Vzorové řešení seminárního úkolu

(do předmětu M6130 Základní statistické metody)

Vypracoval: Bc. Miroslav Moser

<u>Text zadání:</u> V rámci psychologického výzkumu byly u 856 dětí ze základních škol zjišťovány následující údaje:

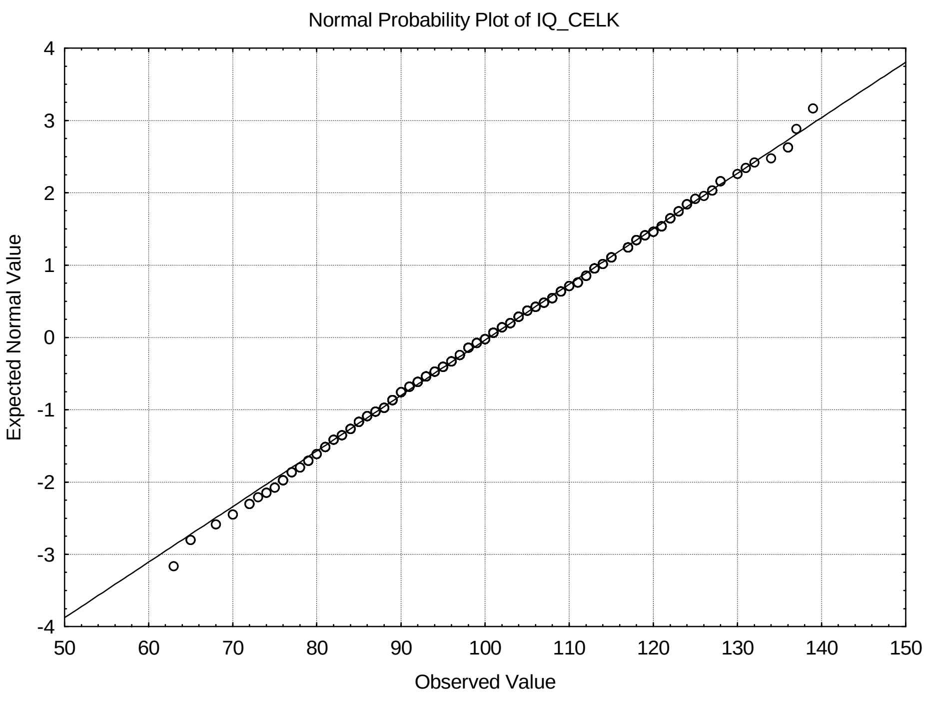
Pohlaví (1 – chlapec, 2 – dívka) – proměnná SEX
IQ verbální – proměnná IQ_VERB
IQ performační – proměnná IQ_PERF
IQ celkové – proměnná IQ_CELK
Třída (1. až 9.) – proměnná TRIDA
Vzdělání matky (1 – základní, 2 – SŠ, 3 – VŠ) – proměnná VZDEL_M
Vzdělání otce (1 – základní, 2 – SŠ, 3 – VŠ) – proměnná VZDEL_O
Sídlo (1 – město, 2 – venkov) – proměnná SIDLO

1. *Sestrojte 95% interval spolehlivosti pro střední hodnotu proměnné IQ_CELK, a to*
   a) *pro všechny děti*
   b) *pro chlapce*
   c) *pro dívky*
   d) *pro městské děti*
   e) *pro venkovské děti.*

ad a) Nejprve ověříme pomocí Kolmogorov-Smirnovova testu, zda data - hodnoty proměnné IQ_CELK, zkoumané u všech dětí, pochází z normálního rozložení.

Variable	Tests of Normality IQ_CELK N	max D	Lilliefors p
IQ_CELK	856	0,037557	p < ,01

Hodnota testové statistiky je d=0,03756 a příslušná modifikovaná kritická hodnota je p<0,01. K-S test zamítá hypotézu o normalitě dat na hladině významnosti 0,05. Sestrojíme Normal-Probability plot a provedeme vizuální posouzení normality.



Z grafu vidíme, že normalita dat (pořízených z rozsáhlého souboru 856 hodnot) je porušena jen mírně. Data tedy dále považujme za normální.

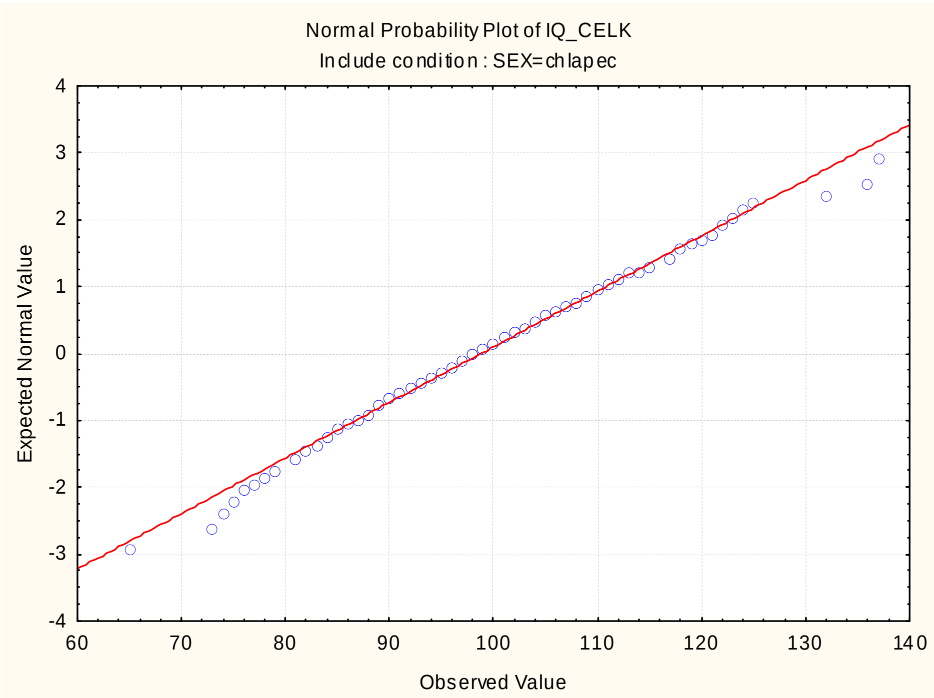
Variable	Descriptive Statistics IQ_CELK Confidence -95,000%	Confidence +95,000%
IQ_CELK	99,58007	101,3195

Střední hodnota proměnné IQ_CELK pro všechny děti leží s pravděpodobností 95% v intervalu (99,58007; 101,3195).

ad b) Ověříme pomocí Kolmogorov-Smirnovova testu, zda data - hodnoty proměnné IQ_CELK, zkoumané u chlapců, pochází z normálního rozložení.

Variable	Tests of Normality IQ_CELK Include condition: SEX=chlapec N	max D	Lilliefors p
IQ_CELK	426	0,045826	p < ,05

Hodnota testové statistiky je d=0,045826 a příslušná modifikovaná kritická hodnota je p<0,05. K-S test zamítá hypotézu o normalitě dat na hladině významnosti 0,05. Sestrojíme Normal-Probability plot a provedeme vizuální posouzení normality.



Z grafu vidíme, že normalita dat (pořízených z rozsáhlého souboru 426 hodnot) je porušena jen mírně. Data tedy dále považujme za normální.

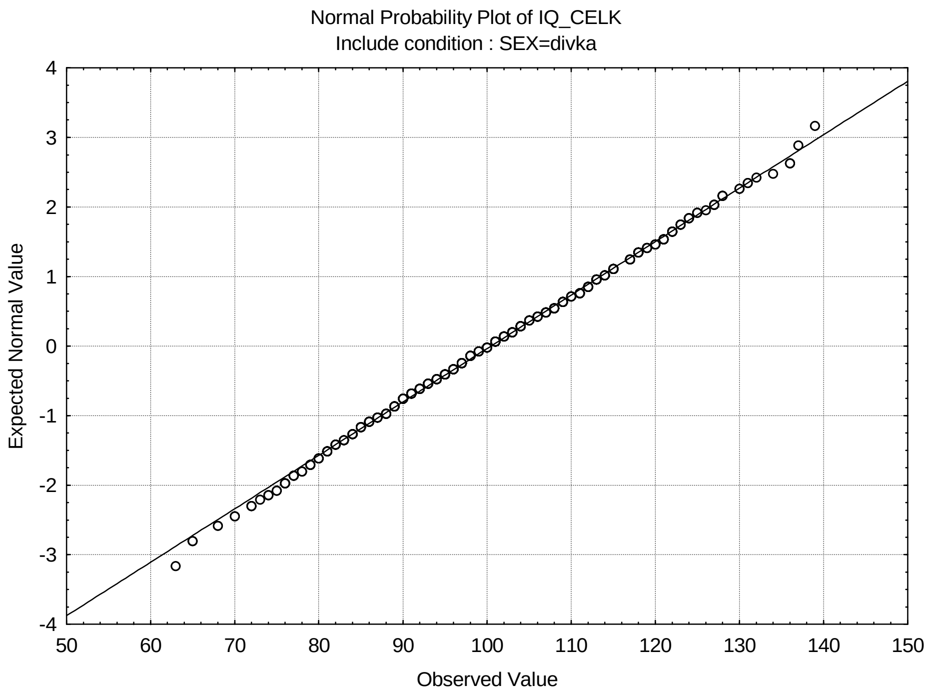
Variable	Descriptive Statistics IQ_CELK Include condition: SEX=chlapec Confidence -95,000%	Confidence +95,000%
IQ_CELK	100,7036	103,2730

Střední hodnota proměnné IQ_CELK pro chlapce leží s pravděpodobností 95% v intervalu (100,7036; 103,2730).

ad c) Ověříme pomocí Kolmogorov-Smirnovova testu, zda data - hodnoty proměnné IQ_CELK, zkoumané u dívek, pochází z normálního rozložení.

Variable	Tests of Normality IQ_CELK Include condition: SEX=divka N	 max D	 Lilliefors p
IQ_CELK	430	0,034964	p > .20

Hodnota testové statistiky je d=0,034964 a příslušná modifikovaná kritická hodnota je p>0,20. K-S test nezamítá hypotézu o normalitě dat na hladině významnosti 0,05. Sestrojíme Normal-Probability plot a provedeme vizuální posouzení normality.



Graf podporuje normalitu dat.

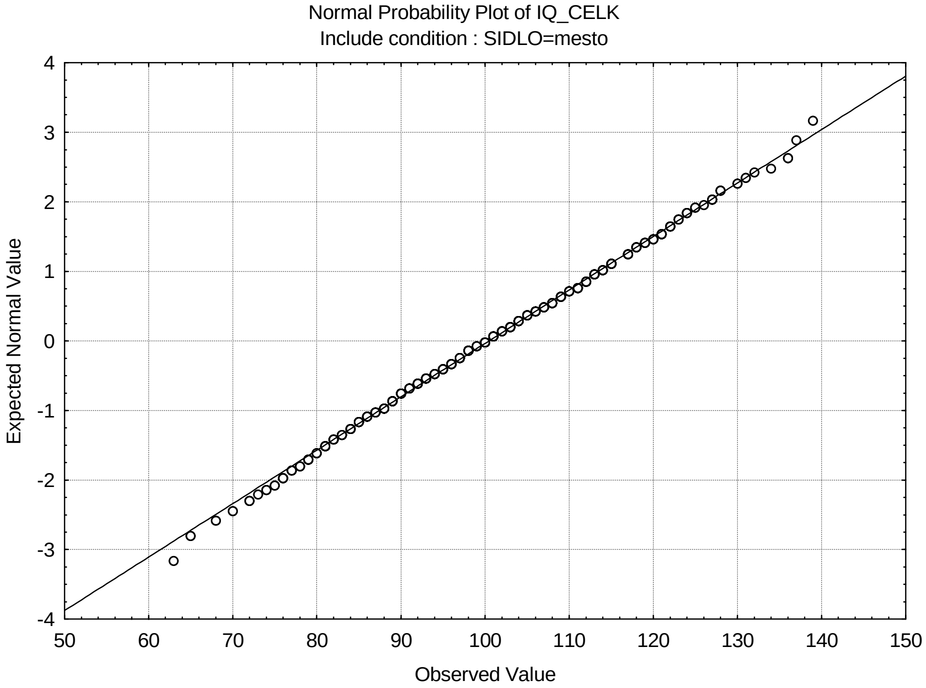
Variable	Descriptive Statistics IQ_CELK Include condition: SEX=divka Confidence -95,000%	 Confidence +95,000%
IQ_CELK	97,76473	100,0864

Střední hodnota proměnné IQ_CELK pro dívky leží s pravděpodobností 95% v intervalu (97,76473; 100,0864).

ad d) Ověříme pomocí Kolmogorov-Smirnovova testu, zda data - hodnoty proměnné IQ_CELK, zkoumané u městských dětí, pochází z normálního rozložení.

Variable	Tests of Normality IQ_CELK Include condition: SIDLO=mesto N	 max D	 Lilliefors p
IQ_CELK	473	0,042267	p < ,05

Hodnota testové statistiky je d=0,042267 a příslušná modifikovaná kritická hodnota je p<0,05. K-S test zamítá hypotézu o normalitě dat na hladině významnosti 0,05. Sestrojíme Normal-Probability plot a provedeme vizuální posouzení normality.



