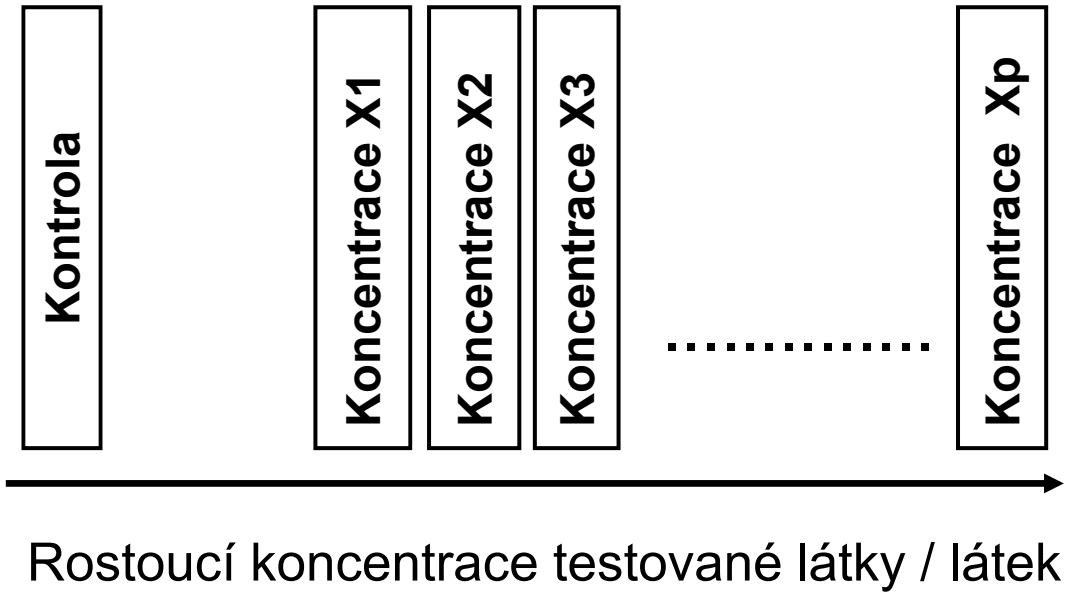
# ANOVA – analýza rozptylu

# *Analýza rozptylu - ANOVA*



Základní technika sloužící k posouzení rozdílů mezi více úrovněmi pokusného zásahu

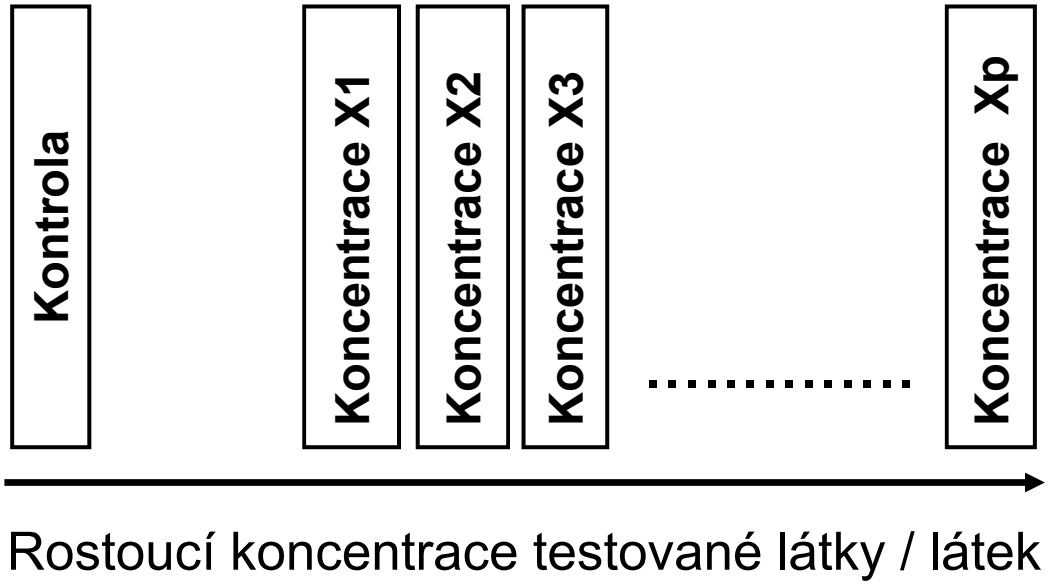



- Celkově významné změny v reakci biologického systému
- Vzájemné rozdíly účinku jednotlivých dávek
- Rozdíly účinku dávek od kontroly

# *Analýza rozptylu - ANOVA*



Významné kroky analýzy, vedoucí k efektivnímu srovnání variant




- Splnění předpokladů analýzy
  Transformace dat
- Relevantnost kontroly
  (vliv vlastní aplikace látek)
- Vhodnost modelu ANOVA pro účely testu
- Vlastní srovnání variant
  Minimalizace chyb při ověřování hypotéz

# *Analýza rozptylu - ANOVA*



***SPLNĚNÍ PŘEDPOKLADŮ ANOVA JE NEZBYTNOU PODMÍNKOU POUŽITÍ TÉTO TECHNIKY***

**ANOVA**
**= parametrická analýza dat**

1. **Předpoklad nezávislosti opakování experimentu**
2. **Homogenita rozptylu v rámci pokusných variant**
3. **Normalita rozložení v rámci pokusných variant**

ALTERNATIVOU JSOU NEPARAMETRICKÉ METODY

# Analýza rozptylu - ANOVA

## Předpoklady analýzy rozptylu jsou nezbytné pro dosažení síly testu

• **Symetrické rozložení hodnot a normalita odchylek** od hodnoceného modelu ANOVA. Velkou část dat lze adekvátně normalizovat použitím logaritmické transformace. Předpoklad lognormální transformace může pochopitelně být teoreticky vyloučen u mnoha datových souborů obsahujících diskrétní parametry, kde je indikována vhodnost jiného typu transformace. U asymetricky rozložených a u diskrétních dat je nutné využít neparametrické alternativy analýzy rozptylu.

• **Homogenita rozptylu** je nutným předpokladem pro smysluplnost vzájemných srovnání pokusných variant. U testů toxicity by splnění tohoto předpokladu mělo být ověřováno (Bartlettův test), neboť vážné rozdíly (až řádové) v jednotkách testovaného parametru mohou nastat v důsledku inhibice dávkami látky. Nehomogenita rozptylu je často ve vztahu k nenormalitě (asymetrii) dat a lze ji odstranit vhodnou normalizující transformací.

• **Statistická nezávislost reziduí** vyhodnocovaného modelu ANOVA. Pokud odhad a posouzení korelačních vztahů mezi pokusnými variantami není přímo předmětem výzkumu, lze jejich vliv na vyhodnocení odstranit znáhodněním dat v rámci pokusných variant - tedy změnou pořadí v náhodné. Rozsah vlivu těchto autokorelačních vztahů musí být ovšem primárně omezen správností experimentálního uspořádání.

• **Aditivita** jako předpoklad týkající se složitějších experimentálních uspořádání. Exaktní otestování aditivity více pokusných faktorů je procedura poměrně náročná na experimentální design vyvážený co do počtu opakování. Je rovněž obtížné testovat interakci na nestandardních datech, neboť případná transformace může změnit charakter odchylek původních dat od hodnoceného modelu ANOVA.

# Analýza rozptylu - ANOVA

## Omezení aplikace ANOVA lze řešit

• **Chybějící data.** Vážným problémem jsou chybějící údaje o celé skupině kombinací testovaných látek, například u faktoriálních pokusů, kdy je znemožněno hodnocení experimentu jako celku.

• **Různé počty opakování** Jde o typický jev pro experimentální datové soubory. Při různých počtech opakování v experimentálních variantách jsou testy ANOVA citlivější na nenormalitu dat. Pokud jsou počty opakování zcela odlišné(až na řádové rozdíly), je nutno použít neparametrické techniky nebo analýzu rozptylu nevyvážených pokusů.

• **Odlehlé hodnoty.** Ojedinělé odlehlé hodnoty musí být před parametrickou analýzou rozptylu vyloučeny.

• **Nedostatek nezávislosti mezi rezidui modelu.** Jde o závažný nedostatek, zkreslující výsledek F-testu. Velmi často je tato skutečnost důsledkem špatného provedení nebo naplánování experimentu.

• **Nehomogenita rozptylu.** Velmi častý nedostatek experimentálních dat, často související s nenormalitou rozložení nebo s odlehlými hodnotami.

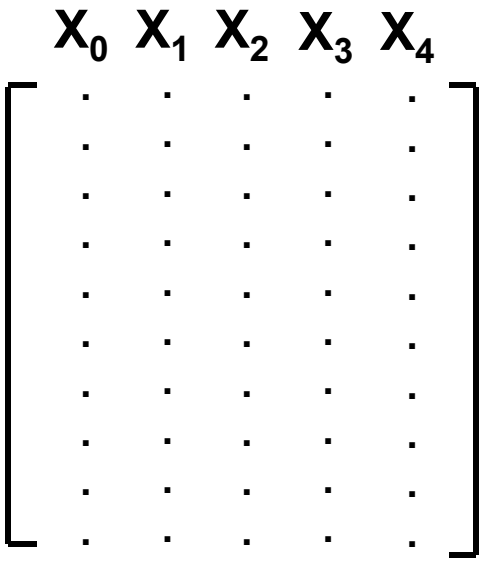
• **Nenormalita dat.** I v tomto případě lz situaci upravit vyloučením odlehlých hodnot nebo normalizující transformací.

• **Neaditivita kombinovaného vlivu více pokusných zásahů.** Tuto situaci lze testovat jednak speciálními testy aditivity nebo přímo F testem kontrolujícím významnost vlivu interakce pokusných zásahů. Při významné interakci je nutné prozkoumat především její charakter ve vhodném experimentálním uspořádání.

