={ČETNOSTI(D2:D1001;F2:F20)}											
	A	B	C	D	E	F	G	H	I	J	K	L	
1	ID studenta	Ústní	Matematika	Celkem									
2	722-4499	418	518	936		600	0						
3	605-3475	465	557	1022		650	0						
4	546-3500	463	549	1012		700	0						
5	655-8550	466	587	1053		750	0						
6	812-6448	520	544	1064		800	0						
7	814-6503	470	537	1007		850	0						
8	332-3453	533	549	1082		900	3						
9	55-0476	476	570	1046		950	48						
10	745-0539	468	548	1016		1000	155						
11	675-1544	441	562	1003		1050	329						
12	836-6513	570	560	1130		1100	291						
13	146-4509	503	554	1057		1150	150						
14	45-9450	551	556	1107		1200	23						
15	924-3538	498	562	1060		1250	1						
16	664-9559	466	525	991		1300	0						
17	596-5425	414	549	963		1350	0						
18	914-1483	497	545	1042		1400	0						
19	266-3469	581	519	1100		1450	0						
20	116-2492	402	568	970		1500	0						

Obrázek 17.12: Pomocí funkce ČETNOSTI popojte analýzu rozložení se vstupními daty.

Tyto úkoly lze provést pomocí nástroje Pořadová statistika, a navíc lze vytvořit i procentuální pořadí každé hodnoty ve vstupní oblasti. Chcete-li pracovat s tímto nástrojem, klepněte na tlačítko Analýza dat na kartě Data, vyberte položku Pořadová statistika a percentily a poté klepněte na tlačítko OK. Na obrázku 17.13 vidíte dialogové okno Pořadová statistika a percentily.

POZNÁMKA

Pokud na kartě Data nevidíte tlačítko Analýza dat, nahlédněte do části „Instalace doplňku Analytické nástroje“ na straně 551.

Výstup nástroje Pořadová statistika a percentily, jenž je znázorněn v dolní části obrázku 17.13, lze vykládat následujícím způsobem. První řádek výstupní tabulky (F2:I2) nám říká, že 285. položka ve vstupní oblasti je celkový výsledek 1206, který je prvním v pořadí a je lepší než 100 procent ostatních výsledků. Níže uvádíme několik faktů, na něž byste měli pamatovat při práci s nástrojem Pořadová statistika a percentily:

- Nejvhodnější je zvolit v dialogovém okně Pořadová statistika a percentily možnost Popisky v prvním řádku a poté zahrnout hlavičku sloupce do vstupní oblasti. Díky tomu použije druhý sloupec ve výstupní tabulce stejný popis. Nezařnete-li popis do vstupní oblasti, bude mít výstupní sloupec popis Sloupec1.

POZNÁMKA

Vyberete-li možnost Popisky v prvním řádku, ale nezahrnete buňku obsahující popis do vstupní oblasti, bude název určen podle první buňky ve vstupní oblasti. Pokud by například vstupní oblast na obrázku 17.13 byla \$D\$2:\$D\$1001, výsledný popis v sloupci G by byl 936 místo Celkem.

- Na obrázku 17.13 jsme analyzovali jeden sloupec s daty, mohli bychom však dohromady analyzovat sloupce Ústní, Matematika a Celkem. V takovém případě bychom určili vstupní oblast B1:D1001 a nástroj by vygeneroval 12 sloupců pro výstup, čtyři pro každý vstupní sloupec.
- Lze také zajistit umístění výstupní tabulky na nový list či do nového sešitu, což je vhodné, jestliže vyberete více sloupců vstupní oblasti a vznikne tak velká výstupní tabulka.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	ID studenta	Ústní	Matematika	Celkem											
2	722-4499	418	518	936											
3	605-3475	465	557	1022											
4	546-3500	463	549	1012											
5	655-8550	466	587	1053											
6	812-6448	520	544	1064											
7	814-6503	470	537	1007											
8	332-3453	533	549	1082											
9	55-0476	476	570	1046											
10	745-0539	468	548	1016											
11	675-1544	441	562	1003											
12	836-6513	570	560	1130											
13	146-4509	503	554	1057											
14	45-9450	551	556	1107											
15	924-3538	498	562	1060											
16	664-9559	466	525	991											
17	596-5425	414	549	963											
18	914-1483	497	545	1042											
19	266-3469	581	519	1100											
20	116-2492	402	568	970											

