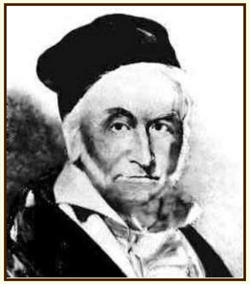
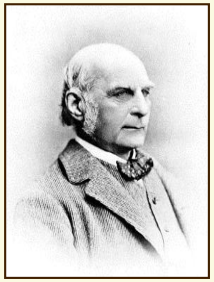
REGRESIJSKA ANALIZA

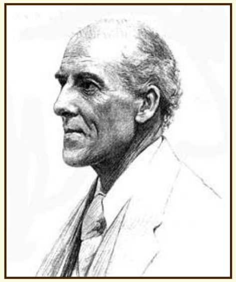
U razvoju regresijske analize najznačajniju ulogu su imali:



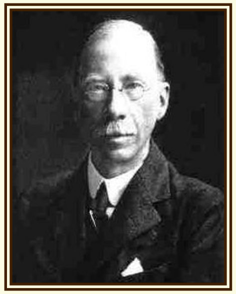
Carl Friedrich Gauss
(1822. – 1911.)



Francis Galton
(1822. – 1911.)



Karl Pearson
(1857. – 1936.)



George Udny Yule
(1871. – 1951.)

REGRESIJSKA ANALIZA

Regresijska analiza je matematičko-statistički postupak kojim se utvrđuje odgovarajuća funkcionalna veza (relacija) između jedne ***zavisne*** ili ***kriterijske varijable*** i jedne ili više ***nezavisnih*** ili ***prediktorskih varijabli.***

Zavisna (kriterijska) varijabla je varijabla čiji se varijabilitet objašnjava putem nezavisnih varijabli.

Nezavisne (prediktorske) varijable su varijable na temelju kojih se objašnjava varijabilitet zavisne varijable.

REGRESIJSKA ANALIZA

Regresijska analiza se u kineziologiji najčešće koristi u svrhu:

- utvrđivanja utjecaja jedne varijable ili skupa varijabli na neku kriterijsku varijablu (npr. utvrđivanje utjecaja karakteristika građe tijela na rezultat u bacanju kugle) i
- utvrđivanje trenda razvoja rezultata u nekom sportu (npr. utvrđivanje trenda razvoja najboljih rezultata u bacanju kugle na svjetskim prvenstvima)

REGRESIJSKA ANALIZA

Funkcionalna veza između prediktorskih varijabli i kriterijske varijable definira se utvrđivanjem odgovarajuće ***regresijske jednadžbe***. Opći oblik regresijske jednadžbe izgleda ovako:

nezavisne (prediktorske) varijable

$$Y = f(X_1, X_2, \ldots\ldots, X_m) + e$$

odgovarajuća funkcija

greška prognoze

zavisna (kriterijska) varijabla

REGRESIJSKA ANALIZA

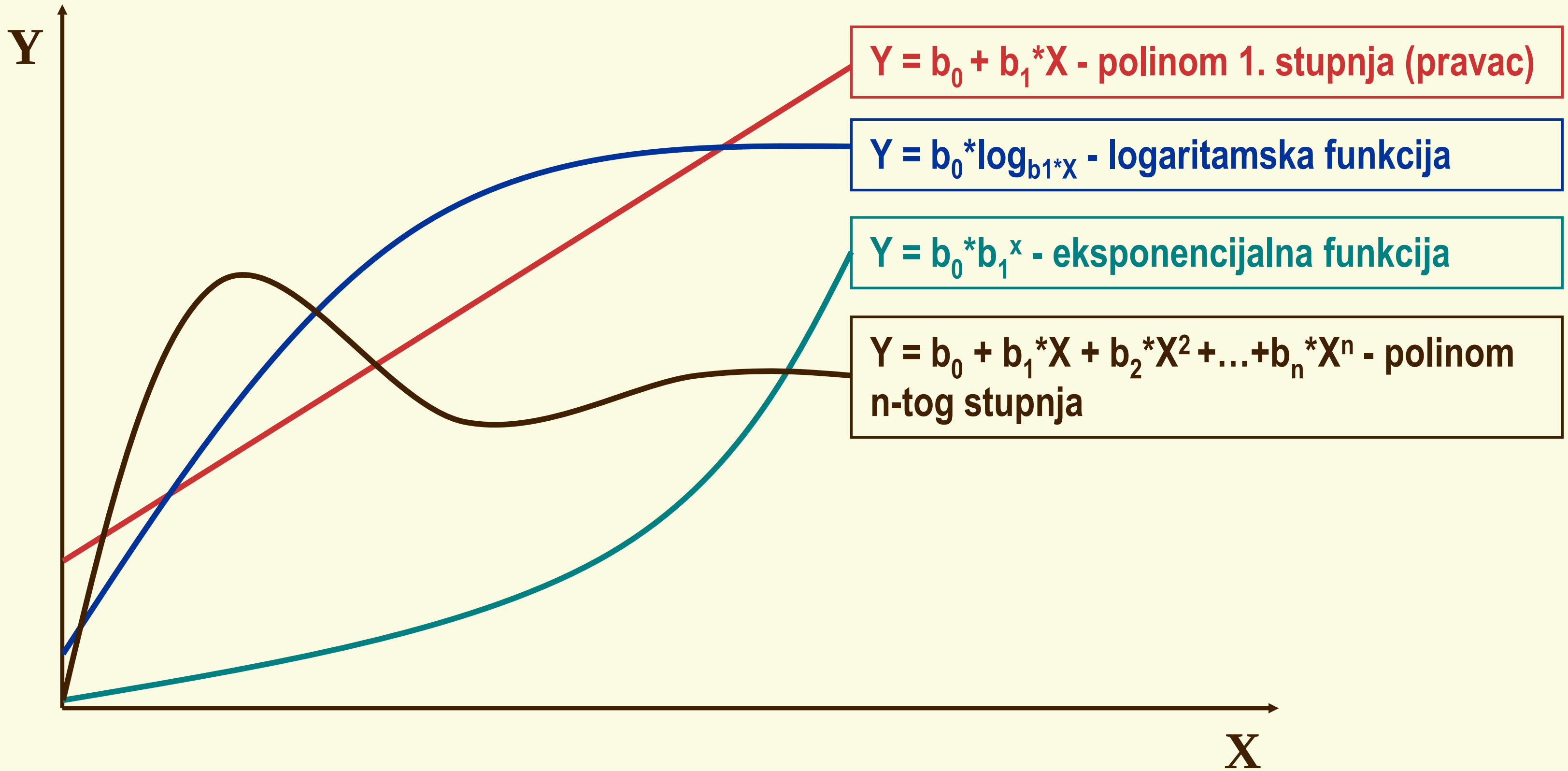
Regresijske modele moguće je generalno podijeliti na temelju dvaju kriterija i to:

- prema broju nezavisnih varijabli na:
  - ***jednostavne (simple) regresijske modele*** i
  - ***višestruke (multiple) regresijske modele,*** te

- prema odnosu između zavisne i nezavisnih varijabli na:
  - ***linearne regresijske modele*** i
  - ***nelinearne regresijske modele***.

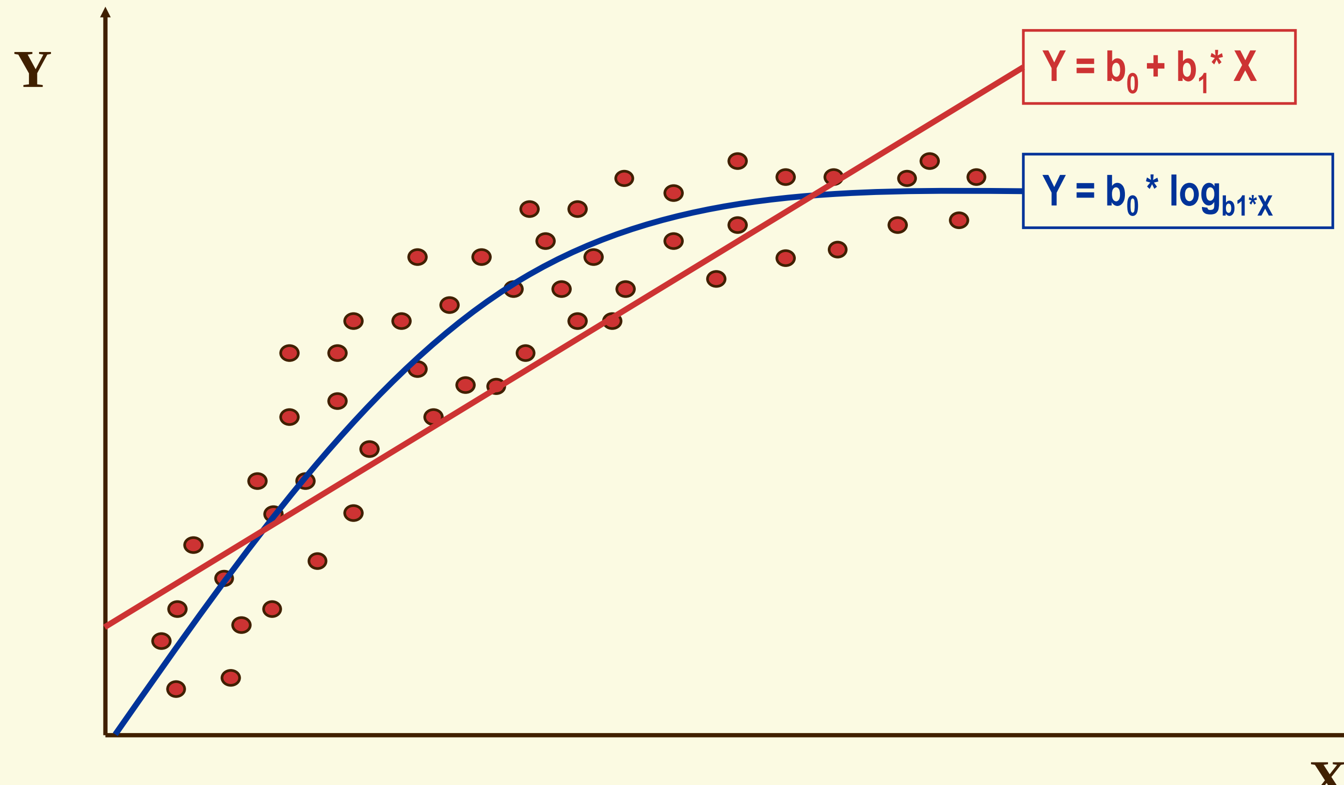
REGRESIJSKA ANALIZA

Linearni i nelinearni modeli jednostavne regresijske analize:



REGRESIJSKA ANALIZA

Odabir modela jednostavne regresijske analize vrši se pomoću korelacijskog dijagrama.



(***Primjer:*** logaritamska funkcija je prikladan model)

REGRESIJSKA ANALIZA

Jednostavna linearna regresijska analiza

Jednostavnom linearnom regresijskom analizom utvrđuje se linearna povezanost između jedne nezavisne (prediktorske) i jedne zavisne (kriterijske) varijable pri čemu regresijska jednadžba ima sljedeći oblik:

$$y_i = b_0 + b_1 x_i + e_i$$

gdje je

- y_i - rezultat entiteta *i* u kriterijskoj varijabli
- b_0 i b_1 - regresijski koeficijenti
- x_i - rezultat entiteta *i* u prediktorskoj varijabli
- e_i - rezidualna vrijednost entiteta *i*
- $i = 1,...,n$

REGRESIJSKA ANALIZA

Jednostavna linearna regresijska analiza

Regresijski koeficijenti omogućavaju prognoziranje rezultata entiteta u kriterijskoj varijabli na temelju rezultata u prediktorskoj varijabli putem sljedeće formule:

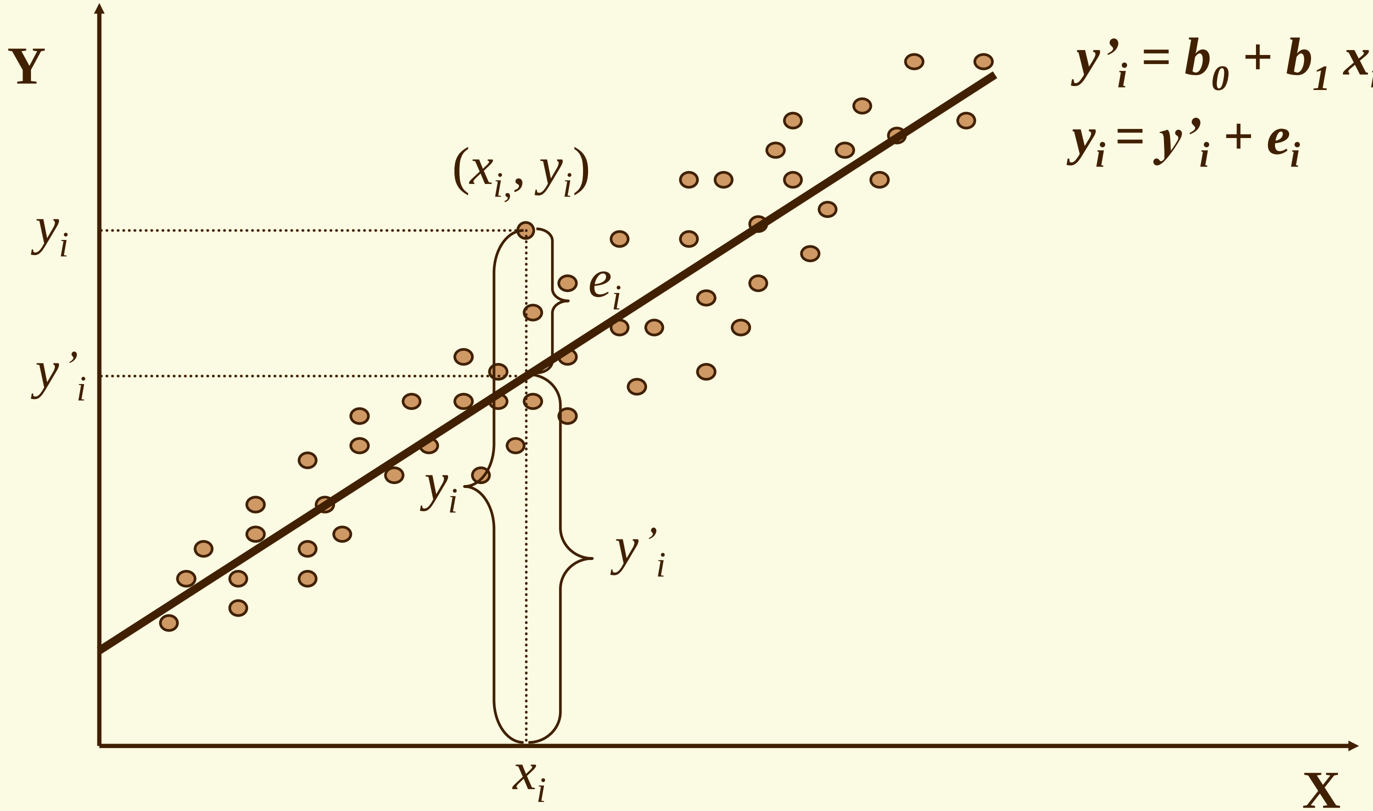
$$y'_i = b_0 + b_1 x_i$$

gdje je

- y'_i - prognozirani rezultat entiteta *i* u kriterijskoj varijabli
- b_0 i b_1 - regresijski koeficijenti
- x_i - rezultat entiteta *i* u prediktorskoj varijabli
- $i = 1,...,n$

REGRESIJSKA ANALIZA

Jednostavna linearna regresijska analiza



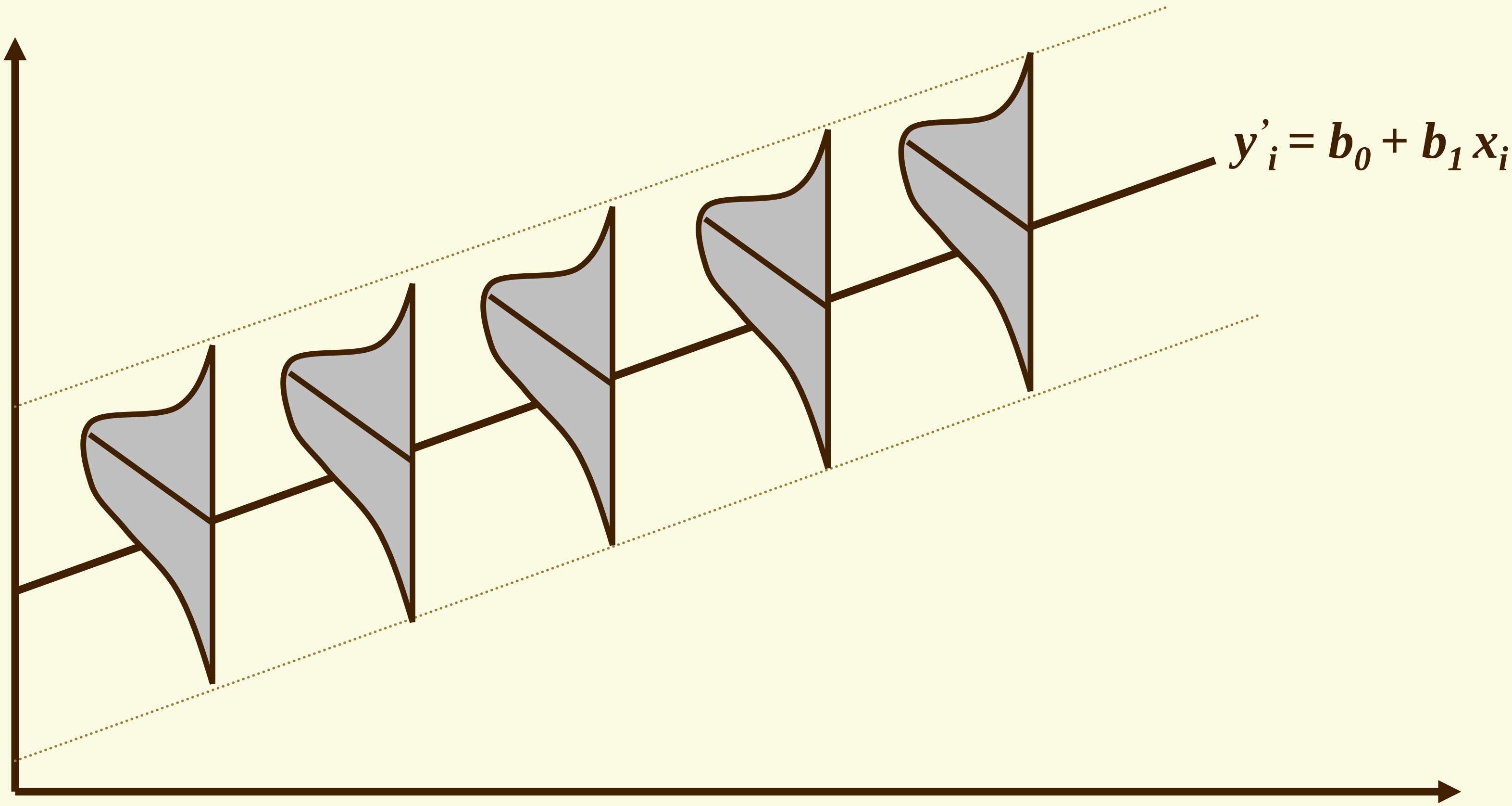
$$y'_i = b_0 + b_1 x_i$$

$$y_i = y'_i + e_i$$

(Prikaz regresijskog pravca, originalnih i prognoziranih rezultata u kriterijskoj varijabli i rezidualnih vrijednosti)

REGRESIJSKA ANALIZA

Jednostavna linearna regresijska analiza



(Distribucija rezidualnih vrijednosti oko regresijskog pravca)

REGRESIJSKA ANALIZA

Jednostavna linearna regresijska analiza

Koeficijenti regresijskog pravca utvrđuju se ***metodom najmanjih kvadrata***.

