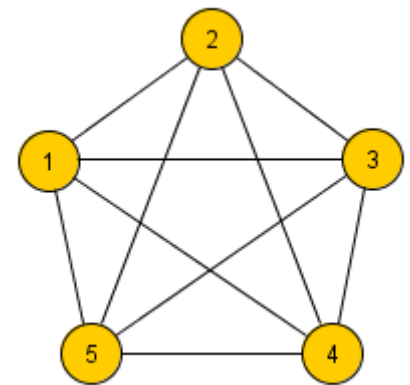
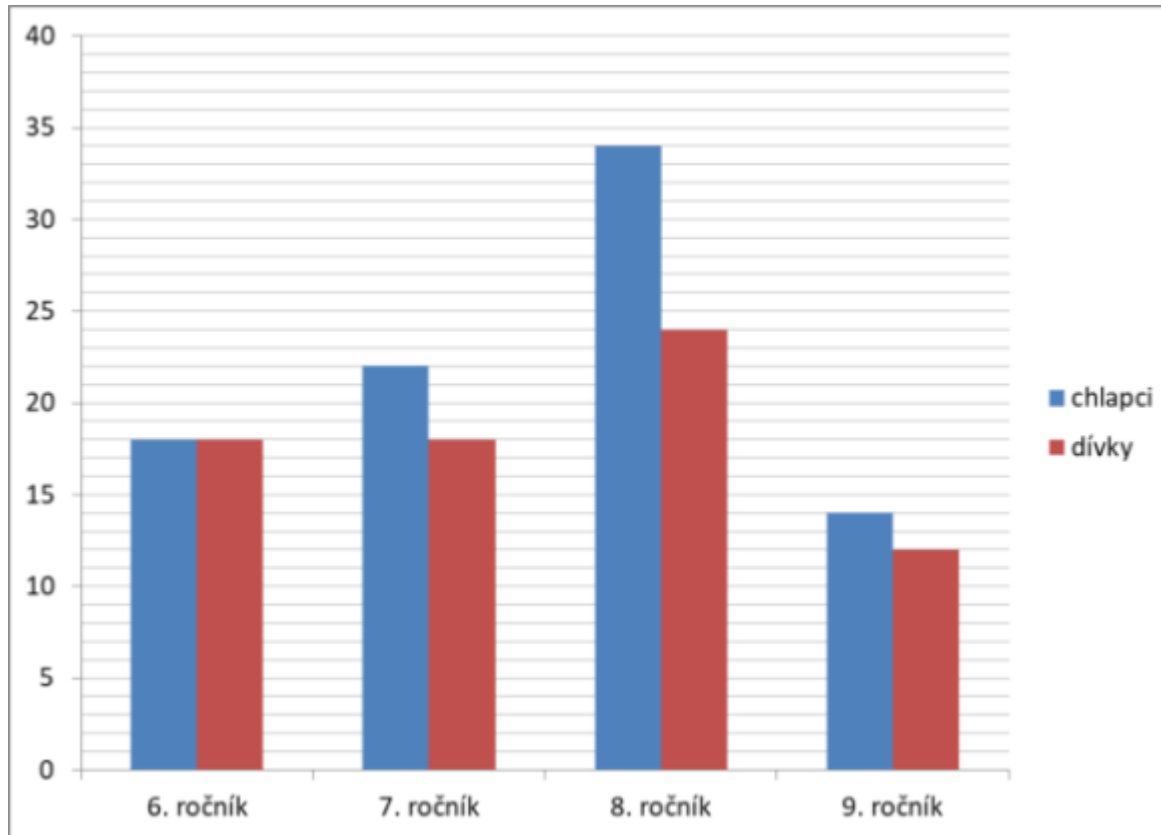


Statistické výpočty v Excelu



Milí studenti, milé studentky,
na dalších snímcích najdete podrobné návody a postupy jak vypočítat v Excelu všechny výpočty, o kterých jsme hovořili v metodologickém semináři.