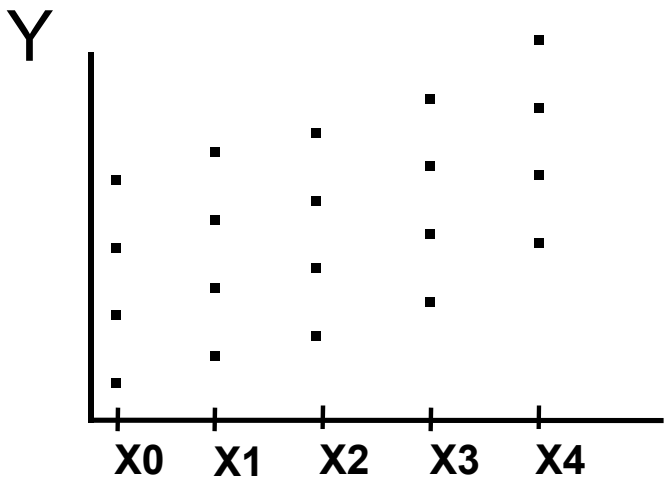
# Modely analýzy rozptylu



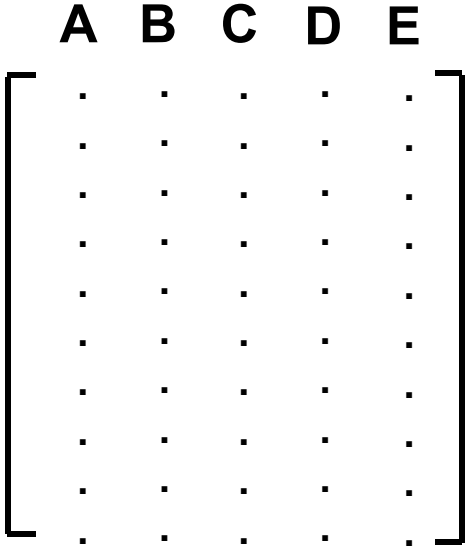
## Model I. Pevný model



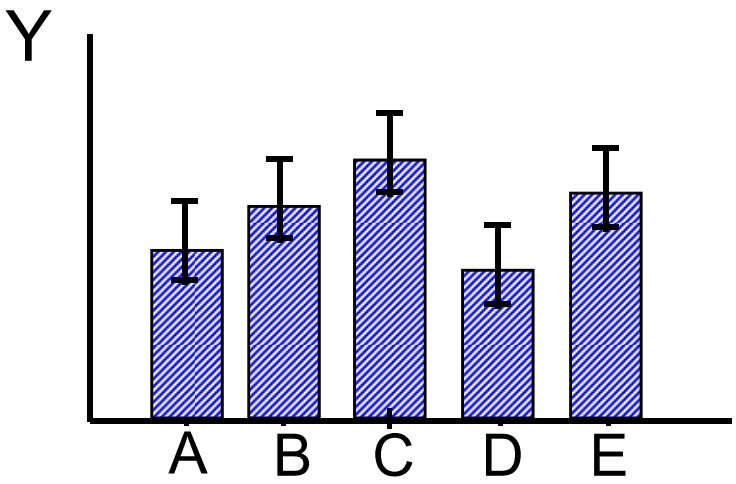

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$




## Model II. Náhodný model




$$y_{ij} = \mu + A_i + \varepsilon_{ij}$$



# ANOVA – základní výpočet

- Základním principem ANOVY je porovnání rozptylu připadajícího na:
  - Rozdělení dat do skupin (tzv. effect, variance between groups)
  - Variabilitu objektů uvnitř skupin (tzv. error, variance within groups), předpokládá se, že jde o náhodnou variabilitu (=error)

1. Variabilita mezi skupinami

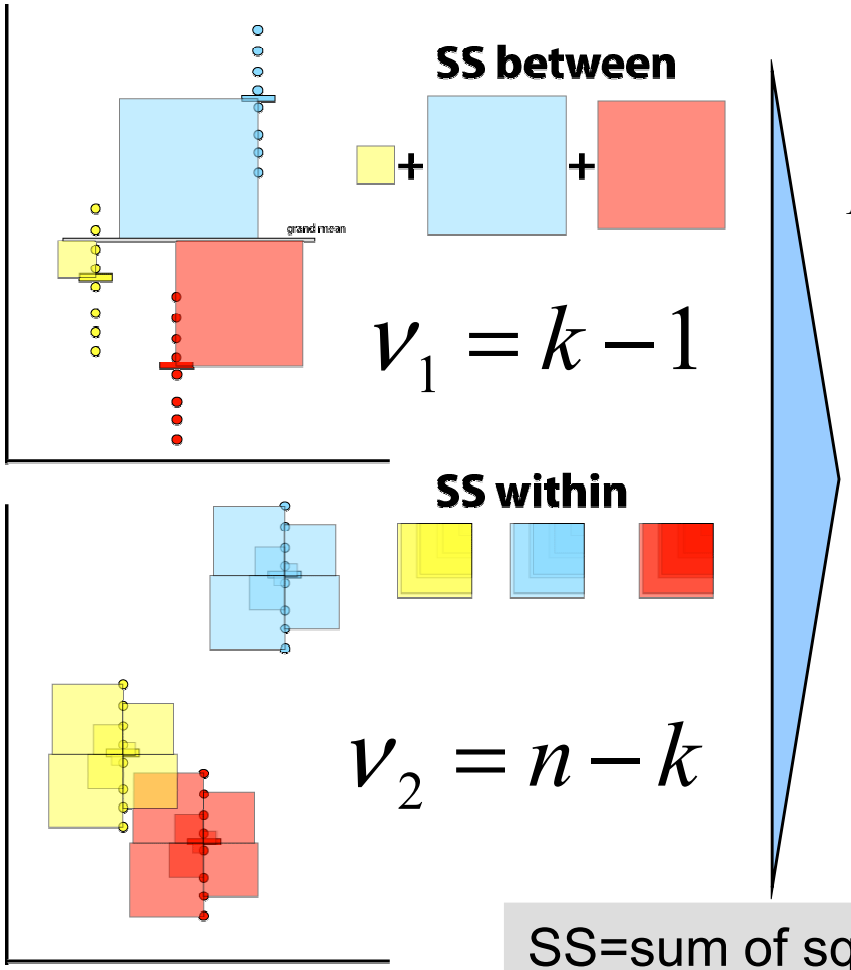
Rozptyl je počítán pro celkový průměr (tzv. grand mean) a průměry v jednotlivých skupinách dat

Stupně volnosti jsou odvozeny od počtu skupin (= počet skupin -1)

2. Variabilita uvnitř skupin

Rozptyl je počítán pro průměry jednotlivých skupin a objekty uvnitř příslušných, celková variabilita je pak sečtena pro všechny skupiny

Stupně volnosti jsou odvozeny od počtu hodnot (= počet hodnot - počet skupin)




$$F = \frac{between\_groups}{within\_groups}$$

Výsledný poměr (F) porovnáme s tabulkami F rozložení pro $v_1$ a $v_2$ stupňů volnosti

SS=sum of squares

# *Modely analýzy rozptylu - základní výstup*



***Základním výstupem analýzy rozptylu je***
***Tabulka ANOVA - frakcionace komponent rozptylu***

| **Zdroj rozptylu** | St. v. | SS | MS | F |
|---|---|---|---|---|
| **Pok. zásah** (mezi skupinami) | **a -1** | **$SS_B$** | **$SS_B/(a-1)$** | **$MS_B/MS_E$** |
| **Uvnitř skupin** | **N - a** | **$SS_E$** | **$SS_E/(N-a)$** | |
| **Celkem** | **N -1** | **$SS_T$** | | |

**$SS_B/SS_T$** → Kvantifikovaný podíl rozdílu mezi pokusnými zásahy na celkovém rozptylu

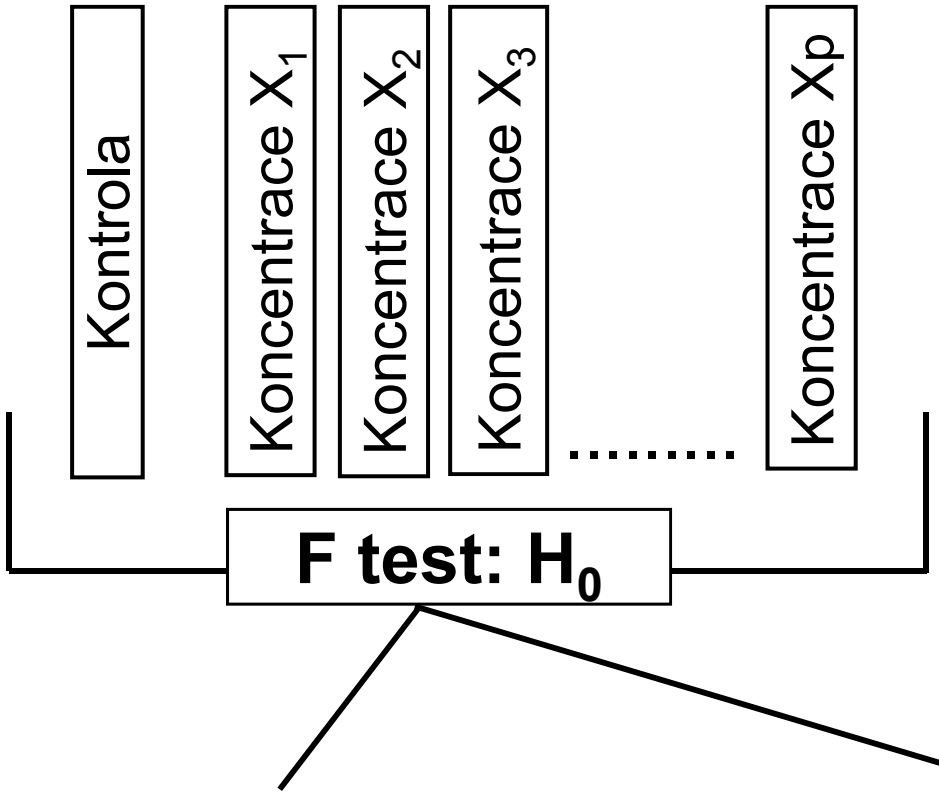
**$MS_B/MS_T$** → Statistická významnost rozdílu