Metoda najmanjih kvadrata temelji se na uvjetu da je suma kvadrata rezidualnih vrijednosti minimalna

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - y'_i)^2 = \min$$

gdje je

- e_i - rezidualna vrijednost entiteta i
- y_i - rezultat entiteta i u kriterijskoj varijabli
- y'_i - prognozirani rezultat entiteta i u kriterijskoj varijabli
- $i = 1,...,n$

REGRESIJSKA ANALIZA

Jednostavna linearna regresijska analiza

Regresijski koeficijent b_0 predstavlja odsječak na osi zavisne varijable *y*, odnosno, vrijednost zavisne varijable *y* ukoliko je vrijednost nezavisne varijable $x = 0$.

$$b_0 = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x^2{}_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n \sum_{i=1}^{n} x^2{}_i - (\sum_{i=1}^{n} x_i)^2}$$

gdje je

- y_i - rezultat entiteta *i* u kriterijskoj varijabli
- x_i - rezultat entiteta *i* u prediktorskoj varijabli
- $i = 1,...,n$

REGRESIJSKA ANALIZA

Jednostavna linearna regresijska analiza

Regresijski koeficijent b_1 određuje nagib pravca, odnosno, pokazuje koliko se u prosjeku linearno mijenja vrijednost zavisne varijable *y* za jedinični porast vrijednosti nezavisne varijable *x*.

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x^2{}_i - (\sum_{i=1}^{n} x_i)^2}$$

gdje je

- y_i - rezultat entiteta *i* u kriterijskoj varijabli
- x_i - rezultat entiteta *i* u prediktorskoj varijabli
- $i = 1,...,n$

REGRESIJSKA ANALIZA

Jednostavna linearna regresijska analiza

Regresijski koeficijenti se također mogu izračunati i rješavanjem regresijske jednadžbe u matričnom obliku:

$$\overbrace{\begin{vmatrix} y_1 \\ . \\ . \\ . \\ y_n \end{vmatrix}}^{\boldsymbol{y}} = \overbrace{\begin{vmatrix} 1 & x_1 \\ . & . \\ . & . \\ . & . \\ 1 & x_n \end{vmatrix}}^{\boldsymbol{X}} \cdot \overbrace{\begin{vmatrix} b_0 \\ b_1 \end{vmatrix}}^{\boldsymbol{b}} + \overbrace{\begin{vmatrix} e_1 \\ . \\ . \\ . \\ e_n \end{vmatrix}}^{\boldsymbol{e}}$$

gdje je

- ***y*** - vektor *n* rezultata entiteta u kriteriju
- ***X*** - matrica reda $n \cdot 2$ rezultata entiteta u prediktoru
- ***b*** - vektor regresijskih koeficijenata
- ***e*** - vektor *n* rezidualnih vrijednosti