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	ID studenta	Ústní	Matematika	Celkem		Místo	Celkem	Hodnocení	Procent						
2	722-4499	418	518	936		1	936	972	2,70%						
3	605-3475	465	557	1022		2	1022	657	33,90%						
4	546-3500	463	549	1012		3	1012	726	26,60%						
5	655-8550	466	587	1053		4	1053	447	54,60%						
6	812-6448	520	544	1064		5	1064	379	61,70%						
7	814-6503	470	537	1007		6	1007	756	23,80%						
8	332-3453	533	549	1082		7	1082	270	72,60%						
9	55-0476	476	570	1046		8	1046	502	49,50%						
10	745-0539	468	548	1016		9	1016	693	29,90%						
11	675-1544	441	562	1003		10	1003	779	21,30%						
12	836-6513	570	560	1130		11	1130	61	93,70%						
13	146-4509	503	554	1057		12	1057	419	57,60%						
14	45-9450	551	556	1107		13	1107	140	85,50%						
15	924-3538	498	562	1060		14	1060	403	59,20%						
16	664-9559	466	525	991		15	991	834	16,40%						
17	596-5425	414	549	963		16	963	921	7,50%						
18	914-1483	497	545	1042		17	1042	526	46,30%						
19	266-3469	581	519	1100		18	1100	175	82,20%						
20	116-2492	402	568	970		19	970	902	9,40%						

Obrázek 17.13: Pomocí nástroje Pořadová statistika a percentily lze vygenerovat výstupní tabulku, jako je ta na dolním obrázku.

Generování náhodných čísel

Vestavěná funkce NÁHČÍSLO vrátí rovnoměrně rozložená náhodná reálná čísla větší nebo rovna 0 a menší než 1. Jinými slovy, všechny hodnoty od 0 do 1 sdílejí stejnou pravděpodobnost, že budou vráceny jako sada vzorců založených na funkci NÁHČÍSLO. Protože se jedná o poměrně malý vzorek, rozložení není dokonale rovnoměrné. Nicméně opakované testy ukazují, že funkce NÁHČÍSLO neupřednostňuje žádnou pozici v rámci svého spektra rozložení. Více informací naleznete v části „Funkce NÁHČÍSLO a RANDBETWEEN“ na straně 492.

KORELAČNÍ TABULKY

Vstupní a výstupní tabulky znázorněné v dolní části obrázku 17.13 sdílejí společný sloupec dat – sloupec Celkem – a stejný počet řádků. Liší se však řazení tabulek, a proto se řádky neshodují. Nejjednodušším řešením je seřadit výstupní tabulku podle sloupce Bod; v tomto kontextu určuje **Bod** pozici odpovídajícího datového bodu ve vstupní oblasti. Díky tomu bude výstupní tabulka po seřazení podle sloupce Bod ve stejném pořadí jako vstupní tabulka:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID studenta	Ústní	Matematika	Celkem		Místo	Celkem	Hodnocení	Procent					
2	722-4499	418	518	936		1	936	972	2,70%					
3	605-3475	465	557	1022		2	1022	657	33,90%					
4	546-3500	463	549	1012		3	1012	726	26,60%					
5	655-8550	466	587	1053		4	1053	447	54,60%					
6	812-6448	520	544	1064		5	1064	379	61,70%					
7	814-6503	470	537	1007		6	1007	756	23,80%					
8	332-3453	533	549	1082		7	1082	270	72,60%					
9	55-0476	476	570	1046		8	1046	502	49,50%					
10	745-0539	468	548	1016		9	1016	693	29,90%					
11	675-1544	441	562	1003		10	1003	779	21,30%					
12	836-6513	570	560	1130		11	1130	61	93,70%					
13	146-4509	503	554	1057		12	1057	419	57,60%					
14	45-9450	551	556	1107		13	1107	140	85,50%					
15	924-3538	498	562	1060		14	1060	403	59,20%					

