

Povzetek

V prvem poglavju najprej konstruiramo model multiple regresije in navedemo predpostavke o linearnosti modela, konstantnosti disperzije ter nekoreliranosti spremenljivk y , potem pa se lotimo ocenjevanja parametrov β in σ^2 . Najprej ju ocenimo z metodo najmanjših kvadratov ter si ogledamo lastnosti dobljenih cenilk in geometrijsko predstavimo to metodo. Definiramo model v centrirani obliki, ki ga uporabimo tudi v kasnejših razdelkih. Pod dodatno predpostavko o normalnosti modela določimo še cenilke po metodi največjega verjetja in si podrobneje ogledamo tudi njihove lastnosti. Za ocenjevanje statističnega pomena odvisnosti med opazovanimi in predvidenimi spremenljivkami y definiramo *determinacijski koeficient* R^2 kot razmerje regresijske in popolne vsote kvadratov. Njegov koren imenujemo *multipli korelacijski koeficient* R . V primeru, ko so spremenljivke y med seboj korelirane ali pa nimajo konstantne disperzije, za ocenjevanje parametrov uporabimo posplošeno metodo najmanjših kvadratov. Na koncu prvega poglavja raziščemo še posledice napačnega izbora modela, torej učinke uporabe predločenega ali poddoločenega modela na pristranskost in disperzije cenilk.

Drugo poglavje se ukvarja s preizkušanjem domnev in intervali zaupanja za parametre β pod predpostavko o normalni porazdelitvi spremenljivk y . Najprej definiramo nekatere pomembne porazdelitve: porazdelitev χ^2 , porazdelitev F in porazdelitev t , potem pa si ogledamo neodvisnost ter porazdelitve kvadratnih form v spremenljivki y . To uporabimo pri testu celotnega regresijskega modela in testu na podmnožici parametrov β , torej testu polnega in reduciranega modela. Nato uporabljeni testni statistiki F izrazimo še s pomočjo determinacijskega koeficienta R^2 . Na podoben način pridemo tudi do testov splošnih linearnih domnev, potem pa s pomočjo tega testa ali s testom polnega in reduciranega modela testiramo enega ali več parametrov β oziroma eno ali več linearnih kombinacij teh parametrov. Definiramo območje zaupanja za vektor β , intervale zaupanja za posamezne parametre β , za linearne kombinacije teh parametrov, za matematično upanje $E(y)$ ter za disperzijo σ^2 , intervale predvidevanja za bodoče meritve in simultane intervale. Ker so bili testi na začetku tega poglavja izpeljani z uporabo neformalnih metod, ki temeljijo na iskanju neodvisnih vsot kvadratov s χ^2 porazdelitvijo, jih na koncu poglavja obravnavamo še z bolj formalnega stališča razmerja verjetij.

V dodatku navedemo še nekatera enostavna dejstva iz linearne algebre, ki so bila uporabljena v diplomski nalogi.

Math. Subj. Class. (MSC 2000): 62J05, 62T99

Ključne besede: Multipla linearna regresija, pojasnjevana (odvisna) spremenljivka, pojasnjevalna (neodvisna) spremenljivka, zasnovna matrika, parcialni regresijski koeficienti, metoda najmanjših kvadratov, normalne enačbe, linearna cenilka, vektor ostankov, najboljša linearna (kvadratna) nepristranska cenilka, prostor spremenljivk, prostor vzorcev, prostor ocen, centrirajoča matrika, cenilke največjega verjetja, funkcija verjetja, zadostna cenilka, popolna vsota kvadratov, regresijska vsota kvadratov, determinacijski koeficient, multipli korelacijski koeficient, prilagojeni determinacijski koeficient, posplošena cenilka po metodi najmanjših kvadratov, poddoločen model, predoločen model, porazdelitev χ^2 , porazdelitev F , porazdelitev t , prostostne stopnje, necentralni parameter, testiranje domnev, ničelna domneva, alternativna domneva, pristop polnega in reduciranega modela, splošna linearna domneva, stopnja značilnosti, Bonferronijev pristop, Schefféjev pristop, območje zaupanja, interval zaupanja, interval predvidevanja, test z razmerjem verjetij.

Keywords: Multiple linear regression, dependent (response) variable, independent (predictor) variable, design matrix, partial regression coefficients, least squares approach, normal equations, linear estimator, vector of residuals, best linear (quadratic) unbiased estimator, variable space, sample space, estimation space, centering matrix, maximum likelihood estimators, likelihood function, sufficient estimator, total sum of squares, regression sum of squares, coefficient of determination, squared multiple correlation, adjusted coefficient of determination, generalized least squares estimator, underfitting, overfitting, χ^2 -distribution, F -distribution, t -distribution, degrees of freedom, non-central parameter, tests of hypotheses, null hypotheses, alternative hypotheses, full-and-reduced-model test, general linear hypotheses, significance level, Bonferroni approach, Scheffé approach, confidence region, confidence interval, prediction interval, likelihood ratio test.

Literatura

- [1] Alvin C. Rencher, *Linear models in statistics*, John Wiley and sons,inc., 2000
- [2] F.A. Greybill, *On quadratic estimates of variance components*, Annals of mathematical statistics, 1954
- [3] F.A. Greybill, A.W. Wortham *A note on uniformly best, unbiased estimators for variance components*, Journal of the American statistical association, 1956
- [4] F.A. Greybill *Theory and application of the linear model*, North Scituate, 1976
- [5] R.V. Hogg, A.T. Craig *Introduction to mathematical statistics (5th edition)*, Englewood Cliffs, 1995
- [6] B.K. Ghosh *Some monotonicity theorems for chi-square, F and t distributions with applications*Journal of the Royal Statistical Society 35, 1973
- [7] S.R. Searle *Linear models*, 1971
- [8] A. Wald *Tests of statistical hypotheses concerning several paramethers when the number of observations is large*, Transactions of the American Mathematical Society 54, 1943