

Datový soubor auta_USA.sta

U 30 značek amerických aut byly zjišťovány hodnoty těchto dvou parametrů: Objem motoru (veličina X, v objemových uncích) a ujetá vzdálenost na galon benzínu (veličina Y, v mílích – zkratka mpg).

Úkoly:

1. Popište datový soubor pomocí číselných charakteristik. Pro obě proměnné sestrojte histogramy.
2. Pro obě proměnné sestrojte N-P plot a pomocí Lillieforsova testu, S-W testu a A-D testu testujte hypotézu o normalitě obou proměnných.
3. Pomocí dvourozměrného tečkového diagramu s proloženou elipsou 95% konstantní hustoty pravděpodobnosti orientačně ověřte dvourozměrnou normalitu dat.
3. Vypočte výběrový koeficient korelace veličin X, Y, interpretujte jeho hodnotu a testujte hypotézu o jeho nevýznamnosti.
4. Předpokládejte, že přímka je vhodným modelem závislosti Y na X. Vypočtete odhady regresních parametrů a napište rovnici regresní přímky. Interpretujte parametry regresní přímky.
5. Najděte odhad rozptylu, vypočtete index determinace a interpretujte ho.
6. Najděte 95% intervaly spolehlivosti pro regresní parametry.
7. Proveďte celkový F-test a dílčí t-testy.
8. Nakreslete dvourozměrný tečkový diagram s proloženou regresní přímkou a 95% pásem spolehlivosti a 95% predikčním pásem.
9. Vypočtete střední absolutní procentuální chybu predikce (MAPE).
10. Proveďte analýzu reziduí.
11. Testujte hypotézu o adekvátnosti modelu přímky a vypočtete poměr determinace.
12. Nyní předpokládejte, že parabola je vhodným modelem závislosti Y na X. Úkoly 4 až 11 nyní proveďte pro model paraboly.
13. Na závěr zhodnoťte, který z obou modelů lépe vystihuje závislost Y na X.