REGRESIJSKA ANALIZA

Jednostavna linearna regresijska analiza

$$y = X b \qquad /X^T$$

$$X^T y = X^T X b \qquad / (X^T X)^{-1}$$

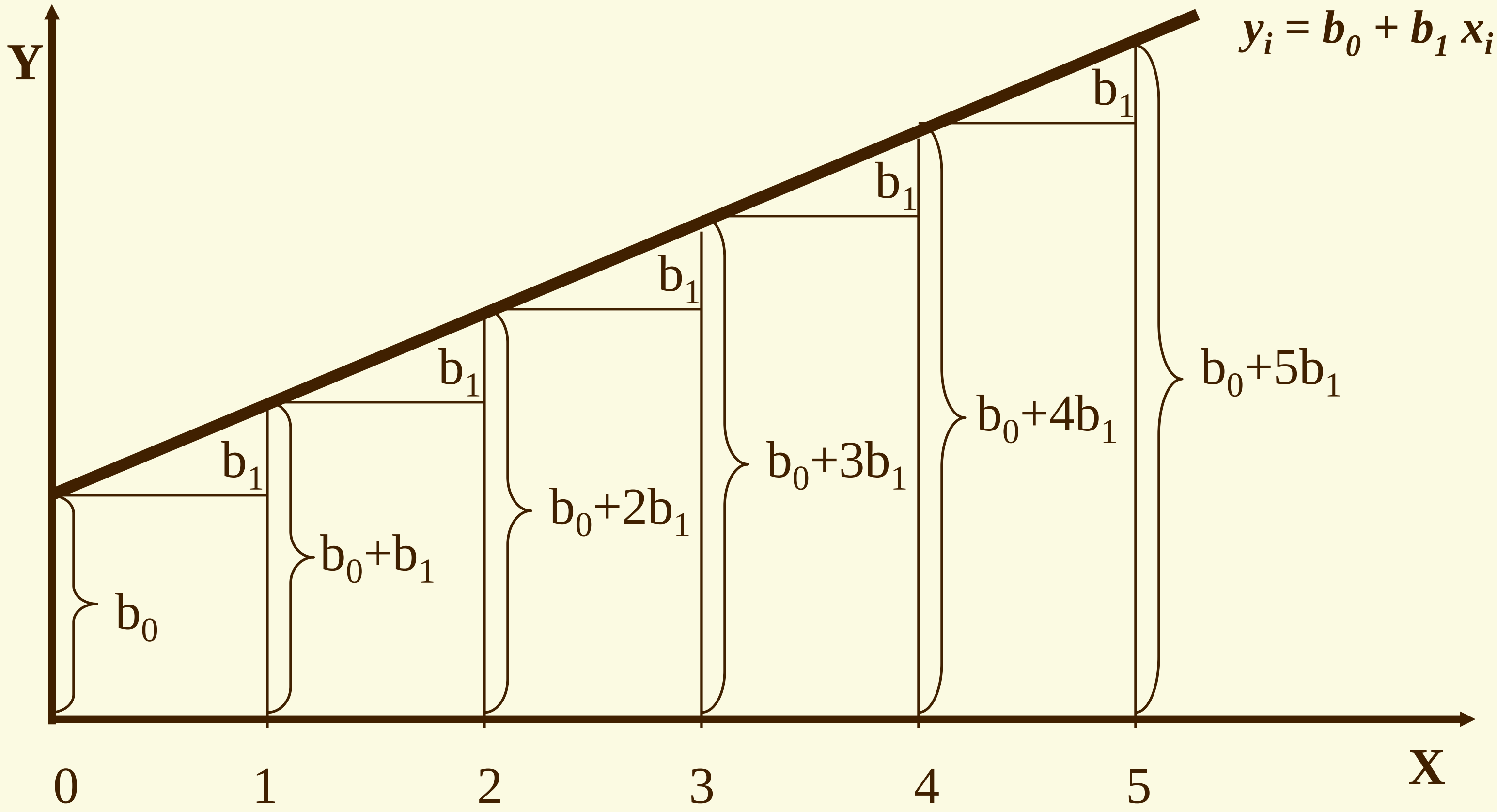
$$b = (X^T X)^{-1} X^T y$$

$$y' = X b$$

$$e = y - y'$$

REGRESIJSKA ANALIZA

Jednostavna linearna regresijska analiza



(Prikaz regresijskih koeficijenata b_0 i b_1)

REGRESIJSKA ANALIZA

Jednostavna linearna regresijska analiza

Standardna pogreška prognoze (σ_e) je drugi korijen iz varijance rezidualnih vrijednosti, a predstavlja mjeru reprezentativnosti regresijskog modela.

$$\sigma_e = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y'_i)^2}{n-2}}$$

gdje je

- y_i - rezultat entiteta *i* u kriterijskoj varijabli
- y'_i - prognozirani rezultat entiteta *i* u kriterijskoj varijabli
- $i = 1,...,n$
- *n* - broj entiteta

REGRESIJSKA ANALIZA

Jednostavna linearna regresijska analiza

Koeficijent korelacije između kriterijske i prediktorske varijable izražava veličinu njihove linearne povezanosti.

Kada je $\boldsymbol{r_{x,y} = 0}$ to znači da nezavisna varijabla *x* nema nikakav utjecaj na varijabilitet kriterijske varijable *y* .

Ako koeficijent korelacije ima maksimalnu vrijednost $\mathbf{r_{x,y} = 1}$, to znači da je cjelokupan varijabilitet varijable *y* moguće pripisati utjecaju varijable *x*.

Kvadrat koeficijenta korelacije (r^2) naziva se ***koeficijent determinacije***, a predstavlja proporciju varijance kriterijske varijable koju je moguće objasniti putem prediktorske varijable.

STATISTICA 7

Jednostavna linearna regresijska analiza

Jednostavna linearna regresijska analiza izvodi se slijedom koraka: padajući izbornik *Statistics* → *Multiple Regression*. U dijaloškom okviru koji se pokreće odabirom opcije *Variables* potrebno je označiti zavisnu varijablu (*Dependent var.*) i nezavisnu varijablu (*Independent variable list*). Nakon odabira varijabli rezultatima regresijske analize se pristupa putem opcije *Summary: Regression results*.

Zadatak - U datoteci *TREND.sta* utvrdite regresijsku jednadžbu kojom je moguće prognozirati tjelesnu masu (*TEZ*) testiranog ispitanika na temelju broja tjedana vježbanja (*BRM*)!

