

6. POROVNÁVÁNÍ PRŮMĚRŮ

6.1 Normalizace proměnné

Pro správné použití většiny statistických metod je potřeba, aby měly proměnné přibližně normální rozložení. Nejjednodušší způsob hrubého zjištění normality je grafické zobrazení pomocí příkazu HISTOGRAM. Pokud proměnná nemá normální rozložení, je vhodné provést nějakou transformaci.

LADDER/GLADDER – zobrazí možnosti transformace proměnné.

```
ladder age
gladder age
```

6.2 Porovnávání průměrů

CI – umožní nastavit interval spolehlivosti (confidence interval) pro výpočet aritmetických průměrů.

```
ci age, level(90)
```

MEAN – výpočet aritmetického průměru v analyzovaném souboru.

```
mean age
mean age, over (v291)
mean age [pw=weight], over (v291)
```

Tím ale nezjistíme, jestli je uvedený rozdíl v průměrech také v základním souboru, tedy jestli je rozdíl tzv. statisticky významný. První náповědou může být pohled na intervaly spolehlivosti. Pokud se nepřekrývají, je pravděpodobné, že i v základní populaci existuje statisticky významný rozdíl. Statisticky korektně se ale test statistické významnosti provádí pomocí t-testu.

TTEST – provede test statistické významnosti rozdílů v průměrech. Testujeme tzv. nulovou hypotézu, která říká, že mezi dvěma průměry **NENÍ** v základní populaci žádný rozdíl. Příkaz TTEST vypisuje pravděpodobnost, s jakou platí alternativní hypotéza H_a (komplementární k nulové hypotéze H_0), platná hypotéza H_a má hodnotu $\Pr(|T| > |t|)$ menší než 0.05. Parametr UNEQUAL říká, že rozložení testované proměnné v základním a výběrovém souboru mají odlišné rozložení. TTEST je možno použít i pro testování jednoho průměru.

```
ttest age, by(v291)
ttest age, by(v291) unequal
test age==48
```

ONEWAY – provede test ANOVA, na rozdíl od t testu je možno zavést váhy. Hodnota Prob > F ukazuje, s jakou pravděpodobností se průměry neodlišují (tj. s jakou pravděpodobností platí nulová hypotéza). Pokud je hodnota menší než 0,05, nulovou hypotézu zamítáme, tj. rozdíly v základním souboru existují. Parametr SIDAK vypíše podrobné informace o rozdílech mezi jednotlivými skupinami v rámci proměnné. Opět platí, že hodnota <0,05 znamená existenci statisticky významného rozdílu.

```
oneway age v306 [w=weight]
oneway age v306 [w=weight], sidak
```