Z grafu vidíme, že normalita dat (pořízených z rozsáhlého souboru 473 hodnot) je porušena jen mírně. Data tedy dále považujme za normální.

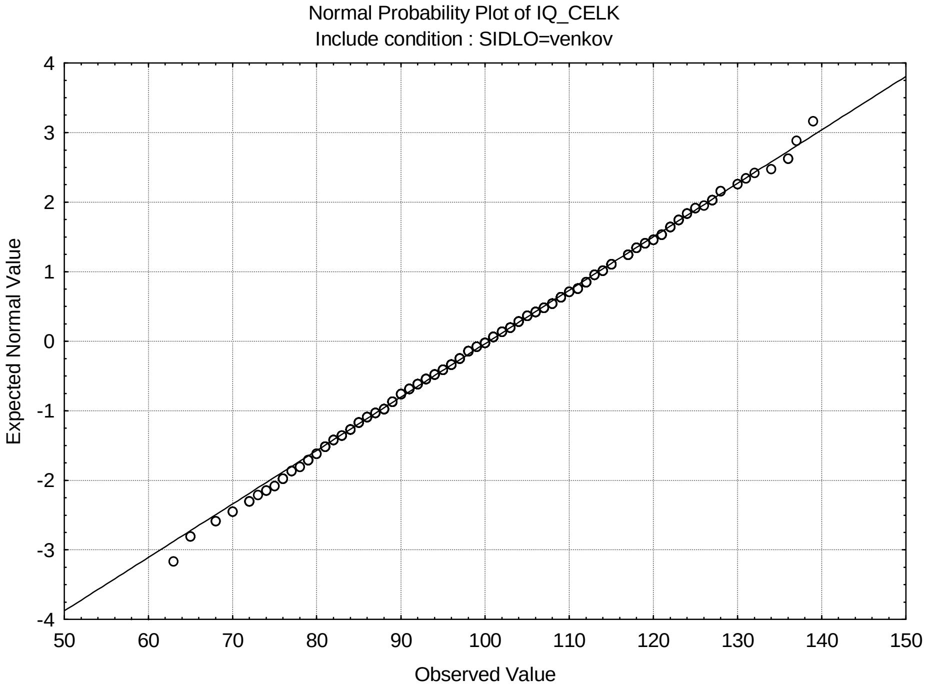
Variable	Descriptive Statistics IQ_CELK Include condition: SIDLO=mesto Confidence -95,000%	Confidence +95,000%
IQ_CELK	100,5672	103,0269

Střední hodnota proměnné IQ_CELK pro městské děti leží s pravděpodobností 95% v intervalu (100,5672; 103,0269).

ad e) Ověříme pomocí Kolmogorov-Smirnovova testu, zda data - hodnoty proměnné IQ_CELK, zkoumané u venkovských dětí, pochází z normálního rozložení.

Variable	Tests of Normality IQ_CELK Include condition: SIDLO=venkov N	max D	Lilliefors p
IQ_CELK	383	0,048465	p < ,05

Hodnota testové statistiky je d=0,059565 a příslušná modifikovaná kritická hodnota je p<0,05. K-S test zamítá hypotézu o normalitě dat na hladině významnosti 0,05. Sestrojíme Normal-Probability plot a provedeme vizuální posouzení normality.



Z grafu vidíme, že normalita dat (pořízených z rozsáhlého souboru 383 hodnot) je porušena jen mírně. Data tedy dále považujme za normální.

Variable	Descriptive Statistics IQ_CELK Include condition: SIDLO=venkov Confidence -95,000%	Confidence +95,000%
IQ_CELK	97,58778	99,98402

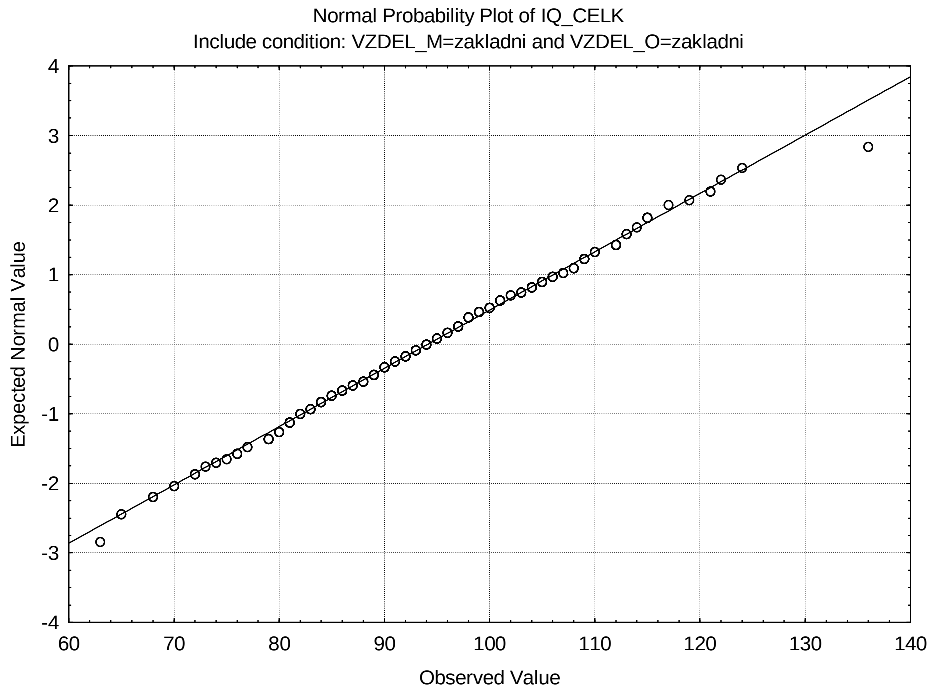
Střední hodnota proměnné IQ_CELK pro venkovské děti leží s pravděpodobností 95% v intervalu (97,58779; 99,98402).

2. *Sestrojte 95% interval spolehlivosti pro rozdíl středních hodnot proměnné IQ_CELK, kde první skupinu tvoří děti, jejichž oba rodiče mají základní vzdělání a druhou skupinu tvoří děti, jejichž oba rodiče mají vysokoškolské vzdělání.*

Nejprve ověříme pomocí Kolmogorov-Smirnovova testu, zda data – hodnoty proměnné IQ_CELK pro děti, jejichž oba rodiče mají základní vzdělání, pochází z normálního rozložení.

Variable	Tests of Normality IQ_CELK Include condition: VZDEL_M=zakladni and VZDEL_O=zakladni N	max D	Lilliefors p
IQ_CELK	296	0,044311	p < ,20

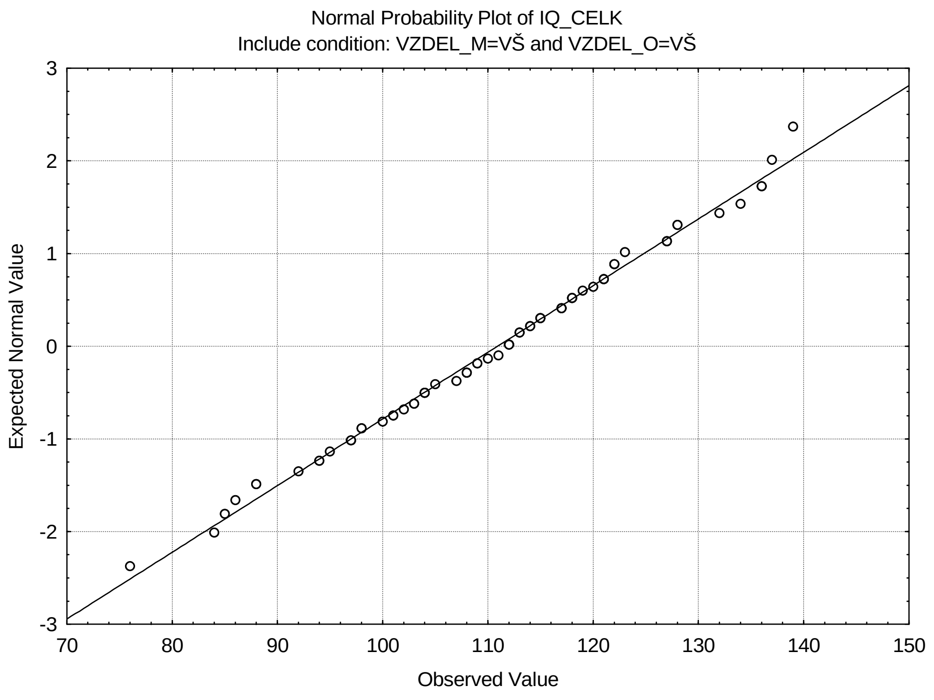
Hodnota testové statistiky je d=0,04431 a příslušná modifikovaná kritická hodnota je p<0,20. K-S test nezamítá hypotézu o normalitě dat na hladině významnosti 0,05. Grafické znázornění pomocí Normal-Probability plotu normalitu dat podporuje.



Dále ověříme pomocí Kolmogorov-Smirnovova testu, zda data – hodnoty proměnné IQ_CELK pro děti, jejichž oba rodiče mají vysokoškolské vzdělání, pochází z normálního rozložení.

Variable	Tests of Normality IQ_CELK Include condition: VZDEL_M=VŠ and VZDEL_O=VŠ N	max D	Lilliefors p
IQ_CELK	75	0,065367	p > .20

Hodnota testové statistiky je d=0,06537 a příslušná modifikovaná kritická hodnota je p>0,20. K-S test nezamítá hypotézu o normalitě dat na hladině významnosti 0,05. Grafické znázornění pomocí Normal-Probability plotu normalitu dat podporuje.



Ověříme předpoklad o shodě rozptylů (Leveneův test).

Variable	Levene Test of Homogeneity of Variances (zadání) Marked effects are significant at p < ,05000							
	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
IQ_CELK	113,5253	1	113,5253	20144,00	369	54,59077	2,079570	0,150130

Hodnota testové statistiky je F=2,0796 a p-hodnota je p=0,15>0,05 – na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů IQ_CELK pro děti, jejichž rodiče mají základní vzdělání a pro děti, jejichž rodiče mají vysokoškolské vzdělání. Vypočteme výběrové průměry a směrodatné odchylky pro obě skupiny.

Breakdown Table of Descriptive Statistics (zadání) N=371 (No missing data in dep. var. list)			
P3	IQ_CELK Means	IQ_CELK N	IQ_CELK Std.Dev.
1	94,1385	296	11,82604
2	110,9067	75	13,60164
All Grps	97,5283	371	13,92766

Ty dosadíme do vzorce pro výpočet 95% intervalu spolehlivosti pro rozdíl středních hodnot.

	DM	HM
1	-19,8702256	-13,6661544

Rozdíl středních hodnot proměnné IQ_CELK skupiny dětí, jejichž oba rodiče mají základní vzdělání a skupiny dětí, jejichž rodiče mají vysokoškolské vzdělání leží s pravděpodobností 95% v intervalu (-19,87;-13,67). Z uvedeného vztahu vyplývá, že vyšší IQ_CELK mají děti, jejichž oba rodiče mají VŠ vzdělání.

3. *Na hladině významnosti 0,05 testujte hypotézu, že se neliší střední hodnota IQ_CELK*
   a) *chlapců a dívek*
   b) *městských a venkovských dětí.*

   Pro obě situace nakreslete krabicové diagramy.

Normalitu dat jsme ověřili již v úkolu číslo jedna.

ad a) Ověříme předpoklad shody rozptylů (Leveneův test).

Variable	Levene Test of Homogeneity of Variances (zadání) Marked effects are significant at p < ,05000							
	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
IQ_CELK	280,8677	1	280,8677	48100,38	854	56,32363	4,986677	0,025802

Hodnota testové statistiky je F=4,986677 a p-hodnota je p=0,025802<0,05 Na hladině významnosti 0,05 zamítáme hypotézu o shodě rozptylů IQ_CELK pro chlapce a dívky. Provedeme dvouvýběrový T-test (pro případ, kdy se rozptyly liší).

Variable	T-tests; Grouping: SEX (zadání) Group 1: CHLAPCI Group 2: DIVKY							
	Mean CHLAPCI	Mean DIVKY	t-value	df	p	t separ. var.est.	df	p 2-sided
IQ_CELK	101,9883	98,92558	3,478263	854	0,000530	3,476696	844,5692	0,000534

Hodnota testové statistiky pro t-test shody rozptylů je t=3,476696 a p-hodnota je p=0,000534<0,05. Na hladině významnosti 0,05 zamítáme hypotézu o shodě rozptylů IQ_CELK pro chlapce a pro dívky.

ad b) Ověříme předpoklad shody rozptylů (Leveneův test).