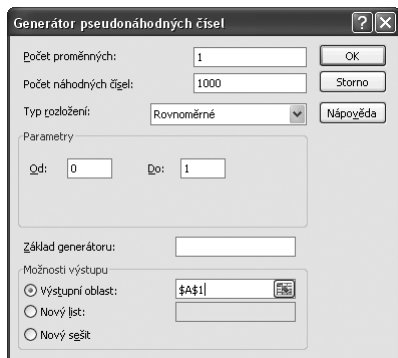
Chcete-li přidat informace z výstupní tabulky do existující vstupní tabulky, můžete odstranit sloupec bod (protože tento sloupec jednoduše určuje číslo řádku), sloupec Celkem (protože vstupní tabulka již sloupec Celkem obsahuje) a prázdný sloupec výstupní tabulky, čímž vytvoříte jednu, korelační tabulku.

ŘEŠENÍ PROBLÉMŮ

NÁHODNÁ ČÍSLA SE NEUSTÁLE MĚNÍ

Funkce NÁHČÍSLO je jednou z *nestálých* funkcí Excelu – to znamená, že se přepočítává při každé změně listu, k čemuž dochází vždy, když zadáte data do buňky. Chcete-li vytvořit sadu náhodných čísel a poté je „zmrazit“, vyberte všechny vzorce NÁHČÍSLO v listu a stisknutím kombinace kláves Ctrl+C je zkopírujte. Poté klepněte na tlačítko Vložit na kartě Domů a zvolte příkaz Vložit hodnoty pro nahrazení nestálých vzorců pevně stanovenými hodnotami. Místo funkce NÁHČÍSLO lze také použít nástroj Generátor pseudonáhodných čísel (popisovaný dále), jenž vytvoří konstanty místo vzorců.

Nástroj Generátor pseudonáhodných čísel vytvoří sadu náhodných čísel, jež nejsou rovnoměrně rozložena. Pomocí nástroje Histogram lze seřadit a načrtnout výsledky pro rozhodovací analýzu Monte-Carlo a jiné druhy simulací. K dispozici je šest typů rozložení: Rovnoměrné, Normální, Bernoulliho, Binomické, Poissonovo a Diskrétní (definované uživatelem). Navíc lze v seznamu Typ rozložení vybrat i položku Vzorky a vytvořit tak nenáhodná čísla s určitými intervaly. Na kartě Data klepněte na tlačítko Analýza dat, vyberte položku Generátor pseudonáhodných čísel a klepnutím na tlačítko OK otevřete dialogové okno, jež vidíte na obrázku 17.14.



Obrázek 17.14: Oblast Parametry v dialogovém okně Generátor pseudonáhodných čísel se mění podle vámi zvoleného typu rozložení.

Následující dva body jsou důležité při práci s Generátorem pseudonáhodných čísel:

- V textových polích Počet proměnných a Počet náhodných čísel lze určit, kolik sloupců s čísly požadujete a kolik čísel má být v každém sloupci. Chcete-li například 10 sloupců, z nichž každý bude obsahovat 100 čísel, zadejte do textového pole Počet proměnných hodnotu 10 a do textového pole Počet náhodných čísel hodnotu 100.
- Lze také určit základ generátoru. Při každém generování sady náhodných čísel s využitím konkrétního typu rozložení se stejnou základní hodnotu však získáte stejnou řadu čísel; z tohoto důvodu byste základ generátoru měli určit pouze v případě, že chcete mít možnost zopakovat sekvenci náhodných čísel.

