
CVIČENÍ 25.2.2013

DATA: Soubor `run100.csv` obsahuje informace výsledcích v běhu na 100 m mužů na všech olympijských hrách od roku 1900 do roku 2008 (dosažený čas v sekundách).

POLYNOMIÁLNÍ TRENDY (LINEÁRNÍ REGRESE)

1. Načtete si data do R Studia a prohlédnete si je. Zejména si všimnete, jaká je frekvence měření a zda se jedná o pravidelně pozorovaná data. Které hodnoty chybí/přebývají do pravidelnosti a proč?
2. Znázorníte si graf vývoje výsledků.

```
attach(run100)
plot(Result~Year)
```

Je v datech viditelný nějaký časový trend?

3. Pro jednoduchost v modelech převed'te veličinu udávající rok na hodnoty od 1900, tj. např.

```
Rok=Year-1900
```

4. Odhadněte model s lineárním trendem.

```
m1=lm(Result~Rok)
```

Na základě odhadnutého modelu proved'te:

- Graficky znázorníte odhadnutý trend.

```
plot(Result~Rok)
abline(m1,col="red")
```

- K jakému zlepšení dochází v průměru každým rokem? Je zlepšování v běhu statisticky významné?

```
summary(m1)
```

- Proveďte predikci času na Olympiádě v Londýně v roce 2012 a na olympiádě v Riu v roce 2016.

```
predict(m1,newdata=data.frame(Rok=c(112,116)))
```

5. Proložte řadu kvadratickým trendem, tj. kvadratickou funkcí v proměnné Rok

```
m2=lm(Result~Rok+I(Rok^2)).
```

Na základě odhadnutého modelu zjistěte:

- Porovnejte do jednoho obrázku proložení lineárním a kvadratickým trendem.

```
plot(Result~Rok)
abline(m1,col="red")
lines(fitted(m2)~Rok,col="blue")
```

- Je kvadratický člen statisticky významný?
 - Proveďte predikci času na Olympiádě v Londýně v roce 2012 a na olympiádě v Riu v roce 2016. Výsledky srovnajte s predikcemi na základě lineárního trendu.
 - Proložená kvadratická funkce je na zkoumaném úseku dat klesající. Spočítejte, kdy přibližně dojde k jejímu minimu a kolik toto minimum bude.
6. Který model bychom vybrali jako vhodnější a proč?
Připomeňte si, jaké číselné charakteristiky znáte pro posouzení kvality modelu. Určete jejich hodnoty pro oba modely a porovnejte je.
7. Mělo by smysl uvažovat vyšší mocniny Rok, tj. kubické a vyšší polynomiální trendy?

NELINEÁRNÍ REGRESE

8. Uvažujte pro data tzv. modifikovaný exponenciální trend, tj.

$$\text{Result}_t = \gamma + \alpha \cdot \beta^{\text{Rok}_t} + e_t.$$

Jedná se o příklad nelineárního regresního modelu. Proč by právě tento model mohl být pro naše data vhodný?

9. Odhadněte výše uvedený trend.

```
m3=nls(Result~c+a*b^Rok,start=list(c=8,a=10,b=0.8))
summary(m3)
```

Jaká je interpretace parametru γ ?

10. Porovnejte graficky proložení tímto trendem s lineárním a kvadratickým.

```
plot(Result~Rok)
abline(m1,col="red")
lines(fitted(m2)~Rok,col="blue")
lines(fitted(m3)~Rok,col="green")
```

11. Porovnejte číselné charakteristiky (MSE, MAE, MAPE, AIC, BIC) pro všechny uvažované modely. Který model bychom zvolili jako nejvhodnější? (S přihlédnutím k danému problému a charakteru dat).
12. Proveďte analýzu reziduí pro výsledný model.