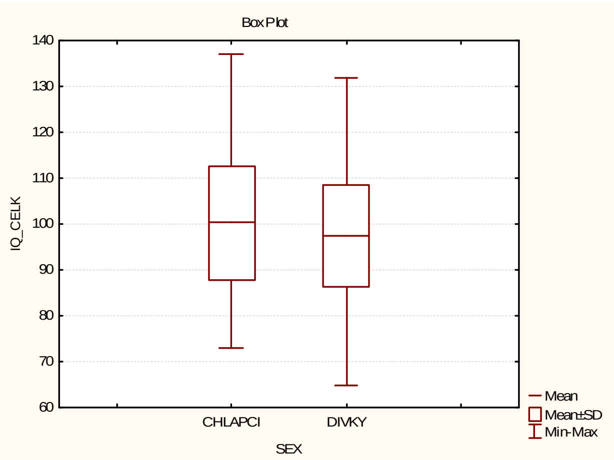
Variable	Levene Test of Homogeneity of Variances (zadání) Marked effects are significant at p < ,05000 SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
IQ_CELK	496,0439	1	496,0439	49200,90	854	57,61230	8,610034	0,003433

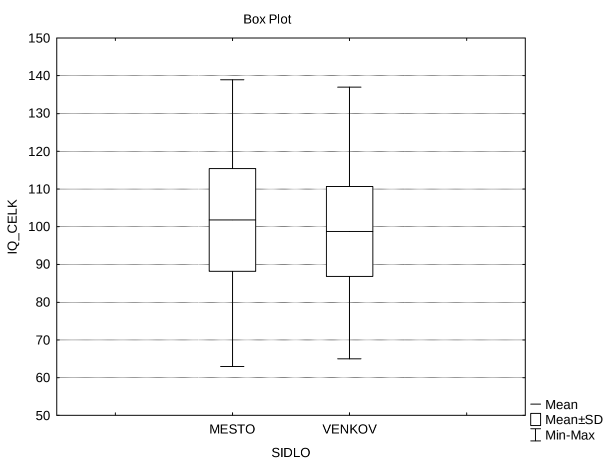
Hodnota testové statistiky je F=8,610034 a p-hodnota je p=0,003433<0,05 Na hladině významnosti 0,05 zamítáme hypotézu o shodě rozptylů IQ_CELK pro městské a venkovské děti. Provedeme dvouvýběrový T-test (pro případ, kdy se rozptyly liší).

Variable	T-tests; Grouping: SEX (zadání) Group 1: CHLAPCI Group 2: DIVKY Mean CHLAPCI	Mean DIVKY	t-value	df	p	t separ. var.est.	df	p 2-sided
IQ_CELK	101,7970	98,78590	3,399750	854	0,000706	3,447138	848,6899	0,000594

Hodnota testové statistiky pro t-test shody rozptylů je t=3,447138 a p-hodnota je p=0,000594<0,05. Na hladině významnosti 0,05 zamítáme hypotézu o shodě rozptylů IQ_CELK pro městské a venkovské děti

Krabicové diagramy podporují zamítnutí hypotézy o shodě středních hodnot pro oba případy.



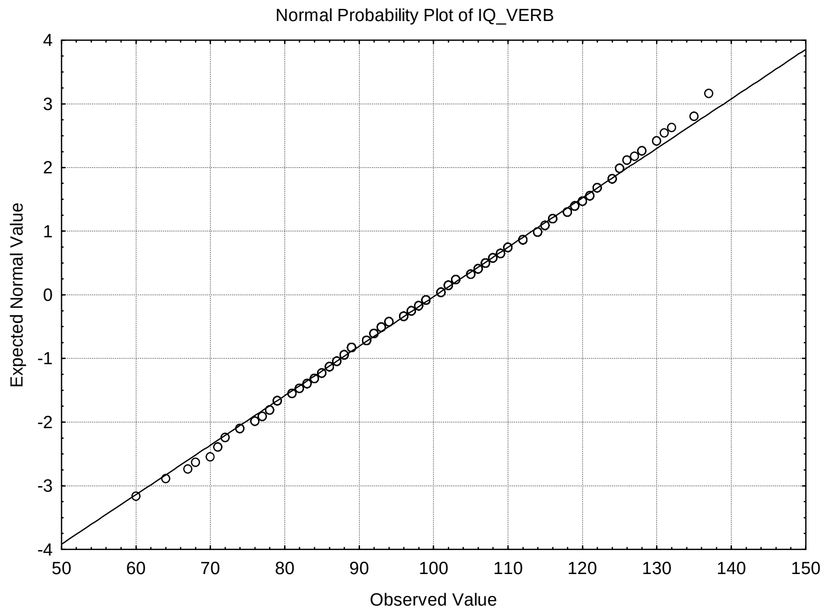


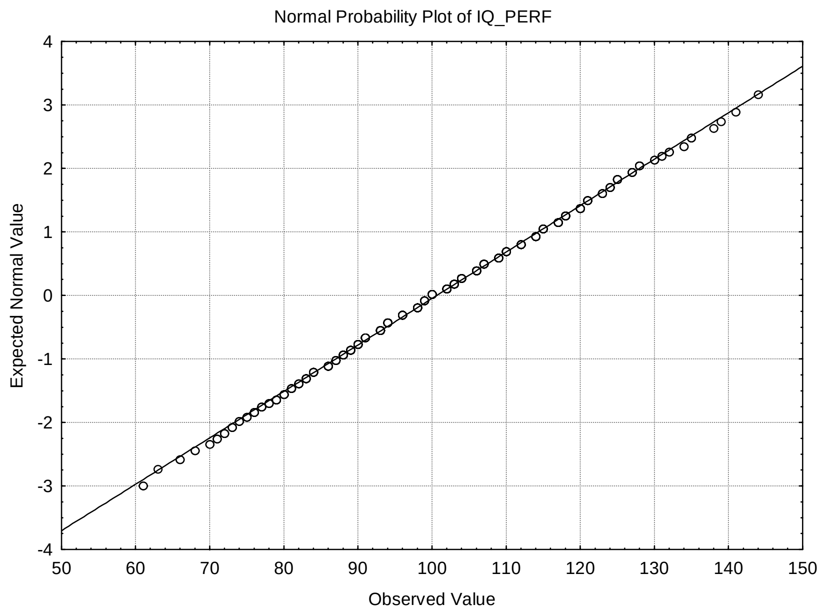
4. *Na hladině významnosti 0,05 testujte hypotézu, že rozdíl středních hodnot proměnných IQ_VERB a IQ_PERF je nulový, a to*
   a) pro všechny děti
   b) pro chlapce
   c) pro dívky
   d) pro městské děti
   e) pro venkovské děti.

ad a) Ověříme normalitu dat IQ_VERB a IQ_PERF a sestrojíme příslušné Normal-Probability ploty.

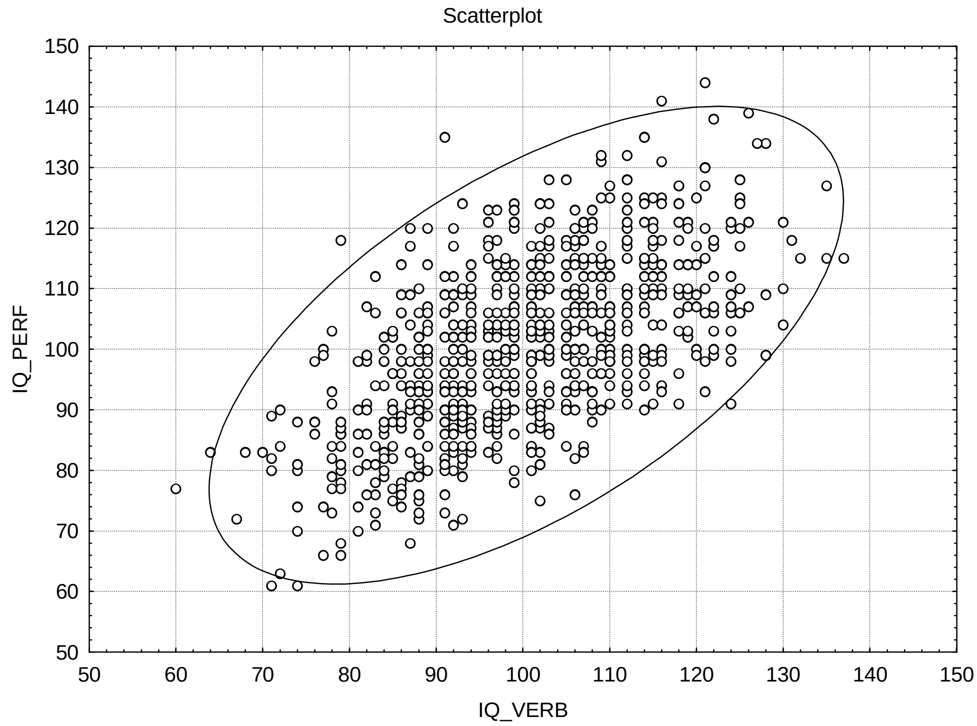
Variable	Tests of Normality N	max D	Lilliefors p
IQ_VERB	856	0,045137	p < ,01

Variable	Tests of Normality N	max D	Lilliefors p
IQ_PERF	856	0,044792	p < ,01





K-S test zamítá obě hypotézy, nicméně z grafů je vidět, že normalita dat je porušena jen mírně. Navíc - soubor je velkého rozsahu, považujme dále data za normální. Posoudíme ještě dvourozměrnou normalitu dat.



Z diagramu vidíme, že většina dat padne do elipsy, která určuje 95% oblast spolehlivosti – data tedy vykazují dvourozměrnou normalitu.

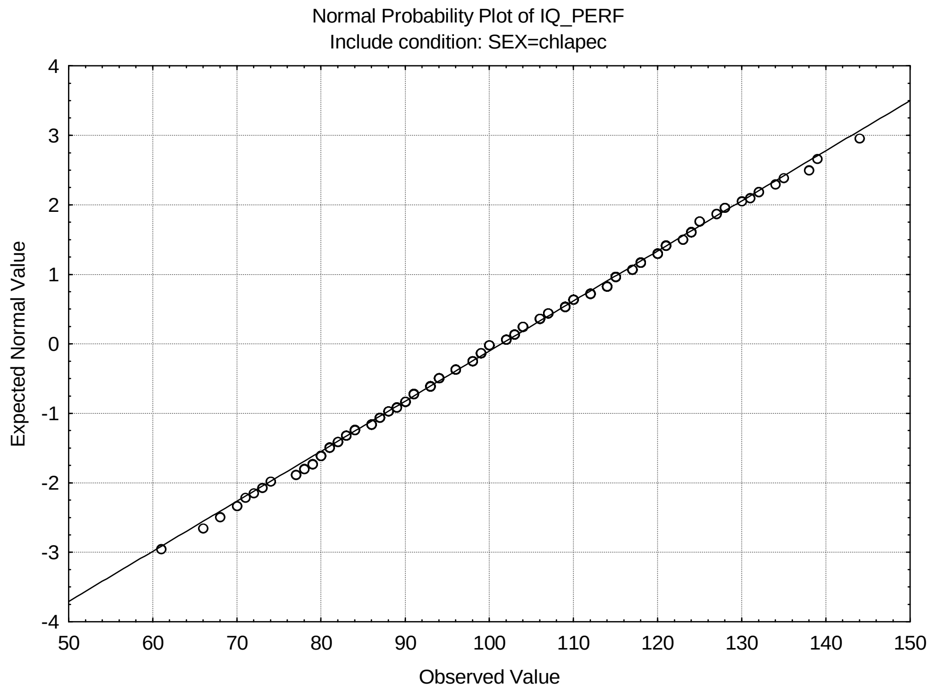
Variable	T-test for Dependent Samples Marked differences are significant at p < ,05000 Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p
IQ_VERB	100,3925	12,80671						
IQ_PERF	100,6519	13,61699	856	-0,259346	11,91993	-0,636565	855	0,524579

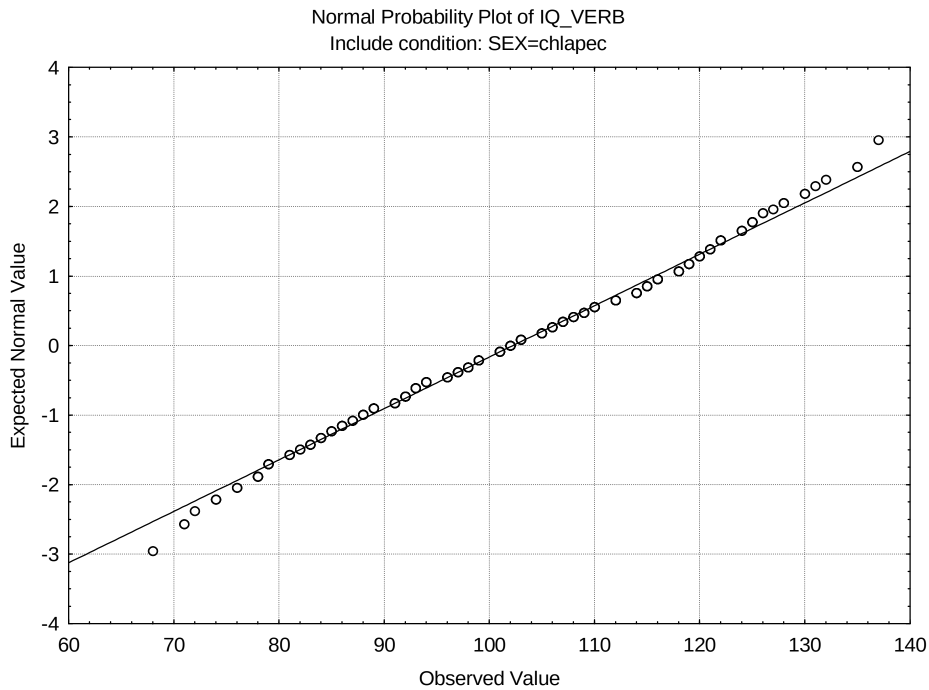
Hodnota testové statistiky t-testu je t=-0,636565 a p-hodnota je p=0,524579>0,05. Párový t-test na hladině významnosti 0,05 nezamítá hypotézu o shodě středních proměnných IQ_VERB a IQ_PERF pro všechny děti.

ad b) Ověříme normalitu dat IQ_VERB a IQ_PERF pro chlapce a sestrojíme příslušné Normal-Probability ploty.

