

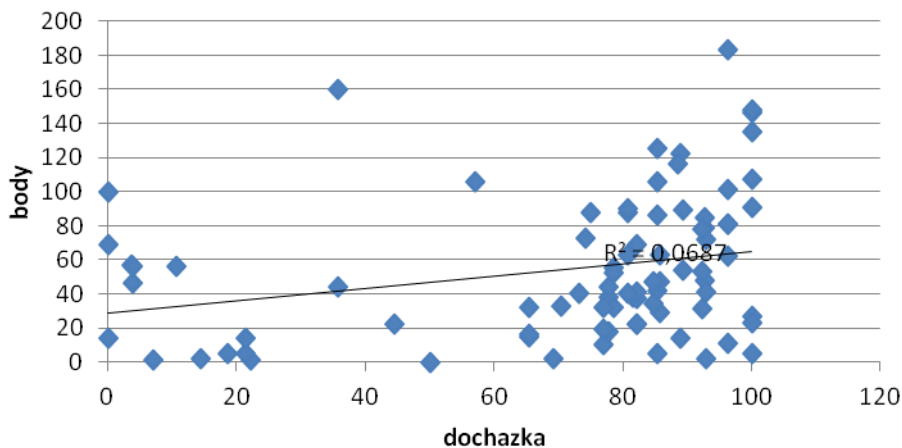
Korelace

Komentované řešení pomocí MS Excel

Vstupní data

	A	B
1	dochazka	body
2	70,4	33
3	100	135
4	92,6	79
5	85,2	42
6	100	27
7
84	96,2	81

Bodový graf



- Tabulka se vstupními daty je umístěna v oblasti A2:B84 (viz. obrázek)
- Prvotní představu o tvaru a síle závislosti docházky a počtu bodů nám poskytne bodový graf, do kterého navíc přidáme graf regresní funkce (detailní postup tvorby viz. úloha lineární regrese)
- Z bodového grafu je patrné, že mezi docházkou a počtem bodů je pozitivní lineární závislost. Tato závislost je však dosti slabá, o čemž svědčí i velmi malý koeficient determinace $R^2 = 0,0687$. Pomocí docházky by se nám tedy podařilo vysvětlit pouze necelých 7 % variability počtu bodů. Z převážné části jsou tedy bodové rozdíly jednotlivých studentů ovlivněny jinými faktory.
- Z bodového grafu je také vidět, že studenty lze rozdělit v zásadě do dvou skupin – na ty, kteří chodí pravidelně (účast nad 70 %), těch je většina, a na ty, kteří nechodí skoro vůbec (účast pod 20 %).

Základní statistiky

	E	F	G
1		dochazka	body
2	Prumer	=PRŮMĚR(A2:A84)	=PRŮMĚR(B2:B84)
3	Směrodatná odchylka (S)	=SMODCH.VÝBĚR(A2:A84)	=SMODCH.VÝBĚR(B2:B84)
4	Počet pozorování (n)	=POČET(A2:A84)	=POČET(B2:B84)
5			
6	korelační koeficient (r)	=CORREL(A2:A84;B2:B84)	

	E	F	G
1		dochazka	body
2	Prumer	71,46	54,65
3	Směrodatná odchylka (S)	29,95	41,41
4	Počet pozorování (n)	83	83
5			
6	korelační koeficient (r)	0,2621	

- Abychom získali základní představu o studovaných datech, spočítáme si průměr (míra polohy), směrodatnou odchylku (míra variability) a Pearsonův korelační koeficient (síla lineární závislosti).

Histogram – docházka

Postup:

- Chceme-li použít test nenulovosti korelačního koeficientu, musíme nejdříve ověřit normalitu obou proměnných (ve skutečnosti bychom měli ověřit dvourozměrnou normalitu, to je však obtížné). K ověření (jednorozměrné) normality použijeme histogram a Q-Q graf. Tvorba obou grafů je detailně popsána v prezentaci k úloze na jednovýběrové testy.

