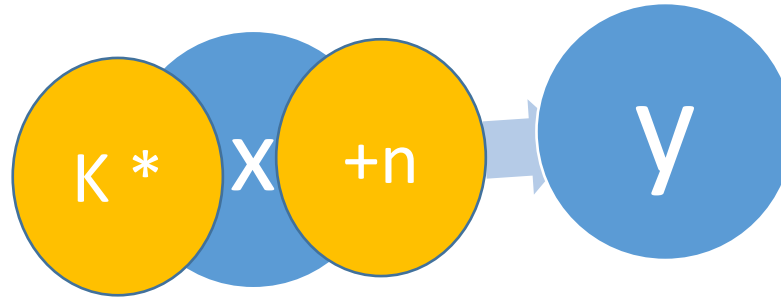


Regresijska analiza

Doc. dr. Petra Povalej Bržan

Univariatna linearna enačba

Vhod- neodvisne
spremenljivke



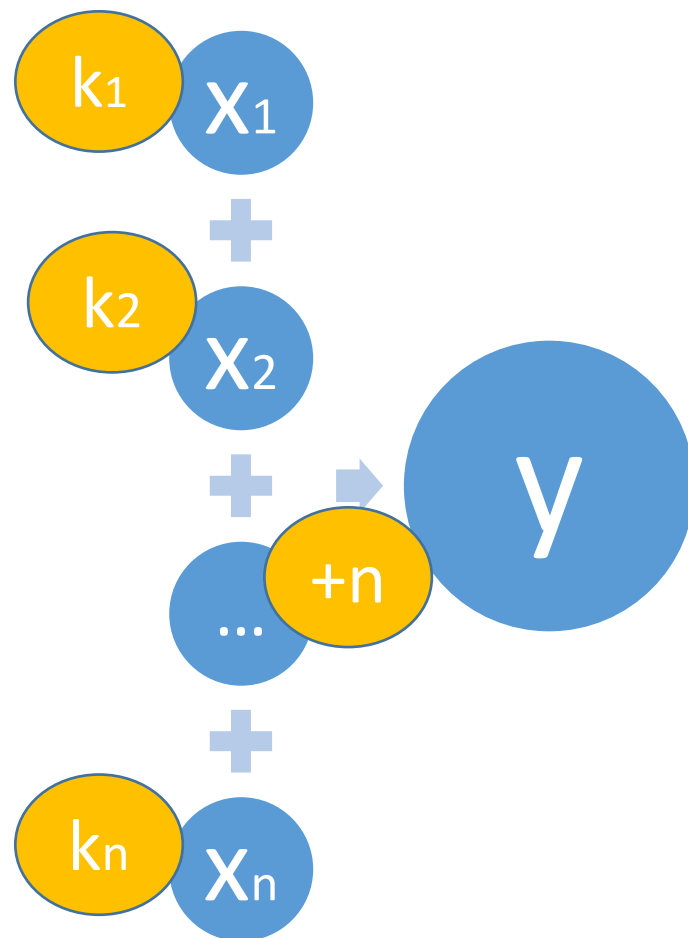
Primer 1: Nivo povprečnega sladkorja v krvi
je povezan s starostjo.

Primer 2: Krvni pritisk s starostjo narašča.

Izhod – odvisna
spremenljivka

Multivariatna linearna enačba

Vhod- neodvisne
spremenljivke



Izhod – odvisna
spremenljivka

Primer: Nivo povprečnega sladkorja v krvi je povezan s starostjo in ITM.