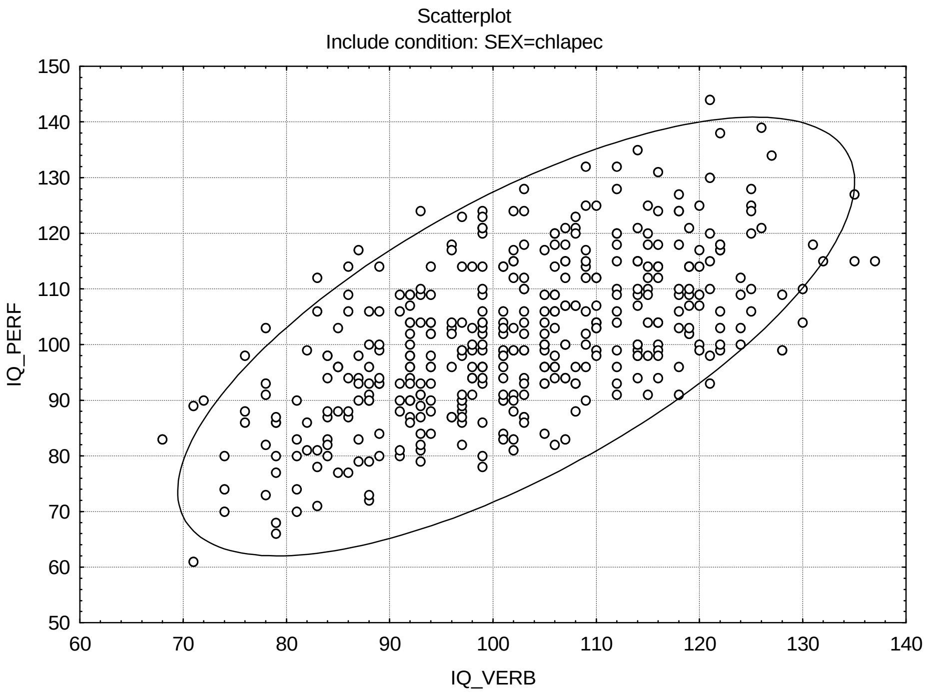
Variable	Tests of Normality N	max D	Lilliefors p
IQ_VERB	426	0,051092	p < ,01

Variable	Tests of Normality N	max D	Lilliefors p
IQ_PERF	426	0,053426	p < ,01





K-S test zamítá obě hypotézy, nicméně z grafů je vidět, že normalita dat je porušena jen mírně. Navíc - soubor je velkého rozsahu, považujme dále data za normální. Posoudíme ještě dvourozměrnou normalitu dat.



Z diagramu vidíme, že většina dat padne do elipsy, která určuje 95% oblast spolehlivosti – data tedy vykazují dvourozměrnou normalitu.

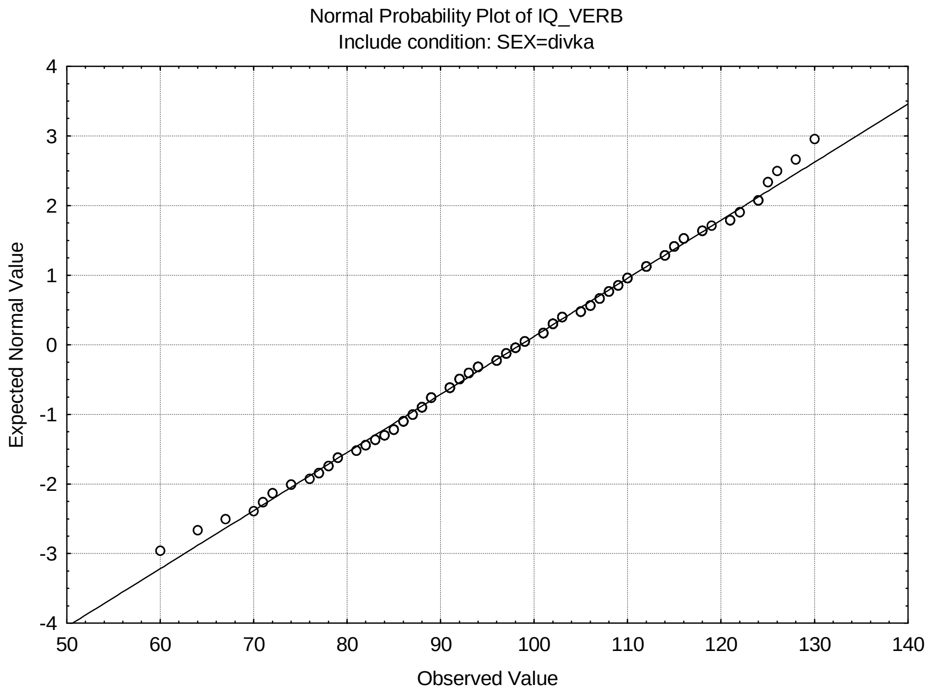
Variable	T-test for Dependent Samples Marked differences are significant at p < ,05000							
	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p
IQ_VERB	102,2559	13,41651						
IQ_PERF	101,4437	13,78978	426	0,812207	12,12202	1,382917	425	0,167416

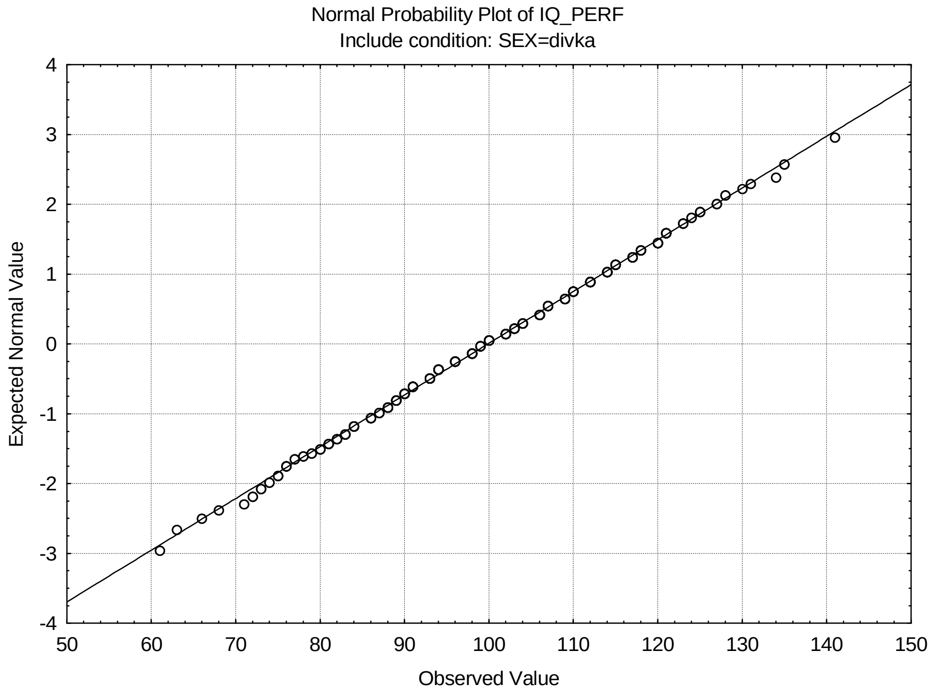
Hodnota testové statistiky t-testu je t=1,382917 a p-hodnota je p=0,167416>0,05. Párový t-test na hladině významnosti 0,05 nezamítá hypotézu o shodě středních proměnných IQ_VERB a IQ_PERF pro chlapce.

ad c) Ověříme normalitu dat IQ_VERB a IQ_PERF pro dívky a sestrojíme příslušné Normal-Probability ploty.

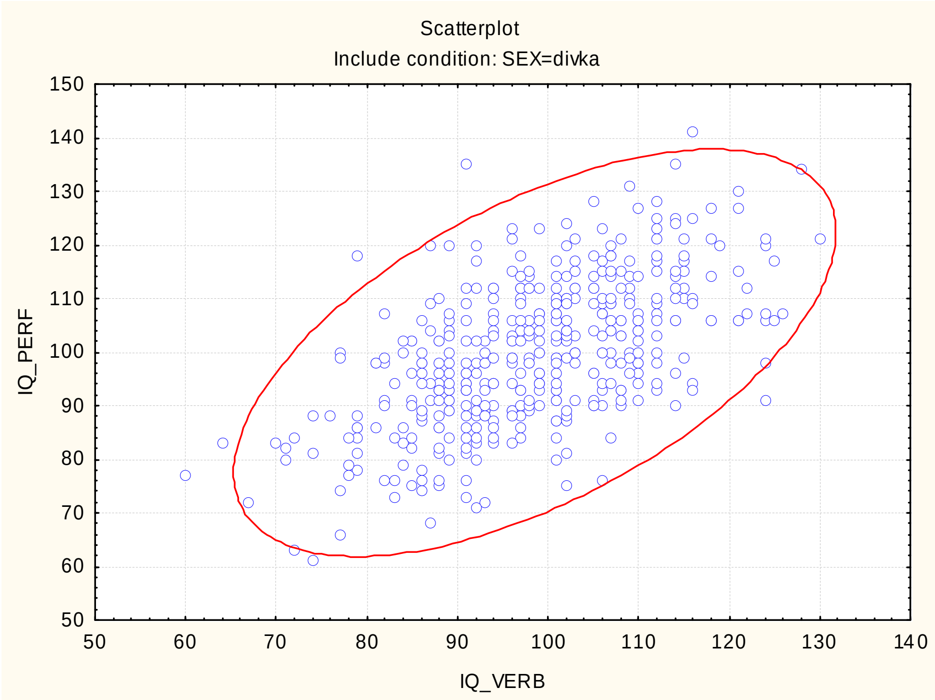
Variable	Tests of Normality		
	N	max D	Lilliefors p
IQ_VERB	430	0,042116	p < ,10

Variable	Tests of Normality		
	N	max D	Lilliefors p
IQ_PERF	430	0,050494	p < ,01





K-S test zamítá obě hypotézy, nicméně z grafů je vidět, že normalita dat je porušena jen mírně. Navíc - soubor je velkého rozsahu, považujme dále data za normální. Posoudíme ještě dvourozměrnou normalitu dat.



Z diagramu vidíme, že většina dat padne do elipsy, která určuje 95% oblast spolehlivosti – data tedy vykazují dvourozměrnou normalitu.

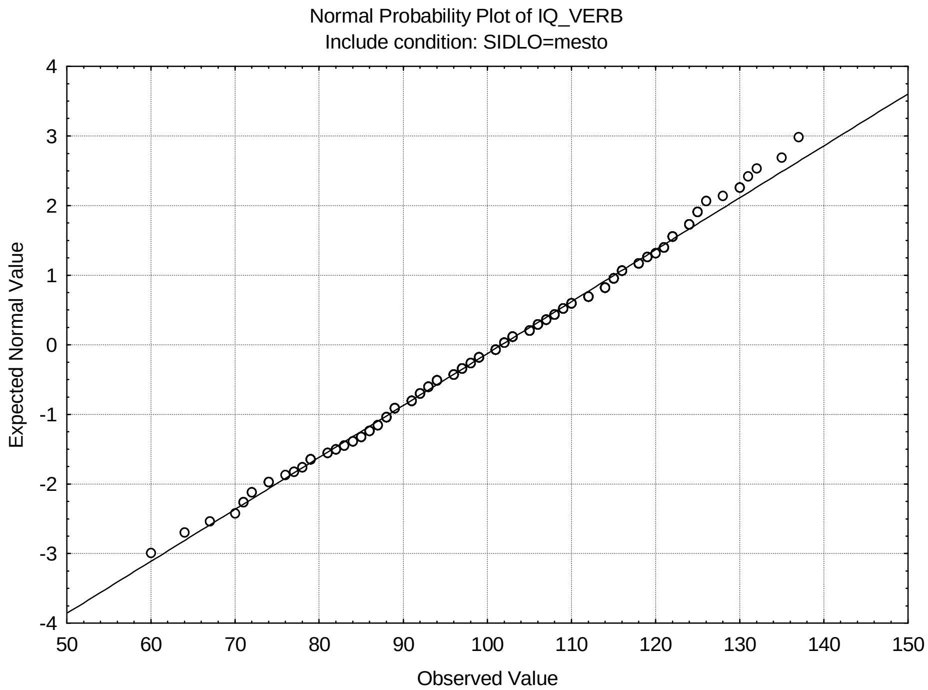
Variable	T-test for Dependent Samples Marked differences are significant at p < ,05000							
	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p
IQ_VERB	98,54651	11,90332						
IQ_PERF	99,86744	13,41358	430	-1,32093	11,63326	-2,35458	429	0,018994

