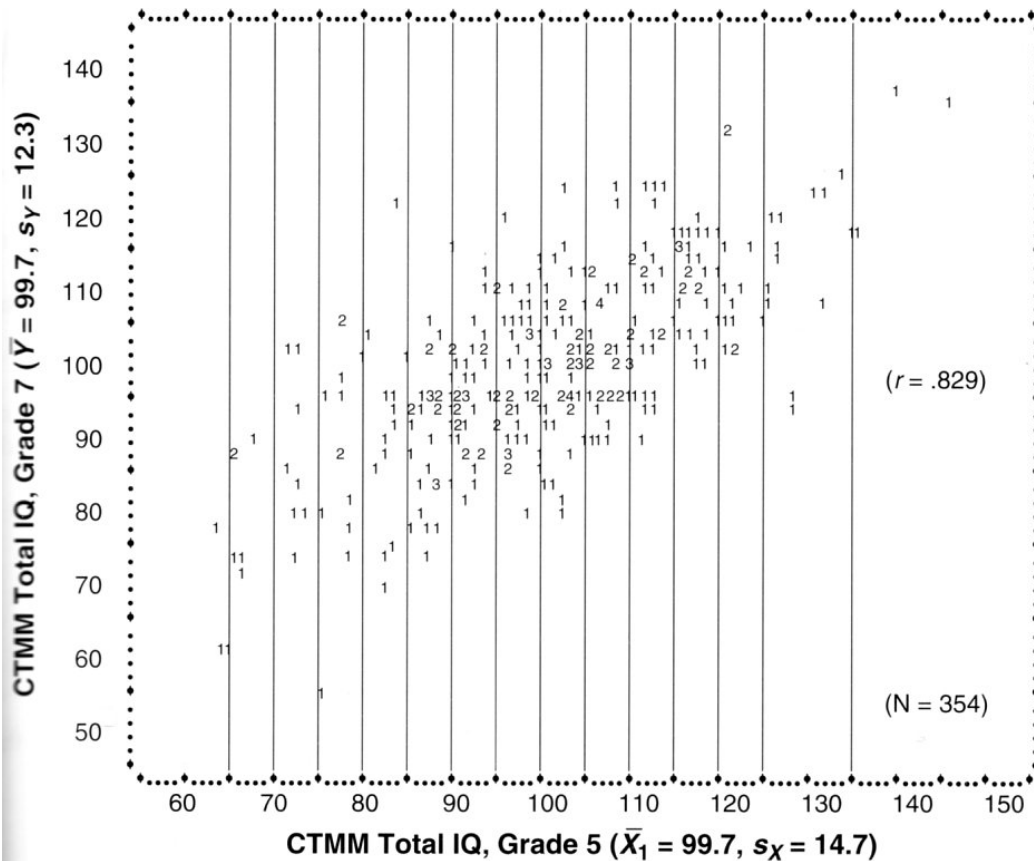


Regrese

Pro označení hodnot predikované proměnné používám níže apostrof (Y')

3. Který termín patří mezi ostatní nejméně?
a) percentil b) korelace c) regrese d) predikce
4. Pokud $r=0,5$ a $z_X=2,0$, kolik je odhad hodnoty Y , tj. $z_{Y'}$?
5. Jakému percentilu odpovídá hodnota $z_{Y'}$ z předchozího příkladu? Předpokládejte, že rozložení proměnných je normální.
6. Pokud $r_{XY}=0,5$, a výkon osoby u proměnné X odpovídá P_2 , odhadněte na úrovni jakého percentilu bude výkon téže osoby u proměnné Y .
a) P_{50} b) P_{75} c) P_{16} d) P_2
7. Pokud $r=-0,6$ a $z_X=-1,5$, kolik je odhad hodnoty Y , tj. $z_{Y'}$?
8. Je-li $r=1$, znamená to, že ve všech párech hodnot korelovaných proměnných jsou obě hodnoty stejné?
12. Platí, že s rostoucí hodnotou r roste i směrodatná odchylka reziduí (reziduální rozptyl)?
13. Když predikujeme Y z X a $s_X=s_Y=15$, platí že korelační koeficient (r) je roven regresnímu koeficientu (b)?
16. Pokud $s_Y = 10$ a $r = 0,6$, jaká je směrodatná odchylka reziduálních hodnot (s_e)?
17. Předpokládáme-li biviační normální rozložení
a) Pokud směrodatná odchylka reziduálních hodnot $s_e = 8$, kolik procent skutečných hodnot závislé proměnné se od předpovězené hodnoty liší o méně než 8 bodů?
b) A kolik procent skutečných hodnot bude o více než 8 bodů vyšší než předpovězené hodnoty?
c) Bude podíl předpovědí, které podhodnotí skutečnou hodnotu o více než 8 bodů stejný jako v otázce b)?
18. Korelace mezi IQ skóry kteréhokoli z rodičů a jejich dětí je přibližně 0,5. Také víme, že průměr IQ rodičů i dětí je stejný – 100 a směrodatné odchylky v obou populacích jsou také stejné – 15.
a) Odhadněte průměrné IQ dětí matek s IQ = 130,
b) Odhadněte průměrné IQ dětí otců s IQ = 90,
c) Odhadněte průměrné IQ dětí matek s IQ = 100.
19. Průměr IQ obou rodičů koreluje s IQ jejich dítěte přibližně 0,6. Jaká bude směrodatná odchylka chyb odhadu (s_e , s_{res}) při predikci IQ dětí?
20. Pokud $s_{res} = 12$, v jakém procentu případů se budou skutečné IQ skóry lišit od předpovězených o více než 12 bodů.