Cilji analize

- Ali obstaja povezava med dvema (večinim) spremenljivkama/mi?
- Kako močna je povezava?
- Katera spremenljivka največ prispeva k izračunu izhodne spremenljivke?
- Napoved Y če poznamo X

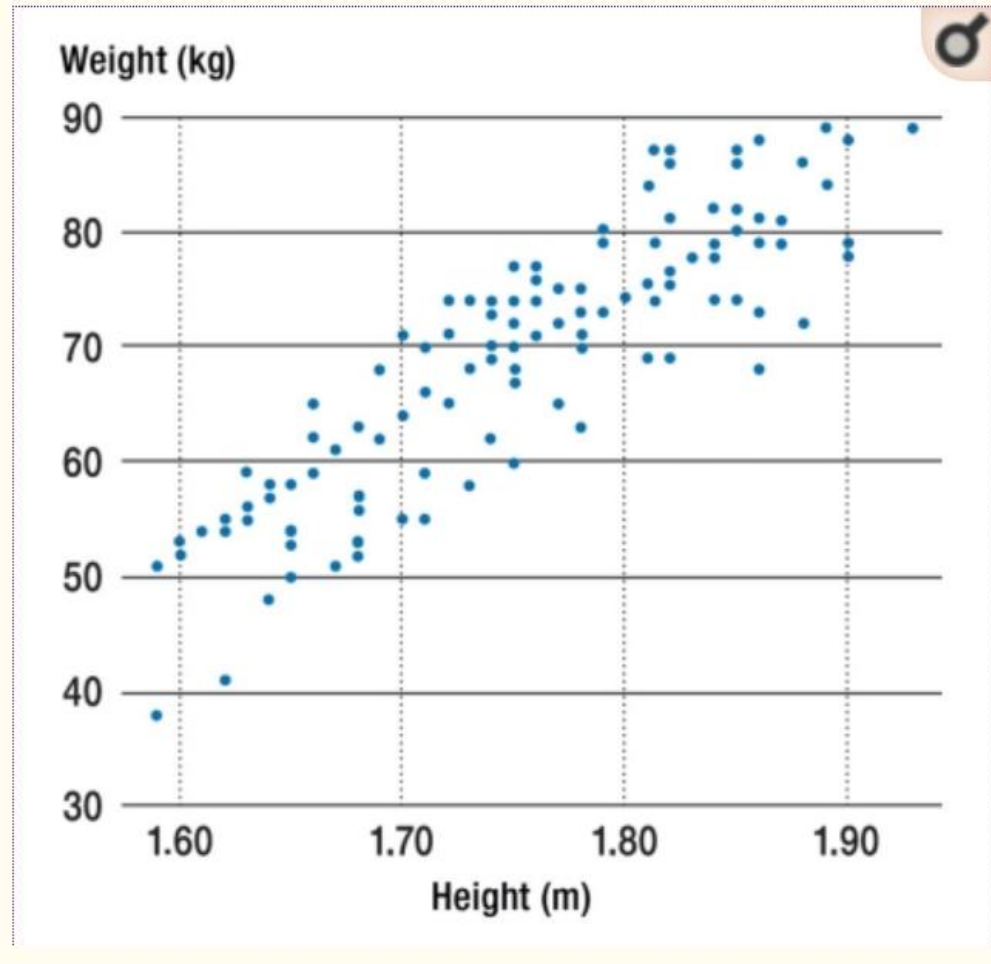
Regresijska analiza omogoča:

- **Opis:** Razmerja med odvisnimi spremenljivkami in neodvisnimi spremenljivkami je mogoče statistično opisati s pomočjo regresijske analize.
- **Ocena:** Vrednosti odvisnih spremenljivk je mogoče oceniti iz opazovanih vrednosti neodvisnih spremenljivk.
- **Prognostika (napovedovanje):** Ugotoviti je mogoče dejavnike tveganja, ki vplivajo na rezultat, in na osnovi regresijskega modela izračunati prognoze za posamezne enote.