Informace pro uklidnění a povzbuzení:

Výpočty t-testu, směrodatných odchylek, průměrů ani korelací nepočítáme „ručně“

Počítáme je **pomocí Excelu nebo jiného statistického programu** (např. na katedře učitelství je pro studenty k dispozici program SPSS, v navazujícím metodologickém semináři se zájemci naučí tyto i pokročilejší výpočty ve volně dostupném software JASP).

V našem kurzu se naučíme provádět výpočet průměru (\bar{x}), směrodatné odchylky (**SD**), **t-testu** a Pearsonova korelačního koeficientu (r) v Excelu, který máme všichni k dispozici.

Pokud máme číselné odpovědi z dotazníků přepsané v Excelu, je to otázka několika sekund.

Nejprve však potřebujeme vědět, JAK ČÍSELNÁ DATA Z DOTAZNÍKŮ PŘEPSAT DO EXCELU.

Při vysvětlování budeme vycházet ze cvičných dat naší absolventky učitelství, která ve své diplomové práci zkoumala, čeho si mladší a starší adolescenti cení u svých učitelů.

Přepis dotazníků do Excelu jako příprava pro statistické výpočty

EFA_Lída_učitelé - Microsoft Excel

Soubor Domů Vložení Rozložení stránky Vzorce Data Revize Zobrazení

Calibri 11 A⁺ A⁻ Zamakovat text

B *I* U Sloučit a zarovnat na střed

Obecný % 000 0,00 0,00

Podmíněné formátování Formátovat jako tabulku

Styly

O6 3

	B	C	D	E	F	G	H	I	J	K	L
	číslo dotazníku	pohlaví studenta (muž=1, žena=0)	věk (starší=1 mladší=0)	třída	škála 1-5 Laskavost	škála 1-5 Přísnost	škála 1-5 Humor	škála 1-5 Odborné znalosti	škála 1-5 Sympatie	škála 1-5 Přípravy na hodinu	škála 1-5 Důsled- nost
1	1	0	0	2	4	3	4	5	4	5	5
2	2	0	0	2	5	3	4	5	4	4	4
3	3	0	0	2	4	2	5	5	3	2	3
4	4	1	0	2	3	2	5	3	4	4	4
5	5	0	0	2	4	3	3	5	3	3	5
6	6	1	0	2	3	3	5	4	5	5	3
7	7	0	0	2	4	2	2	4	4	5	5
8	8	1	0	2	3	3	3	4	3	4	4
9	9	1	0	2	5	3	5	5	5	4	4
10	10	1	0	2	4	4	3	3	4	2	3
11	11	1	0	2	4	2	5	4	5	3	3
12	12	1	0	2	4	2	5	4	4	3	3
13	12	1	0	2	4	2	5	4	4	3	3

Komentář – viz další snímek

Komentář k předchozímu snímku č. 3 - jak se přepisují číselné a škálové odpovědi z dotazníků do Excelu

Autorka tohoto výzkumu zkoumala, **čeho si mladší a starší adolescenti cení u svých učitelů**. Na předchozím snímku máme ukázkou listu s jejími daty v Excelu.

Údaje od jednoho respondenta uvádíme v jednom řádku:

- začínáme **uvedením čísla dotazníku/respondenta** (viz **červená buňka** a pod ní první sloupec s čísly dotazníků od respondentů).
- Pak většinou následují validizační údaje, jako je **pohlaví, věk apod.** (Viz **zelené položky**.) Pohlaví je **dichotomní položka** (může mít jen dvě hodnoty). Muže ve statistice značíme číslem 1 a ženy=0. Autorka této studie vytvořila dichotomní položku i z věku, protože měla soubor mladších a starších adolescentů. (Ve své hypotéze předpokládala, že najde určité rozdíly v hodnocení učitelů mezi mladšími a staršími adolescenty). Mladším žákům ze ZŠ přiřadila nulu, starším ze SŠ jedničku.
- *(pro jistotu si také zaznamenala **číslo třídy**, odkud byl konkrétní respondent – sloupec 4. S touto proměnnou však ve statistických výpočtech nepracovala)*
- V dalších sloupcích následují **přepisy čísel odpovědí týkajících se vlastního výzkumu** (viz **modré položky**). Zde měli žáci a studenti **na škále 1-5** vyjádřit, jak moc je pro ně důležitá u učitele určitá vlastnost (*1 = zcela nedůležitá, 5 = velmi důležitá*).