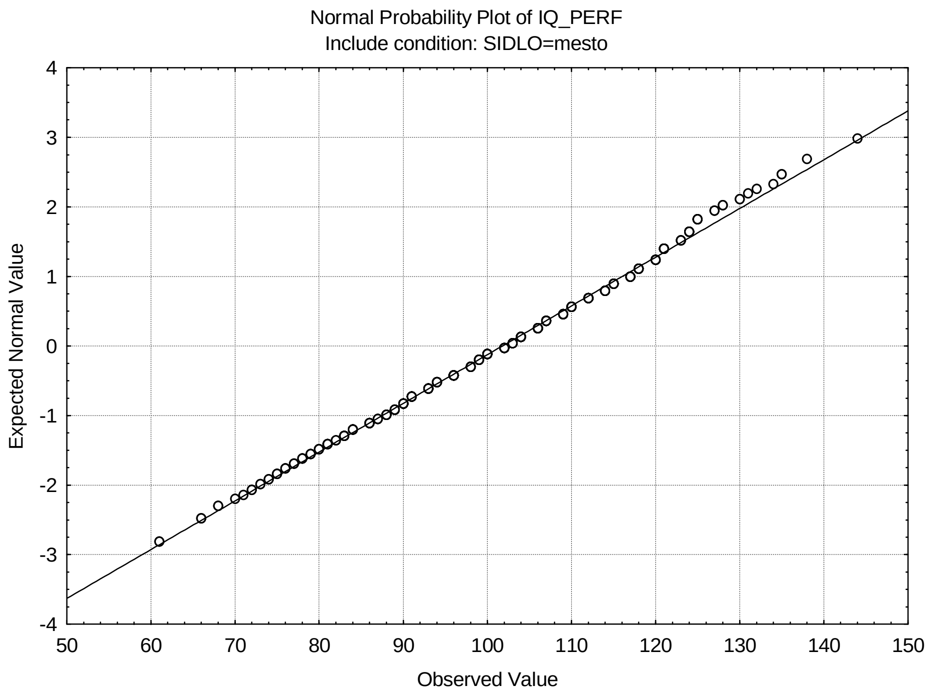
Hodnota testové statistiky t-testu je t=-2,35458 a p-hodnota je p=0,018994<0,05. Párový t-test na hladině významnosti 0,05 zamítá hypotézu o shodě středních proměnných IQ_VERB a IQ_PERF pro dívky.

ad d) Nejprve ověříme normalitu dat IQ_VERB a IQ_PERF pro městské děti a sestrojíme příslušné Normal-Probability ploty.

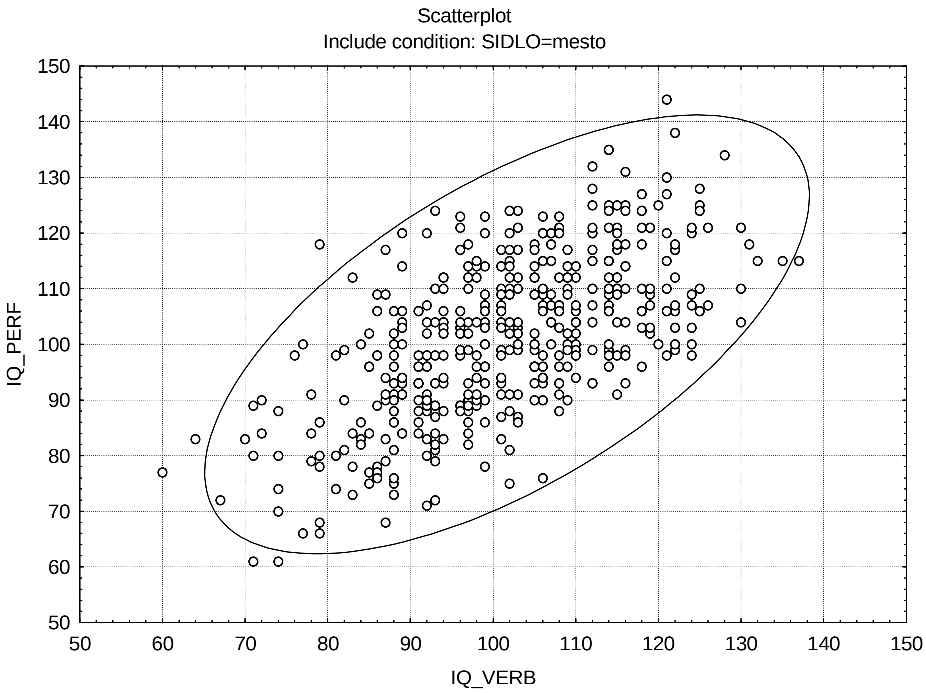
Variable	Tests of Normality		
	N	max D	Lilliefors p
IQ_VERB	473	0,048642	p < ,01

Variable	Tests of Normality		
	N	max D	Lilliefors p
IQ_PERF	473	0,039873	p < ,10





K-S test zamítá obě hypotézy, nicméně z grafů je vidět, že normalita dat je porušena jen mírně. Navíc - soubor je velkého rozsahu, považujme dále data za normální. Posoudíme ještě dvourozměrnou normalitu dat.



Z diagramu vidíme, že většina dat padne do elipsy, která určuje 95% oblast spolehlivosti – data tedy vykazují dvourozměrnou normalitu.

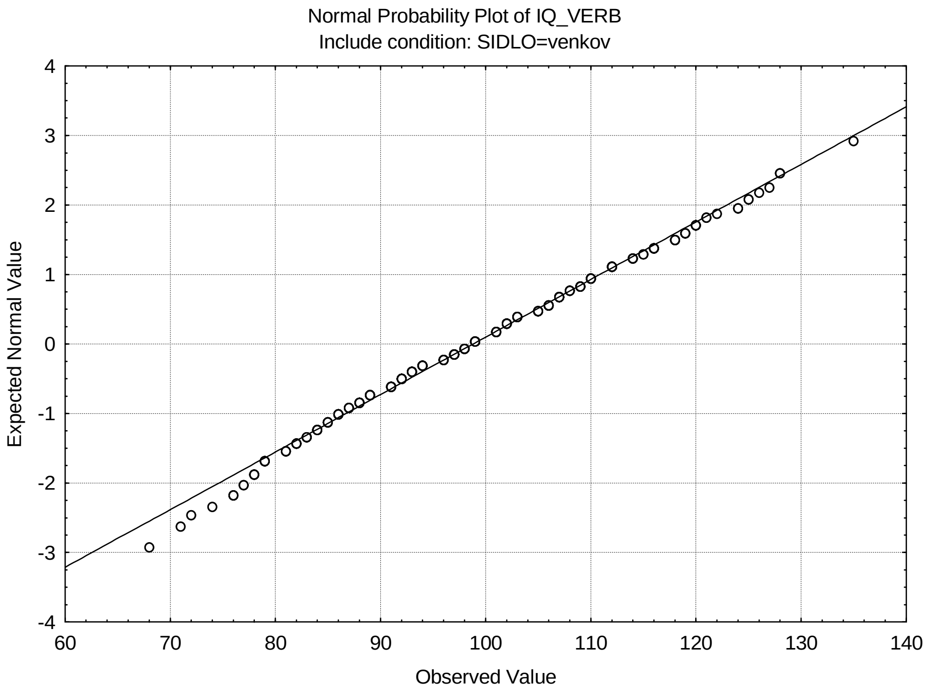
Variable	T-test for Dependent Samples Marked differences are significant at p < ,05000 Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p
IQ_VERB	101,6913	13,31169						
IQ_PERF	101,7780	14,18131	473	-0,086681	11,97574	-0,157417	472	0,874984

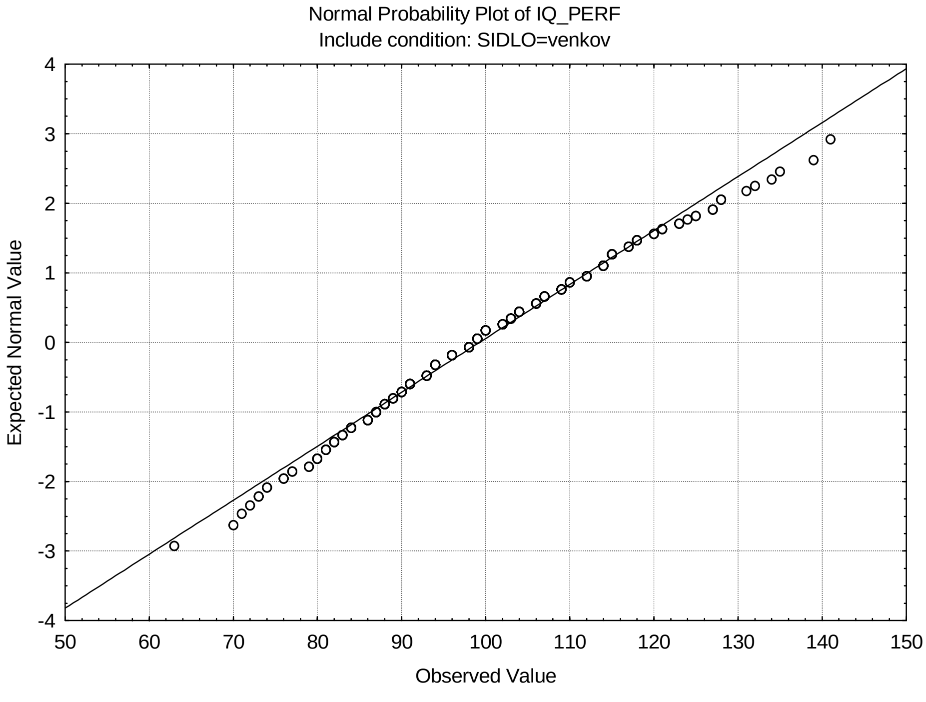
Hodnota testové statistiky t-testu je t=-0,157417 a p-hodnota je p=0,874984>0,05. Párový t-test na hladině významnosti 0,05 nezamítá hypotézu o shodě středních proměnných IQ_VERB a IQ_PERF pro městské děti.

ad e) Ověříme normalitu dat IQ_VERB a IQ_PERF pro venkovské děti a sestrojíme příslušné Normal-Probability ploty.

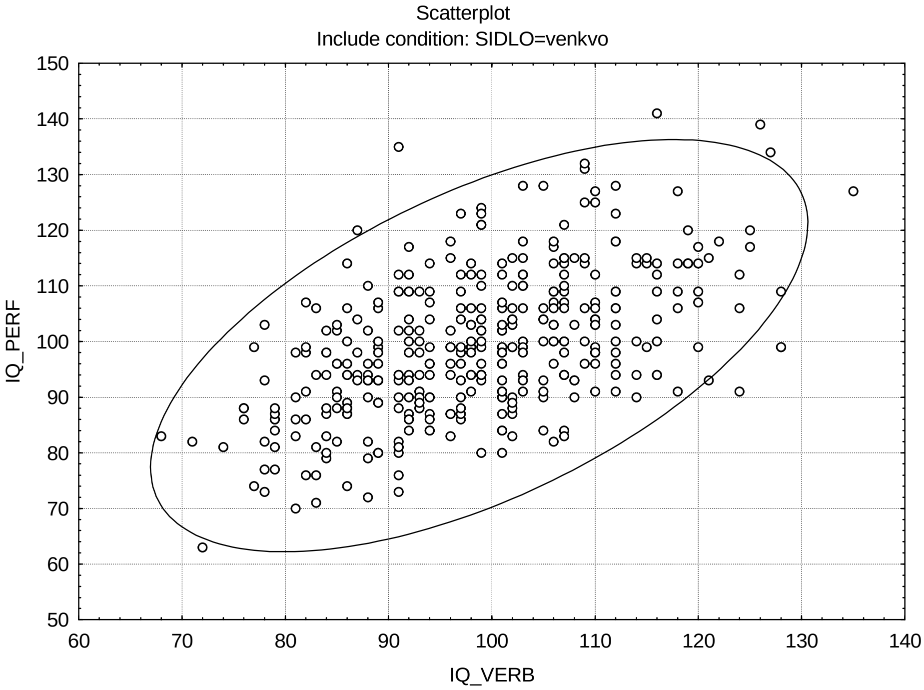
Variable	Tests of Normality N	max D	Lilliefors p
IQ_VERB	383	0,052204	p < ,05

Variable	Tests of Normality N	max D	Lilliefors p
IQ_PERF	383	0,067163	p < ,01





K-S test zamítá obě hypotézy, nicméně z grafů je vidět, že normalita dat je porušena jen mírně. Navíc - soubor je velkého rozsahu, považujme dále data za normální. Posoudíme dvourozměrnou normalitu dat.



Z diagramu vidíme, že většina dat padne do elipsy, která určuje 95% oblast spolehlivosti – data tedy vykazují dvourozměrnou normalitu.

Variable	T-test for Dependent Samples Marked differences are significant at p < ,05000 Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p
IQ_VERB	98,78851	11,97815						
IQ_PERF	99,26110	12,76775	383	-0,472585	11,86281	-0,779635	382	0,436088

Hodnota testové statistiky t-testu je t=-0,779635 a p-hodnota je p=0,436088>0,05. Párový t-test na hladině významnosti 0,05 nezamítá hypotézu o shodě středních proměnných IQ_VERB a IQ_PERF pro venkovské děti.