Vrste regresije

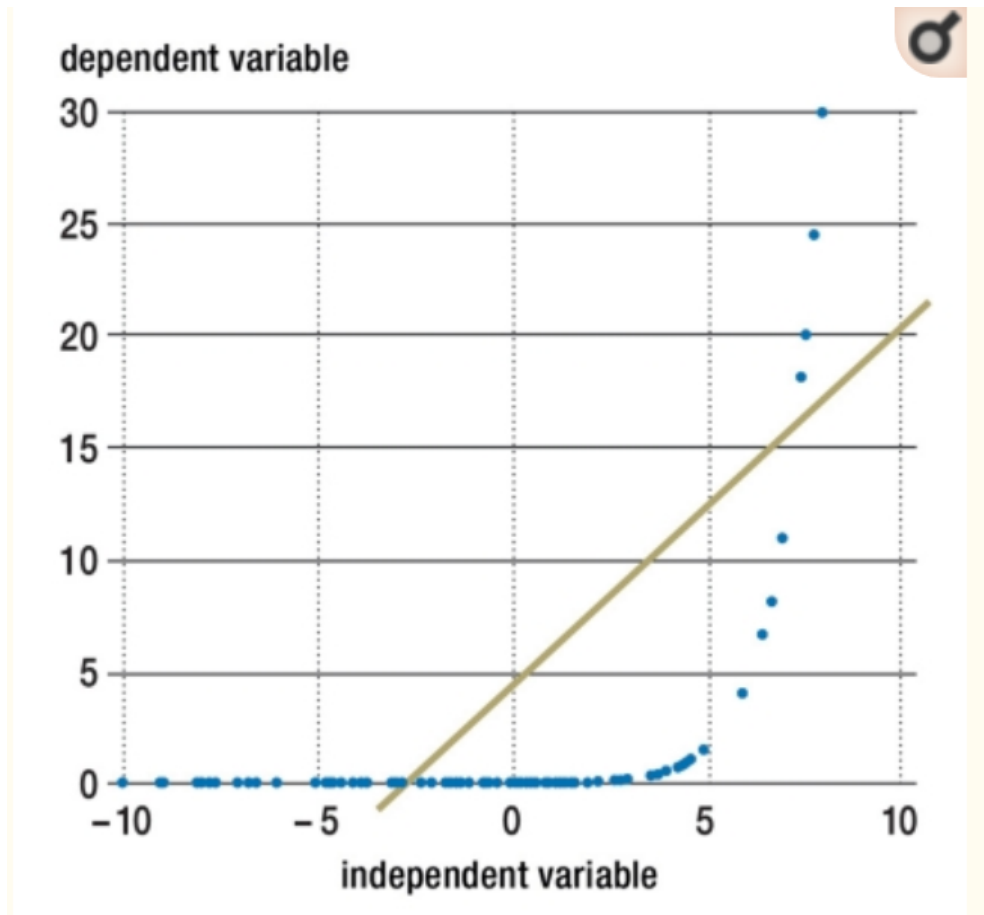
	Uporaba	Odvisna spremenljivka	Neodvisna spremenljivka	
Linearna regresija	Opis linearne povezanosti med spremenljivkami	Zvezna (teža, starost, pritisk,...)	Zvezna ali / in nominalna (slamnata oblika)	
Logistična regresija	Predikcija verjetnosti za pripadnost skupini (izhod: da/ne)	Dihotomna (uspešnost zdravljenja da /ne, bolezen da/ne...)	Zvezna ali / in nominalna (slamnata oblika)	
Cox-ova regresija	Analiza preživetja	Čas preživetja (od diagnoze do terminacijskega dogodka)	Zvezna ali / in nominalna (slamnata oblika)	