# *Analýza rozptylu - obecný F test*



**obecný F test**

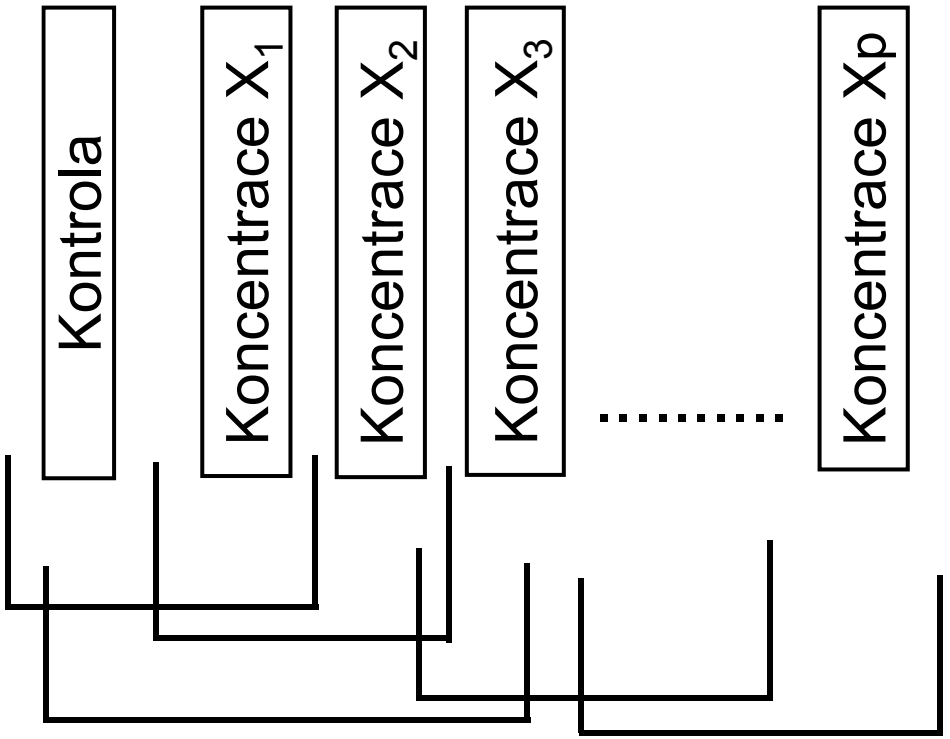
$H_0: m_1 = m_2 = m_3 = .... = m_p$



# Analýza rozptylu - Testy kontrastů



ANOVA:$H_0$ zamítnuta
Testy kontrastů




Plánované

Neplánované

Pro srovnání variant s kontrolou

Testování kontrastů "Multiple range testy"

Parametrické

Neparametrické

**Rozdíly v smysluplných kombinacích ?**

# Příklad: Anova - One way



Dávka rostlinného stimulátoru (0, 4, 8, 12 mg/l)
A = 4 ; n = 8

## I. ANOVA

Bartlett's test: P = 0,9847
K-S test: P = 0,482 - 0,6525 pro jednotlivé kategorie

| Source | D. f. | SS | MS | F |
|---|---|---|---|---|
| **Between Groups** | 3 | 305,8 | 101,9 | 8,56 |
| **Within Groups** | 28 | 322,2 | 11,9 | |
| **Total (corr.)** | 31 | 638,0 | | |

## II. Multiple Range Test

NKS -test

| Level | Average | Homogenous Groups | | |
|---|---|---|---|---|
| 0 | 34,8 | x | | |
| 4 | 41,4 | | x | |
| 12 | 41,8 | | x | |
| 8 | 52,6 | | | x |

# Příklad: Anova - One way



I. Zásah: 4 klinická stadia virové choroby (napadá kr. buňky)

*Sledovaná veličina: aktivita enzymu v těchto krevních buňkách*

$$H_o: \quad \mu_1 = \mu_2 = \mu_3 = \mu_4$$

**n = 3**
**MODEL = ?**

| | I | II | III | IV |
|---|---|---|---|---|
| | 22,8 | 16,4 | 11,2 | 14,2 |
| | 19,4 | 17,8 | 18,2 | 10,1 |
| | 12,5 | 19,1 | 15,8 | 12,8 |
| $\Sigma$ | 65,7 | 53,3 | 45,2 | 37,1 |
| průměr | 21,9 | 17,8 | 15,1 | 12,4 |

II.

| Source | D.f. | MS | F | P |
|---|---|---|---|---|
| **Between groups** | 3 | 49,6 | 8,39 | 0,0075 |
| **Within groups** | 8 | 5,9 | | |
| **Total (corr.)** | 11 | - | | |

III. Komponenta rozptylu:

$$\sigma_A^2 \sim S_A^2 = \frac{MS_A - MS_e}{n} = \frac{49{,}6 - 5{,}9}{3} = 14{,}57$$

$$S_A^2 = 2{,}5 \cdot S_e^2$$

IV.

$$\rho_I \sim r_I = \frac{S_A^2}{S_A^2 + S_e^2} = 0{,}7142$$

# Srovnání variant v testech

## Srovnáváni variant po celkovém testu ANOVA

Mnoho existujících algoritmů není vhodných pro konkrétní případ

**Day and Quin**
**Ecological Monographs,1989**

| Test | Využití | Poznámka |
|---|---|---|
| Dunnett Williams | Srovnání s kontrolou | Ex. i modifikace pro různá n. |
| ANOVA testy (F) | Orthogonální kontrasty | Plánovaná srovnání |
| Ryan Q test | Jednoduché kontrasty | Vyhodnocen jako nejlepší test |

**Testy pro jednoduché kontrasty**

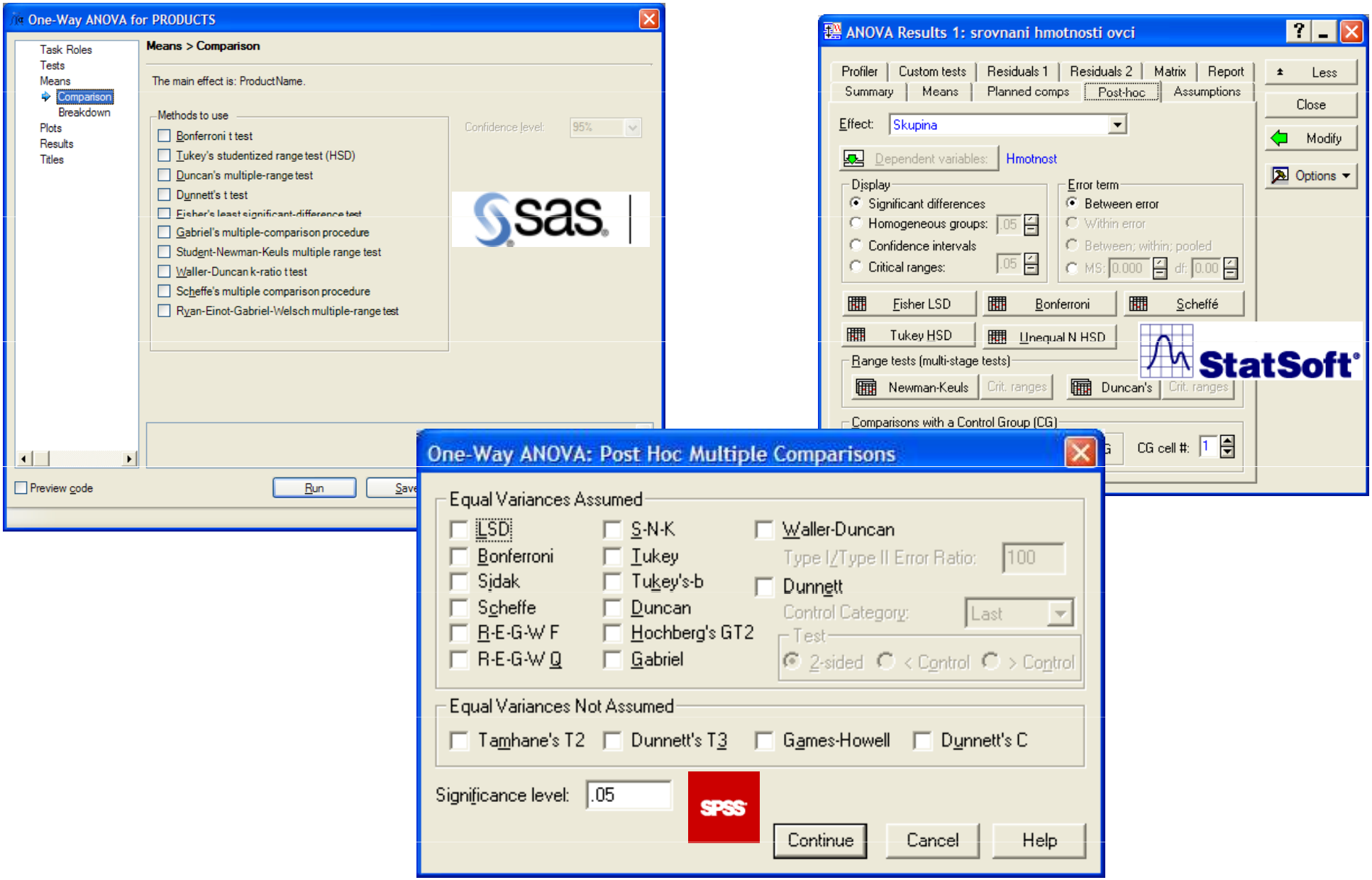
| | | |
|---|---|---|
| Scheffe | Tukey | LSD |
| Bonferroni | Dunn-Sidák | Kramer |