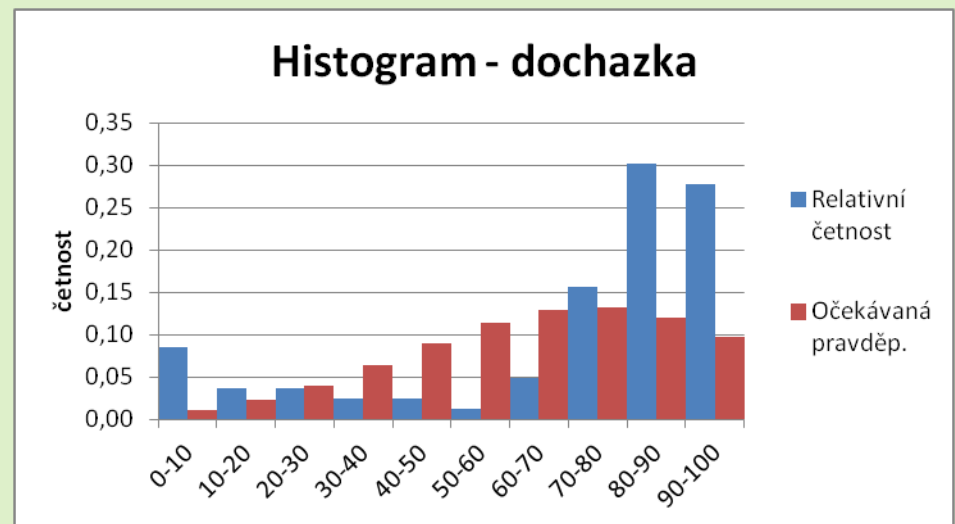
	E	F	G	H	I
20	Interval - dochazka				
21	Dolní mez	Horní mez	Popis	Relativní četnost	Očekávaná pravděpodobnost
22	0	10	0-10	0,08	0,01
23	10	20	10-20	0,04	0,02
24	20	30	20-30	0,04	0,04
25	30	40	30-40	0,02	0,06
26	40	50	40-50	0,02	0,09
27	50	60	50-60	0,01	0,11
28	60	70	60-70	0,05	0,13
29	70	80	70-80	0,16	0,13
30	80	90	80-90	0,30	0,12
31	90	100	90-100	0,28	0,10
32	Celkem			1	

22 =COUNTIFS(\$A\$2:\$A\$84;">="&\$E22;\$A\$2:\$A\$84;"<="&\$F22)/POČET(\$A\$2:\$A\$84)

22 =NORMDIST(F22;\$F\$2;\$F\$3;1)-NORMDIST(E22;\$F\$2;\$F\$3;1)

Interpretace výsledků

- Z histogramu je patrné, že empirické rozdělení docházky (modré sloupce) se výrazně liší od normálního rozdělení (červené sloupce).
- Empirické rozdělení je bimodální, čímž se potvrdila naše domněnka o existenci dvou skupin studentů, s pravidelnou účastí a s minimální účastí.
- Z výše uvedeného vyplývá, že **normalitu v případě docházky předpokládat nemůžeme.**

