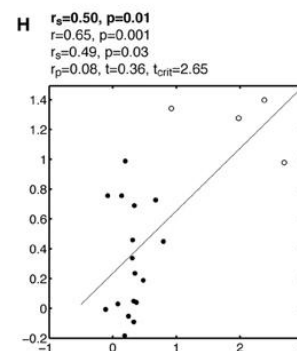
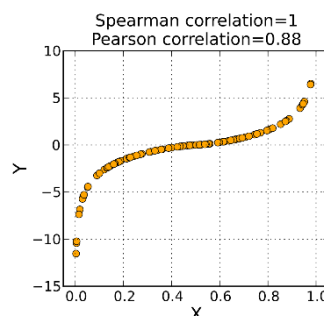
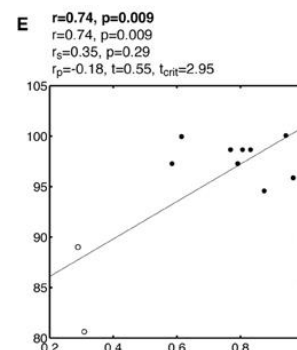
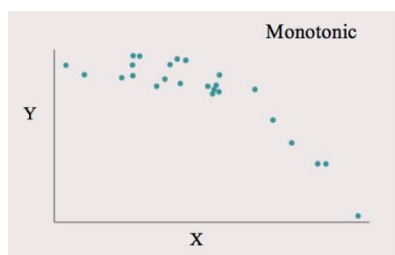


**Korelační analýza** popisuje vztah spojitéch proměnných (obvykle velikost vlivu nezávisle proměnné na závisle proměnnou).

Parametrická korelační analýza předpokládá **lineární vztah** mezi proměnnými a normální rozložení obou proměnných (tedy neslouží jako důkaz linearity). Mírou tohoto lineárního vztahu je **Pearsonův korelační koeficient  $r$** . Leží v rozmezí od -1 do 1. Absolutní hodnota větší než 0,5 značí korelaci (a větší než 0,9 silnou korelaci). I malé  $r$  nicméně může být statisticky významně odlišné od nuly, což neznamená silnou korelaci, jen značí, že korelace není způsobena náhodou (stává se to u velkých souborů dat).

U korelační analýzy je potřeba data vizualizovat pomocí bodového grafu, který odhalí případy „falešně velkého“  $r$ : odlehlé hodnoty a dvě různé skupiny dat.



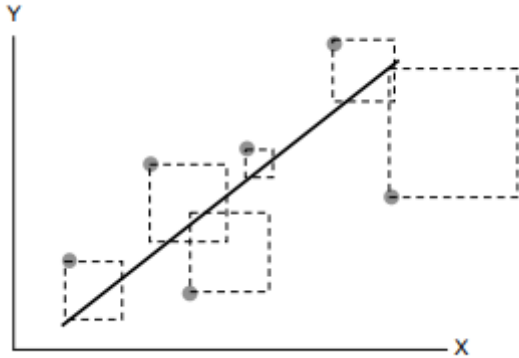
Neparametrická korelační analýza – pro jiné než lineární, ale jen **monotónní** závislosti – využívá **Spearmanův korelační koeficient  $r_s$** . Spearmanův koeficient není citlivý na odlehlé hodnoty (vyjde malý a nevýznamný).

**$R^2$  (koeficient determinace)** značí podíl celkové variability, který je vysvětlený zvoleným (lineárním nebo jiným) modelem. Pro lineární model je  $R^2$  rovno  $r^2$ .

Kalibrační závislosti a jiné predikce

**Regresní analýza** predikuje hodnotu závisle proměnné ( $y$ ) na základě známých hodnot nezávisle proměnných (prediktorů,  $x$ ). Principem je pomocí sady vzorků se známými hodnotami  $x$  i  $y$  (např. kalibračních vzorků) nalézt nejlepší predikční funkci, tu nazveme **regresní funkce** nebo regresní **model**. Následně nám stačí měřit  $x$  a z regresní funkce dopočítáme  $y$ . Model by se ale měl používat jen v tom rozmezí obou veličin, kde se pohybovaly ony známé vzorky. Nejčastější modely pro spojitá data jsou lineární, exponenciální, případně polynomický. (Pozn.: exponenciální závislost je možno transformovat na lineární tím, že vyneseme logaritmy hodnot.) Vybíráme vždy nejjednodušší použitelný model (ten, který má dostatečně velký koeficient determinace a zároveň má nejmenší počet koeficientů).

