

Do ISu odevzdáváte písemné odpovědi v tomto dokumentu a output s výpočty.

1. Jaké metody posouzení normality rozdělení znáte (3 body)?

Matematicky:

- Kolmogorov-Smirnovův test
- Šikmost a špičatost

Graficky:

- Q-Q Plot
- Histogram

2. Otestujte, zda jde proměnná DCactNM (počet online aktivit, které respondent vykonává) normálně rozložená. Vložte tabulku s výsledkem testu a okomentujte ji. (2 body)

Testujeme a zamítáme nulovou hypotézu – rozložení počtu online aktivit není normální. Naznačuje to i histogram, k průměrným sedmi aktivitám počet postupně narůstá a naopak prudce klesá k sedmnácti – říkáme tomu rozložení zešikmené zprava.

Tests of Normality

	Kolmogorov-Smirnov ^a		
	Statistic	df	Sig.
DCactNM Number of online activities, TOTAL (17)	.076	25130	.000

a. Lilliefors Significance Correction

3. Vykonávají čeští (country) chlapci (sexchild) více online aktivit (DCactNM) než dívky? A můžeme takový výsledek očekávat i v populaci českých dětí? Vložte tabulky s výsledky testu a okomentujte je. (5 bodů)

Nejprve je potřeba vybrat pouze české respondenty pomocí SELECT CASES, a poté provést (Independent Samples) t-test a testovat nulovou hypotézu o shodě populačních průměrů. Napovědět by Vám mělo, že nezávislá proměnná je dichotomická (nominální) a závislá proměnná kardinální.

Z výsledků vyplývá, že čeští chlapci vykonávají v průměru o 0,3 aktivity více než dívky, což není příliš velký rozdíl. Tabulku s výsledky t-testu jsem transponovala, aby se vešla na stránku. Výsledek čteme z druhého sloupce (Equal variances not assumed) a vidíme, že nelze nulovou hypotézu o shodě průměru chlapců a dívek zamítnout – tedy se populační průměry neliší.

Group Statistics

	sexchild Child gender	N	Mean	Std. Deviation	Std. Error Mean
DCactNM Number of online activities, TOTAL (17)	1.00 Male	493	8.30	3.743	.169
	2.00 Female	516	8.00	3.255	.143

Independent Samples Test

		DCactNM Number of online activities, TOTAL (17)	
		Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F	16.660	
	Sig.	.000	
t-test for Equality of Means	t	1.362	1.358
	df	1007	974.097
	Sig. (2-tailed)	.174	.175
	Mean Difference	.300	.300
	Std. Error Difference	.221	.221
95% Confidence Interval of the Difference	Lower	-.132	-.134
	Upper	.733	.735

4. **Máme hypotézu, že děti z rodin s nižším vzděláním budou častěji problémové, resp. že se vzděláním rodičů bude klesat míra problémů dítěte. Otestujte ji, vložte tabulky s výsledky a okomentujte je. Pracujte s dětmi ze všech zemí. (5 bodů). DCSDQcpMN – škála conduct problems, vyšší skóre znamená více problémů; DPEDUHH4 – nejvyšší dosažené vzdělání v domácnosti.**

Nejprve je třeba vypnout filtr z předchozí úlohy, opět pomocí SELECT CASES. A poté pomocí ANOVY (nezávislá je kategorizovaná ordinální proměnná a má více než dvě kategorie, závislá proměnná je kardinální) opět testujeme hypotézu o shodě populačních průměrů.