5. *Na hladině významnosti 0,05 proveďte analýzu rozptylu proměnné IQ_CELK pro faktor vzdělání matky. Nezapomeňte ověřit předpoklady o datech. V případě zamítnutí nulové hypotézy aplikujte Scheffého metodu mnohonásobného porovnávání. Pro všechny úrovně faktoru nakreslete krabicové diagramy.*
   Nepovinný úkol: Tentýž úkol proveďte pro faktor vzdělání otce.

Povinný úkol:

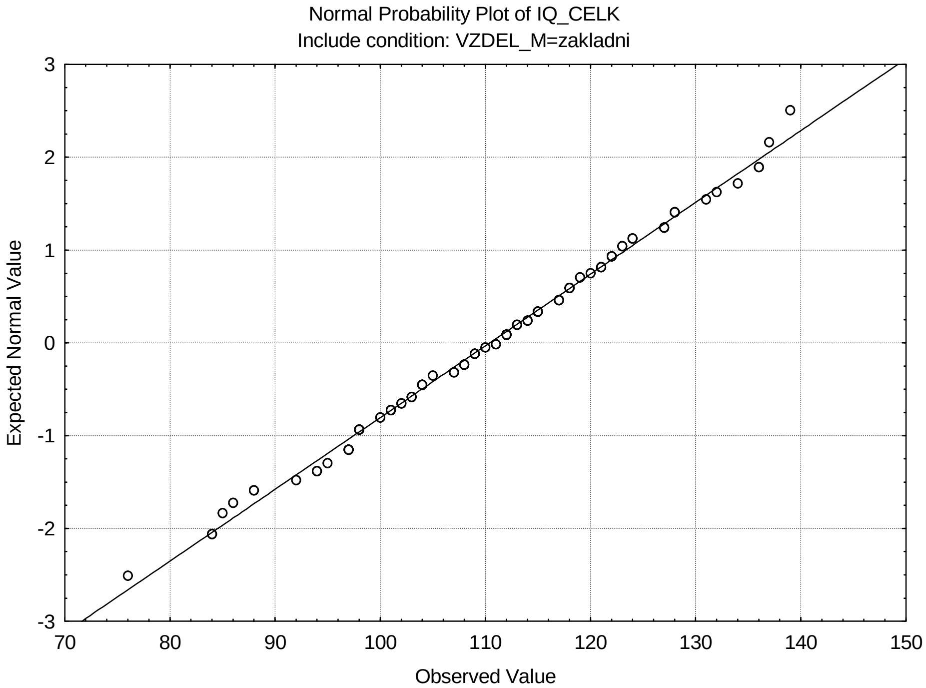
Ověříme normalitu dat pro jednotlivé úrovně faktoru VZDEL_M.

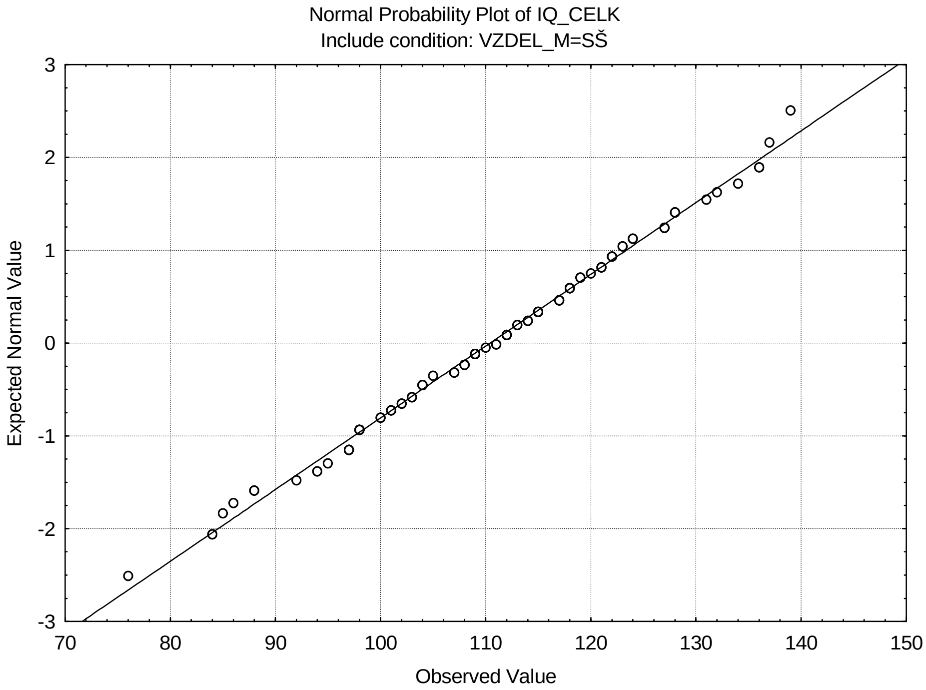
Variable	Tests of Normality Include condition: VZDEL_M=zakladni		
	N	max D	Lilliefors p
IQ_CELK	361	0,040499	p < ,15

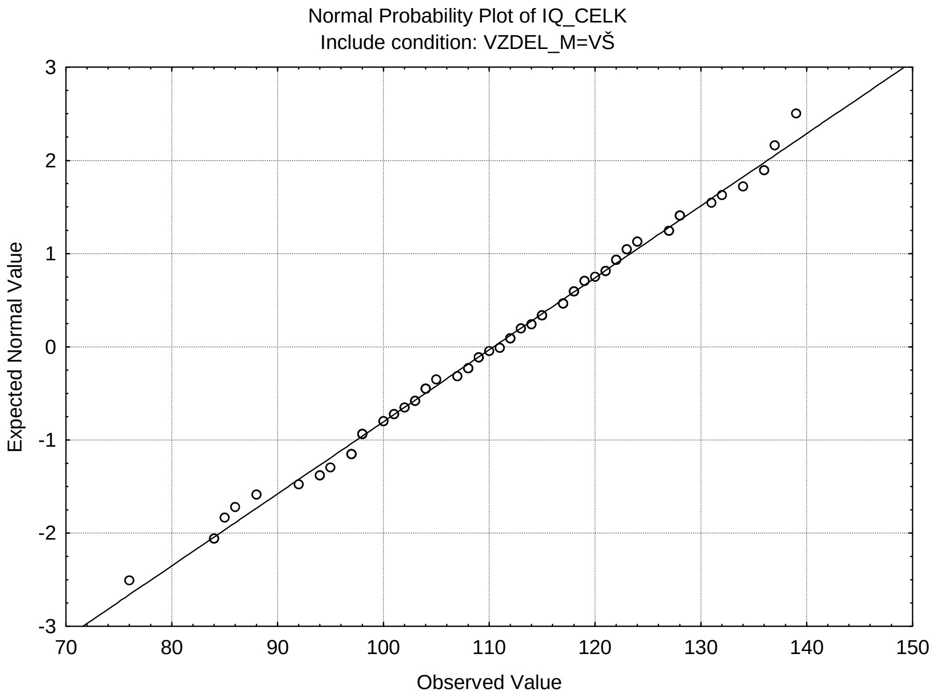
Variable	Tests of Normality Include condition: VZDEL_M=SŠ		
	N	max D	Lilliefors p
IQ_CELK	386	0,040237	p < ,15

Variable	Tests of Normality Include condition: VZDEL_M=VŠ		
	N	max D	Lilliefors p
IQ_CELK	109	0,051004	p > .20

Pro VZDEL_M ZŠ a SŠ je p-hodnota K-S testu p< 0,15 a pro VZDEL_M VŠ je p-hodnota K-S testu p>0,2. K-S test tedy ani jednu z hypotéz o normalitě dat na hladině významnosti 0,05 nezamítá. Grafy, Normal-Probability ploty, tyto hypotézy rovněž podporují.



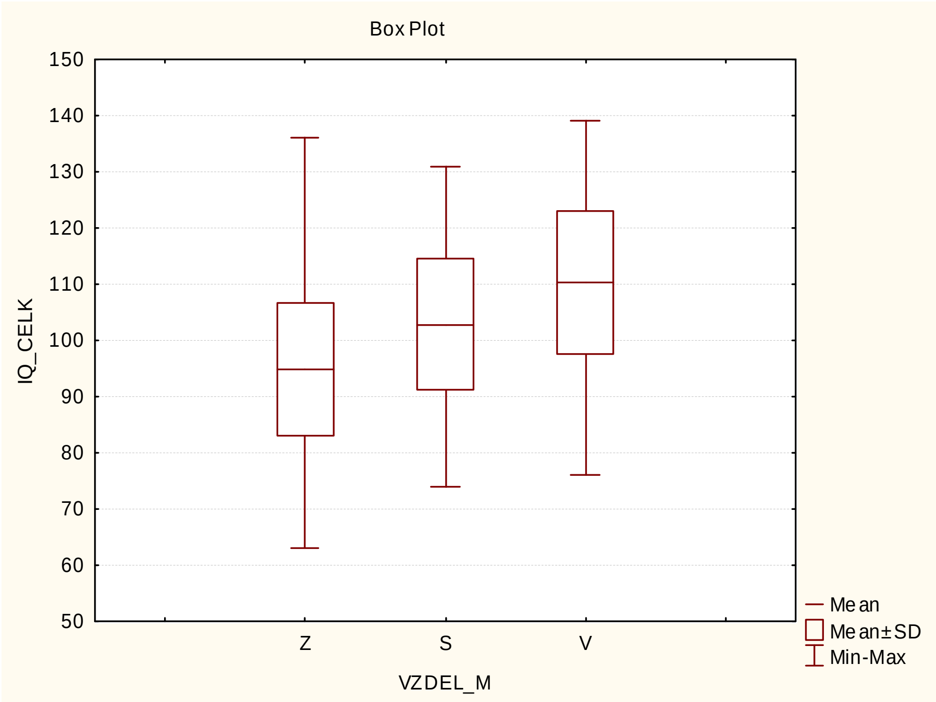




Přistoupíme k ověření hypotézy. Nejdříve zjistíme výběrové průměry a směrodatné odchylky pro různé úrovně faktoru VZDEL_M.

Breakdown Table of Descriptive Statistics			
VZDEL_M	IQ_CELK Means	IQ_CELK N	IQ_CELK Std.Dev.
Z	94,8837	361	11,70981
S	102,8394	386	11,69283
V	110,4220	109	12,71795
All Grps	100,4498	856	12,96409

Krabicové diagramy pro všechny úrovně faktoru VZDEL_M.



Ověříme předpoklad o shodě rozptylů (Leveneův test).

Variable	Levene Test of Homogeneity of Variances Marked effects are significant at p < ,05000 SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
IQ_CELK	69,73055	2	34,86528	42130,50	853	49,39097	0,705904	0,493950

Hodnota testové statistiky je F=0,705904 a p-hodnota je p=0,49395>0,05 – na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů IQ_CELK pro různé úrovně faktoru VZDEL_M. Provedeme analýzu rozptylu.

Variable	Analysis of Variance Marked effects are significant at p < ,05000 SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
IQ_CELK	24228,10	2	12114,05	119469,7	853	140,0583	86,49289	0,00

Protože je p-hodnota (téměř nulová) menší, než 0,05, zamítáme na hladině významnosti 0,05 hypotézu o shodě středních hodnot. Scheffého metoda mnohonásobného porovnávání ukazuje, že na hladině významnosti 0,05 se liší všechny 3 dvojice výběrů.

VZDEL_M	Scheffe Test; Variable: IQ_CELK Marked differences are significant at p < ,05000 {1} M=94,884	{2} M=102,84	{3} M=110,42
Z {1}		0,000000	0,000000
S {2}	0,000000		0,000000
V {3}	0,000000	0,000000	

Nepovinný úkol:

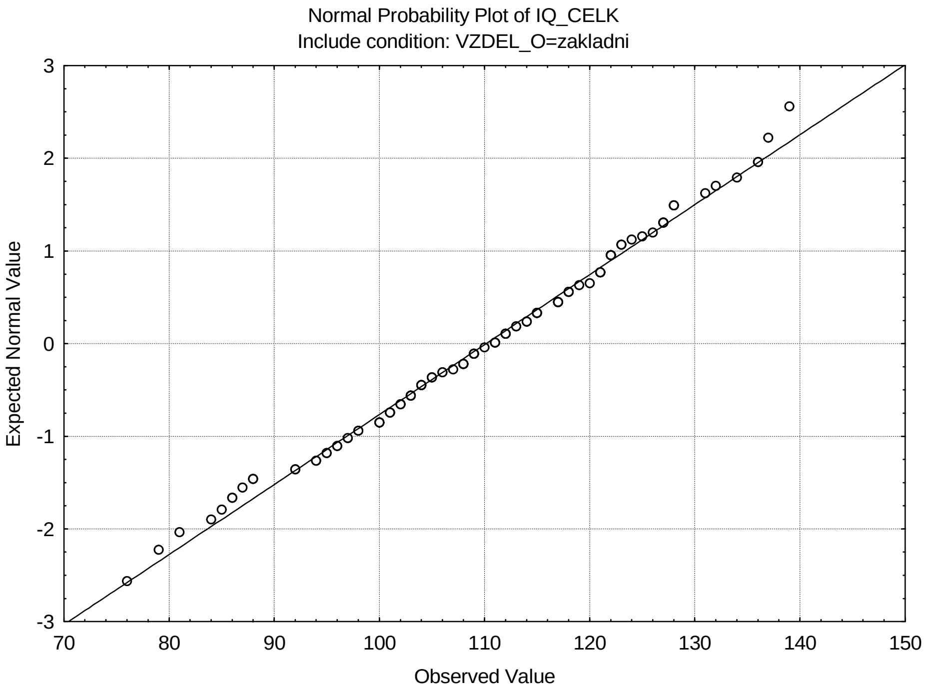
Ověříme normalitu dat pro jednotlivé úrovně faktoru VZDEL_O.