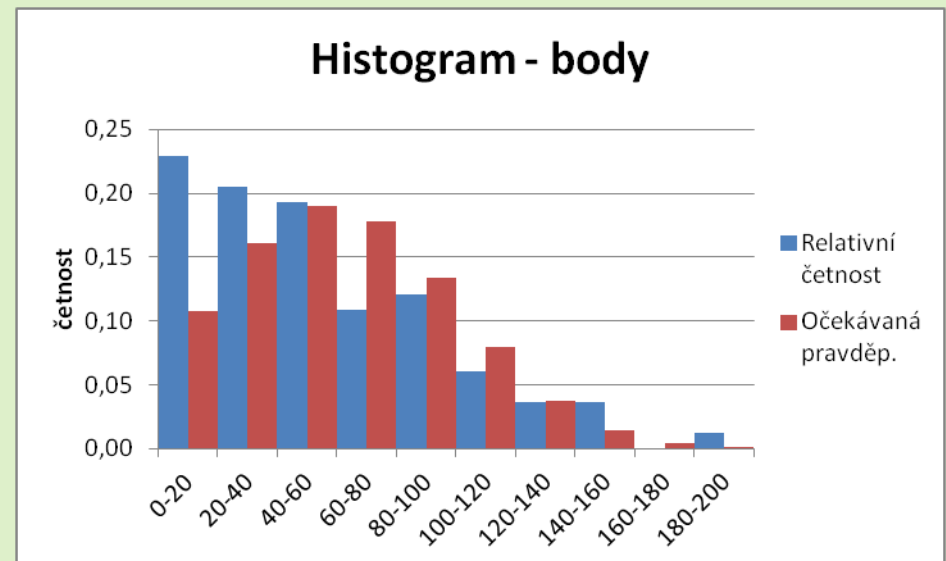


Histogram – body

Interpretace výsledků

- Ve srovnání s docházkou je rozdělení počtu bodů blíže normálnímu rozdělení. Ale i v případě počtu bodů jsou jisté rozdíly oproti normálnímu rozdělení patrné. Hlavní rozdíl je zjevné zešíkmení empirického rozdělení, zatímco normální rozdělení je symetrické.
- Pro lepší představu se ještě podíváme na Q-Q graf (viz. další snímek).

	E	F	G	H	I
36	Interval - body				
37	Dolní mez	Horní mez	Popis	Relativní četnost	Očekávaná pravděpodobnost
38	0	20	0-20	0,23	0,11
39	20	40	20-40	0,20	0,16
40	40	60	40-60	0,19	0,19
41	60	80	60-80	0,11	0,18
42	80	100	80-100	0,12	0,13
43	100	120	100-120	0,06	0,08
44	120	140	120-140	0,04	0,04
45	140	160	140-160	0,04	0,01
46	160	180	160-180	0,00	0,00
47	180	200	180-200	0,01	0,00
48	Celkem			1	