Metody rozložení náhodných čísel

Parametry umístěné přímo pod seznamem Typ rozložení v dialogovém okně Generátor pseudonáhodných čísel se mění v závislosti na vámi zvoleném typu rozložení. Jak vidíte na obrázku 17.14, vyberete-li v seznamu Typ rozložení položku Rovnoměrné, můžete v textových polích Od a Do určit počáteční a koncové body rozložení.

Rovnoměrné rozložení náhodných čísel

Při volbě této možnosti určujete dvě čísla, mezi nimiž (a včetně nich) se má vygenerovat sada náhodných čísel. Výsledek je v podstatě totožný jako u funkce RANDBETWEEN, přičemž se vygeneruje rovnoměrně rozložená sada reálných čísel. Chcete-li nastavit jiné koncové body než 0 a 1 nebo mají být sady čísel založené na stejné základní hodnotě, je využití této možnosti pohodlnější než práce s funkcí NÁHČÍSLO.

Normální rozložení náhodných čísel

Normální rozložení má následující vlastnosti:

- Jedna konkrétní hodnota – střední – se vyskytne pravděpodobněji než kterákoli jiná hodnota.
- Hodnoty větší než střední hodnota se budou vyskytovat pravděpodobněji než hodnoty menší.
- Hodnoty blíže ke střední hodnotě se budou vyskytovat pravděpodobněji než hodnoty vzdálené od střední hodnoty.

Chcete-li vygenerovat normálně rozložená náhodná čísla, určujete dva parametry: střední hodnotu a směrodatnou odchylku. Směrodatná odchylka je průměrný absolutní rozdíl mezi náhodnými čísly a střední hodnotou. (Přibližně 68 procent hodnot v normálním rozložení bude spadat do jedné standardní odchylky od střední hodnoty.)

Generování náhodných čísel pomocí Bernoulliho rozložení

Možnost Bernoulliho rozložení je určeno pravděpodobností výskytu (hodnota p) v daném pokusu za předpokladu, že všechny pokusy mají stejnou pravděpodobnost úspěchu a úspěch jednoho pokusu nemá žádný vliv na úspěch následných pokusů. (Úspěch v tomto kontextu nemá za následek žádnou hodnotu. Jinými slovy lze pomocí tohoto rozložení simulovat selhání stejně jako úspěch.) Všechny hodnoty ve výstupu Bernoulliho rozložení jsou buď 0, nebo 1.

Pravděpodobnost, že každá buňka vrátí 1, je dána jediným parametrem – hodnotou P – za níž dosahujete čísla od 0 do 1. Chcete-li například řadu 100 náhodných Bernoulliho hodnot, jejichž nejpravděpodobnější součet je 27, definujete výstupní oblast se 100 buněk a určíte hodnotu P jako 0,27.

Generování náhodných čísel pomocí binomického rozložení

Možnost binomického rozložení simuluje počet výskytů hodnoty ve stanoveném počtu výběrů, přičemž je určena pravděpodobnost výskytu. Stejně jako u možnosti Bernoulliho rozložení jsou výběry nezávislé; to znamená, že jednotlivé výsledky na sebe nemají vliv. Chcete-li vygenerovat binomicky rozložená čísla, určíte počet výběrů a hodnotu P (pravděpodobnost), že některý z výběrů bude úspěšný. (Ani zde nemá úspěch v tomto kontextu za následek žádnou hodnotu. Jinými slovy lze pomocí tohoto rozložení simulovat selhání stejně jako úspěch.)

Předpokládejme například, že máte 10 prodejních prezentací za týden, přičemž ve 20 procentech případů svůj produkt prodáte a chtěli byste zjistit, jaká by mohla být pravděpodobnost úspěchu v příštím roce. Do textového pole Počet náhodných čísel zadejte hodnotu 50 (50 pracovních týdnů v roce), do textového pole Pravděpodobnost výskytu zadejte hodnotu 0,2 a do textového pole Počet výběrů zadejte hodnotu 10. Dozvíte se, že v osmi týdnech z celého příštího roku pravděpodobně nedosáhnete žádných prodejů.