Výpočet průměrné hodnoty ze sloupce čísel:

PRŮMĚR x ✓ f_x =PRŮMĚR(E2:E118)

A	B	C	D	E	F	G	H
číslo dotazníku	pohlaví studenta (muž=1, žena=0)	věk (starší=1 mladší=0)	třída	škála 1-5 Laskavost	škála 1-5 Přísnost	škála 1-5 Humor	škála 1-5 Odborné znalosti
110	1	1	5	5	3	4	4
111	1	1	5	3	3	4	3
112	1	1	5	5	3	3	4
113	1	1	5	3	3	5	4
114	0	1	5	3	3	4	4
115	0	1	7	4	3	3	5
116	1	1	7	2	3	4	5
117	1	1	7	5	3	4	5

=PRŮMĚR(E2:E118)

List1 List2 List3 CH-D starší_mladší FA-otázka 5 Faktorovka-motivy Faktorovka demotivace rozdíly k fakt|

Komentář k předchozímu snímku č. 5 - výpočet průměru

Pro jakékoli výpočty je nutné, **aby žádná hodnota ve sloupcích čísel nechyběla a aby mezi nimi nebyly prázdné buňky.**

Výpočet průměru, směrodatné odchylky a t-testu počítáme z hodnot u jedné proměnné, které jsou uvedené v jednom sloupci.

VÝPOČET PRŮMĚRU:

- Podíváme se, **jak je pro adolescenty v průměru důležitá např. „laskavost učitele“** (sloupec E)
- Zaznamenejme si názvy buněk, kde začínají a končí číselné údaje u proměnné „laskavost“. Údaje od prvního respondenta začínají na řádce č. **2**, údaje od posledního respondenta jsou na řádce **118**.
- Klikneme dolů ve stejném sloupci E do některé prázdné buňky pod daty od respondentů.
- Do této buňky (nebo nahoru do příkazového řádku) vložíme výpočetní formuli – viz červeně zakroužkovaná místa na obrázku.
- Formule pro průměr zní: **=PRŮMĚR(první buňka:poslední buňka)**
- V našem případě: **=PRŮMĚR(E2:E118)**
- Enter vypočítá hodnotu průměru. Ten vyšel $x' = 4,0$
- Na škále 1-5 je tedy laskavost učitelů od žáků a studentů velmi oceňovaná.

Stejné výpočty u proměnných ve vedlejších sloupcích

A	B	C	D	E	F	G	H
číslo dotazníku	pohlaví studenta (muž=1, žena=0)	věk (starší=1 mladší=0)	třída	škála 1-5 Laskavost	škála 1-5 Přísnost	škála 1-5 Humor	škála 1-5 Odborné znalosti
115	0	1	7	4	3	3	5
116	1	1	7	2	3	4	5
117	1	1	7	5	3	4	5
				4,017094	2,897436	3,96581	

VYCHYTÁVKA:

Chceme-li vypočítat stejný výpočet (průměr, směrodatnou odchylku nebo t-test) i pro vedlejší sloupce čísel vpravo od původního výpočtu, najedeme kurzorem na pravý dolní roh buňky, kde už takový výpočet máme, a protáhneme buňku doprava do dalších sloupců s proměnnými.

Formule vzorce se takto sama zkopíruje a provede stejný výpočet i pro tyto vedlejší sloupce čísel. (Zde je zkopírovaný výpočet průměru pro „laskavost“ ($x' = 4,0$) z předchozího snímku. Průměrná hodnota pro vedlejší „přísnost“ je mnohem nižší ($y' = 2,9$). Pro žáky a studenty je tedy důležitější laskavý učitel než učitel přísný.)