Primer linearne regresije

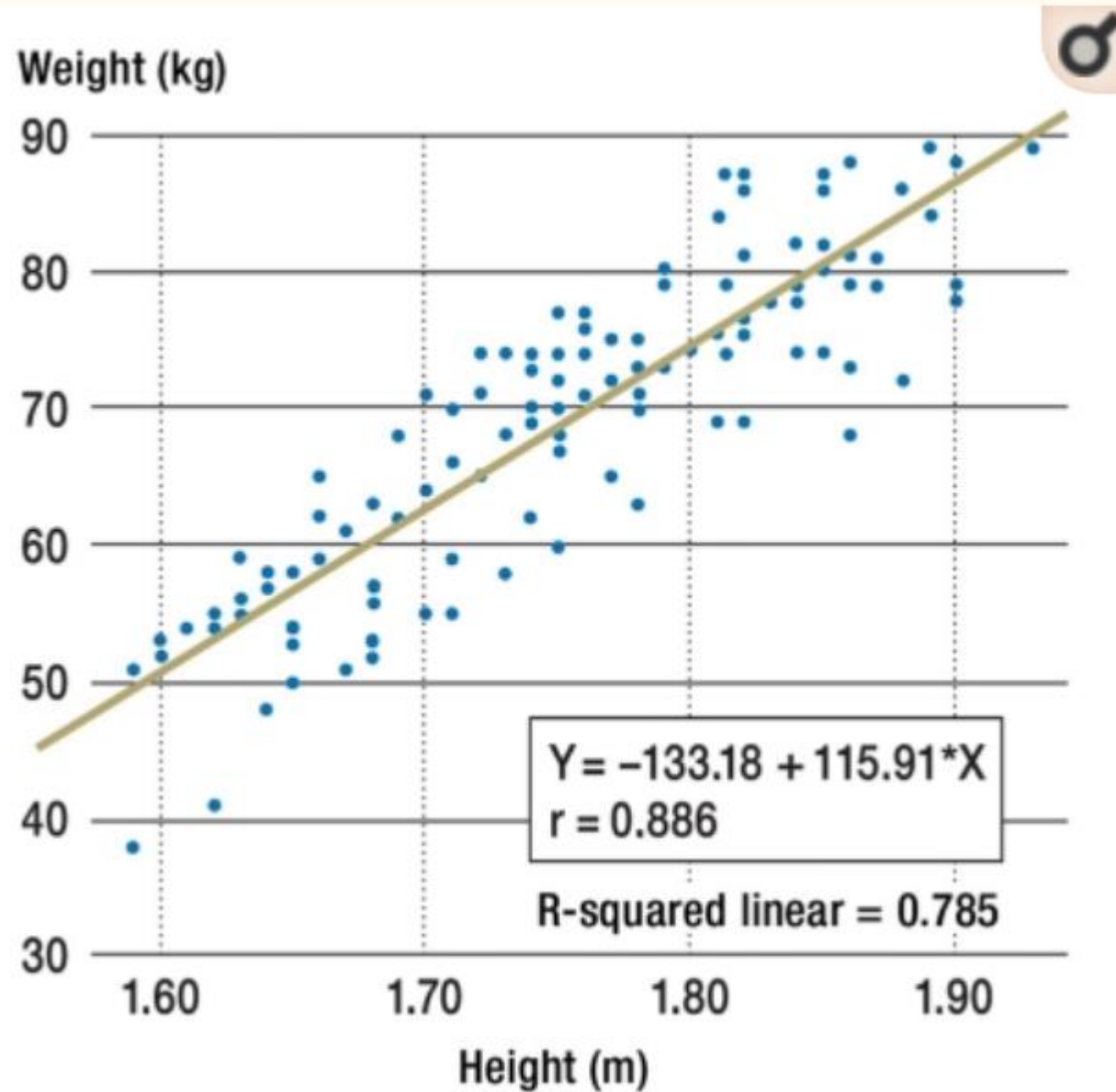


Neodvisna spremenljivka: zvezna ali binarna ali kategorična

Odvisna spremenljivka: zvezna



Eksponentna
povezava med
spremenljivkama –
neprimerno za
linearno regresijo



$$y = k \cdot x + n$$

n – kje seka premica
os x

K – nagib premice

Primer 1: Izdelajmo čim bolj natančen model za izračun teže otroka pri starosti 1 mesec na osnovi porodne teže otroka

- Baza: `dojenje_Urejena.csv`

Y – zvezna spremenljivka – teža otroka pri starosti 1 mesec

X – zvezna spremenljivka – porodna teža otroka

- Uporabimo Linearno regresijo (ker je Y zvezna spremenljivka)

Osnovna enačba

Vhodna spremenljivka -> Izhodna spremenljivka
(Vhodne spremenljivke -> Izhodna spremenljivka)

$$\text{Izhod}_i = (\text{Model}_i) + \text{napaka}_i$$

$$\text{Izhod} = (c + b * x) + \varepsilon$$

c- konstanta (vrednost odvisne spremenljivke, ko je vhodna spremenljivka X enaka 0)

b – regresijski koeficient – nagib premice (smer in moč povezave z odvisno spremenljivko - efekt)