Generování náhodných čísel pomocí Poissonova rozložení

Poissonovo rozložení je určeno hodnotou λ , která je rovna převrácené hodnotě střední hodnoty. Poissonovo rozložení se často používá k určení pravděpodobného počtu událostí, k nimž dojde v určitém časovém úseku. Události jsou na sobě nezávislé; to znamená, že jednotlivé události nemají žádný vliv na pravděpodobnost jiných událostí.

Předpokládejme, že přijmete průměrně 10 služebních hovorů za den. Chcete vědět, jak často lze očekávat přijetí 18 či více služebních hovorů za den během jednoho roku. Abyste tuto informaci získali, zadejte do pole Počet náhodných čísel hodnotu 260 (52 týdnů s 5 pracovními dny) a do pole λ napište 10 (očekávaný průměr). Následně lze pomocí funkce COUNTIF spočítat, kolikrát se 18 objeví ve výstupní oblasti. Více informací naleznete v části „Funkce SUMIF, SUMIFS a COUNTIF“ na straně 540.

Generování náhodných čísel pomocí diskrétního rozložení

Pomocí diskrétního rozložení lze vytvářet vlastní vzorky rozložení určením tabulky možných výsledných hodnot spolu s pravděpodobností spojenou s každou z výsledných hodnot. Hodnoty pravděpodobností musí být od 0 do 1 a součet všech pravděpodobností v tabulce

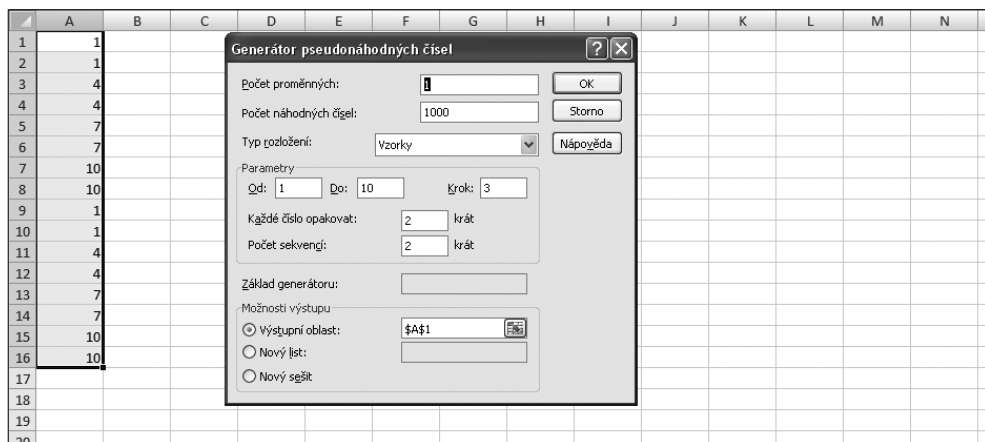
musí být roven 1. Chcete-li použít možnost diskrétního rozložení, určete možné výsledné hodnoty a jejich pravděpodobnosti jako oblast se dvěma sloupci, přičemž odkaz na ni je jediným zadávaným parametrem pro tuto možnost.

Mohli byste například vytvořit vlastní vzorek rozložení k vygenerování náhodných vzorků prodejů lopat na sníh založených na vstupní oblasti se dvěma sloupci: Číslo měsíce a Pravděpodobnost sněhu.

Generování polonáhodných čísel pomocí typu rozložení Vzorky

Zvolíte-li v seznamu Typ rozložení možnost Vzorky, vygenerují se čísla, jež jsou zároveň náhodná i částečně nenáhodná. Po výběru možnosti Vzorky se otevře dialogové okno, jež vidíte na obrázku 17.15.

Možnost Vzorky vytváří jednu či více aritmetických řad s volitelným vnitřním opakováním. Chcete-li například vytvořit řady hodnot z obrázku 17.15, vyplňte dialogové okno podle obrázku, přičemž budete požadovat dvě řady čísel 1 až 10, v poli krok určíte interval 3 a každé číslo se bude opakovat dvakrát v rámci každého cyklu.



