

## Korelace

Menu:	QCExpert	Korelace
-------	----------	----------

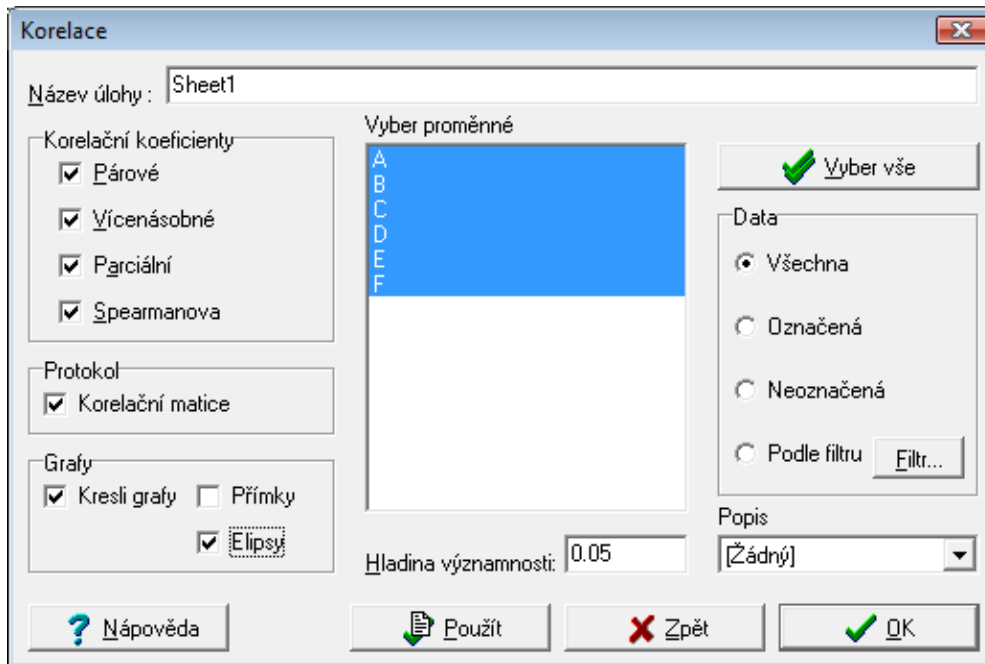
Korelační analýza je jednou z důležitých informací při analýze možných vztahů mezi měřenými veličinami. Pomocí výběrových korelačních koeficientů  $r$  se snažíme především rozhodnout, zda příslušný „skutečný“ (populační) korelační koeficient  $\rho$  je nenulový, což by znamenalo statistické prokázání (na zvolené hladině významnosti) souvislosti mezi příslušnými proměnnými  $r_{A,B}$ . Je doporučena jako jedna ze základních pomocných metod pro řízení jakosti. Korelace vyjadřuje v podstatě míru lineární závislosti (korelace) mezi proměnnými. Korelace se vyjadřuje pomocí korelačních koeficientů  $r_{A,B}$ . Indexy  $A, B$  říkají, že se jedná o korelaci mezi proměnnými  $A$  a  $B$ . QCExpert™ počítá tři druhy korelačních koeficientů: párové, parciální a vícenásobné. Jejich podrobnější význam je uveden níže v odstavci 0. Párový a parciální korelační koeficient může obecně nabývat hodnot od -1 do +1. Hodnoty blízké +1 ukazují na silnou lineární závislost mezi dvěma proměnnými. Kladné znaménko  $r_{A,B}$  znamená, že s rostoucím  $A$  má tendenci růst i  $B$ , záporné znaménko  $r_{A,B}$  znamená, že s rostoucím  $A$  má  $B$  tendenci klesat. Záporné znaménko korelačního koeficientu nesmí být zaměňováno s nepřímou úměrou. I v případě negativního korelačního koeficientu se jedná spíše o přímou úměru, jen se záporným znaménkem u konstanty úměrnosti, tedy  $B = -k.A$ , na rozdíl od nepřímé úměry  $B = k/A$  ( $k > 0$ ). Hodnoty vyhodnocené jako statisticky nevýznamné znamenají, že nelze rozhodnout, zda se proměnné  $A$  a  $B$  ovlivňují. O tom, jak velkou hodnotu korelačního koeficientu je již možné považovat za statisticky významnou závislost, lze rozhodnout na základě testu významnosti korelačního koeficientu (jeho statistické odlišnosti od nuly), podobně jako u autokorelačního koeficientu v modulu Základní statistice. Korelační koeficient mezi stejnými proměnnými (např.  $r_{A,A}$ ) je vždy 1 a ve výsledcích se neuvádí. Pořadí  $A$  a  $B$  lze u párových a parciálních koeficientů zaměnit, proto se výsledcích uvádí pro každou dvojici jen jeden koeficient. Vícenásobný korelační koeficient vyjadřuje míru závislosti zvolené proměnné na všech ostatních. S rostoucím počtem proměnných vždy roste (nebo alespoň neklesá). Spearmanovy korelační koeficienty se orientačně používají v případě přítomnosti vybočujících měření nebo nelineárních závislostí, kdy bývají spolehlivější než klasické koeficienty.

