

Experimental Design and Statistics - AGA46E

M. Maciak (Czech University of Life Sciences, Prague)

Lab Session 9 - Summer Term 2015

1 Jednoduchá lineární regrese v programu R

- pomocí helpu zjistíte, jak pracuje funkce (příkaz) `lm()` který počítá jednoduchý lineární regresní model;
- použijete příkaz `lm()` a spočítáte několik jednoduchých lineárních regresních modelů - například:

```
> testData <- data.frame(y = rnorm(20), x = rnorm(20), z = rnorm(20))
> lm(y ~ x, data = testData)
```

Detailnější výsledky získáme pomocí příkazu `summary()`:

```
> summary(lm(y ~ x, data = testData))
```

- Interpretujte jednotlivé numerické hodnoty, které tvoří výstup příkazu `summary()`;
- Nakreslete graf závislosti pro x a y a pomocí příkazu `abline()` zakreslete odhadnutou regresní přímku;
- Analogickým způsobem je možné přidat to modelu více proměnných a overit jejich důležitost (statistickou významnost) pro daný model:

```
> summary(lm(y ~ x + z, data = testData))
```

- Opět pomocí příkazu `summary()` dostaneme základní hodnoty ohledem statistické inference;
- Použijete datový soubor *Orange* který je k dispozici v programu R a odhadnete regresní přímku pro závislost průměru (circumference) na věku stromu (age);

```
> attach(Orange)
> summary(lm(circumference ~ age))
```

- použijete aj proměnnou pro typ stromu - *Tree* a přidejte tuto proměnnou do již existujícího modelu:

```
> summary(lm(circumference ~ age + Tree))
```

- zakreslete jednotlivé regresní přímky do xy grafu;

2 Jednoduchá lineární regrese a jednoduché interakce

- někdy se také stane, že dvě různé vysvětlující proměnné vzájemně působí ať jedna na druhou a společně ovlivňují závislou proměnnou; V případě vzájemného vlivu vysvětlujících proměnných na sebe navzájem je dobře overit významnost interakce v modelu.

- v lineárním regresním modelu implementujeme interakci pomocí znaménka násobení - *:

```
> summary(lm(circumference ~ age + Tree + age*Tree))
```

- jaký je hlavní rozdíl mezi předchozím modelem a modelem s interakcí mezi proměnnými *age* a *Tree*?
- jednotlivé regresní přímky odhadnuté modelem zakreslíte do *xy* grafu a porovnejte s předchozími grafy;

3 ANOVA vs. Regrese

- uvažujte pouze 5 druhů stromů a aplikujte metodu analýzy rozptylu pomocí příkazu `anova()`:

```
> anova(lm(circumference ~ Tree))
```

- porovnejte výsledky s výsledky lineárního regresního modelu, který získáme pomocí příkazu `lm()`:

```
> summary(lm(circumference ~ Tree))
```