H4: Lineaire Regressie met 1 regressor

1. LINEAR REGRESSION MODEL

Populaion regression line:

Test score Yi = ß0 + ß1 Xi + ui

X = independent variable/ regressor

Y = dependent variable

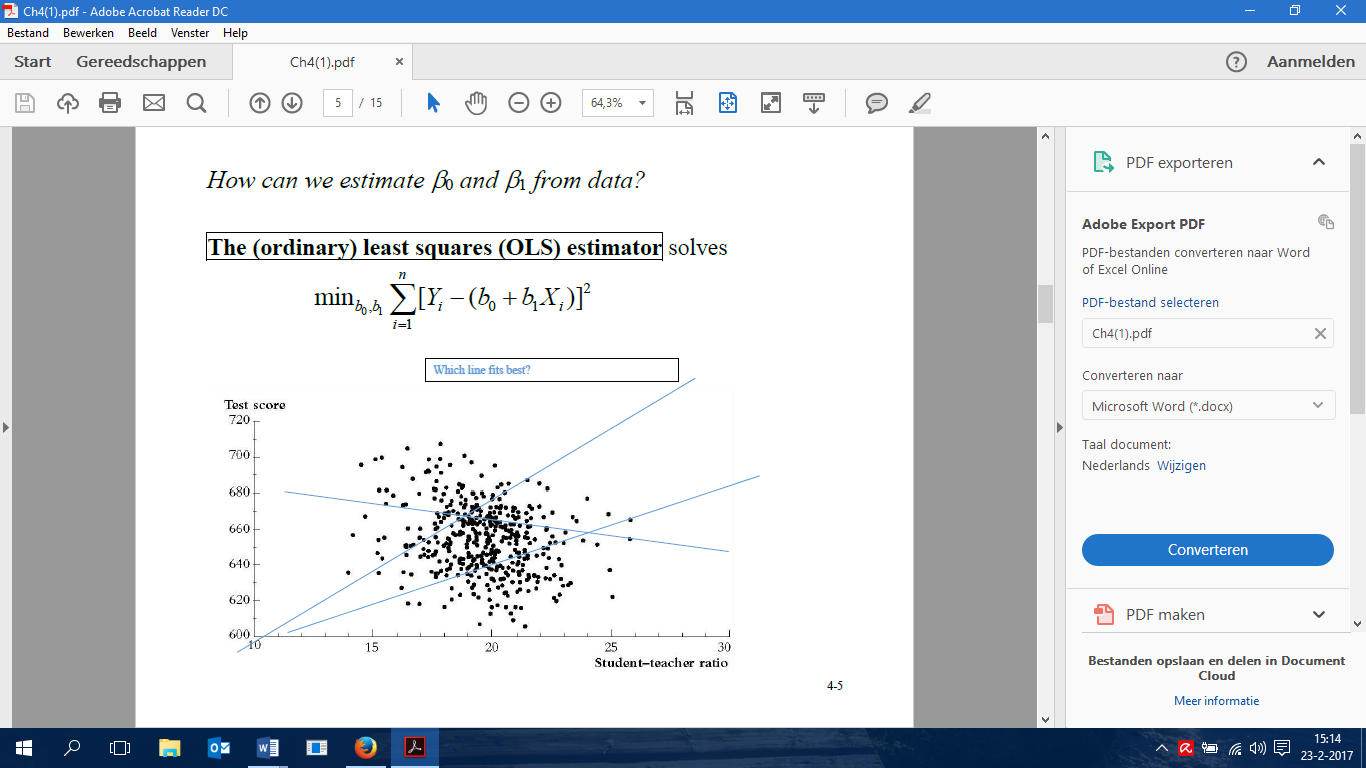
ß0 = intercept

ß1 = slope

ui = error term (need E(ui)=0 )