### Poznámka:

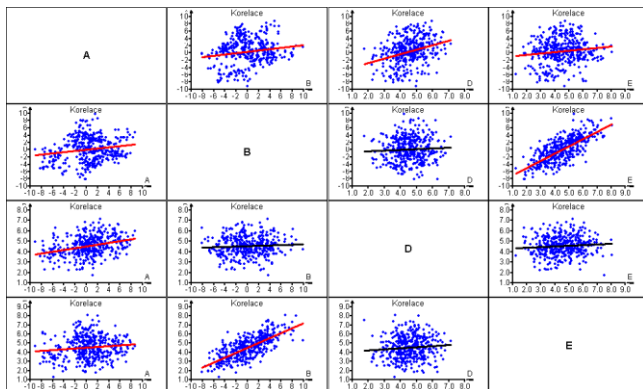
Korelační analýzu je třeba provést, měří-li se několik kvantitativních znaků jednoho výrobku, z nichž je pak třeba konstruovat regulační diagramy. Významně korelované znaky nelze vynášet do samostatných (např. Shewhartových) diagramů  $x$ -průměr, neboť ty předpokládají nezávislost rozdělení znaku a neposkytovaly by správné výsledky. Pro sledování vzájemně korelovaných znaků jakosti je třeba použít Hotellingův diagram pro více proměnných.

## Data a parametry

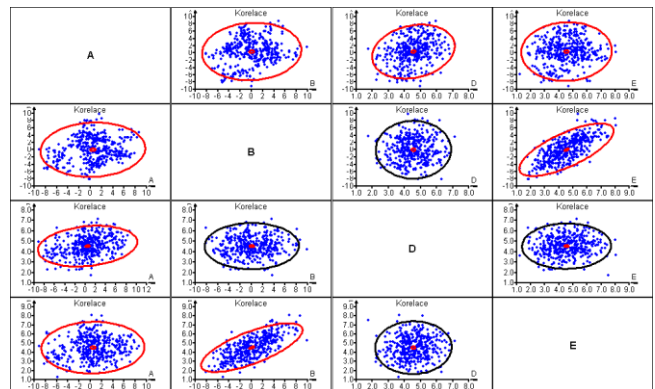
Data jsou ve sloupcích, každý sloupec odpovídá jedné proměnné. Počet dat v jednotlivých sloupcích musí být stejný. Řádky s chybějícími hodnotami budou z výpočtu vyloučeny. Minimální počet sloupců je 2. Minimální počet řádků je 3. Názvy sloupců by měly odpovídat názvům analyzované proměnné, např. Obsah\_Cr, Obsah\_Mn, Obsah\_S. Výběr sloupců z aktuálního listu se může provést v poli *Vyber sloupce* dialogového panelu *Korelační analýza*, Obrázek 1. Implicitně jsou vybrány všechny sloupce obsažené v aktuálním listu. Podle zaškrtnutých položek *Korelační koeficienty* se vypočítají párové, parciální nebo vícenásobné korelační koeficienty. Hladina významnosti je použita při testování významnosti korelačních koeficientů. (Nezaměňovat hladinu významnosti s významností korelace). Podmínkou (předpokladem) pro správné určení korelačních koeficientů (mimo Spearmanových) je normalita dat a nepřítomnost vybočujících měření. Tyto předpoklady je třeba ověřit v modulu základní statistika. Políčko *Kreslit grafy* určuje, zda se vytvoří matice grafů párových závislostí. Je-li zatrhnuto políčko *Přímky*, zobrazí se v grafech regresní přímky. Je-li zatrhnuto políčko *Elipsy*, zobrazí se v grafech konfidenční elipsy pro data a pro aritmetický průměr.