Výpočet směrodatné odchylky ze sloupce čísel:

The image shows a screenshot of an Excel spreadsheet. The formula bar at the top displays the formula `=SMODCH.VÝBĚR(E3:E119)`, which is circled in red. The spreadsheet contains a table with the following data:

A	B	C	D	E	F	G
číslo dotazníku	pohlaví studenta (muž=1, žena=0)	věk (starší=1, mladší=0)	třída	škála 1-5 Laskavost	škála 1-5 Přísnost	škála 1-5 Humor
115	0	1	7	4	3	3
116	1	1	7	2	3	4
117	1	1	7	5	3	4
				4,017094		

At the bottom of the spreadsheet, the formula `=SMODCH.VÝBĚR(E3:E119)` is entered in a cell, also circled in red. The spreadsheet interface includes tabs for 'List1', 'List2', 'List3', 'CH-D', 'starší_mladší', 'FA-otázka 5', 'Faktorovka-motivy', 'Faktorovka demotivace', and 'rozdíly k fakti'.

Komentář k předchozímu snímku č. 8 - výpočet směrodatné odchylky

Podíváme se, **jaká je směrodatná odchylka u významu „laskavost učitele“ pro adolescenty** (pořád sloupec E, stejné buňky jako u průměru).

- Klikneme ve stejném sloupci E do jiné prázdné buňky pod záznamy respondentů.
- Do této buňky (nebo nahoru do příkazového řádku) vložíme výpočetní formuli pro výpočet směrodatné odchylky.
- Formule pro výpočet směrodatné odchylky zní **=SMODCH.VÝBĚR(první buňka:poslední buňka)**
- V našem případě: =SMODCH.VÝBĚR(E2:E118)
- Enterem vypočítáme její hodnotu. Ta vyšla 0,9.

Informace nepovinná, navíc:

- *Můžeme tedy předpokládat, že přibližně 68 % (čili 2/3) respondentů uvedlo u vlastnosti „laskavost učitele“ hodnotu v rozmezí $x' - 0,9$ až $x' + 0,9$. Jejich odpovědi se tedy pohybovaly v rozmezí 3,1-4,9. Při zaokrouhlení na celá čísla, která respondenti museli volit v rámci dotazníkové hodnotící škály 1-5, tedy dvě třetiny respondentů z tohoto souboru zakroužkovaly hodnoty 3, 4 a 5.*

Výpočet t-testu ze dvou polovin sloupců čísel:

PRŮMĚR x fx =TTEST(E2:E73;E74:E118;2;2)

A	B	C	D	E	F	G	H
číslo dotazníku	pohlaví studenta (muž=1, žena=0)	věk (starší=1 mladší=0)	třída	škála 1-5 Laskavost	škála 1-5 Přísnost	škála 1-5 Humor	škála 1-5 Odborné znalosti
112	1	1	5	5	3	3	4
113	1	1	5	3	3	5	4
114	0	1	5	3	3	4	4
115	0	1	7	4	3	3	5
116	1	1	7	2	3	4	5
117	1	1	7	5	3	4	5

=TTEST(E2:E73;E74:E118;2;2)

List1 List2 List3 CH-D starší_mladší FA-otázka 5 Faktorovka-motivy Faktorovka demotivace rozdíly k fakti

Komentář k předchozímu snímku č. 10 - výpočet t-testu – 1. část

Výpočtem t-testu vyhodnocujeme, zda se liší dva podsoubory respondentů v určité proměnné. Můžeme se podívat, **zda se liší mladší a starší adolescenti v tom, jak je pro ně důležitá laskavost učitelů.**