**Testy nevhodné**

| | | |
|---|---|---|
| Duncan | Student - Newmann-Keuls | Waller-Duncan k ratio |

# Řada post-hoc testů v různých SW





# Hypotetické příklady - Multiple Range Tests



| | Level | Homogenous Group |
|---|---|---|
| 15 | 1 | x |
| 18 | 2 | xx |
| 22 | 3 | xx |
| 26 | 4 | x |
| 38 | 5 | x |

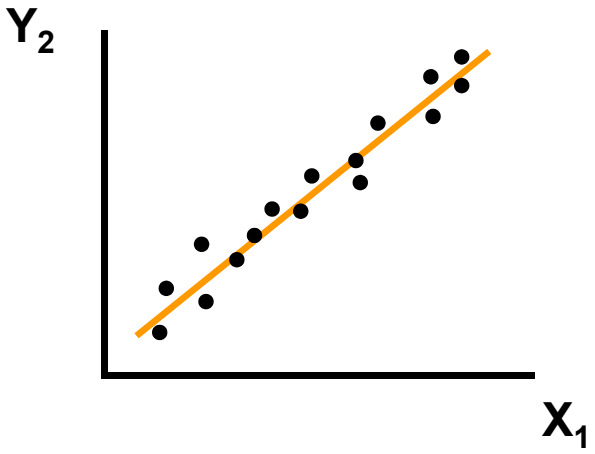
| | Level | Homogenous Group |
|---|---|---|
| 15 | 1 | x |
| 22 | 2 | x |
| 24 | 3 | xx |
| 29 | 4 | x |
| 30 | 5 | x |

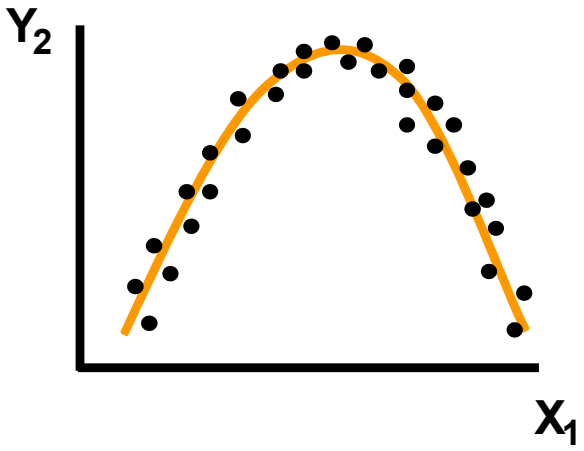
| | Level | Homogenous Group |
|---|---|---|
| 15 | 1 | x |
| 18 | 2 | xx |
| 22 | 3 | x |
| 29 | 4 | x |
| 36 | 5 | x |

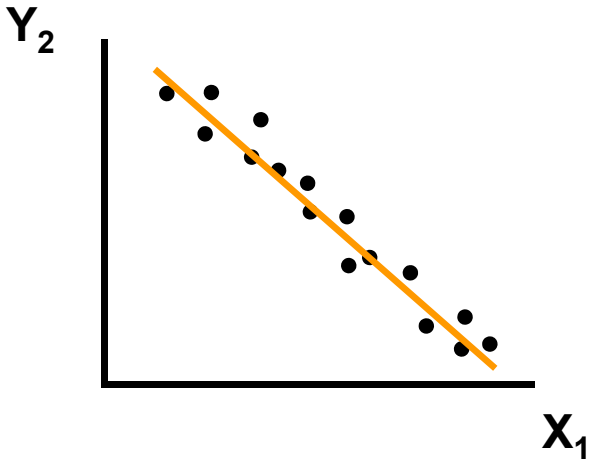
# Korelace a regrese

# Základy korelační analýzy - I.

## Korelace - vztah (závislost) dvou znaků (parametrů)




| $X_2$ \ $X_1$ | ANO | NE |
|---|---|---|
| ANO | a | b |
| NE | c | d |

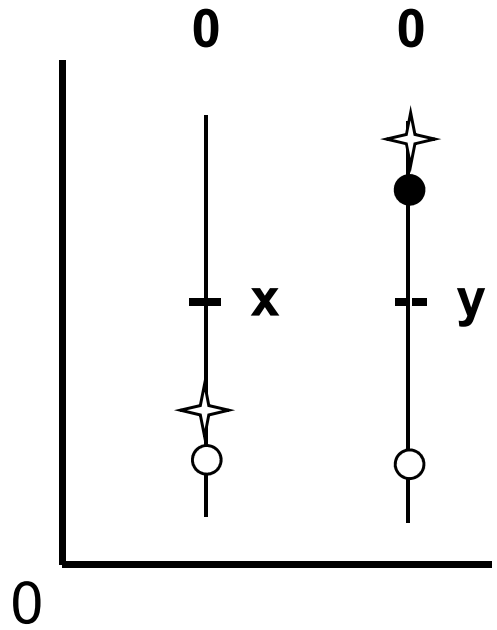
# Základy korelační analýzy - II.

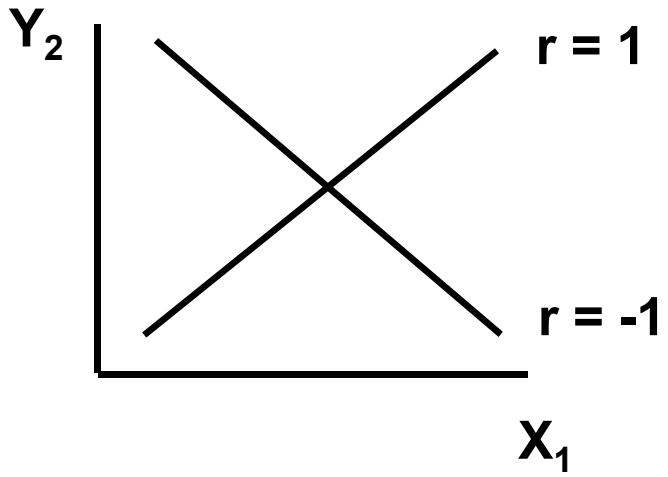
## Parametrické míry korelace

Kovariance

$$Cov(x, y) = E(x_i - \bar{x}).(y_i - \bar{y})$$

Pearsonův koeficient korelace



# Základy korelační analýzy - III.



| $P_I$ (zem) | 10 | 14 | 15 | 32 | 40 | 20 | 16 | 50 |
|---|---|---|---|---|---|---|---|---|
| $P_I$ (rostl.) | 19 | 22 | 26 | 41 | 35 | 32 | 25 | 40 |

$$I = 1,.....,n; n = 8; v = 6$$

$$r = \frac{Cov\,(x,y)}{S_x.S_y} = \frac{\sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i}{\sqrt{\left[\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2\right]\left[\sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2\right]}} = 0,7176$$

I. $H_0 : \rho = \phi : \alpha = 0,05$

**tab :** $r(v=6) = 0,7076$

---

II. $H_0 : \rho = \phi$ $\qquad t = \left[\frac{r}{\sqrt{1-r^2}}\right] \cdot \sqrt{n-2} \qquad v = n-2$

$$t = \frac{0,7176}{0,6965} \cdot \sqrt{6} = 2,524$$

**tab :** $t_{0,975}^{(n-2)} = 2,447$

$\leq$

# Základy korelační analýzy - IV.
# Srovnání dvou korelačních koeficientů (r)

**1.** $n_1 = 1258$, $r_1 = 0{,}682$ **2.** $n_2 = 462$, $r_2 = 0{,}402$

Krevní tlak **x** koncentrace kysl. radikálů

$$Z_i = 1.1513 \cdot \log \frac{(1 + r_i)}{(1 - r_i)}$$

$Z_1 = 0{,}833$ $Z_2 = 0{,}426$

**Test:** $H_0 : \rho_1 = \rho_2$ ; $\alpha = 0{,}05$

$$Z = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} = \frac{0{,}407}{0{,}0545} = 7{,}461$$

**tabulky** : $Z_{0{,}975} = 1{,}96$

**7,461 >> 1,96 => P << 0,01**

# Základy korelační analýzy - V. Neparametrická korelace (rs)

| $P_I$ v půdě | 1 | 2 | 3 | 6 | 7 | 5 | 4 | 8 |
|---|---|---|---|---|---|---|---|---|
| $P_I$ v rostl. | 1 | 2 | 4 | 8 | 6 | 5 | 3 | 7 |
| $d_I$ | 0 | 0 | 1 | 2 | -1 | 0 | -1 | -1 |