X – vrednost neodvisne spremenljivke

ε – napaka modela

Ocena modela

Model Summary - OtrokTeza1Mesec

Model	R	R ²	Adjusted R ²	RMSE	Durbin-Watson		
					Autocorrelation	Statistic	p
H ₀	0.000	0.000	0.000	548.028	0.019	1.950	0.633
H ₁	0.598	0.357	0.355	440.045	0.039	1.919	0.439

R – korelacijski koeficient

R² – odstotek variance v odvisni spremenljivki, ki je pojasnjen s tem modelom – giblje se med 0 in 1. večji kot je R², bolj natančen je model (manjša je napaka modela)

Adjusted R² – prilagojen R² na celotno populacijo.

35,7% variabilnosti v teži otroka pri starosti 1 mesec lahko pojasnimo na osnovi otrokove porodne teže.

R² znaša 0,357 kar pomeni, da je s tem modelom pojasnjene skoraj 36% variance v teži otroka pri starosti 1 mesec.

Opis modela (1)

Coefficients ▼

Model		Unstandardized	Standard Error	Standardized	t	p
H ₀	(Intercept)	4340.755	28.568		151.945	< .001
H ₁	(Intercept)	1826.707	177.843		10.271	< .001
	OtrokTeza	0.735	0.052	0.598	14.255	< .001

Porodna teža otroka je pozitivno povezana s težo otroka pri starosti 1 mesec. Višja porodna teža torej pomeni višjo težo pri starosti 1 mesec ($p < 0,001$).

- Nestandardizirani koeficienti (Unstandardized) - povedo enačbo modela
- Standardizirani koeficienti (Standardized) – primerjamo spremenljivke med seboj na isti ravni (standardizirane vrednosti) in povedo katera spremenljivka ima močnejši vpliv na težo otroka pri starosti en mesec ter smer vpliva (predznak) – najmočnejše vpliva porodna teža

Opis modela (2)

Coefficients ▼

Model		Unstandardized	Standard Error	Standardized	t	p
H ₀	(Intercept)	4340.755	28.568		151.945	< .001
H ₁	(Intercept)	1826.707	177.843		10.271	< .001
	OtrokTeža	0.735	0.052	0.598	14.255	< .001

Enačba modela:

Teža otroka pri starosti 1 mesec = $1826,707 + 0,735 \cdot \text{OtrokTeža}$

INTERPRETACIJA:

- Če bi bila porodna teža 0 g bi bila teža otroka pri starosti 1 mesec 1826,707 g. (malce nesmiselno)
- Teža pri starosti 1 mesec se poveča za 0,735 g za vsak gram večje porodne teže.

Opis rezultatov

- *S pomočjo linearne regresije smo preverili povezavo med težo otroka pri starosti 1 mesec in porodno težo otroka. Ugotavljamo, da lahko težo otroka pri starosti 1 mesec statistično značilno napovemo s pomočjo porodne teže ($\beta=0,598$; $p<0,001$) ($F(1, 366)=203,217$; $p<0,001$). Z modelom lahko pojasnimo skoraj 36% variance v teži otroka pri starosti 1 mesec. Regresijska enačba se glasi: **Teža otroka pri starosti 1 mesec = $1826,707 + 0,735 * \text{OtrokTeža}$***

Primer 2: Izdelajmo čim bolj natančen multivariatni model za izračun teže otroka pri starosti 1 mesec

- Baza: dojenje_Urejena.csv
- Y: teža otroka pri starosti 1 mesec
- X: porodna teža in dolžina, obseg glavice

Y – zvezna spremenljivka

X – tri zvezne spremenljivke

- Uporabimo Linearno regresijo

Ocena modela

Model Summary - OtrokTeza1Mesec ▼

Model	R	R ²	Adjusted R ²	RMSE	Durbin-Watson		
					Autocorrelation	Statistic	p
H ₀	0.000	0.000	0.000	548.028	0.019	1.950	0.633
H ₁	0.602	0.363	0.358	439.218	0.028	1.941	0.563

R – korelacijski koeficient

R² – odstotek variance v odvisni spremenljivki, ki je pojasnjen s tem modelom.