REGRESIJSKA ANALIZA

Višestruka linearna regresijska analiza

Višestrukom linearnom regresijskom analizom utvrđuje se linearna povezanost između dviju ili više nezavisnih (prediktorskih) i jedne zavisne (kriterijske) varijable pri čemu regresijska jednadžba ima sljedeći oblik:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_m x_{im} + e_i$$

gdje je

- y_i - rezultat entiteta i u kriterijskoj varijabli
- $b_0, \ldots, b_m$ - regresijski koeficijenti
- $x_{i1}, \ldots, x_{im}$ - rezultati entiteta i u m prediktorskih varijabli
- e_i - rezidualna vrijednost entiteta i
- $i = 1, \ldots, n$ (n - broj entiteta), a m - broj prediktora

REGRESIJSKA ANALIZA

Višestruka linearna regresijska analiza

Regresijski koeficijenti mogu se izračunati rješavanjem regresijske jednadžbe u matričnom obliku:

$$\boldsymbol{y} = \boldsymbol{X}\,\boldsymbol{b} + \boldsymbol{e}$$

$$\begin{vmatrix} y_1 \\ . \\ . \\ y_n \end{vmatrix} = \begin{vmatrix} 1 & x_{11} & . & . & x_{1m} \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & x_{n1} & . & . & x_{nm} \end{vmatrix} \cdot \begin{vmatrix} b_0 \\ b_1 \\ . \\ b_m \end{vmatrix} + \begin{vmatrix} e_1 \\ . \\ . \\ e_n \end{vmatrix}$$

REGRESIJSKA ANALIZA

Višestruka linearna regresijska analiza

$$y = X b \qquad /X^T$$

$$X^T y = X^T X b \qquad /(X^T X)^{-1}$$

$$b = (X^T X)^{-1} X^T y$$

$$y' = X b$$

$$e = y - y'$$

gdje je

- ***y*** - vektor *n* rezultata entiteta u kriteriju
- ***X*** - matrica reda $n \cdot m+1$ rezultata entiteta u *m* prediktora
- ***b*** - vektor regresijskih koeficijenata
- ***y'*** - vektor *n* prognoziranih vrijednosti u kriteriju
- ***e*** - vektor *n* rezidualnih vrijednosti

REGRESIJSKA ANALIZA

Višestruka linearna regresijska analiza

Regresijski koeficijent b_0 predstavlja vrijednost zavisne varijable *y* ukoliko je vrijednost svih nezavisnih varijabli jednaka *0* .

Regresijski koeficijenti $b_1,...,b_m$ pokazuju koliko se u prosjeku linearno mijenja vrijednost zavisne varijable *y* za jedinični porast vrijednosti odgovarajuće nezavisne varijable ($x_1,...,x_m$) uz uvjet da su vrijednosti ostalih nezavisnih varijabli konstantne.

REGRESIJSKA ANALIZA

Višestruka linearna regresijska analiza

Ako se kriterijska i prediktorske varijable prethodno standardiziraju regresijska jednadžba poprima sljedeći oblik:

$$k_i = \beta_1 z_{i1} + \beta_2 z_{i2} + \ldots + \beta_m z_{im} + \varepsilon_i$$

gdje je

- k_i - standardizirani rezultat entiteta i u kriterijskoj varijabli
- $\beta_1, \ldots, \beta_m$ - standardizirani regresijski koeficijenti
- $z_{i1}, \ldots, z_{im}$ - standardizirani rezultati entiteta i u m prediktorskih varijabli
- ε_i - standardizirana rezidualna vrijednost entiteta i
- $i = 1, \ldots, n$ (n - broj entiteta), a m - broj prediktora

REGRESIJSKA ANALIZA

Višestruka linearna regresijska analiza

Standardizirani regresijski koeficijenti mogu se izračunati rješavanjem sljedeće jednadžbe u matričnom obliku:

$$\mathbf{k} = \mathbf{Z}\ \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\begin{vmatrix} k_1 \\ . \\ . \\ k_n \end{vmatrix} = \begin{vmatrix} z_{11} & . & . & z_{1m} \\ . & . & . & . \\ . & . & . & . \\ z_{n1} & . & . & z_{nm} \end{vmatrix} \cdot \begin{vmatrix} \beta_1 \\ . \\ . \\ \beta_m \end{vmatrix} + \begin{vmatrix} \varepsilon_1 \\ . \\ . \\ \varepsilon_n \end{vmatrix}$$

REGRESIJSKA ANALIZA

Višestruka linearna regresijska analiza

$\boldsymbol{k} = \boldsymbol{Z}\,\boldsymbol{\beta} \qquad / \boldsymbol{Z}^T n^{-1}$

$\underbrace{\boldsymbol{Z}^T \boldsymbol{k} n^{-1}} = \underbrace{\boldsymbol{Z}^T \boldsymbol{Z} n^{-1}}\ \boldsymbol{\beta}$

$$\begin{vmatrix} r_1 \\ . \\ . \\ r_m \end{vmatrix} \quad \begin{vmatrix} 1 & r_{12} & . & r_{1m} \\ r_{21}. & . & . & . \\ . & . & . & . \\ r_{m1} & . & . & 1 \end{vmatrix}$$

$\boldsymbol{r} = \boldsymbol{R}\,\boldsymbol{\beta} \qquad / \boldsymbol{R}^{-1}$