**Lineární regresní model** se používá pro spojité výsledkové proměnné. Regresní koeficienty neboli **parametry** regresní funkce v lineárním modelu (což může být přímka, ale i např. kvadratická funkce) se hledají tzv. **metodou nejmenších čtverců**, která minimalizuje druhé mocniny (aby se chyby obou směrů navzájem nevyvulovaly) rozdílů mezi jednotlivými pozorováními a regresní funkcí.



Lineární regresní funkce:  $Y = \beta_0 + \beta_1x (+ \beta_2x_2\dots) + \varepsilon$

Parametry: Y je hodnota výsledku,  $\beta_0$  je posun neboli absolutní člen (intercept),  $\beta_1$  je směrnice (slope), x je hodnota prediktoru (nebo více prediktorů) a  $\varepsilon$  je odchylka (reziduum).

**Odlehlost bodů** při regresní analýze zjistíme tak, že spočítáme rozdíly experimentálních hodnot a jim odpovídajících hodnot vypočítaných z regresní funkce. Hodnoty rozdílů pak testujeme např. Grubbsovým testem. Odlehlé body vyloučíme a regresní funkci spočítáme znovu bez nich.

Užitečné je zjistit **intervaly spolehlivosti odhadů parametrů**. Pokud se např. překrývají intervaly spolehlivosti směrnic dvou přímek, mohou být tyto přímky rovnoběžné. Pokud se překrývají IS všech parametrů, mohou být závislosti shodné. Pokud IS pro posun zahrnuje nulu, je většinou vhodné najít novou funkci procházející počátkem (obzvláště pokud jde např. o kalibrační závislost pro hodnoty y blízké nule, kdy dává smysl, že nulová hodnota y odpovídá nulové hodnotě x – tj. nemáme šum).

Ve vyšších hodnotách budou pravděpodobně větší i hodnoty čtverců odchylek. Pokud sestrojujeme kalibrační závislost přes několik řádů, měli bychom dát větší váhu menším hodnotám x (a y), jinak by kalibrace byla relativně přesnější ve vysokých hodnotách než v nízkých. Je možné to vyřešit např. sestrojením kalibrační křivky pro logaritmicky transformovaná data.

**Logistický regresní model** se používá pro binární výsledkové proměnné (a spojité prediktory). Místo metody nejmenších čtverců se využívá metoda maximální věrohodnosti. S použitím regresních koeficientů a hodnoty nezávisle proměnné pak spočítáme  $y = \text{logit}(p)$  a z něj výsledek p, tedy pravděpodobnost nastání jevu při dané hodnotě prediktoru (prediktorů).  $\text{logit}(p) = \ln(p/(1-p))$ , tedy  $p = \exp(y) / (1 + \exp(y))$ .

Mez detekce a stanovitelnosti analytických metod

**Mez detekce (limit of detection, LOD)** je nejnižší koncentrace analytu, kterou je metoda schopná spolehlivě zaznamenat (říct, že analyt je ve vzorku přítomný).  $\text{LOD} = (3,33 \times \text{směrodatná odchylka odezvy slepého vzorku}) / \text{směrnice regresní funkce}$ .

**Mez stanovitelnosti (limit of quantitation/quantification, LOQ)** je nejnižší koncentrace analytu, kterou je metoda schopná spolehlivě určit (říct, kolik analytu ve vzorku je).  $\text{LOQ} = (10 \times \text{směrodatná odchylka odezvy slepého vzorku}) / \text{směrnice regresní funkce}$ .

Pokud nemá metoda šum (odezva slepého vzorku je nulová), je možné použít chybu odezvy nejnižší koncentrace, která se statisticky významně liší od nuly, případně chybu parametru  $\beta_0$ .