Obrázek 1 Korelace, vstupní dialogový panel s vybranými sloupci



Obrázek 2 Grafy párových korelací s regresní přímkou



Obrázek 3 Grafy párových korelací s konfidenční elipsou

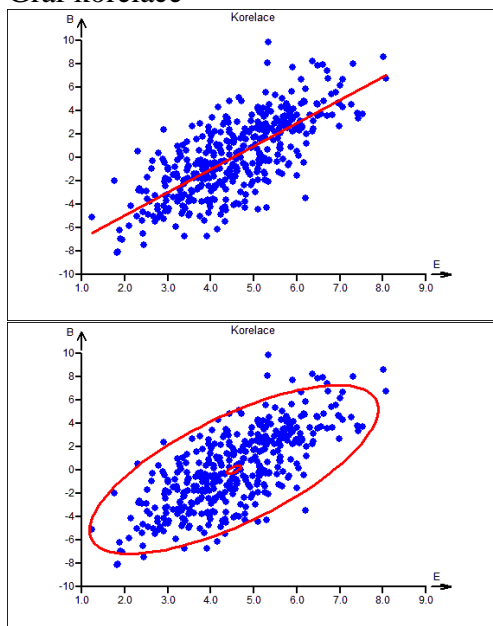
## Protokol

Počet řádků Počet sloupců	Počet platných řádků. Vybraný počet sloupců pro výpočet, tedy proměnných.																									
Korelační matice	Tabulka klasických (Pearsonových), parciálních a Spearmanových korelací ve formě korelační matice. Statisticky významné korelační koeficienty jsou označeny červeně a tučně. Tyto tři tabulky jsou generovány jen pokud je zaškrtnutý box <i>Korelační matice</i> , vždy na začátku v příslušném odstavci.																									
	<table border="1"> <thead> <tr> <th>Proměnná</th> <th>A</th> <th>B</th> <th>D</th> <th>E</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>1</td> <td><b>0.17361</b></td> <td>0.32615</td> <td><b>0.13051</b></td> </tr> <tr> <td>B</td> <td><b>0.17361</b></td> <td>1</td> <td>0.05236</td> <td><b>0.73488</b></td> </tr> <tr> <td>D</td> <td><b>0.32615</b></td> <td>0.05236</td> <td>1</td> <td>0.08633</td> </tr> <tr> <td>E</td> <td><b>0.13051</b></td> <td><b>0.73488</b></td> <td>0.08633</td> <td>1</td> </tr> </tbody> </table>	Proměnná	A	B	D	E	A	1	<b>0.17361</b>	0.32615	<b>0.13051</b>	B	<b>0.17361</b>	1	0.05236	<b>0.73488</b>	D	<b>0.32615</b>	0.05236	1	0.08633	E	<b>0.13051</b>	<b>0.73488</b>	0.08633	1
Proměnná	A	B	D	E																						
A	1	<b>0.17361</b>	0.32615	<b>0.13051</b>																						
B	<b>0.17361</b>	1	0.05236	<b>0.73488</b>																						
D	<b>0.32615</b>	0.05236	1	0.08633																						
E	<b>0.13051</b>	<b>0.73488</b>	0.08633	1																						
Párové korelace	Hodnoty párových korelačních koeficientů. Koeficienty statisticky významné na zvolené hladině jsou vyznačeny tučně. V levém sloupci jsou dvojice proměnných, mezi kterými se koeficient počítá. Kladná významná hodnota																									

	koeficientu znamená, že s rostoucí jednou proměnnou roste i druhá proměnná. Záporná významná hodnota znamená, že roste-li jedna proměnná, druhá proměnná klesá. Pořadí ve dvojici lze zaměnit. Statisticky významné koeficienty jsou vyznačeny červeně a tučně.
Parciální korelace	Jsou podobné párovým koeficientům, ale vyjadřují vzájemnou korelaci s vyloučením vlivu všech ostatních proměnných. Mají tedy smysl jen v případě více než dvou proměnných. Parciální koeficienty obvykle věrněji vyjadřují skutečnou závislost mezi proměnnými nezkreslenou vlivem závislostí mezi ostatními proměnnými.
Vícenásobné korelace	Vyjadřují míru lineární závislosti zvolené proměnné na všech ostatních dohromady. Tyto koeficienty jsou vyšší než největší odpovídající párový koeficient a vždy rostou s počtem proměnných, i když jsou párové korelace nevýznamné a obvykle nadhodnocují skutečnost. Statisticky významné koeficienty jsou vyznačeny červeně a tučně.
Spearmanovy koeficienty	Neparametrické odhady párových korelací založené na pořadí místo na hodnotách dat. Tyto koeficienty jsou robustní a doporučují se v případě přítomnosti vybočujících měření. Jsou analogií mediánu nebo kvantilových odhadů Hornové v Základní statistice. Statisticky významné koeficienty jsou vyznačeny červeně a tučně.

## Graf

### Graf korelace



Grafy představují vzájemné párové vztahy mezi proměnnými. Nazývají se rovněž rozptylovými grafy a korespondují s párovými korelačními koeficienty. Výhodou grafů je schopnost odhalit případnou nelineární závislost, která se neprojeví v korelačních koeficientech. Jednotlivé grafy reprezentují jednotlivé prvky korelační matice dat. Je-li před výpočtem zatrženo příslušné políčko *Přímky* nebo *Elipsy*, znázorní se v grafu také orientačně regresní přímka, nebo konfidenční elipsa. Větší elipsa odpovídá zvolenému  $(1-\alpha)$ -kvantilu dat a menší, vnitřní elipsa označuje párový  $(1-\alpha)$ -interval spolehlivosti aritmetického průměru příslušných dvou proměnných,  $\alpha$  je hladina významnosti zadaná v dialogovém panelu. Při významné párové korelaci je přímka, resp. elipsa červená, při nevýznamné korelaci je černá.