

# Vanliga frågor och svar FAQ

Peter M. Dahlgren, JMG

2016-11-30

Här sammanställs frågor som dykt upp från studenter under årens lopp, inklusive svar med tips på vidare läsning. Svaren fungerar som komplement till kurslitteraturen, inte ersättning.

## Grundläggande

### Vad är sambandsmått?

Sambandsmått, eller korrelationsmått, är ett mått på sambandet mellan två variabler. Det innebär att två variabler samvarierar med varandra. Till exempel samvarierar vikt och längd. Ju längre man är, desto tyngre är man.

Videor om korrelation:

- What Is Correlation? (6 min)
- Statistics 101: Understanding Correlation (27 min)

### Vilket sambandsmått ska jag välja?

Det är variabelns skalnivå som avgör vilket sambandsmått 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.

Om du har variablerna kön (man/kvinna) och inkomst (i kronor) har kön den minsta skalnivån. Kön är då på nominalskalenivå och inkomst på intervallskalenivå. Därmed bör du välja Cramérs V.

## Hur tolkar jag sambandsmått?

Det är ämneskunskap som framför allt bör guida tolkningen av sambandsmått. Om du saknar ämneskunskap har Cohen (1988) gett riktlinjer för Pearsons  $r$  och Cramérs  $V$ :

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

Pearsons  $r$  varierar från -1 till +1. Cramérs  $V$  kan bara variera från 0 till 1 eftersom det är på nominalskalenivå.

## Vad är skillnaden mellan Phi och Cramérs $V$ ?

Phi kan du ignorera helt och hållet. Titta bara på Cramérs  $V$ .

Phi är ett sambandsmått som endast fungerar för korstabeller med storleken  $2 \times 2$  med två binära variabler. Cramérs  $V$  fungerar däremot på alla storlekar av korstabeller.

## Vad är p-värden?

P-värden kommer från ett statistiskt test som man genomför på en datamängd. P-värdet är sannolikheten att få ett resultat som är minst lika extremt, under förutsättning att det inte finns något samband i populationen.

P-värdet går från 0 till 1. Ju närmare 0 det är, desto mindre sannolikhet att du har fått resultatet ifråga, under förutsättning att det inte finns något samband i populationen.

Vanliga missförstånd kring p-värden:

- P-värdet visar *inte* sannolikheten att resultatet beror på slumpen.
- P-värdet visar *inte* sannolikheten att få samma resultat om man gör om analysen.
- P-värdet visar *inte* storleken på sambandet.
- P-värdet 0,049 är *inte* bättre än ett p-värde på 0,051.
- P-värdet bör *inte* betraktas svartvit.
- P-värdet visar *inte* vad som är relevant, viktigt eller vetenskapligt intressant.
- Stort p-värde betyder *inte* att nollhypotesen är sann.

Bli inte vemodig om du har svårt att förstå p-värden. Till och med forskare brottas med frågan: Not Even Scientists Can Easily Explain P-values. Det är ett krångligt koncept som är lätt att få om bakfoten. P-värden har kritiserats i nästan hundra år, kanske starkast av psykologen Paul Meehl som sa att det är

“one of the worst things that ever happened in the history of psychology” (Meehl, 1978, s. 817).

Fler videor om p-värden:

- What is a p-value? (6 min)
- Hypothesis testing and p-values - Khan Academy (12 min)

### Hur tolkar jag p-värdet?

Ett p-värde på 0,012 betyder exempelvis att det är 1,2 % sannolikhet att få det resultat du fått, eller mer extremt, under förutsättning att det inte finns något samband i populationen. Med andra ord betyder det att det är väldigt osannolikt att sambandet kunde uppstå om det faktiskt inte fanns i populationen. När du tittat på p-värdet tittar du därefter på sambandsmättet för att se hur stort sambandet är.

Vi kan ta ett enkelt 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. Då finns det inget samband mellan korgarna och bollar (detta är nollhypotesen).

Men om det är 30 bollar 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.

Det är sampelstorleken och effektstorleken som påverkar p-värdet. Det innebär att man får mindre p-värden om man har många analysenheter. I en surveyundersökning är det därför lättare att få lägre p-värden om man frågar fler personer. Det är också lättare att få låga p-värden ju större sambandet är.

### Vad är statistisk signifikans?

Man brukar traditionellt säga att p-värden under 0,05 ( $p < 0,05$ ) är *statistiskt signifikanta*, medan p-värden över 0,05 ej är statistiskt signifikanta. Detta är en godtycklig gräns och används av konvention. Det finns alltså inget magiskt med denna gräns. Gränsen brukar kallas alfanivå eller signifikansnivå, och den bestämmer man själv *innan* man gör analysen.

### Vad är nollhypotesen?

Nollhypotesen är antagandet att det inte finns någon skillnad mellan två variabler. Om du signifikanteststar ett sambandsmått, som Pearsons  $r$ , är nollhypotesen att  $r = 0$ , det vill säga att det inte finns något samband mellan variablerna.

Ju längre ifrån 0 sambandet är, desto lägre kommer p-värdet att bli.

Statistiska tester utgår alltid från nollhypotesen.

### 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 den kausala riktningen (om du inte genomför experiment eller dylikt).

## SPSS

### Var hittar jag p-värdet i SPSS?

Det står ofta *approximate significance* eller bara *Sig.*

### Hur analyserar jag en fråga som är kodad på flera variabler?

Låt säga att du har kodat en nyhetsartikel för att se vilka aktörer som förekommer i texten. Varje aktör kan då kodas så att de blir en egen variabel, exempelvis:

- Förekommer politiker? Ja=1 Nej=0
- Förekommer journalister? Ja=1 Nej=0
- Förekommer medborgare? Ja=1 Nej=0

Hur analyserar man detta med en frekvenstabell eller korstabell?

I SPSS kan du använda verktyget **Analyze > Tables > Custom Tables**. Du kan också använda verktyget **Analyze > Multiple Response**. Se länkarna nedan.

- Analysis Multiple response question (dichotomies)
- Analysis Multiple response question (categories)

### Vilka viktiga begrepp bör jag känna till?

- felmarginal

- hypotes
- kausalitet
- korrelation
- korstabell
  - $\text{Chi}^2$
- deskriptiv statistik
  - typvärde
  - medelvärde
  - median
  - standardavvikelse
  - modalprocent
  - kvartil
  - variationsvidd
  - sambandsmått
    - \* Cramér's V
    - \* Spearman's rho
    - \* Kendall's tau-b/tau-c
    - \* Pearson's 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.
- Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology* 46, 806–834.