**i = 1, ….. n; n = 8 => v = 6**

$$r_s = 1 - \frac{6 \cdot \sum di^2}{n\left(n^2 - 1\right)} = 0,9048$$

**tab :** $r_s(v = 6) = 0,89$

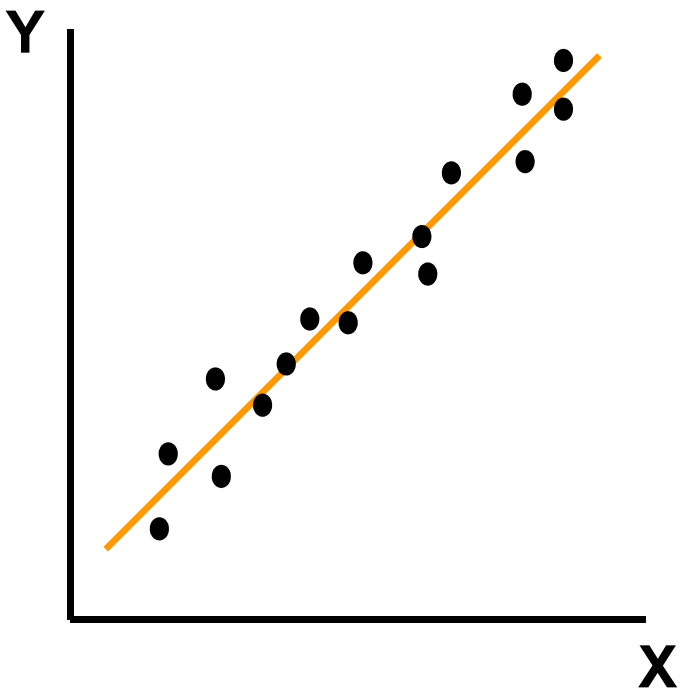
| Pacient č. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Lékař 1 | 4 | 1 | 6 | 5 | 3 | 2 | 7 |
| Lékař 2 | 4 | 2 | 5 | 6 | 1 | 3 | 7 |
| $d_I$ | 0 | -1 | 1 | -1 | 2 | -1 | 0 |

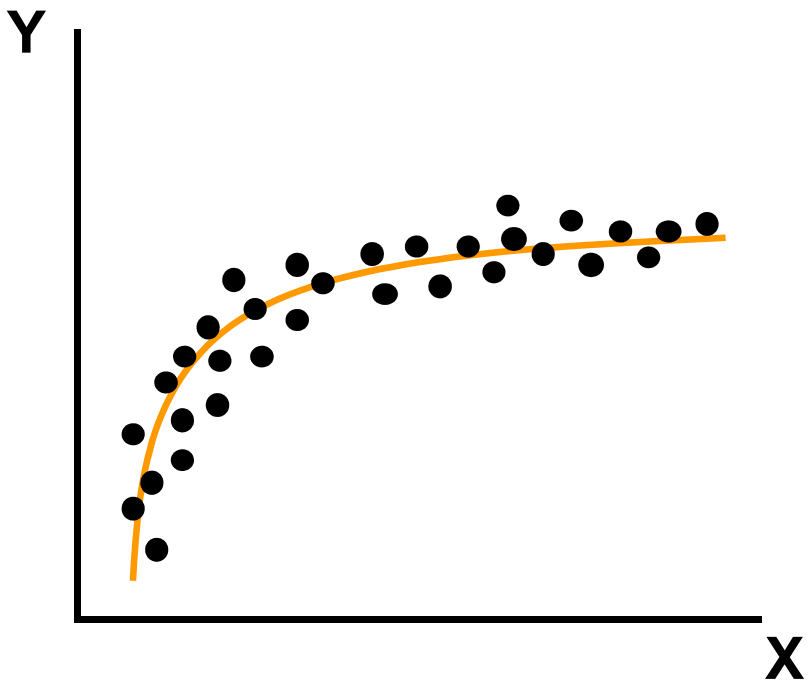
$$r_s = 1 - \frac{6 \cdot 8}{7\left(49 - 1\right)} = 0,857$$

**P = 0,358**

# Korelace v grafech I.






**Vztahy velmi často implikují funkční vztah mezi Y a X.**

$$Y = a + b \,.\, X$$

$$Y = a + b_1 \,.\, X_1 + b_2 \,.\, X_2 + b_3 \,.\, X_3$$

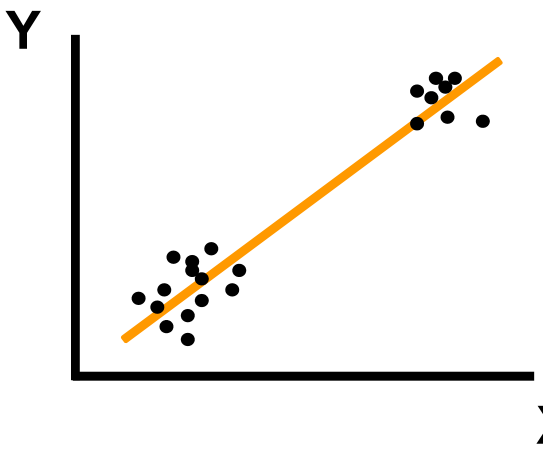
$$Y = a + b_1 \,.\, X_1 + b_2 \,.\, X_2$$

$$Y = a + b_1 \,.\, X_1 + b_2 \,.\, X_2 + b_3 \,.\, X_1 \,.\, X_2$$

# Korelace v grafech II.

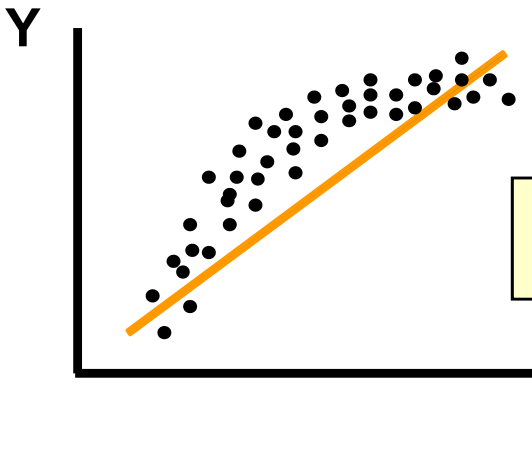


## Problém rozložení hodnot




r = 0,981
(p < 0,001)

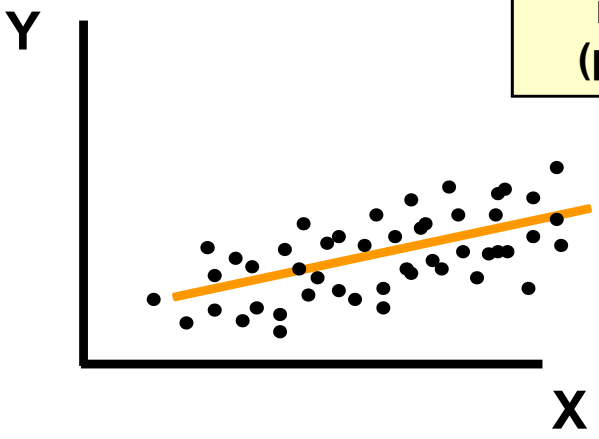
## Problém typu modelu




r = 0,761
(p < 0,032)

## Problém velikosti vzorku


r = 0,891
(p < 0,214)




r = 0,212
(p < 0,008)

# Základy regresní analýzy

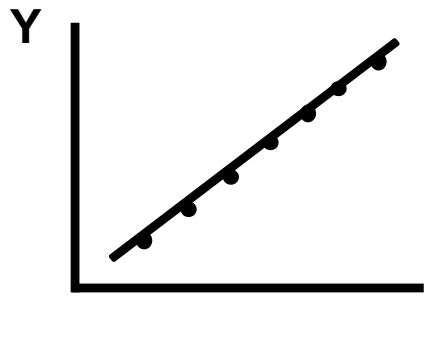
## Regrese - funkční vztah dvou nebo více proměnných

**Jednorozměrná**
**y = f(x)**

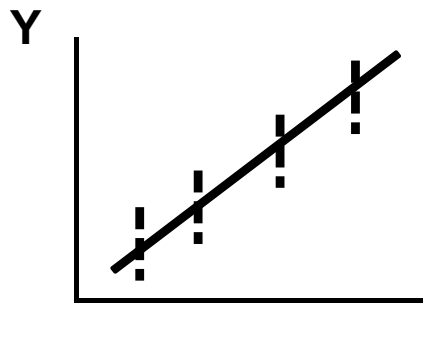
**Vícerozměrná**
**y = f(x1, x2, x3, ……xp)**

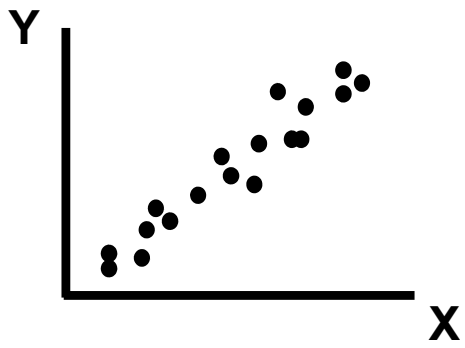
**Vztah x, y**

- **Deterministický**




- **Regresní, stochastický**




**Pro každé x existuje pravděpodobnostní rozložení y**

# Příklady lineární nebo "linearizovatelné" regrese



I.

**Y …… koncentrace antigenů**
**X …… čas**

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 \quad \approx \quad Y = \beta_0 + \beta_1 (\check{c}as) + \beta_2 (\check{c}as)^2$$

$\beta_0 : 0{,}014 \qquad P = 0{,}328$

$\beta_1 : 0{,}182 \qquad P = 0{,}000$

$\beta_2 : 0{,}089 \qquad P = 0{,}001$

---

II.

**Y …… koncentrace $O_2$ ve vodě**
**X …… koncentrace org. C ve vodě**

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2$$

---

III.

