

2.1 en 2.2

We zullen nu gaan kijken naar het basis model van de meervoudige regressie. Het model gaat over: DATA = MODEL + ERROR

Wanneer mensen een oorzaak toeschrijven aan een bepaalde gebeurtenis dan zijn dit causale (oorzakelijke) attributies. Echter, er zijn meestal meerdere oorzaken die voor het resultaat hebben gezorgd. Het feit dat mensen meestal maar één oorzaak toeschrijven aan de gebeurtenis wordt *monocausality* genoemd. Om meerdere oorzaken mee te kunnen nemen moeten we een complexer model maken. Wanneer we twee of meer onafhankelijke variabelen hebben, gebruiken we de volgende formule:

$$\hat{Y} = b_1x_1 + b_2x_2 + c$$

In deze formule is b_1 de helling van de eerste onafhankelijke variabele, en b_2 is de helling van de tweede onafhankelijke variabele.

Soms gebruiken we in de statistiek de term controleren voor, maar wat betekent dat eigenlijk? In de experimentele psychologie zal gecontroleerd worden voor variabelen waarin we niet geïnteresseerd zijn door deze te standaardiseren. Maar het is moeilijk om voor alles te controleren, wanneer dit niet mogelijk is gebruiken we statistische controle.

Wanneer we een onafhankelijke variabele hebben, kunnen we een bepaald deel van de variantie in de afhankelijke variabele verklaren. Wanneer we een tweede onafhankelijke variabele toevoegen, willen we weten hoeveel variantie deze variabele verklaard. Maar er is een probleem, deze twee variabelen zullen hoogstwaarschijnlijk correleren. Wanneer ze correleren, zullen ze een deel van de variantie delen. Het meervoudige regressie model rekent deze delen uit, door deze correlaties mee te nemen in de berekening. Het model rekent ook het effect van elke onafhankelijke variabele apart uit, wanneer de andere variabele zou zijn verwijderd.

2.3

Wanneer we beide onafhankelijke variabelen in de regressie analyse invoeren, krijgen we schattingen van de helling coëfficiënten voor elke variabele, waarbij wordt gecontroleerd voor de andere variabelen.

2.4

R is de meervoudige correlatie, ook wel de determinatiecoëfficiënt genoemd. Het laat de totale correlatie tussen alle onafhankelijke variabelen en de afhankelijke variabele zien. R^2 is de waarde van R in het kwadraat. R^2 staat voor de totale variantie in de afhankelijke variabele verklaard door de onafhankelijke variabelen. De waarde van R^2 kan worden gezien als het deel van de verklaarde variantie dat wordt verklaard door het verschuiven van de komma twee plaatsen naar rechts, en dit laat de waarde in een percentage zien.

2.5

De aangepaste R^2 is de R^2 maar kleiner, deze zorgt ervoor dat er een goede schatting wordt gemaakt van de waarde van R^2 in de populatie. De reden dat deze aanpassing wordt gedaan is dat wanneer een andere onafhankelijke variabele wordt toevoegt, het niet erg waarschijnlijk is dat de correlatie tussen die onafhankelijke variabele en de afhankelijke variabele gelijk is aan 0, zelfs wanneer het 0 is in de populatie. R^2 zal altijd een beetje omhoog gaan wanneer er een andere onafhankelijke variabele wordt toegevoegd. Dus de aangepaste R^2 is naar beneden aangepast om te compenseren voor de toename van R^2 . Hoe kleiner de steekproef, hoe groter de random variantie van 0 zal zijn, dus zal er een grotere aanpassing naar beneden nodig zijn van R^2 . De berekening van R^2 gaat als volgt:

$$\text{Adj. } R^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1} \quad (102)$$

Waarin n het aantal mensen is en k het aantal onafhankelijke variabelen. Wanneer n groter wordt, is de mate waarin R^2 naar beneden moet worden aangepast kleiner. En wanneer k groter wordt, neemt de mate waarin R^2 kleiner wordt toe.