Descriptives

DCSDQcpMN SDQ CONDUCT PROBLEMS (MEAN)

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
▶ 1.00 ISCED 1; primary or less	1862	1.4294	.35807	.00830	1.4131	1.4457	1.00	3.00
2.00 ISCED 2; lower secondary	4203	1.4018	.32321	.00499	1.3921	1.4116	1.00	3.00
3.00 ISCED 3/4; upper and post-secondary	11517	1.4108	.31068	.00289	1.4051	1.4165	1.00	3.00
4.00 ISCED 5/6; tertiary	7440	1.3608	.29544	.00343	1.3540	1.3675	1.00	2.80
Total	25022	1.3958	.31307	.00198	1.3919	1.3997	1.00	3.00

Z tabulky Descriptives vidíme, že děti z rodin z nejnižším vzděláním mají nejvíce problémů a naopak, ale se nejedná o lineární vztah, neboť děti ze třetí vzdělanostní kategorie mají více problémů než děti z druhé, což bychom nejspíš čekali naopak. Můžeme se nicméně pomocí ANOVY podívat, jestli jsou rozdíly v průměrech statisticky významné.

Nejprve provedeme test homogenity rozptylu.

Test of Homogeneity of Variances

DCSDQcpMN SDQ CONDUCT PROBLEMS (MEAN)

Levene Statistic	df1	df2	Sig.
28.274	3	25018	.000

A poté přistoupíme k analýze. Z tabulky ANOVA vyplývá, že můžeme zamítnout nulovou hypotézu o shodě průměrů – mezi proměnnými je tedy vztah a některé vzdělanostní kategorie se budou zřejmě statisticky významně odlišovat.

ANOVA

DCSDQcpMN SDQ CONDUCT PROBLEMS (MEAN)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	13.981	3	4.660	47.813	.000
Within Groups	2438.426	25018	.097		
Total	2452.406	25021			

Abychom zjistili které, musíme provést post hoc test (ideálně Bonferonniho, který se nejsnáze intrpretuje, neboť označuje rozdíly hvězdičkami). Z tabulky jsou patrné statisticky významné rozdíly mezi všemi vzdělanostními kategoriemi s výjimkou první a třetí kategorie a druhé a třetí kategorie.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: DCSDQcpMN SDQ CONDUCT PROBLEMS (MEAN)

Bonferroni

(I) DPEDUHH4 EDUCATION - HIGHEST IN HOUSEHOLD (4)	(J) DPEDUHH4 EDUCATION - HIGHEST IN HOUSEHOLD (4)	Mean Difference (I- J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00 ISCED 1; primary or less	2.00 ISCED 2; lower secondary	.02756*	.00869	.009	.0046	.0505
	3.00 ISCED 3/4; upper and post-secondary	.01862	.00780	.102	-.0020	.0392
	4.00 ISCED 5/6; tertiary	.06865*	.00809	.000	.0473	.0900
2.00 ISCED 2; lower secondary	1.00 ISCED 1; primary or less	-.02756*	.00869	.009	-.0505	-.0046
	3.00 ISCED 3/4; upper and post-secondary	-.00895	.00563	.670	-.0238	.0059
	4.00 ISCED 5/6; tertiary	.04108*	.00602	.000	.0252	.0570
3.00 ISCED 3/4; upper and post-secondary	1.00 ISCED 1; primary or less	-.01862	.00780	.102	-.0392	.0020
	2.00 ISCED 2; lower secondary	.00895	.00563	.670	-.0059	.0238
	4.00 ISCED 5/6; tertiary	.05003*	.00464	.000	.0378	.0623
4.00 ISCED 5/6; tertiary	1.00 ISCED 1; primary or less	-.06865*	.00809	.000	-.0900	-.0473
	2.00 ISCED 2; lower secondary	-.04108*	.00602	.000	-.0570	-.0252
	3.00 ISCED 3/4; upper and post-secondary	-.05003*	.00464	.000	-.0623	-.0378

*. The mean difference is significant at the 0.05 level.

