

Korrelation och regression

Samband mellan variabler

Mathias Johansson

Kristofer Söderström

2025-12-15

Innehållsförteckning

1	Korrelation	1
2	Tolkning av korrelation	2
3	Beräkna korrelation (Pearson)	2
4	Enkel linjär regression	2
5	Beräkna regressionskoefficienter	2
6	Residualer (fel)	3
7	Determinationskoefficienten R^2	3
8	Inferentiell statistik för korrelation	3
9	Konfidensintervall för lutning	3
10	Antaganden för regression	3
11	Korrelation vs Kausalitet	4
12	Exempel på falska samband	4
13	Praktisk tillämpning	4
14	Sammanfattning	4
15	Interaktivt verktyg	4
16	Referenser	5

1 Korrelation

- Mäter styrkan och riktningen på ett linjärt samband mellan två variabler
- Korrelationskoefficienten (r) varierar från -1 till +1
- $r = +1$: Perfekt positivt samband

- $r = 0$: Inget linjärt samband
- $r = -1$: Perfekt negativt samband

2 Tolkning av korrelation

Värde på r	Tolkning
0.7 – 1.0	Stark positiv
0.3 – 0.7	Måttlig positiv
-0.3 – 0.3	Svag/ingen
-0.7 – -0.3	Måttlig negativ
-1.0 – -0.7	Stark negativ

3 Beräkna korrelation (Pearson)

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}$$

- \bar{x} och \bar{y} är medelvärden
- Mäter hur väl punkterna följer en rät linje

4 Enkel linjär regression

Förutsäger värdet på Y baserat på X:

$$\hat{y} = a + bx$$

- a = intercept (skärningspunkt med y-axeln)
- b = lutning (slope)
- \hat{y} = förutsagt värde

5 Beräkna regressionskoefficienter

Lutning:

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

Intercept:

$$a = \bar{y} - b\bar{x}$$

6 Residualer (fel)

- **Residual** = Observerat värde - Förutsagt värde
- $e_i = y_i - \hat{y}_i$
- Visar hur långt varje punkt ligger från regressionslinjen
- Summan av residualerna = 0 (i minsta-kvadratmetoden)

7 Determinationskoefficienten R^2

$$R^2 = r^2$$

- Andelen av variansen i Y som förklaras av X
- Varierar från 0 till 1 (0% till 100%)
- $R^2 = 0.64$ betyder att 64% av variationen förklaras

8 Inferentiell statistik för korrelation

Nollhypotes: $H_0 : \rho = 0$ (ingen korrelation i populationen)

Teststatistika:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

- Frihetsgrader: $df = n - 2$
- Om $p < 0.05$: Korrelationen är statistiskt signifikant

9 Konfidensintervall för lutning

$$b \pm t_{\alpha/2} \cdot SE_b$$

där standardfelet för b:

$$SE_b = \frac{s_e}{\sqrt{\sum (x_i - \bar{x})^2}}$$

- s_e = standardfelet för residualerna
- Ger ett intervall för den "sanna" lutningen

10 Antaganden för regression

1. **Linjäritet** – Sambandet är linjärt
2. **Oberoende** – Observationerna är oberoende
3. **Normalitet** – Residualerna är normalfördelade
4. **Homoskedasticitet** – Konstant varians i residualerna

11 Korrelation vs Kausalitet

- **Viktigt!** Korrelation innebär inte orsakssamband
- “Correlation is not causation”
- Kan finnas:
 - Omvänd kausalitet
 - Bakomliggande variabel (confounding)
 - Slumpmässigt samband

12 Exempel på falska samband

- Glasskonsumtion korrelerar med drunkningsolyckor
 - Bakomliggande variabel: Sommarväder
- Antalet filmer med Nicolas Cage korrelerar med drunkningar i pooler
 - Slumpmässigt samband

13 Praktisk tillämpning

1. Visualisera alltid data först (scatterplot)
2. Beräkna r och R^2
3. Testa om korrelationen är signifikant
4. Tolka resultaten i kontext
5. Var försiktig med extrapolering

14 Sammanfattning

Begrepp	Beskrivning
r	Korrelationskoefficient (-1 till +1)
R^2	Förklarad varians (0 till 1)
a	Intercept (y när $x=0$)
b	Lutning (ändring i y per enhet x)
Residual	Fel = observerat - förutsagt

15 Interaktivt verktyg

Prova det interaktiva regressionsverktyget för att:

- Klicka och skapa egna datapunkter
- Se regressionslinjen uppdateras i realtid
- Visualisera residualer (fel)

Se: [Interaktiv regression](#)

16 Referenser

- [Pearson correlation](#)
- [Linear regression](#)
- [Correlation vs causation](#)