Obrázek 17.15: Možnost Vzorky v seznamu Typ rozložení vytvoří aritmetické řady s operačním opakováním.

ODKAZ

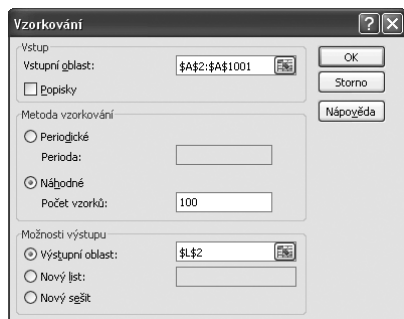
Další informace naleznete v části „Vyplňování a vytváření datových řad“ na straně 226.

Pokud krokový interval vede k tomu, že řady jsou nad určenou horní hodnotou, obsahuje výstupní oblast horní hodnotu, protože poslední interval je zkrácený. Určíte-li například krokový interval 4 a čísla 1 až 10, Excel vytvoří řady hodnot 1, 5, 9 a 10.

Vzorkování populace čísel

Nástroj Vzorkování extrahuje podmnožinu čísel z větší skupiny (či populace) čísel. Ze vstupní oblasti lze vzorkovat určený počet hodnot náhodně nebo každou n -tou hodnotu. Nástroj Vzorkování zkopíruje extrahovaná čísla do vámi určené výstupní oblasti. Chcete-li otevřít dialogové okno Vzorkování (viz obrázek 17.16), klepněte na tlačítko Analýza dat na kartě Data.

Hodnoty ve vstupní oblasti musí být číselné. Mohou obsahovat prázdné hodnoty a datum v případě, že jej zadáváte jako čísla, a ne jako text. Například pro zjednodušení grafu s denními cenami zboží lze pomocí nástroje Vzorkování extrahovat každý n -tý datový bod a poté vytvořit nový graf z extrahovaných dat.



Obrázek 17.16: Nástroj Vzorkování extrahuje náhodnou nebo periodickou množinu číselné populace.

POZNÁMKA

Chcete-li provádět stejné vzorkování z oblasti obsahující textové hodnoty, nastavte řady vzestupných celých čísel počínaje 1 ve sloupci podél textových hodnot a poté pomocí nástroje Vzorkování extrahujte čísla z těchto řad. Následně lze s využitím výsledných čísel jako argumentů ve funkci INDEX sestavit seznam vzorkových textových hodnot. Více informací naleznete v části „Funkce INDEX“ na straně 506.

Výpočty klouzavých průměrů

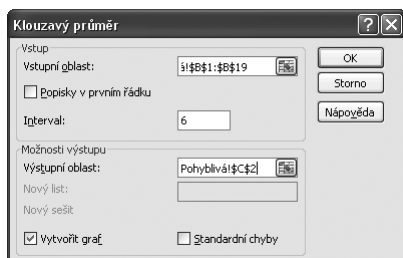
Klouzavý průměr je metoda předpovědi, jež zjednodušuje tendenční analýzu vyladěním výkyvů, k nimž postupem času dochází v měřených hodnotách. Tyto výkyvy mohou být způsobeny náhodným šumem, jenž je často vedlejším produktem měřicí metody. Například měření výšky rostoucího dítěte se může lišit kvůli přesnosti pravítka a v tom, zda dítě stojí rovně nebo se hrbí. Lze však použít různá měření a s postupem času je vyladovat, přičemž výsledkem bude křivka, jež odráží skutečnou rychlost růstu dítěte. Výkyvy v měřeních mohou být důsledkem jiných přechodných podmínek. Například měsíční prodeje se mohou lišit s počtem pracovních dnů v měsíci nebo v případě, že je nejlepší prodejce firmy na dovolené.