5. Změřte sílu vztahu mezi následujícími dvojicemi proměnných, vyberte správný koeficient asociace, výsledek okomentujte s ohledem na těsnost (sílu) vztahu, zobecnitelnost na základní soubor a smyslnost takové úvahy. (12 bodů)

- Čas denně strávený online (*DCtimeuse*) a počet online dovedností (*DCskillsNM*).
- Pohlaví dítěte (*sexchild*) a přesvědčení, že je zkušenější uživatel/ka internetu než rodiče (*DCwebableA*).
- Vzdělání rodiče, který poskytl rozhovor (*DPPTEDUA4*), a jeho sebejistota při používání internetu (*QP218*).
- Počet přátel na sociálních sítích (*QC316*) a míra problémů s vrstevníky (*DCSDQppMN*).

V této otázce šlo ve všech případech o koeficienty asociace – to je jediný způsob, jak změřit sílu vztahu.

- Obě proměnné jsou kardinální, použijeme proto Pearsonův korelační koeficient, vypočteme i Spearmana, abychom získali představu o linearitě vztahu. Oba koeficienty se příliš neliší, není tedy třeba pochybovat o linearitě vztahu. Hodnota $r=0,32$, ($p=0,000$) naznačuje, že s časem stráveným online rostou i dovednosti, naopak by to moc nedávalo smysl, hodnota r naznačuje středně těsný vztah (na sociální vědy dobrý výsledek), který lze navíc podle hodnoty p /sig. očekávat i v populaci.

Correlations

		Dctimeuse Estimated minutes online each day	DCskillsNM Number of skills, TOTAL (8)
Dctimeuse Estimated minutes online each day	Pearson Correlation	1	.318**
	Sig. (2-tailed)		.000
	N	24548	18211
DCskillsNM Number of skills, TOTAL (8)	Pearson Correlation	.318**	1
	Sig. (2-tailed)	.000	
	N	18211	18594

** . Correlation is significant at the 0.01 level (2-tailed).

Correlations

			Dctimeuse Estimated minutes online each day	DCskillsNM Number of skills, TOTAL (8)
Spearman's rho	Dctimeuse Estimated minutes online each day	Correlation Coefficient	1.000	.338**
		Sig. (2-tailed)	.	.000
		N	24548	18211
	DCskillsNM Number of skills, TOTAL (8)	Correlation Coefficient	.338**	1.000
		Sig. (2-tailed)	.000	.
		N	18211	18594

** . Correlation is significant at the 0.01 level (2-tailed).

- b. Závislá proměnná je ordinální (3 kategorie) a nezávislá proměnná je nominální (ale dichotomická) – můžeme tedy použít koeficienty pro ordinální proměnné. Použijeme Somersovo D (pro asymetrické vztahy. Výsledek čteme z druhého řádku (znalost ve srovnání s rodiči je závislou proměnnou) – záporná hodnota naznačuje, že dívky si budou důvěřovat méně než chlapci, zároveň se však hodnota blíží nule, tedy vztah mezi proměnnými je zanedbatelný (byť zobecnitelný, což nám ale nijak

Directional Measures

			Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Ordinal by Ordinal	Somers' d	Symmetric	-.035	.006	-5.816	.000
		DCwebableI know more about the internet than my parents (9-16) Dependent	-.040	.007	-5.816	.000
		sexchild Child gender Dependent	-.030	.005	-5.816	.000

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

nepomůže).

- c. Obě pro měnné jsou ordinální, podobně jako v předchozím případě můžeme použít Somersovo D – opět čteme výsledek ze druhého řádku – se vzděláním roste i jistota rodiče při používání internetu, ačkoli vztah není příliš silný (na sebejistotu mají totiž vliv i mnohé jiné faktory), můžeme jej očekávat i

v základním souboru.

Directional Measures

			Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Ordinal by Ordinal	Somers' d	Symmetric	.176	.006	28.735	.000
		QP218 How confident are you in using the internet? Dependent	.172	.006	28.735	.000
		DPPTEDUA4 EDUCATION - PARENT INTERVIEWEE (4) Dependent	.181	.006	28.735	.000

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

- d. Počet přátel měříme na ordinální škále, problémy s vrstevníky na kardinální – opět tedy použijeme koeficient pro ordinální proměnné. Tentokrát však není tak jisté, která proměnná je závislá a která nezávislá (nízký počet online přátel může být výsledkem konfliktů, konflikty ale také mohou být způsobeny právě vyloučením z kolektivu) – použijeme tedy koeficient pro symetrické vztahy: buď Kendalovo tau_c nebo symetrického Somerse. Oba koeficienty jsou podobně nízké, naznačují zanedbatelnou souvislost, byť zobecnitelnou.