- Budeme tedy pracovat (a) s půlkou sloupce odpovědí od mladších adolescentů (b) a s druhou půlkou sloupce odpovědí od starších adolescentů.
- Zda se jedná o mladší, nebo starší adolescenty, je uvedené ve třetím sloupečku čísla 0 (=mladší) a 1 (=starší).
- Nejdříve si zkontrolujeme, že máme respondenty seřazeny tak, že nejdříve jsou za sebou uvedeny odpovědi jen ode všech respondentů z mladší skupiny a poté od skupiny starší. Pokud to tak není, seřadíme je pomocí horní ikony „DATA“, kde v příkazu SEŘADIT vybereme příslušnou proměnnou, zde „věk“. Víme, že mladší respondenti zde mají hodnotu 0 a starší hodnotu 1.
- Podíváme se, ve kterém řádku končí údaje od mladších respondentů (=kde končí nuly ve sloupci „věk“). Je to řádek 73. Data od starších respondentů začínají hned v navazujícím řádku, tam kde začínají ve sloupci „věk“ jedničky, tedy řádek 74).
- Teď už jen dosadíme do výpočetní formule pro t-test pod sloupcem měřícím „laskavost u učitele“ zvlášť rozmezí buněk pro mladší skupinu a rozmezí buněk pro starší skupinu.
- Výpočetní formule pro t-test je **=TTEST(rozmezí buněk pro první podskupinu; rozmezí buněk pro druhou podskupinu;2;2).**
- V našem případě bude mít výpočetní formule pro t-test u laskavosti tuto podobu: **=TTEST(E2:E73;E74:E118;2;2).**
- Výsledek je 0,87.

- **Výpočet t-testu v Excelu** nepočítá hodnotu t , ale **rovnou ji převádí na statistickou významnost p** . Ta se odvozuje od dvou nám známých hladin: (1) pětiprocentní hladina významnosti, zkratkou 5% hl.v. (neboli $p < 0,05$) a přísnější jednaprocentní hladina významnosti, zkratkou 1% hl.v. (neboli $p < 0,01$).
- Výsledek $p=0,87$ je mnohem vyšší než obě hraniční hodnoty pro statistickou významnost. Musel by být naopak menší než $p=0,05$ (viz směr zobáčků v závorkách o řádek výše).
- Můžeme tedy konstatovat, že **„mladší a starší adolescenti se neliší v tom, jak je pro ně důležitá laskavost u jejich učitelů ($p=0,87$)“**.

V tuto chvíli Vás asi napadne, že předchozí výpočty průměru a směrodatné odchylky (které mají být v diplomkách a bakalářkách uváděny v tabulce popisné statistiky) jsme počítali pro všechny respondenty dohromady. Pokud jsme ale stanovili hypotézu, že se bude lišit význam určitých vlastností učitele u mladších a starších respondentů, pak nemá smysl uvádět průměry a odchylky za celý soubor, ale musíme je vypočítat zvlášť pro obě podskupiny.

(Při jejich výpočtu vycházíme ze stejného rozhraní buněk s údaji pro (a) mladší a (b) starší podskupiny respondentů jako při výpočtu rozdílů mezi nimi pomocí t-testu.)

Podobně můžeme provést i výpočet t-testu pro porovnání dívek a chlapců. Samozřejmě také počítáme průměry a směrodatné odchylky zvlášť pro dívky a chlapce.

-nejdříve si seřadíme dotazníky podle pohlaví pomocí horní ikony „**DATA – seřadit**“. Tentokrát v nabídce vybíráme druhý sloupec s proměnnou „**pohlaví**“.

- Dívky s hodnotou 0 pak budou seskupeny v horní části Excelového listu, chlapci s hodnotou 1 budou níže pod nimi. Opět si zaznamenáme, na kterém řádku končí dívčí dotazníky a že na dalším řádku pokračují chlapecké.
- T-test vypočítáme stejně, pouze ve výpočetní formuli upravíme rozmezí buněk podle podsouborů dívek a chlapců ve sloupci pro „laskavost“.

Komentář k předchozímu snímku č. 13 - výpočet Pearsonova korelačního koeficientu – 1. část

Výpočtem korelačního koeficientu vyhodnocujeme, **jak silný vztah existuje mezi dvěma nezávislými proměnnými**. Autorka těchto výzkumných dat si takovou hypotézu nestanovila. Její data však tento druh výpočtu umožňují.