2. ORDINARY LEAST SQUARES (OLS) ESTIMATOR

= How can we estimate ß0  and ß1 from data

=

= kleinste kwadraat schatter, zoek bo en b1 die bij beste rechte horen

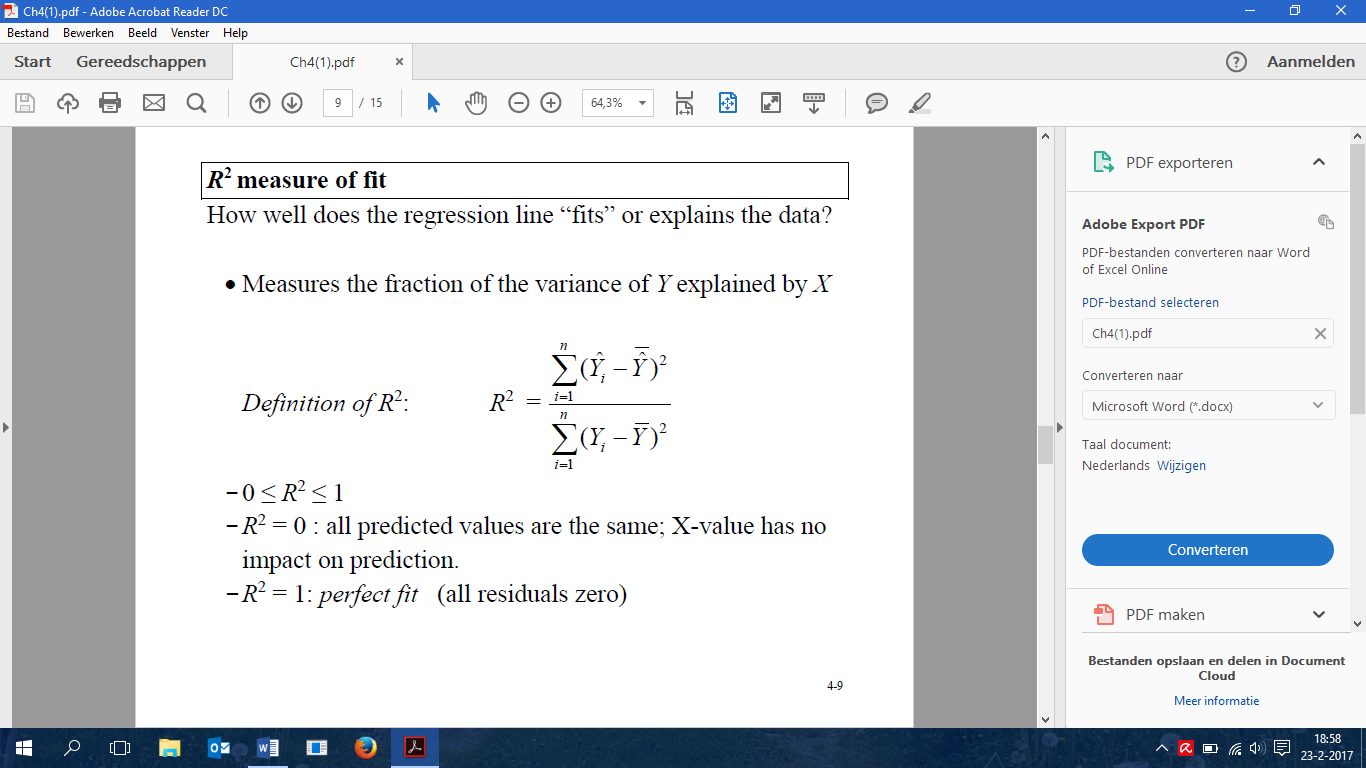
= minimaliseert som v kwadraat fouten v alle schatters m = b+b1X



3. R2 MEASURE OF FIT

= Hoeveel voorspellende/ verklarende kracht heeft model

= Fractie v steekproefvariantie Yi verklaard door Xi



0 ≤ R2≤ 1

R2 = 0: geen voorspellende kracht (elke variabele gelijk)

R2 = 1: perfect fit (elk residu = 0)

! Je weet niet vanaf wanneer R2 ‘groot’ is (hangt af v moeilijkheid voorspelling)

4. STATA OUTPUT

Zie slide 4-10

**Test score**

= Hoe goed je het doet op toets

**BI zelf berekenen**

= Schatter ± 1.96 x SE

Testen op hypothese nul

🡪 Want als die hypothese waar is heeft de hele berekening geen zin (effect variabele = 0)

5. LEAST SQUARES ASSUMPTIONS

Yi = ß0 + ß1 Xi + ui

3 Assumpties voor OLS

* **E(ui|Xi) = 0 = E(ui)**  
  = cruciale conditie  
  = Voor elke waarde v X, verwachte waarde u = 0   
  (weggelate variabele zelfde of je nu met grote of kleine klassen zit, gemiddelde 0)  
  🡪 Als voldaan: u en X ongecorreleerd  
  🡪 u en X onafhankelijk: conditie voldaan  
  Helaas niet altijd voldaan
* **(Xi, Yi), i=1,…,n are independent x identically distributed (i.i.d.)**= Onafhankelijk v andere variabelen x hun waardenverdeling  
  ! I.i.d. is niet altijd vervuld maar wel grootste deel vd tijd= we hebben een willekeurige steekproef
* **X and Y have finite fourth moments**  
  = Geen grote uitschieters in steekproef  
  OLS kan echter wel gevoelig zijn aan uitschieters, PLOT YOUR DATA!

6. SAMPLING DISTRIBUTION OF THE OLS ESTIMATOR

Onder de 3 OLS assumpties geldt:

E(^ß1) = ß1 (dus ß1 unbiased)

^ß1 🡪 ß1 als n stijgt

N large: 