

LINEÁRNÍ REGRESE — POKRAČ.

23.11.2012

ÚVODNÍ NASTAVENÍ.

– Otevřete si R Studio. Z internetu si stáhněte data `olympics.csv` a `blood.csv`.

POLYNOMIÁLNÍ REGRESE

1. Načtěte si data `olympics.csv`. Datový soubor obsahuje výsledky zlatých medalistů z letních olympiád od roku 1896 ve skoku do výšky, do dálky a hodu diskem.

<code>Year</code>	rok (počítáno od 1900)
<code>Long.Jump</code>	skok daleký (v inch)
<code>High.Jump</code>	skok vysoký (v inch)
<code>Discus</code>	hod diskem (v inch)

Bude nás zajímat, jakým způsobem se výkony na olympiádě mění v čase.

2. Převeďte skok do výšky a skok do dálky na centimetry a hod diskem na metry.
3. Prohlédněte si vývoj každé z disciplín v čase zvlášť. Zdá se být závislost na čase lineární?
4. Podíváme se na skok do výšky.

- (a) Odhadněte model regresní přímky.

```
m1=lm(High.Jump~Year)
summary(m1)
```

- (b) Připomeňte si interpretaci jednotlivých parametrů a koeficientu determinace.

- (c) Proložení přímkou si znázorníme graficky:

```
plot(High.Jump~Year,pch=19)
lines(fitted(m1)~Year,col="red")
```

- (d) Zkusíme uvažovat kvadratickou závislost na čase.

```
m2=lm(High.Jump~Year+I(Year^2))
summary(m2)
```

Přináší zapojení kvadratického členu významné zlepšení? Jak se změnil koeficient determinace?

- (e) Obě proložení si znázorníme do jednoho obrázku:

```
plot(High.Jump~Year,pch=19)
lines(fitted(m1)~Year,col="red")
lines(fitted(m2)~Year,col="blue")
```

Který model bychom zvolili?

- (f) Na základě obou modelů předpovězte výkon ve skoku do výšky na olympiádě v Sydney.

- (g) Zkuste se sami podívat, zda by zapojení vyšší mocniny (polynomu třetího stupně) přineslo významné zlepšení.
5. Ještě se podíváme (neformálně) na předpoklady lineární regrese v lineárním i kvadratickém modelu.

```
r=resid(m1)

library(car)
qqPlot(r,dist="norm")

plot(r,type="b")

plot(r~fitted(m1))
```

Proveďte totéž pro kvadratický model.

6. Podívejte se samostatně na skok do dálky: Opět uvažujte lineární a kvadratickou (příp. složitější) závislost a na základě nejlepšího modelu předpovězte výkon na olympiádě v Sydney.

VÍCENÁSOBNÝ REGRESE

1. Načtěte si data `blood.csv`. Datový soubor obsahuje informace o 11 pacientech: výši jejich systolického krevního tlaku, věk a hmotnost.

```
systolic  systolický krevní tlak
age       věk v letech
weight    hmotnost (v librách)
```

2. Převeďte hmotnost pacientů na kilogramy.
3. Zajímá nás, jak závisí výše krevního tlaku na věku a hmotnosti.

- (a) Prohlédněte si závislost (graf) výše systolického tlaku na věku a hmotnosti zvlášť.
- (b) Odhadneme model, ve kterém bude závislost systolického tlaku na obou veličinách:

```
m=lm(systolic~age+weight)
summary(m)
```

Jak budeme interpretovat odhadnuté parametry? Závisí systolický tlak na věku i hmotnosti?

- (c) Kdybychom chtěli dostat interpretaci i pro absolutní člen, můžeme si posunout obě proměnné např. do jejich průměrné hodnoty

```
m=lm(systolic~I(age-mean(age))+I(weight-mean(weight)))
summary(m)
```

- (d) Předpovězte výši systolického tlaku pro 50 letého pacienta, který váží 90 kg.