Můžeme se podívat, **zda a jak moc souvisí požadavek, aby byl učitel laskavý, s požadavkem, aby byl přísný**. Obě tyto vlastnosti (přísnost a laskavost) bychom mohli považovat za poměrně protichůdné: pokud nějaký student vyžaduje učitelskou laskavost, pak by patrně nemusel chtít, aby byl učitel zároveň přísný. Mohli bychom tedy předpokládat, že vztah mezi nimi bude nepřímo úměrný. Nepřímo úměrnému (resp. zápornému) vztahu mezi oběma proměnnými by měl odpovídat výsledek korelačního koeficientu se zápornou hodnotou (se znaménkem „mínus“).

Obě proměnné – **laskavost** a **přísnost** – máme uvedeny ve sloupcích vedle sebe (viz obrázek). Tentokrát počítejme korelace z celého souboru respondentů:

Komentář k předchozímu snímku č. 15 - výpočet Pearsonova korelačního koeficientu – 2. část

Výpočetní formule pro Pearsonův korelační koeficient je **=PEARSON(rozmezí buněk pro první proměnnou;rozmezí buněk pro druhou proměnnou)**

V našem případě **=PEARSON(E2:E118;F2:F118)**

Tentokrát je jedno, do které z prázdných buněk pod uloženými daty od respondentů zapíšeme formuli.

Výsledek udávající hodnotu Pearsonova **korelačního koeficientu** (r) mezi požadavkem **učitelské laskavosti** a **učitelské přísnosti** vyšel $r = 0,02$. Výsledek není záporný, jak jsme předpokládali. Jde však o zcela mizivou hodnotu, která se blíží nule. Ve Wordové příloze máme orientační údaje o tom, jaká hodnota u korelačního koeficientu r by měla být statisticky významná. V případě 100 respondentů je tato hodnota $r = 0,2$. Počet našich respondentů je jen o málo vyšší než sto a r dosahuje pouze desetiny požadované velikosti.

Je tedy možné říci, že to, zda studenti preferují u učitelů laskavost, vůbec nezávisí na tom, zda u nich preferují i přísnost.

(pro zvědavé – ostatní korelační vztahy mezi jednotlivými proměnnými požadovaných učitelských vlastností dosahovaly vyšších kladných čísel. Např. mezi požadovanou učitelskou „přísností“ a „připraveností učitele na hodiny“ vyšlo $r = 0,4$. Tato hodnota je nepochybně statisticky významná. Čím víc chtějí žáci a studenti, aby byl učitel přísný, tím víc také požadují, aby se i on sám důkladně připravoval do svých hodin. To nám dává smysl 😊)

Výpočet Chí kvadrátu v Excelu

Ani Chí kvadrát není nutné počítat ručně. Lze si vytvořit výpočetní podklad a pak už jen dosazovat údaje do čtyřpolní tabulky.

Předdefinovaný výpočet najdete v Excelové příloze v Moodle .

Znalost ručního počítání Chí kvadrátu Vám pomůže zkontrolovat, zda se dopočítáte stejného výsledku a zda se někde „neukliknete“ v zadávání výpočtu v Excelu. Také je dobré vědět, co jste se vlastně dopočítali.

Pomocí Excelového výpočtu však můžete mít vyhodnocené otázky během několika minut.

Výpočet Chí kvadrátu v Excelu

	B	C	D	E
		vyhodnocovaná odpověď	ostatní odpovědi	kontrolní součet
	1.skupina	19	11	30
	2.skupina	21	4	25
		Chí kvadrát (χ^2)	2,936389	

V čtyřpolní tabulce v Listě 1:

Ve sloupci „**Vyhodnocovaná odpověď**“ uvádíme počty respondentů z obou porovnávaných podsouborů, kteří uvedli konkrétní variantu odpovědi, kterou vyhodnocujeme.

Do vedlejšího sloupce „ostatní odpovědi“ uvádíme počty zbylých respondentů z těchto podsouborů, kteří tuto variantu neuvedli.

- (kontrolní součty v krajním sloupci by měly odpovídat počtům respondentů v obou podsouborech).



Děkuji Vám za Vaši pozornost