Následující otázky se vztahují k obrázku výše.

P1. Obrázek výše je vygenerovaný scatter IQ skóre 354 dětí testovaných v 5. (X) a 7. (Y) třídě. Ze scatterplotu určete následující:

- Jaký je nejvyšší a nejnižší skóre v 5. třídě?
- Jaký je nejvyšší a nejnižší skóre v 7. třídě?
- Vypadá regrese lineárně?
- Splňuje podle scatteru závislá proměnná podmínku homoscedasticity?
- Spočítejte regresní koeficient $b_{Y,X}$.
- Spočítejte průsečík a .
- Vytvořte regresní rovnici s dosazenými koeficienty.
- Bob získal v 5. třídě skóre IQ 140. Předpovězte jeho skóre IQ v 7. třídě.
- Sam získal v 5. třídě skóre IQ 70. Předpovězte jeho skóre IQ v 7. třídě.
- Zakreslete do obrázku regresní přímku.
- Spočítejte směrodatnou odchylku chyb odhadu s_e (s_{reg}).
- Jaká část predikovaných skóre se bude lišit od skutečných hodnot o méně než 7 bodů.

o) S přibližně dvoutřetinovou pravděpodobností leží Bobovo IQ v sedmé třídě mezi ____ a ____ a Samovo mezi ____ a ____.

P2. Konkrétní test inteligence koreluje s testem čtení 0,82. Víte-li, že $m_{IQ}=100$, $s_{IQ}=15$, $m_C=8$ a $s_C=2$ a předpokládáme-li, že obě proměnné jsou normálně rozložené...

- určete regresní rovnici na predikci čtení ze skóre IQ.
- Jaký je průměrný výkon ve čtení při IQ=100?
- Jaký je průměrný výkon ve čtení při IQ=90?
- Porovnejte percentilové ekvivalenty skóre X a Y' z úlohy c)
- Kolik % dětí s IQ=90 bude mít nadprůměrný skóre ve čtení? [nápověda: cesta vede přes reziduální(chybový) rozptyl]

Odpovědi

- a)
- $z_Y' = 1,0$
- P_{84}

6. c)
7. 0,9
8. Ne nutně, ale ve standardizovaných skórech ano (tj. $z_X = z_Y$).
12. ne, klesá
13. ano
16. 8
17. a) 68%
b) 16%
c) ano
18. a) 115
b) 95
c) 100
19. $15 \cdot (0,8) = 12$
20. 68%

- P1 a) 64 a 146
b) 55 a 138
c) ano
d) ano
e) ano
f) 0,694
g) 30,5
h) $Y' = 0,694X + 30,5$
i) 128
j) 79
l) 6,9
m) cca 68%

- P2. a) $Y' = 0,11X - 3$
b) 8
c) 6,9
d) predikovaný skór má vyšší percentilový ekvivalent (P_{29}) než hodnota prediktoru (P_{25})
e) $s_e = 1,1$ a predikovaný skór pro IQ=90 je 6,9. $m_c = 8$, takže chyby o více než jednu s_e směrem nahoru budou nad 8. Nad $z=1$ je 16% rozložení. Tj. cca 16%.