Directional Measures

			Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Ordinal by Ordinal	Somers' d	Symmetric	-.072	.007	-10.918	.000
		DCSDQppMN SDQ PEER PROBLEMS (MEAN) Dependent	-.073	.007	-10.918	.000
		QC316 Roughly how many people are you in contact with when using [social networking profile]? Dependent	-.071	.006	-10.918	.000

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Symmetric Measures

		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-c	-.071	.006	-10.918	.000
N of Valid Cases		14805			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

6. Pro vztah mezi vzděláním rodiče, který poskytl rozhovor (DPPTEDUA4), a jeho sebejistotou při používání internetu (QP218) vytvořte kontingenční tabulku a upravte ji pro publikování (musíte se správně rozhodnout pro řádková/sloupcová/total procenta). Otestujte, zda je mezi proměnnými vztah. Výsledek okomentujte. (8 bodů)

Nezávislou proměnnou (vzdělání) máme v řádcích, proto použijeme řádková procenta – ta se pak interpretují ve sloupcích. Vidíme například, že se vzděláním jednoznačně roste podíl rodičů, kteří jsou si velmi jistí. Test chí-kvadrát naznačuje existenci vztahu (zamítáme nulovou hypotézu). Níže najdete výstup z SPSS a úplně na

konci, jak by měla v ideálním případě vypadat tabulka pro publikování ve Vaší bakalářské práci (musí být jasné, kde je 100%, velikost vzorku, zdroj a výsledek statistického testu, tabulka je v češtině, očíslovaná a opatřená titulkem).

DPPTEDUA4 EDUCATION - PARENT INTERVIEWEE (4) * QP218 How confident are you in using the internet? Crosstabulation

% within DPPTEDUA4 EDUCATION - PARENT INTERVIEWEE (4)

		QP218 How confident are you in using the internet?				Total
		1 Not at all confident	2 Not very confident	3 Fairly confident	4 Very confident	
DPPTEDUA4 EDUCATION - PARENT INTERVIEWEE (4)	1.00 ISCED 1; primary or less	7.0%	30.9%	44.8%	17.3%	100.0%
	2.00 ISCED 2; lower secondary	4.7%	22.3%	51.7%	21.3%	100.0%
	3.00 ISCED 3/4; upper and post-secondary	3.1%	20.3%	52.8%	23.8%	100.0%
	4.00 ISCED 5/6; tertiary	1.6%	9.8%	49.6%	39.0%	100.0%
Total		3.2%	18.4%	51.2%	27.2%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	933.847 ^a	9	.000
Likelihood Ratio	930.942	9	.000
Linear-by-Linear Association	769.101	1	.000
N of Valid Cases	20127		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 39.25.

Tabulka 1. Sebejistota rodičů při používání internetu podle vzdělání v % (Zdroj: EU Kids Online).

	Vůbec nejistý	Ne moc jistý	Docela jistý	Velmi jistý	Celkem	N
Základní a nižší vzdělání	7.0%	30.9%	44.8%	17.3%	100.0%	1219
Střední bez maturity	4.7%	22.3%	51.7%	21.3%	100.0%	3784
Střední s maturitou	3.1%	20.3%	52.8%	23.8%	100.0%	9535
Vyšší odborné nebo vysokoškolské vzdělání	1.6%	9.8%	49.6%	39.0%	100.0%	5589
Celkem	3.2%	18.4%	51.2%	27.2%	100.0%	20127

* $\chi^2(9, N=20127) = 933,85, p < .01$