$Y = \exp(a + b \cdot x)$ ..... exponenciální

$Y = a \cdot x^b$ ..... multiplikativní

$\frac{1}{Y} = a + b \cdot x$ .....**reciproční**

# Regresní analýza přímky - "Simple regression"



$$Y = a + b \cdot x + e \quad \approx \quad \alpha + \beta \cdot X + \varepsilon$$

$y$ —

- $\alpha \approx a$ **(intercept):** $a = \bar{y} - b \cdot \bar{x}$
- $\beta \cdot X \approx b \cdot x$ **(sklon; slope)**
- $\varepsilon \approx e$ **- náhodná složka:** $\mathrm{N}(0; \sigma_e^2) = N(0; \sigma_y^2 x)$

} **Komponenty tvořící y se sčítají**

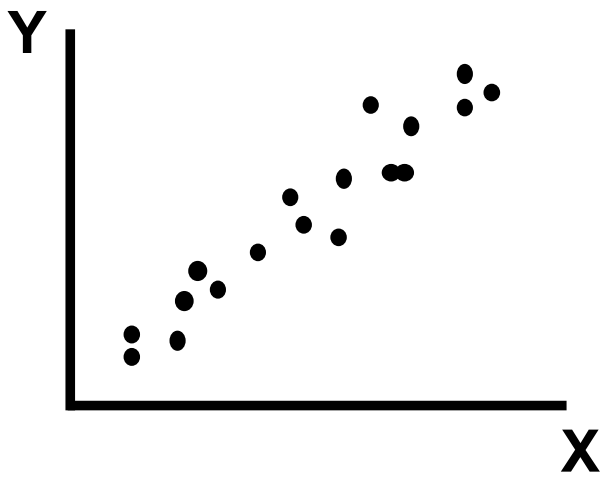
**ε - náhodná složka modelu přímky = rezidua přímky**

$$\sigma_e^2 \left(\sigma_{y \cdot x}^2\right) \Rightarrow \textbf{ rozptyl reziduí}$$

# Základní regresní analýzy: model přímky v datech



$$\begin{matrix}1\\\vdots\\n\end{matrix}\quad\begin{bmatrix}x\\\vdots\end{bmatrix}\qquad\begin{matrix}1\\\vdots\\n\end{matrix}\quad\begin{bmatrix}y\\\vdots\end{bmatrix}$$




$$\begin{matrix}1\\\vdots\\n\end{matrix}\quad\begin{bmatrix}\hat{y}\\\vdots\end{bmatrix}\quad = a + b\,.\quad\begin{bmatrix}x\\\vdots\end{bmatrix}\quad\Rightarrow\quad\begin{bmatrix}y\\\vdots\end{bmatrix}\quad -\quad\begin{bmatrix}\hat{y}\\\vdots\end{bmatrix}\quad =\quad\begin{bmatrix}e\\\vdots\end{bmatrix}$$

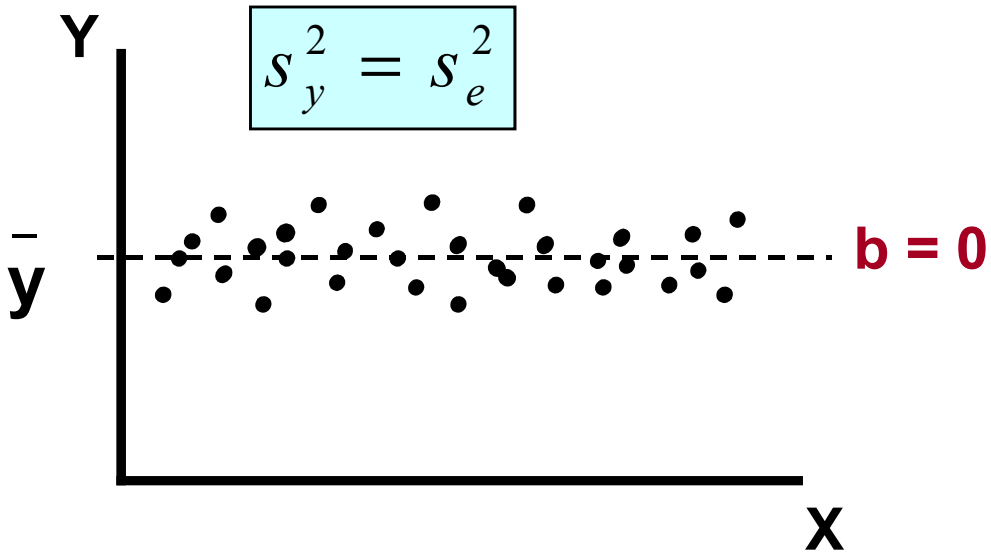
# Základní regresní analýzy: model přímky v datech

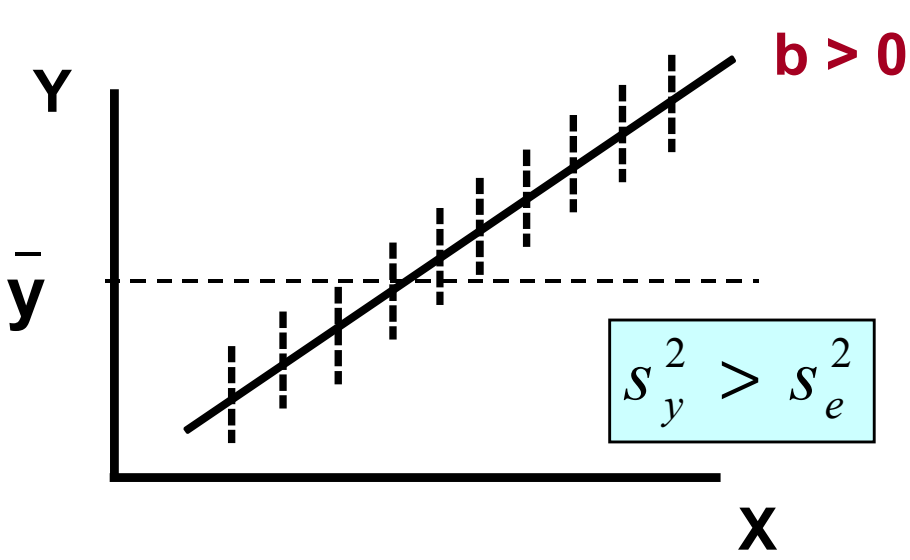


$$\mathbf{x} \quad \mathbf{y} \quad \hat{\mathbf{y}} \quad \mathbf{e}$$

$$\bar{x} \quad \bar{y} \quad \bar{\hat{y}} \quad \bar{e} = 0$$

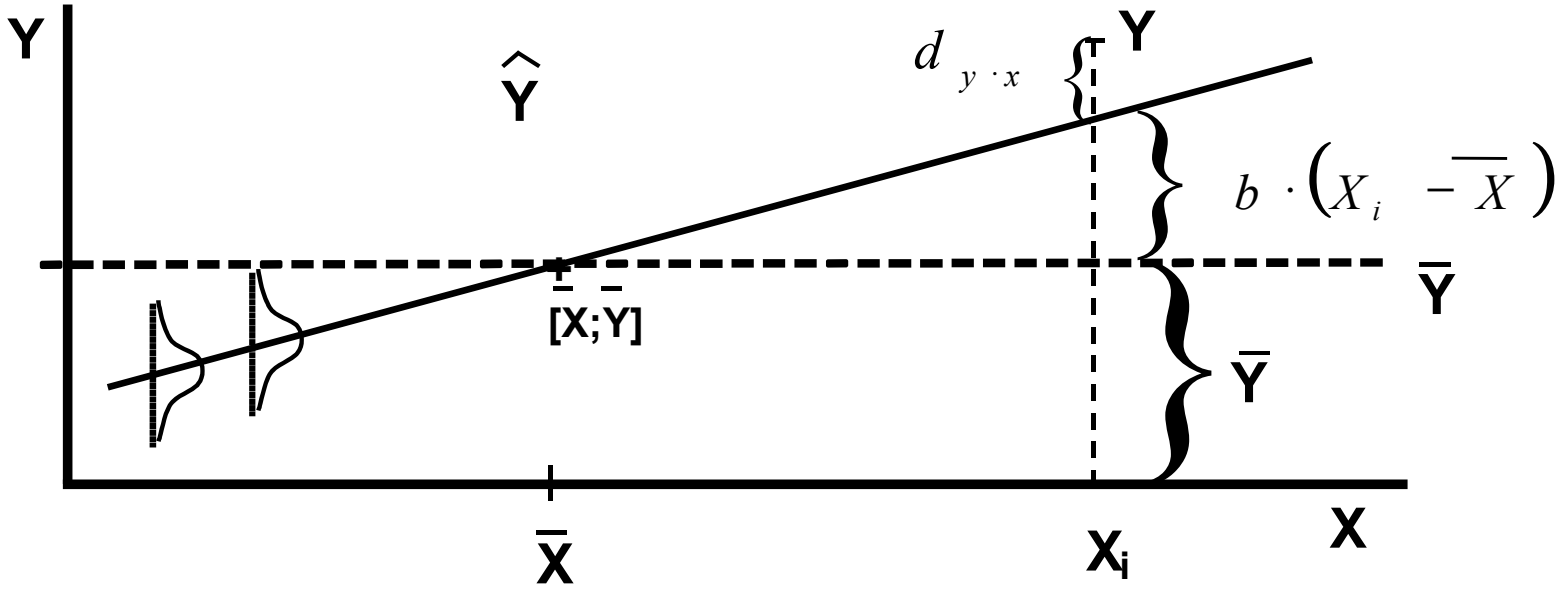
$$s_y^2 \qquad s_e^2$$



# Základní regresní analýzy: model přímky v datech






$$d_{y \cdot x} = y - \hat{y} \qquad \boxed{d_{y \cdot x} = y - \bar{y} - b(X_i - \overline{X})} \qquad \hat{y} = \bar{y} + b(X_i - \overline{X})$$

**Smysl proložení přímky**
minimalizace odchylek

$$d^2_{y \cdot x} \rightarrow \sum \left[ y - \hat{\alpha} - \hat{\beta}(X_i - \overline{X}) \right]$$

## Metoda nejmenších čtverců

1) X: Pevná, nestochastická proměnná
2) Rozložení hodnot y pro každé x je normální
3) Rozložení hodnot y pro každé x má stejný rozptyl
4) Rezidua jsou navzájem nezávislá a mají normální rozložení: $N(0; \sigma_e^2)$