36% variance v teži otroka pri starosti 1 mesec lahko pojasnimo na osnovi otrokove porodne dolžine, porodne teže in obsega glave.

R² znaša 0,363 kar pomeni, da je s tem modelom pojasnjene 36% variance v teži otroka pri starosti 1 mesec.

Opis modela (1)

Coefficients ▼

Model		Unstandardized	Standard Error	Standardized	t	p
H ₀	(Intercept)	4340.755	28.568		151.945	< .001
H ₁	(Intercept)	509.959	779.282		0.654	0.513
	OtrokTeza	0.647	0.078	0.526	8.300	< .001
	OtrokDolzina	12.811	15.573	0.051	0.823	0.411
	OtrokObsegGlavice	27.775	18.109	0.073	1.534	0.126

Vse tri neodvisne spremenljivke so pozitivno povezane s težo otroka pri starosti 1 mesec. Porodna dolžina ne prispeva statistično značilno k natančnosti modela ($p=0,411$) in bi bilo smiselno preskusiti kaj se zgodi, če jo iz modela izključimo (gledamo spremembo v R^2) – poskus optimizacije modela. Višja porodna teža torej pomeni višjo težo pri starosti 1 mesec. Tudi obseg glavice ne prispeva statistično značilno k natančnosti modela ($p=0,126$)..

- Nestandardizirani koeficienti (Unstandardized) - povedo enačbo modela
- Standardizirani koeficienti (Standardized) – primerjamo spremenljivke med seboj na isti ravni (standardizirane vrednosti) in povedo katera spremenljivka ima močnejši vpliv na težo otroka pri starosti en mesec ter smer vpliva (predznak) – najmočnejše vpliva porodna teža

Opis modela (2)

Coefficients ▼

Model		Unstandardized	Standard Error	Standardized	t	p
H ₀	(Intercept)	4340.755	28.568		151.945	< .001
H ₁	(Intercept)	509.959	779.282		0.654	0.513
	OtrokTeza	0.647	0.078	0.526	8.300	< .001
	OtrokDolzina	12.811	15.573	0.051	0.823	0.411
	OtrokObsegGlavice	27.775	18.109	0.073	1.534	0.126

Enačba modela:

Teža otroka pri starosti 1 mesec = 509,959 + 0,647 * OtrokTeža + 12,811 * OtrokDolzina + 27,775 * OtrokObsegGlavice

INTERPRETACIJA:

- Če bi bila porodna teža 0 g, porodna dolžina 0 cm in obseg glavice 0 cm bi bila teža otroka pri starosti 1 mesec 509,959 g. (malce nesmiselno)
- Teža pri starosti 1 mesec se poveča za 0,647g za vsak gram večje porodne teže.
- Teža pri starosti 1 mesec se poveča za 12,811 za vsak centimeter večje porodne dolžine.
- Za vsak cm obsega glavice se teža pri starosti 1 mesec poveča za 27,775g.

Opis rezultatov

- *S pomočjo multivariatne linearne regresije smo preverili povezavo med težo otroka pri starosti 1 mesec in porodno težo in dolžino ter obsegom glavice. Ugotavljamo, da lahko težo otroka pri starosti 1 mesec statistično značilno napovemo s pomočjo porodne teže ($\beta=0,526$; $p<0,001$), porodne dolžine ($\beta=0,051$; $p=0,411$) in obsegom glavice otroka ($\beta=0,073$; $p=0,126$) ($F(3, 364)=69,121$; $p<0,001$). Z modelom lahko pojasnimo več kot 36% variance v teži otroka pri starosti 1 mesec. Regresijska enačba se glasi:*
- ***Teža otroka pri starosti 1 mesec = $509,959 + 0,647 * \text{OtrokTeža} + 12,811 * \text{OtrokDolzina} + 27,775 * \text{OtrokObsegGlavice}$***