

Vanliga frågor och svar FAQ

Peter M. Dahlgren, JMG

2016-11-24

Här sammanställs frågor och svar som brukar dyka upp under kursens gång.

Grundläggande

Vilket sambandsmått ska jag välja?

Det är variabelns skalnivå som avgör vilken korrelation du ska använda.

- Om båda variablerna är på nominalskalenivå: Cramérs V.
- Om båda variablerna är på ordinalskalenivå: Spearmans rho, Kendalls tau-b/tau-c.
- Om båda variablerna är på intervallskalenivå: Pearsons r.
- Om variablerna har olika skalnivåer: Utgå från den minsta skalnivån.

Låt säga att du har två variabler, kön (nominalskala) och inkomst (intervallskala). Då är könsvariabeln den minsta skalnivån. Därmed kan du välja Cramérs V.

Är sambandet stort eller litet?

Det är ämneskunskap som framför allt bör guida tolkningen av sambandsmått.

Om du saknar ämneskunskap har Cohen (1988) gett följande tips:

Tolka Pearsons r och Cramérs V

- 0,10 = litet samband
- 0,30 = medelstort samband
- 0,50 = stort samband

Vad är p-värden?

P-värden definieras som sannolikheten att erhålla ett resultat (eller mer extremt), givet att nollhypotesen är sann.

Det betyder att p-värdet visar sannolikheten att få det resultatet man har fått under förutsättningen att det inte finns någon skillnad.

Enkelt uttryckt visar p-värdet huruvida ett samband, skillnad eller effekt i ett urval faktiskt finns i populationen. Det är en skala som går från 0 till 1. Ett p-värde kommer tillsammans med ett statistiskt test.

Hur tolkar jag p-värdet?

P-värden kan anta vilket värde som helst mellan 0 och 1. Ett p-värde på 0,042 betyder exempelvis att det är 4,2 % sannolikhet att man fått det resultat man fått (givet att nollhypotesen är sann).

Man brukar traditionellt säga att p-värden under 0,05 är statistiskt signifikant, medan p-värden över 0,05 ej är statistiskt signifikanta. Om resultatet är statistiskt signifikant skriver man ofta att $p < 0,05$. Detta är en godtycklig gräns. Gränsen brukar kallas alfanivå eller signifikansnivå, och den bestämmer man själv innan man gör analysen.

Vi kan ta ett exempel. Du har 100 stycken bollar och två korgar. Om bollarna placeras helt slumpmässigt i korgarna kommer det sannolikt att vara 50 bollar i ena korgen och 50 i den andra. Det är nollhypotesen, det vill säga att bollarna är jämnt fördelade.

Låt säga att bollarna inte är jämnt fördelade. Det kan vara 30 i ena korgen och 70 i den andra korgen. Då kan vi uttrycka sannolikheten att vi fått detta resultat, givet att bollarna är fördelade 50/50. Kom ihåg att p-värdet definieras som sannolikheten att erhålla ett resultat (eller mer extremt), givet att nollhypotesen är sann.

Det är sampelstorleken och effektstorleken som påverkar p-värdet.

Vanliga missförstånd kring p-värden

- P-värdet är *inte* sannolikheten att nollhypotesen är sann.
- P-värdet är *inte* sannolikheten att få samma resultat om man gör om analysen.
- P-värdet visar *inte* storleken på sambandet/skillnaden/effekten.
- P-värdet 0,049 är *inte* bättre än ett p-värde på 0,051.
- P-värdet bör *inte* betraktas svartvit.

Hur ska jag rapportera sambandsmått och p-värden?

P-värden rapporteras alltid tillsammans med ett statistiskt test, till exempel sambandsmått. Skriv därför p-värdet efter sambandsmättet. Rapportera alltid p-värdet med 3 decimaler.

Exempel:

Det fanns en medelstor korrelation mellan tidningsläsning och politisk kunskap ($r = 0,31$; $p = 0,002$).

Här syftar r på Pearsons r . Notera att texten i princip skrivs i följande format: Det fanns en [liten/medelstor/stor] korrelation mellan [X] och [Y] ($r = 0,000$; $p = 0,000$).

Du kan inte skriva att tidningsläsning *påverkar* politisk kunskap eftersom vi inte vet kausalriktningen.

Vilka viktiga begrepp bör jag känna till?

- felmarginal
- hypotes
- kausalitet
- korrelation
- korstabell
 - Chi^2
- deskriptiv statistik
 - typvärde
 - medelvärde
 - median
 - standardavvikelse
 - modalprocent
 - kvartil
 - variationsvidd
 - sambandsmått
 - * Cramérs V
 - * Spearmans ρ
 - * Kendalls $\tau\text{-b}/\tau\text{-c}$
 - * Pearsons r
- normaldistribution
 - population
 - sampel
- operationalisering
 - teoretisk definition
 - operationell definition
- reliabilitet
- p-värde
 - statistisk signifikans
- validitet
 - begreppsvaliditet
 - extern validitet
 - intern validitet
 - resultatvaliditet (eller *statistisk inferensvaliditet*)
- variabel

- beroende variabel
- oberoende variabel
- skalnivå

Referenser

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 edition). Hillsdale, N.J: Routledge.