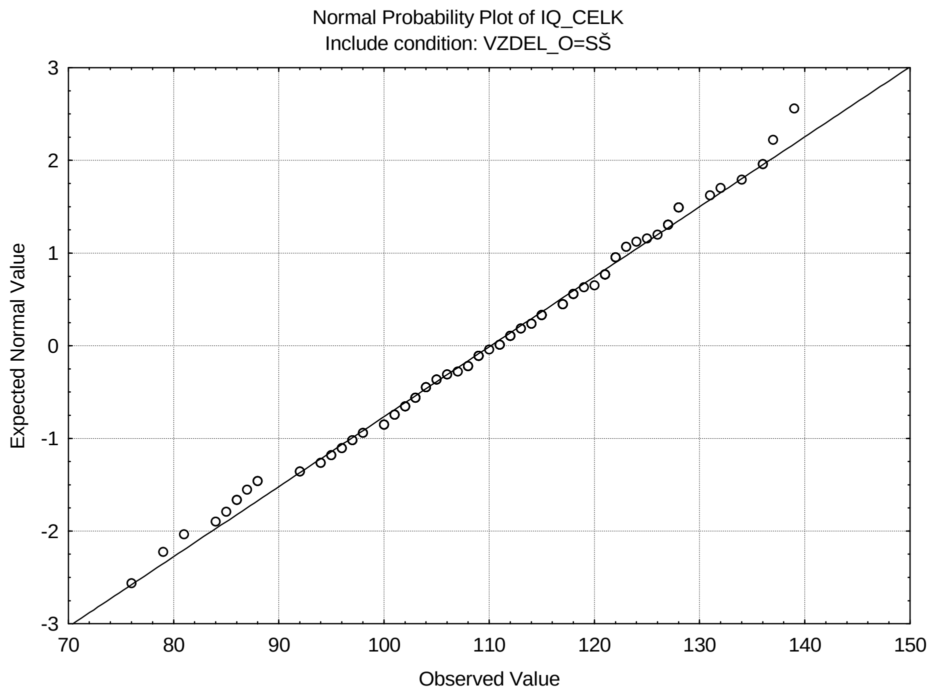
Variable	Tests of Normality Include condition: VZDEL_O=zakladni N	 max D	 Lilliefors p
IQ_CELK	438	0,036196	p < ,20

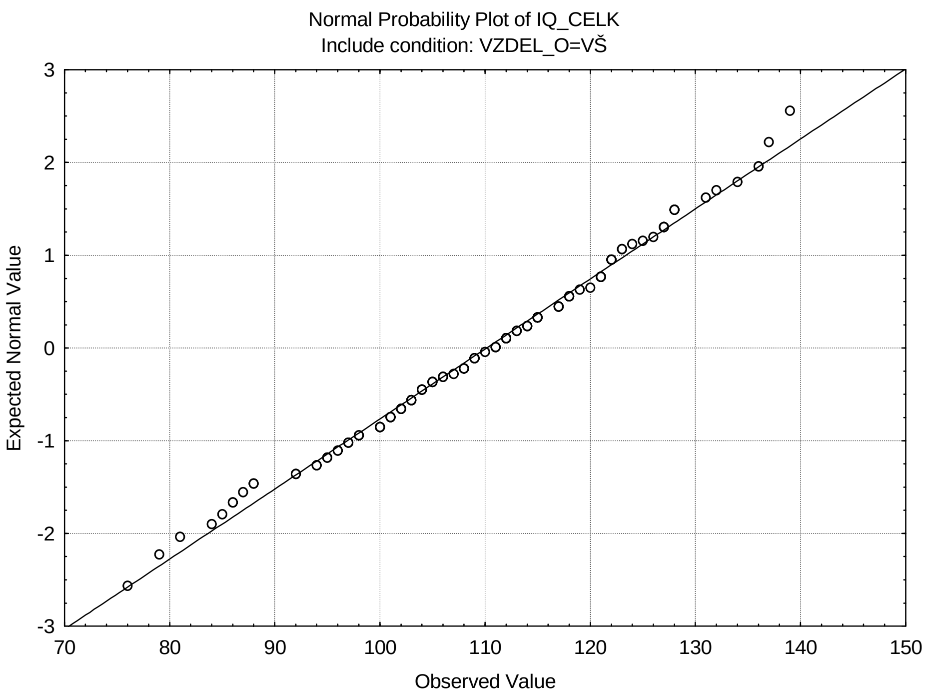
Variable	Tests of Normality Include condition: VZDEL_O=SŠ N	 max D	 Lilliefors p
IQ_CELK	291	0,044058	p < ,20

Variable	Tests of Normality Include condition: VZDEL_O=VŠ N	 max D	 Lilliefors p
IQ_CELK	127	0,049435	p > .20

Pro VZDEL_O ZŠ a SŠ je p-hodnota K-S testu p< 0,20 a pro VZDEL_O VŠ je p-hodnota K-S testu p>0,2. K-S test tedy ani jednu z hypotéz o normalitě dat na hladině významnosti 0,05 nezamítá. Grafy, Normal-Probability ploty, tyto hypotézy rovněž podporují.



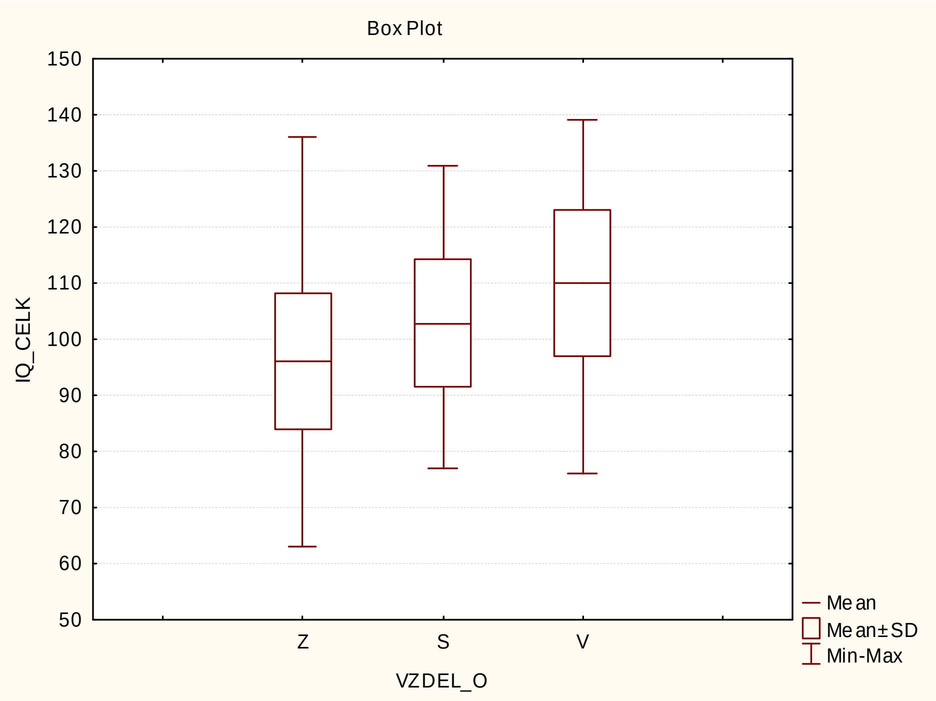




Přistoupíme k ověření hypotézy. Nejdříve zjistíme výběrové průměry a směrodatné odchylky pro různé úrovně faktoru VZDEL_O.

Breakdown Table of Descriptive Statistics			
VZDEL_O	IQ_CELK Means	IQ_CELK N	IQ_CELK Std.Dev.
Z	96,0525	438	12,03787
S	102,8385	291	11,27760
V	110,1417	127	13,04128
All Grps	100,4498	856	12,96409

Krabicové diagramy pro všechny úrovně faktoru VZDEL_O.



Ověříme předpoklad o shodě rozptylů (Leveneův test).

Variable	Levene Test of Homogeneity of Variances Marked effects are significant at p < ,05000 SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
IQ_CELK	146,2338	2	73,11688	41883,35	853	49,10123	1,489105	0,226160

Hodnota testové statistiky je F=1,489105 a p-hodnota je p=0,226160>0,05 – na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů IQ_CELK pro různé úrovně faktoru VZDEL_M. Provedeme analýzu rozptylu.

Variable	Analysis of Variance Marked effects are significant at p < ,05000 SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
IQ_CELK	22059,19	2	11029,59	121638,6	853	142,6010	77,34585	0,00

Protože je p-hodnota (téměř nulová) menší, než 0,05, zamítáme na hladině významnosti 0,05 hypotézu o shodě středních hodnot. Scheffého metoda mnohonásobného porovnávání ukazuje, že na hladině významnosti 0,05 se liší všechny 3 dvojice výběrů.

VZDEL_O	Scheffe Test; Variable: IQ_CELK Marked differences are significant at p < ,05000 {1} M=96,053	{2} M=102,84	{3} M=110,14
Z {1}		0,000000	0,000000
S {2}	0,000000		0,000000
V {3}	0,000000	0,000000	

6. *Vypočtěte Spearmanův koeficient pořadové korelace proměnných VZDEL_M a VZDEL_O. Na hladině významnosti 0,05 testujte hypotézu, že proměnné VZDEL_M a VZDEL_O jsou pořadově nezávislé. Úkol proveďte*
   a) pro všechny děti
   b) pro chlapce
   c) pro dívky
   d) pro městské děti
   e) pro venkovské děti

ad a)

Pair of Variables	Spearman Rank Order Correlations MD pairwise deleted Marked correlations are significant at p <,05000			
	Valid N	Spearman R	t(N-2)	p-level
VZDEL_M & VZDEL_O	856	0,612198	22,62595	0,00

Spearmanův koeficient pořadové korelace proměnných VZDEL_M a VZDEL_O je 0,612198, p-hodnota je téměř nulová. Na hladině významnosti zamítáme hypotézu o pořadové nezávislosti proměnných VZDEL_M a VZDEL_O zkoumané pro všechny děti. Proměnné jsou do určité míry kladně pořadově závislé.

ad b)

Pair of Variables	Spearman Rank Order Correlations MD pairwise deleted Marked correlations are significant at p <,05000 Include condition: SEX=chlapec			
	Valid N	Spearman R	t(N-2)	p-level
VZDEL_M & VZDEL_O	426	0,620294	16,28396	0,00

Spearmanův koeficient pořadové korelace proměnných VZDEL_M a VZDEL_O je 0,620294, p-hodnota je téměř nulová. Na hladině významnosti zamítáme hypotézu o pořadové nezávislosti proměnných VZDEL_M a VZDEL_O zkoumané pro chlapce. Proměnné jsou do určité míry kladně pořadově závislé.

ad c)

Pair of Variables	Spearman Rank Order Correlations MD pairwise deleted Marked correlations are significant at p <,05000 Include condition: SEX=divka			
	Valid N	Spearman R	t(N-2)	p-level
VZDEL_M & VZDEL_O	430	0,604566	15,70183	0,00

Spearmanův koeficient pořadové korelace proměnných VZDEL_M a VZDEL_O je 0,604566, p-hodnota je téměř nulová. Na hladině významnosti zamítáme hypotézu o pořadové nezávislosti proměnných VZDEL_M a VZDEL_O zkoumané pro dívky. Proměnné jsou do určité míry kladně pořadově závislé.

ad d)

Pair of Variables	Spearman Rank Order Correlations MD pairwise deleted Marked correlations are significant at p <,05000 Include condition: SIDLO=mesto			
	Valid N	Spearman R	t(N-2)	p-level
VZDEL_M & VZDEL_O	473	0,639079	18,03263	0,00

Spearmanův koeficient pořadové korelace proměnných VZDEL_M a VZDEL_O je 0,639079, p-hodnota je téměř nulová. Na hladině významnosti zamítáme hypotézu o pořadové nezávislosti proměnných VZDEL_M a VZDEL_O zkoumané pro městské děti. Proměnné jsou do určité míry kladně pořadově závislé.

ad e)