Předpokládejme, že jste vytvořili 18měsíční křivku poptávky (viz obrázek 17.17). K vygenerování tendenční křivky s menším šumem lze sestavit šestiměsíční klouzavý průměr. Prvním bodem na přímkě klouzavého průměru je průměr výsledků z prvních šesti měsíců (leden až červen 2008). Další body určuje průměr výsledků v druhém až sedmém měsíci (únor až červenec 2008) atd. Tuto analýzu za vás může provést nástroj Klouzavý průměr. Na kartě Data klepněte na tlačítko Analýza dat, vyberte položku Klouzavý průměr a klepněte na tlačítko OK. Otevře se dialogové okno Klouzavý průměr (viz obrázek 17.18).

Nástroj Klouzavý průměr vyžaduje tři informace: vstupní oblast obsahující data, jež chcete analyzovat; výstupní oblast, v níž se zobrazí výsledná data; a interval, na jehož základě je vytvořen průměr dat. Chceme-li například určit tříměsíční klouzavý průměr, určíme interval 3.



Obrázek 17.17: Na této 18měsíční křivce poptávky si ukážeme práci s nástrojem Klouzavý průměr.

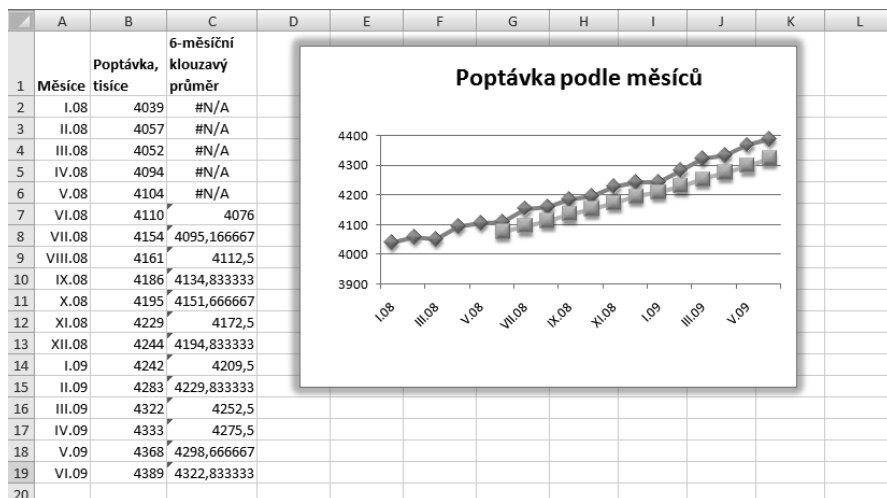


Obrázek 17.18: Nástroj Klouzavý průměr pomáhá vyladit křivky při získávání tendenčního vývoje.

Na obrázku 17.9 vidíte šestiměsíční klouzavý průměr přidáný nad původní křivku poptávky z obrázku 17.8. Nástroj Klouzavý průměr vložil data do sloupce C a použil je k vytvoření příměšší křivky v grafu. Všimněte si, že prvních pět buněk ve výstupní oblasti obsahuje chybové hodnoty #N/A. Protože interval má hodnotu n , bude na začátku výstupu vždy $n-1$ chybových hodnot #N/A. Zahrnutím těchto hodnot do grafu nevznikne žádný problém, protože Excel ponechá počáteční oblast křivky prázdnou.

Na obrázku 17.19 jsme provedli formátování ve spíše prostém grafu vytvořeném nástrojem Klouzavý průměr. Více informací o grafech naleznete v kapitole 6.

Všimněte si, že v každé buňce obsahující hodnotu klouzavého průměru na obrázku 17.19 se v levém horním rohu zobrazuje příznak. Jedná se o příznak označující chybu; poté, co buňku vyberete, zobrazí se inteligentní nabídka s výstrahou, že vzorec nezahrnuje sousední buňky. V tomto případě je vše v pořádku. Chcete-li příznaky odstranit, vyberte všechny označené buňky, klepněte na značku a v zobrazené nabídce zvolte příkaz Ignorovat chybu.



Obrázek 17.19: Nástroj Klouzávy průměr nabízí lepší náhled na celkový tendenční vývoj.