

6 Analýza kovariance (ANCOVA)

Příklad 1. Datový soubor `cneck.txt` obsahuje antropometrické údaje mladých dospělých lidí (převážně studentů vysokých škol z Brna a Ostravy). Zajímá nás, zda se muži a ženy (proměnná `sex`) liší ve střední hodnotě tělesné hmotnosti (proměnná `body.W`, v *kg*), pokud kontrolujeme obvod boků (proměnná `hip.C`, v *mm*). Ostatní proměnné v souboru nás nezajímají.

1. Vykreslete bodový diagram, ve kterém odlišíte muže a ženy. Vidíte už teď nějaká podezřelá pozorování?
2. Zjistěte počty pozorování, výběrové průměry a výběrové směrodatné odchylky proměnných tělesná váha a obvod boků zvlášť pro ženy a pro muže, i pro celý datový soubor.
3. Ověřte předpoklad rovnoběžnosti regresních přímk pro ženy a pro muže.
4. Ověřte zbývající předpoklady analýzy kovariance.
5. Pokud v diagnostických grafech vidíte některá pozorování jako problematická, podívejte se, jestli odpovídají těm, která se vám zdála podezřelá v grafu z prvního úkolu. Tj. přidejte do prvního grafu tato pozorování jako body jinou značkou. Proč se tato pozorování zdají podezřelá?
6. Odstraňte tato pozorování z datového souboru a zopakujte úkoly 3 a 4. Dále pracujte s tímto upraveným souborem a s modely, které jsou sestaveny z něj.
7. Otestujte hypotézu, že regresní přímk pro ženy a pro muže jsou shodné.
8. Výsledný model graficky znázorněte společně s pásem spolehlivosti. Pokud váš výsledný model obsahuje faktor pohlaví, znamená to vykreslení dvou regresních přímk s pásy spolehlivosti (barevně je odlišíte). Ve funkci `predict()` je v argumentu `newdata` potřeba nastavit kategorii proměnné `sex`, např. pro muže získáme hodnoty nastavením `newdata=data.frame(sex='m', hip.C=xx)`, kde `xx` je vektor hodnot obvodu boků, na němž chceme regresní přímk a pás spolehlivosti vykreslit. Obdobně pak postupujeme i pro ženy.

Veškeré testování hypotéz provádějte na hladině významnosti 0.05.