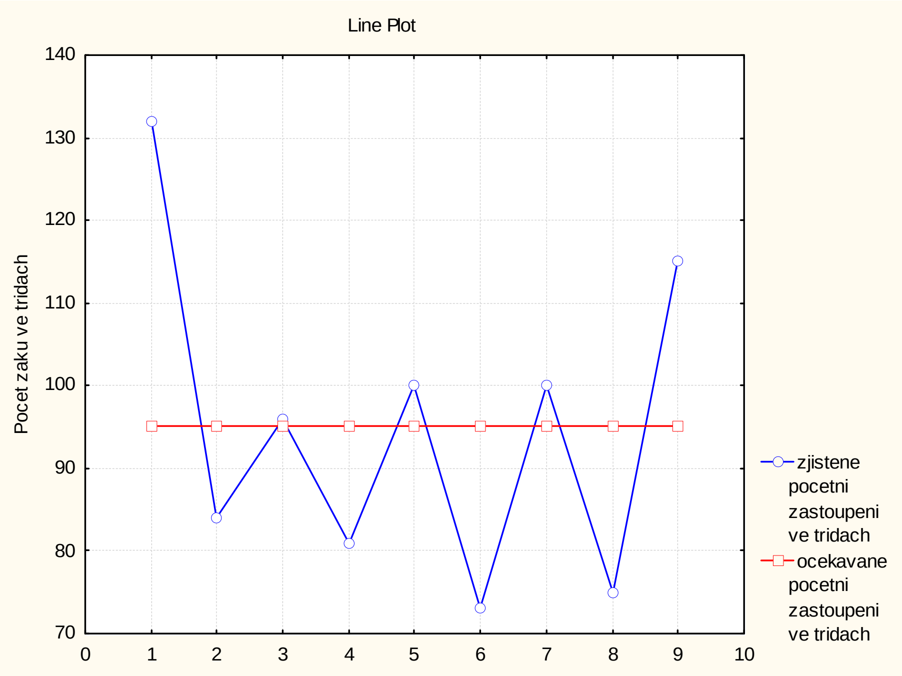
Pair of Variables	Spearman Rank Order Correlations MD pairwise deleted Marked correlations are significant at p <,05000 Include condition: SIDLO=venkov			
	Valid N	Spearman R	t(N-2)	p-level
VZDEL_M & VZDEL_O	383	0,567509	13,45371	0,00

Spearmanův koeficient pořadové korelace proměnných VZDEL_M a VZDEL_O je 0,567509, p-hodnota je téměř nulová. Na hladině významnosti zamítáme hypotézu o pořadové nezávislosti proměnných VZDEL_M a VZDEL_O zkoumané pro venkovské děti. Proměnné jsou do určité míry kladně pořadově závislé.

7. *Na hladině významnosti 0,05 testujte hypotézu, že početní zastoupení dětí v 1. až 9. třídě se řídí rovnoměrným rozložením.*

Case	Observed vs. Expected Frequencies Chi-Square = 31,76168 df = 8 p < ,000103			
	observed Var1	expected Var2	O - E	(O-E)**2 /E
C: 1	132,0000	95,1111	36,8889	14,30737
C: 2	84,0000	95,1111	-11,1111	1,29803
C: 3	96,0000	95,1111	0,8889	0,00831
C: 4	81,0000	95,1111	-14,1111	2,09359
C: 5	100,0000	95,1111	4,8889	0,25130
C: 6	73,0000	95,1111	-22,1111	5,14032
C: 7	100,0000	95,1111	4,8889	0,25130
C: 8	75,0000	95,1111	-20,1111	4,25247
C: 9	115,0000	95,1111	19,8889	4,15901
Sum	856,0000	856,0000	-0,0000	31,76168

Hodnota testové statistiky je 31,76168, p-hodnota p<0,000103. Tedy na hladině významnosti 0,05 hypotézu o rovnoměrném rozložení počtu dětí v jednotlivých třídách zamítáme. Že se početní zastoupení dětí neřídí rovnoměrným rozložením, je patrné z grafu.



8. *Sestrojte 95% asymptotický interval spolehlivosti pro podíl městských dětí a s jeho pomocí testujte na asymptotické hladině významnosti 0,05 hypotézu, že podíl dětí z města a venkova je stejný.*

Ověříme podmínku dobré aproximace $n\theta(1-\theta)>9$. θ neznáme, musíme ho nahradit výběrovým průměrem $m=\frac{473}{856}$. Podmínka $856.\frac{473}{856}(1-\frac{473}{856})\cong 211{,}6343>9$ je však splněna. Dosazením do vzorce pro asymptotický interval spolehlivosti pro parametr alternativního rozložení obdržíme horní a dolní mez intervalu.

	DM	HM
1	0,51926	0,58588

S pravděpodobností 95% leží podíl městských dětí k celkovému počtu dětí v intervalu (0,51926;0,58588).

Kdyby byl podíl dětí z města a venkova stejný, pak by v tomto intervalu musela ležet i hodnota 0,5. Proto na asymptotické hladině významnosti 0,05 zamítáme hypotézu, že podíl dětí z města a venkova je stejný.

9. *Pro celý soubor 856 dětí vypočtěte průměr proměnné IQ_CELK. Na asymptotické hladině významnosti 0,05 testujte hypotézu, že podíl dětí s nadprůměrným IQ_CELK je stejný*
   a) *mezi městskými a venkovskými dětmi*
   b) *mezi chlapci a dívkami*
   c) *mezi dětmi, jejichž oba rodiče mají VŠ vzdělání a dětmi, jejichž oba rodiče mají ZŠ vzdělání.*

Vyjádříme si průměr proměnné IQ_CELK.

ad a) Spočítáme absolutní četnost městských a venkovských dětí (N) a absolutní četnost městských a venkovských dětí s nadprůměrným IQ_CELK (N nad). Dále určíme příslušné relativní četnosti (m) a relativní četnost dětí s nadprůměrným IQ_CELK v rámci celého souboru (m*). Ověříme podmínky dobré aproximace $n_1\theta_1(1-\theta_1)>9$ a $n_2\theta_2(1-\theta_2)>9$. θ_1 a θ_2 neznáme, musíme ho nahradit výběrovými průměry $m_1=0{,}54334$ a $m_2=0{,}43342$. Obě podmínky jsou však splněny (473.0,54334.(1-0,54334)≅117,36>9 a 383.0,43342.(1-0,43342)≅94,05>9). Otestujeme hypotézu rovnosti rozdílu podílu městských a podílu venkovských dětí s nadprůměrným IQ_CELK k nule na asymptotické hladině významnosti 0,05. Výsledek zapíšeme do tabulky (t je realizace testového kritéria a interval (∞ ; -u⟩ ∪ ⟨u ; ∞) je kritický obor).

SIDLO	Tabulka					
	N	N nad	m	m*	t	u
mesto	473	257	0,54334	0,494159	3,198375	1,959964
venkov	383	166	0,43342	0,494159	3,198375	1,959964

Protože realizace testového kritéria patří do kritického oboru, zamítáme na asymptotické hladině významnosti 0,05 hypotézu, že podíl dětí s nadprůměrným IQ_CELK je stejný mezi městskými a venkovskými dětmi.

ad b) Podobně jako za a) sestavíme tabulku, kde figurují podíly dětí s nadprůměrným IQ_CELK mezi chlapci a dívkami a ověříme podmínky dobré aproximace. Obě podmínky jsou však splněny (473.0,528169.(1-0,528169)≅117,87>9 a 373.0,460465.(1-0,460465)≅92,7>9).

SEX	Tabulka					
	N	N nad	m	m*	t	u
chlapci	426	225	0,528169	0,494159	1,980962	1,959964
divky	430	198	0,460465	0,494159	1,980962	1,959964

Protože realizace testového kritéria patří do kritického oboru, zamítáme na asymptotické hladině významnosti 0,05 hypotézu, že podíl dětí s nadprůměrným IQ_CELK je stejný mezi chlapci a dívkami. Realizace testového kritéria je však blízká hranici kritického oboru, zvolíme-li jinou hladinu významnosti, je možné, že na ní hypotézu nezamítneme.

ad c) Podobně jako za a) sestavíme tabulku, kde figurují podíly dětí s nadprůměrným IQ_CELK mezi dětmi, jejichž oba rodiče mají vysokoškolské vzdělání a dětmi, jejichž oba rodiče mají základní vzdělání dívkami a ověříme podmínky dobré aproximace. Obě podmínky jsou však splněny (296.0,283784.(1-0,23784)≅64,02>9 a 75.0,786667.(1-0,786667)≅12,59>9).

VZDEL	Tabulka					
	N	N nad	m	m*	t	u
ZŠ oba	296	84	0,283784	0,385445	-7,99273	1,959964
VŠ oba	75	59	0,786667	0,385445	-7,99273	1,959964

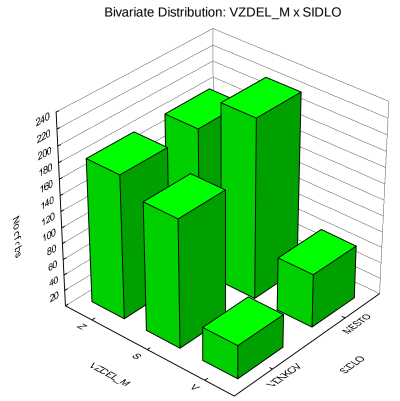
Protože realizace testového kritéria patří do kritického oboru, zamítáme na asymptotické hladině významnosti 0,05 hypotézu, že podíl dětí s nadprůměrným IQ_CELK je stejný mezi dětmi, jejichž oba rodiče mají vysokoškolské vzdělání a dětmi, jejichž oba rodiče mají základní vzdělání.

10. Pro proměnné VZDEL_M a SIDLO sestavte kontingenční tabulku a simultánní četnosti znázorněte též graficky,. Na asymptotické hladině významnosti testujte hypotézu, že proměnné VZDEL_M a SIDLO jsou nezávislé.
Nepovinný úkol: Tentýž úkol proveďte pro proměnnou VZDEL_O.

Povinný úkol:

Kontingenční tabulka a graf simultánní četnosti pro proměnné VZDEL_M a SIDLO:

Summary Frequency Table			
VZDEL_M	SIDLO MESTO	SIDLO VENKOV	Row Totals
Z	181	180	361
S	225	161	386
V	67	42	109
All Grps	473	383	856



Ověříme podmínku dobré aproximace (tzn., že teoretické četností mají být aspoň v 80% případů větší než 5 a ve zbylých 20% případů nemají klesnout pod 2).

Summary Table: Expected Frequencies			
VZDEL_M	SIDLO MESTO	SIDLO VENKOV	Row Totals
Z	199,4778	161,5222	361,0000
S	213,2921	172,7079	386,0000
V	60,2301	48,7699	109,0000
All Grps	473,0000	383,0000	856,0000

Všechny hodnoty jsou větší než 5, podmínka je tedy splněna. Testujme hypotézu, že proměnné VZDEL_M a SIDLO jsou nezávislé.

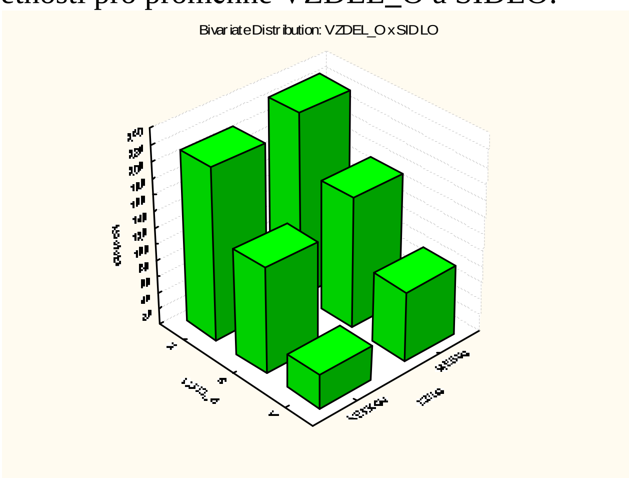
	Statistics: VZDEL_M(3) x SIDLO(2)		
Statistic	Chi-square	df	p
Pearson Chi-square	6,962463	df=2	p=,03077

p-hodnota testové statistiky je menší než 0,05, proto na asymptotické hladině významnosti 0,05 zamítáme hypotézu, že proměnné VZDEL_M a SIDLO jsou nezávislé.

Nepovinný úkol:

Kontingenční tabulka a graf simultánní četnosti pro proměnné VZDEL_O a SIDLO:

Summary Frequency Table			
VZDEL_O	SIDLO MESTO	SIDLO VENKOV	Row Totals
Z	227	211	438
S	161	130	291
V	85	42	127
All Grps	473	383	856



Ověříme podmínku dobré aproximace (tzn., že teoretické četností mají být aspoň v 80% případů větší než 5 a ve zbylých 20% případů nemají klesnout pod 2).

Summary Table: Expected Frequencies			
VZDEL_O	SIDLO MESTO	SIDLO VENKOV	Row Totals
Z	242,0257	195,9743	438,0000
S	160,7979	130,2021	291,0000
V	70,1764	56,8236	127,0000
All Grps	473,0000	383,0000	856,0000

Všechny hodnoty jsou větší než 5, podmínka je tedy splněna. Testujme hypotézu, že proměnné VZDEL_O a SIDLO jsou nezávislé.

	Statistics: VZDEL_O(3) x SIDLO(2)		
Statistic	Chi-square	df	p
Pearson Chi-square	9,083735	df=2	p=,01066

p-hodnota testové statistiky je menší než 0,05, proto na asymptotické hladině významnosti 0,05 zamítáme hypotézu, že proměnné VZDEL_O a SIDLO jsou nezávislé.