# Odhady parametrů pro lineární regresi



I. $b \sim \beta$ : $b = \dfrac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$ $\qquad S_b^2 \sim \sigma_\beta^2$ : $\dfrac{1}{\sum (X_i - \bar{X})^2} \cdot S_{y \cdot x}^2$

$S_{y \cdot x}^2$ = **mean squared deviation from regression**

$S_{y \cdot x}$ = **sample standard deviation from regression**

$$S_{y \cdot x}^2 = \frac{\sum d_{y \cdot x}^2}{n-2} = \frac{\sum Y_i^2 - \dfrac{\sum Y_i^2}{n} - b^2 \cdot \sum (X_i - \bar{X})^2}{n-2}$$

---

II. $a \sim \alpha$ : $a = \bar{Y} - b \cdot \bar{X}$ $\qquad S_a^2 \sim \sigma_\alpha^2$ $\qquad S_\alpha^2 = \left[ \dfrac{1}{n} + \dfrac{\bar{X}^2}{\sum X^2} \right] \cdot S_{y \cdot x}^2$

**intercept**

---

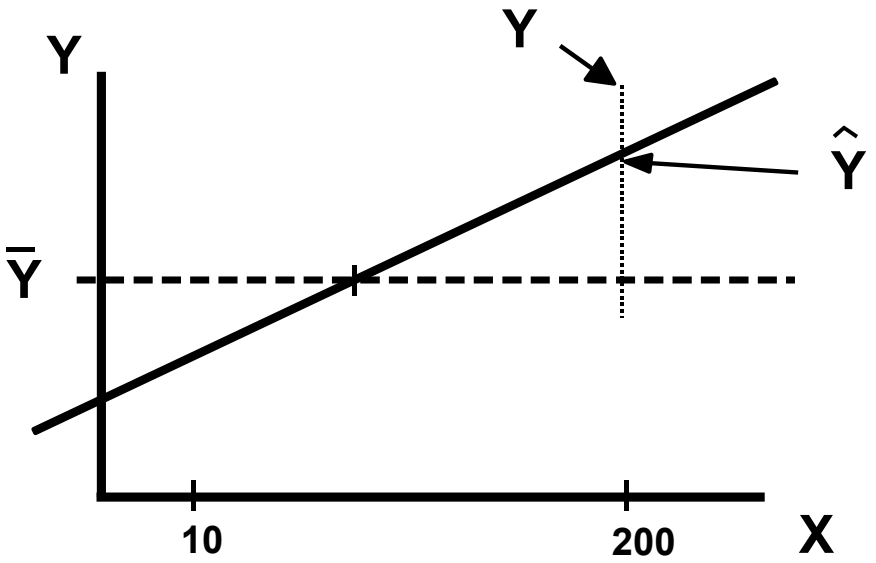
III. $\hat{Y}$ : **modelová hodnota**

$$\hat{Y}_i = a - b \cdot X_i \qquad S_{\hat{y}_i} = (S_{y \cdot x}) \cdot \sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum X^2}}$$

# Smysl lineární regrese



**X: Množství spáleného odpadu (tuny)**
**Y: Koncentrace kovu ve vzduchu(ng/m3)**




$$\left.\begin{aligned}\hat{Y} &= \bar{Y} + b\cdot(X-\bar{X})\\ \hat{Y} &= a + b\cdot X\end{aligned}\right\} a = \bar{Y} - b\cdot\bar{X}$$

**Platí: X = 0; 10; 100; 150; 200; 250; 300 tun**

**Model: Y = a + b . X**

**Výsledek:** $\hat{Y} = 14 + 0{,}123\cdot X; \quad \hat{Y} \rightarrow \left[\frac{\text{ng kov}}{m^3}\right]$

**Např. :** Skutečná data pro X = 200 t:

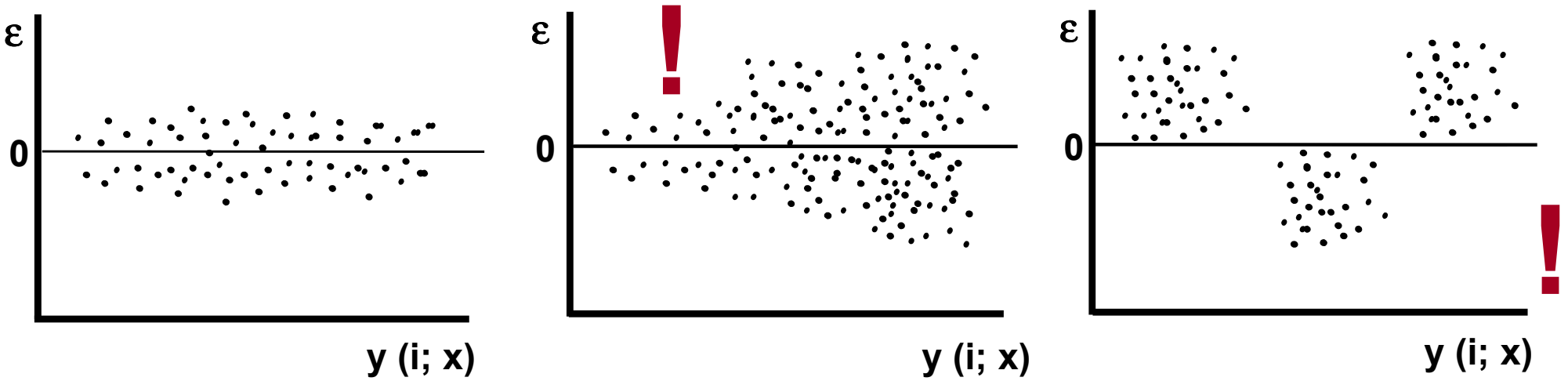
$Y_i$ = 16; 25; 41; 28; 31; 20 => $Y_i$ = 26.8