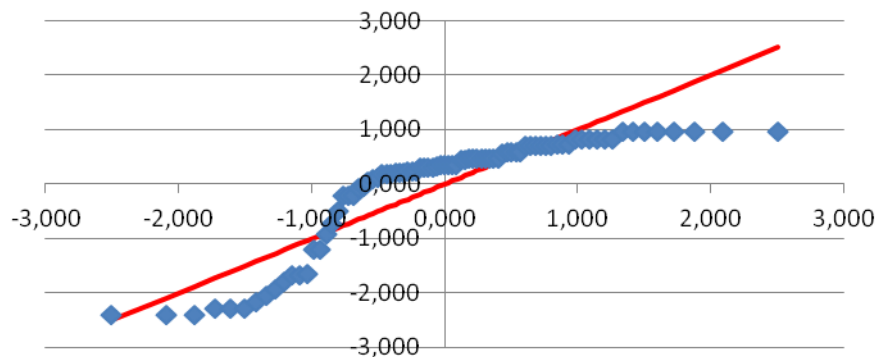


Q-Q grafy

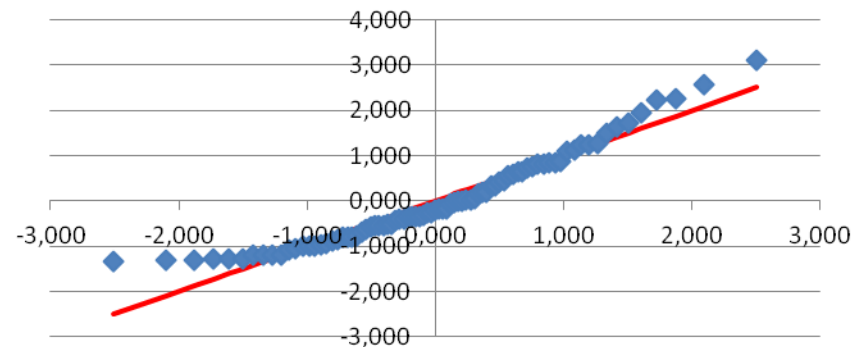
	E	F	G	H	I	J	K
54	Pořadí	Uspořádané hodnoty		Kumulativní	Kvantil N(0,1)	Stadnardizované hodnoty	
55	hodnoty	dochazka	body	pravděp.		dochazka	body
56	1	0	0	0,006	-2,511	-2,400	-1,328
57	2	0	1	0,018	-2,095	-2,400	-1,304
58	3	0	1	0,030	-1,879	-2,400	-1,304
59	4	3,6	2	0,042	-1,726	-2,279	-1,279
60	5	3,8	2	0,054	-1,605	-2,273	-1,279
61
138	83	100	183	0,994	2,511	0,959	3,119

	H	I	J	K
56	$=(E56-0,5)/POČET(ŠEŠ56:ŠEŠ138)$	$=NORMSINV(H56)$	$=(F56-PRŮMĚR(FŠ56:FŠ138))/SMODCH(FŠ56:FŠ138)$	$=(G56-PRŮMĚR(GŠ56:GŠ138))/SMODCH(GŠ56:GŠ138)$

Q-Q graf - dochazka



Q-Q graf - body



Interpretace výsledků

- Q-Q grafy potvrzují naše předchozí zjištění. Rozdělení docházky se výrazně liší od normálního. Rozdělení bodů se více podobá normálnímu rozdělení, ale i tak je zde patrný systematický rozdíl.
- Předpoklad normality tedy přijmout nemůžeme** a nadále tedy musíme s **testem korelace** pracovat jako s **asymptotickým testem** a výsledky musíme interpretovat opatrněji.

Asymptotický test nezávislosti

Postup

- Dosazením do vzorce spočítáme testovou statistiku a jelikož nemáme předem zadánu požadovanou hladinu významnosti, dopočítáme p-hodnotu pro obě alternativy - oboustrannou pro úlohu A) a jednostrannou pro úlohu B) . K tomu se použije funkce **TDIST**.

	E	F
8	Testova statistika	= $\$F\$6 * \text{ODMOCNINA}(\$F\$4 - 2) / \text{ODMOCNINA}(1 - \$F\$6^2)$
9		
10		p-hodnota
11	Úloha A)	= TDIST ($\$F\$8; \$F\$4 - 2; 2$)
12	Úloha B)	= TDIST ($\$F\$8; \$F\$4 - 2; 1$)

	E	F
8	Testova statistika	2,44450
9		
10		p-hodnota
11	Úloha A)	0,01668
12	Úloha B)	0,00834

Závěr

- **Úloha A)** Na základě spočítané **p-hodnoty 0,0167** můžeme zamítnout na hladině významnosti 5 % nulovou hypotézu o nezávislosti ve prospěch oboustranné alternativy. **Prokázali jsme tedy existenci lineární závislosti**. Připomeňme ovšem, že na hladině významnosti 1 % už bychom nulovou hypotézu o nezávislosti nezamítali. Navíc spočítaná p-hodnota je jen přibližná, protože se jedná o asymptotický. Závěr o existenci lineární závislosti je tedy na hranici prokazatelnosti.
- **Úloha B)** V případě jednostranné alternativy vychází **p-hodnota 0,0083**. Tvrzení o **existenci pozitivní lineární závislosti** docházky a počtu bodů je tedy **prokázáno i na hladině významnosti 1 %**. I v tomto případě se však jedná pouze o přibližnou p-hodnotu. Prokázali jsme tedy, že dobrá docházka má pozitivní vliv na výsledky studentů v testech. Tento vliv je však relativně slabý.