$\boldsymbol{\beta} = \boldsymbol{R}^{-1}\,\boldsymbol{r}$

gdje je

- ***k*** - vektor *n* standardiziranih rezultata entiteta u kriteriju
- ***Z*** - matrica reda $n \cdot m$ standardiziranih rezultata entiteta u *m* prediktora
- ***β*** - vektor standardiziranih regresijskih koeficijenata
- ***r*** - vektor korelacija *m* prediktora s kriterijem
- ***R*** - matrica međusobnih korelacija *m* prediktora

REGRESIJSKA ANALIZA

Višestruka linearna regresijska analiza

Standardizirani regresijski koeficijenti $\boldsymbol{\beta_1,\ldots,\beta_m}$ su relativni koeficijenti utjecaja, a predstavljaju veličinu promjene zavisne varijable izraženu u dijelovima standardne devijacije za jedinični porast standardizirane vrijednosti odgovarajuće nezavisne varijable ($z_1,\ldots,z_m$) uz uvjet da su vrijednosti preostalih nezavisnih varijabli konstantne.

REGRESIJSKA ANALIZA

Višestruka linearna regresijska analiza

Statistička značajnost svakog pojedinog regresijskog koeficijenta se testira putem Studentove t-distribucije.

Pri tome je za svaki regresijski koeficijent moguće postaviti sljedeću alternativnu (H1), odnosno nultu (H0) hipotezu:

H1: $b_j \neq 0$ - Utjecaj prediktora *j* na kriterijsku varijablu je statistički značajan uz pogrešku *p*.

H0: $b_j = 0$ - Uz pogrešku *p* ne možemo tvrditi da je utjecaj prediktora *j* na kriterijsku varijablu statistički značajan.

REGRESIJSKA ANALIZA

Višestruka linearna regresijska analiza

Standardna pogreška prognoze (σ_e) je drugi korijen iz varijance rezidualnih vrijednosti, a predstavlja mjeru reprezentativnosti regresijskog modela.

$$\sigma_e = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y'_i)^2}{n-(m+1)}}$$

gdje je

- y_i - rezultat entiteta *i* u kriterijskoj varijabli
- y'_i - prognozirani rezultat entiteta *i* u kriterijskoj varijabli
- $i = 1,...,n$
- *n* - broj entiteta, *m* - broj prediktorskih varijabli

REGRESIJSKA ANALIZA

Višestruka linearna regresijska analiza

Koeficijent multiple korelacije **(*ρ*)** je korelacija između kriterijske varijable i varijable prognoziranih rezultata, a izražava veličinu linearne povezanosti skupa prediktorskih varijabli s kriterijem.

Koeficijent multiple korelacije se kreće u intervalu od *0* do *1* pri čemu *0* označava nikakavu, a *1* potpunu zavisnost kriterijske varijable o skupu prediktorskih varijabli.

Kvadrat koeficijenta korelacije (ρ^2) naziva se ***koeficijent multiple determinacije***, a predstavlja proporciju varijance kriterijske varijable koju je moguće objasniti putem skupa prediktorskih varijabli.

REGRESIJSKA ANALIZA

Višestruka linearna regresijska analiza

Statistička značajnost koeficijenta multiple korelacije (ρ) se testira putem Snedecorove F-distribucije.

Pri tome je moguće postaviti sljedeću alternativnu (H1), odnosno nultu (H0) hipotezu:

H1: $\boldsymbol{\rho \neq 0}$ - Povezanost između skupa prediktora i kriterijske varijable je statistički značajna uz pogrešku *p*.

H0: $\boldsymbol{\rho = 0}$ - Uz pogrešku *p* ne možemo tvrditi da je povezanost između skupa prediktora i kriterijske varijable statistički značajna.

STATISTICA 7

Višestruka linearna regresijska analiza

Višestruka linearna regresijska analiza izvodi se slijedom koraka: padajući izbornik *Statistics* → *Multiple Regression*. U dijaloškom okviru koji se pokreće odabirom opcije *Variables* potrebno je označiti zavisnu varijablu (*Dependent var.*) i dvije ili više nezavisnih varijabli (*Independent variable list*). Nakon odabira varijabli rezultatima regresijske analize pristupa se putem opcije *Summary: Regression results*.

Zadatak - U datoteci *KOSARKA.sta* utvrdite regresijsku jednadžbu kojom je moguće prognozirati uspješnost ekipe (*K2*) na temelju skupa situacijskih parametara (*v1*-*v12*)!

REGRESIJSKA ANALIZA

Literatura za pripremanje kolokvija

Dizdar, D. (2006). *Kvantitativne metode*. Zagreb: Kineziološki fakultet, str. 182-213.

Petz, B. (2002). *Osnovne statističke metode za nematematičare*. Jastrebarsko: str. Naklada Slap, 237-247.

Mejovšek, M. (2003). Uvod u metode znanstvenog istraživanja u društvenim i humanističkim znanostima. Jastrebarsko: Naklada Slap, str. 182-200.