**Odhadnuto z modelu pro X = 200 t:**

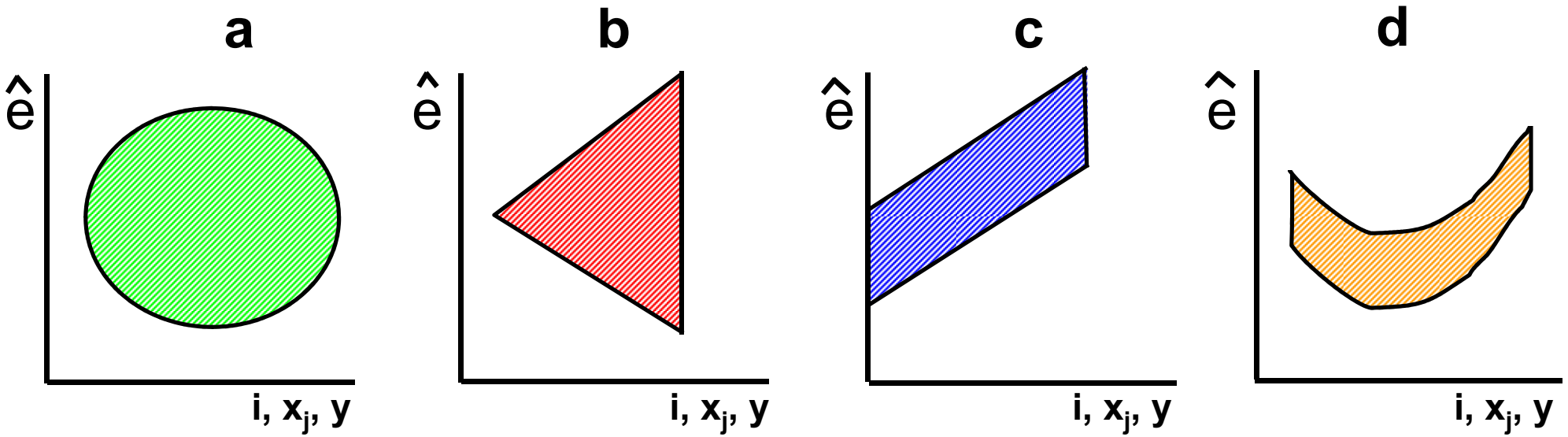
$\hat{Y}$ **= 14 + 0,123 . 200 = 38,6**

# Regresní analýza v grafech

## 3) Grafy residuí modelů (příklady)



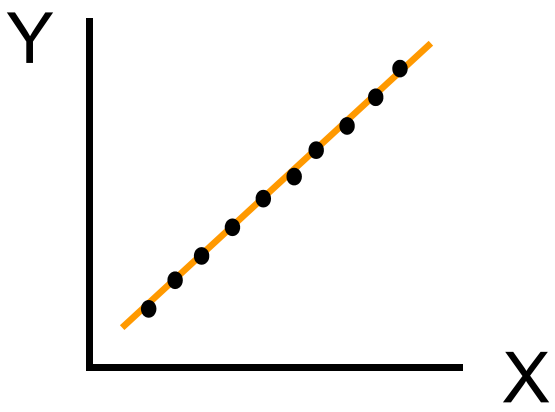

## Obecné tvary residuí modelů (schéma)

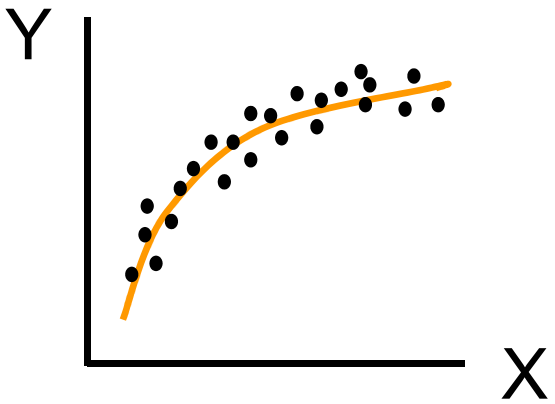


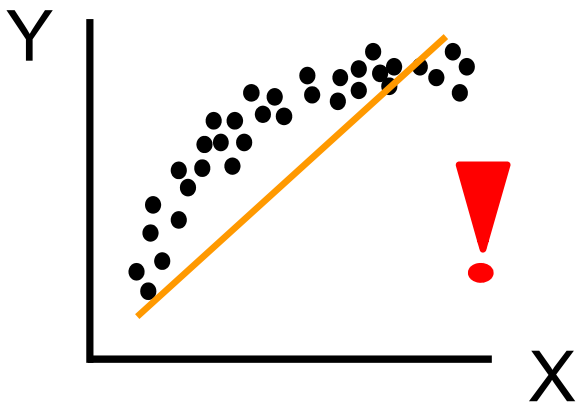
# Regresní analýza v grafech



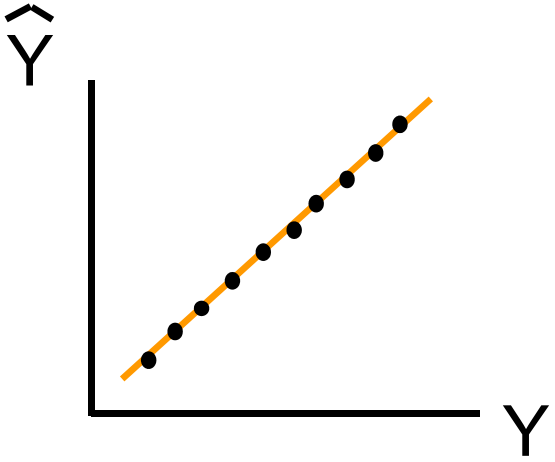
**1) Y vs. X**

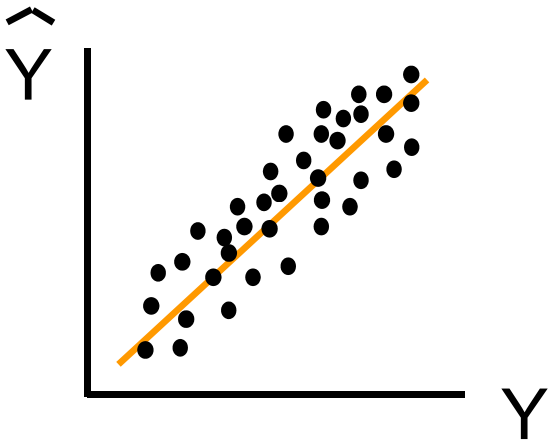


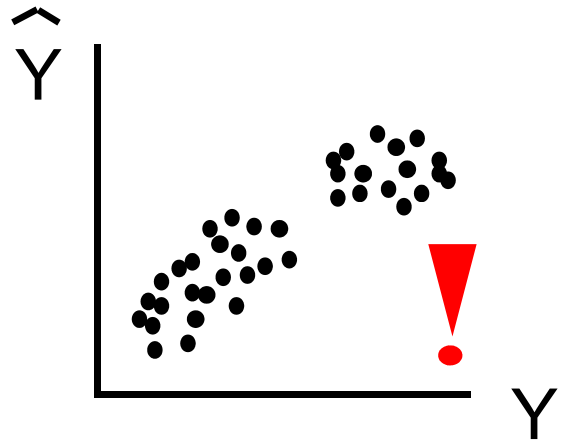




**2) Y vs. $\hat{Y}$**



# Lineární regrese - příklad



**X: Koncentrace drogy: 0; 2; 6; 8; 10; 12; 15 mg/ml krve**
**Y: Koncentrace volných metabolitů**

**Pro každé X: 3 opakování Y**

---

**Model: Y = a + b . x** ➡ **Y = 0,11 + 0,092 . X**

$$t_{0,975}^{(v=19)} = 2,093$$

**I.** $H_0 : \beta = 0; \alpha = 0,05$
$b = 0,092 ; s_b = 0,023$ $\Big\} t = \dfrac{b}{S_b} = 4,00$

**P < 0,01**

$$\beta : b \pm t_{1-\alpha/2}^{(n-2)} \cdot S_b$$

$$P\left(0,044 \le \beta \le 0,140\right) = 0,95$$

---

**II.** $H_0 : \alpha = 0; \alpha = 0,05$
$a = 0,11; s_a = 0,029$ $\Big\} t = \dfrac{a}{S_a} = 3,793$

$$t_{0,975}^{(v=19)} = 2,093$$

$$\alpha : \alpha \pm t_{1-\alpha/2}^{(n-2)} \cdot S_a$$

$$P\left(0,049 \le \alpha \le 0,171\right) = 0,95$$

# Analýza rozptylu jako nástroj analýzy regresních modelů - příklad na modelu přímky



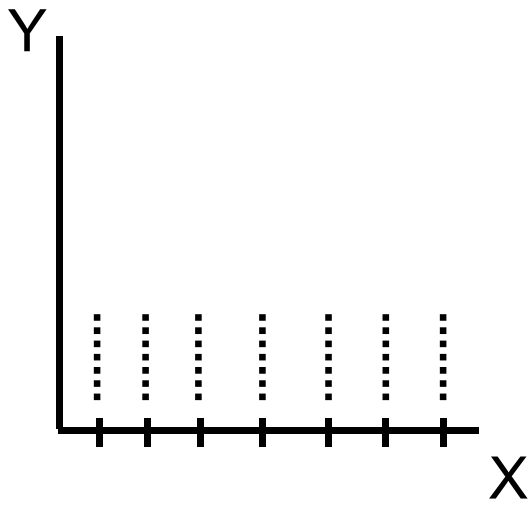
## 1) Experimentální data

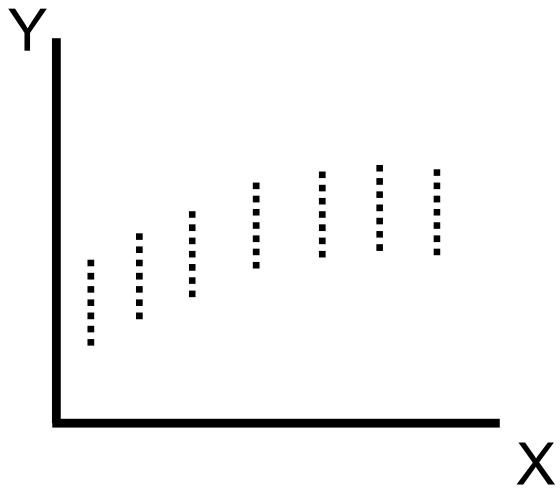
| $y_1$ | $x_0$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|---|
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| $y_n$ | $x_0$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
| | $s_0^2$ | $s_1^2$ | $s_2^2$ | $s_3^2$ | $s_4^2$ |

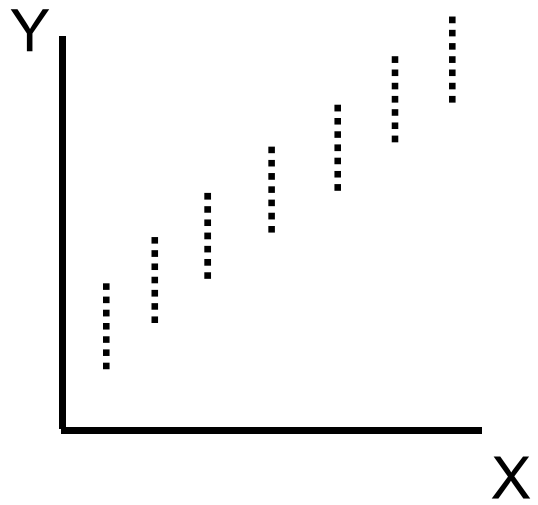
## 2) Celková ANOVA "one way"

| Zdroj rozptylu | St.v. | SS | MS | F |
|---|---|---|---|---|
| Mezi skupinami | a-1 | $SS_B$ | $SS_B/(a-1)$ | MSB/MSE |
| Uvnitř skupin | na-a | $SS_E$ | $SS_E/(na-a)$ | |
| Celkem | na-1 | $SS_T$ | $s_y^2$ | |

$$= \frac{SS_T}{na-1}$$



# Analýza rozptylu jako nástroj analýzy regresních modelů - příklad na modelu přímky

**3)** → **Celková ANOVA** → $SS_B/SS_T$ (variance ratio)

→ $MS_B/MS_E = F$

**4) Analýza rozptylu regresního modelu (zde přímky)**

| Zdroj rozptylu | st.v. | SS | MS | F |
|---|---|---|---|---|
| Model (přímka) | 1 | $SS_{MOD}$ | $MS_{MOD}$ | $MS_{MOD} / MS_R$ |
| Residuum | na - 2 | $SS_R$ | $MS_R$ | |
| celkem | na - 1 | $SS_T$ | | |

$(SS_{MOD}/SS_T) . 100 =$ % rozptylu Y "vyčerpaného" přímkou = koeficient determinace ($R^2$)

# Lineární regrese - příklad



**X:** konc.Cd: 1,2,3,4,5,6 ng/ml
**Y:** absorb: 0,23; 0,49; 0,72; 0,90; 1,16; 1,39

**b=0,228** $S_b=4{,}99.10^{-3}$ **P = 0,000**
**a=0,016** $S_a=0{,}019$ **P = 0,457**

r = 0,999
$R_2$ = 99,81%

**St. Error of est: 0,021**

**ANOVA**

| Source | D.f. | SS | MS | F | P |
|---|---|---|---|---|---|
| Model | 1 | 0,912 | 0,912 | 2086,3 | 0 |
| Residual | 4 | 0,0017 | 0,000425 | | |
| Total ( c ) | 5 | 0,9138 | | | |

$s^2_{y.x} = 4{,}25 \cdot 10^{-4}$
$s^2_y = 0{,}18275$