- deskriptive (beschreibende) Statistik

  Ziel: Daten aus einer Stichprobe beschreiben

  Wert des Korrelationskoeffizienten aus einer Stichprobe, z.B. $r = 0{,}8$

- induktive (schließende) Statistik

  Ziel: von einer Stichprobe auf (andere Stichproben aus der) Grundgesamtheit schließen

  Kann ein Wert dieser Größe nur durch die Zusammensetzung der Zufalls-Stichprobe zustande kommen, auch wenn in der Grundgesamtheit gar kein Zusammenhang vorliegt? D.h.: $H_0 : \rho = 0$

- signifikant = „überzufällig"

  D.h. wenn in der Grundgesamtheit gar kein Zusammenhang vorliegen würde, wäre es sehr unwahrscheinlich*, dass ein Wert dieser Größe nur durch die Zusammensetzung der Zufalls-Stichprobe zustande kommt.

  Man würde also erwarten, auch in den meisten anderen Stichproben aus derselben Grundgesamtheit einen Zusammenhang zu finden.

- signifikant heißt aber nicht notwendigerweise: starker, inhaltlich relevanter Zusammenhang

  * d.h. Wahrscheinlichkeit kleiner als Signifikanzniveau α

Signifikanz vs. Effektstärke

zweiseitiger Test für $H_0 : \rho = 0$ auf 5%-Niveau

- Stärke des Zusammenhangs: $r = 0{,}8$
  Stichprobengröße: $n = 100$
  $\Rightarrow$ Zusammenhang signifikant
- Stärke des Zusammenhangs: $r = 0{,}8$
  Stichprobengröße: $n = 5$
  $\Rightarrow$ Zusammenhang nicht signifikant

- Stärke des Zusammenhangs: $r = 0{,}08$
  Stichprobengröße: $n = 100$
  $\Rightarrow$ Zusammenhang nicht signifikant
- Stärke des Zusammenhangs: $r = 0{,}08$
  Stichprobengröße: $n = 5000$
  $\Rightarrow$ Zusammenhang signifikant

Rangkorrelation

Rang-Korrelation nach Spearman

Der Korrelationskoeffizient nach Bravais-Pearson r:

- setzt ein metrisches Skalenniveau voraus,
- unterschätzt den Zusammenhang von x und y, wenn dieser nicht-linear ist,
- reagiert empfindlich auf Ausreißer.

Der Rang-Korrelationskoeffizient nach Spearman r_s

- wird auch bei ordinalskalierten Daten angewendet,
- spricht auch auf nicht-lineare monotone Zusammenhänge an,
- ist robust gegen Ausreißer.

$r_s = +1$ bedeutet, dass y mit steigendem x streng monoton steigt.
$r_s = -1$ bedeutet, dass y mit steigendem x streng monoton fällt.
$r_s = 0$ bedeutet, dass kein monotoner Zusammenhang besteht.

Die Rang-Korrelation r_s ist definiert als die Korrelation der Rangplätze der Messwerte, d.h.:

- Rang-Korrelationskoeffizient nach Spearman $r_s =$ Korrelationskoeffizient nach Bravais-Pearson r angewendet auf die Ränge:

$$r_s = \frac{n \cdot \sum_i rg(x_i) \cdot rg(y_i) - \left(\sum_i rg(x_i) \right) \cdot \left(\sum_i rg(y_i) \right)}{\sqrt{\left[n \cdot \sum_i rg(x_i)^2 - \left(\sum_i rg(x_i) \right)^2 \right] \cdot \left[n \cdot \sum_i rg(y_i)^2 - \left(\sum_i rg(y_i) \right)^2 \right]}}$$

Dabei ist $rg(x_i)$ der Rang von x_i und $rg(y_i)$ der Rang von y_i (die Ränge werden für die x-Werte und die y-Werte getrennt vergeben).

- Vereinfachte Berechnung (nur verwenden sofern keine Bindungen vorliegen):

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} (rg(x_i) - rg(y_i))^2}{n \cdot (n^2 - 1)}$$

Messwerte	x_i	33	15	17	11	40
Rangplätze	$rg(x_i)$	4	2	3	1	5
Messwerte	y_i	47	32	29	28	56
Rangplätze	$rg(y_i)$	4	3	2	1	5

Damit ergibt sich für die Rang-Korrelation:

$$r_s = 1 - \frac{6 \cdot [0^2 + (-1)^2 + 1^2 + 0^2 + 0^2]}{5 \cdot (25 - 1)} = 0.90$$

- Treten mehrere gleiche Werte auf, spricht man von *Bindungen* ("ties").
- Die Rangvergabe bei Bindungen wird mit Durchschnittsbildung über die in Frage kommenden Ränge gelöst, d. h. es wird jedem der Werte der durchschnittliche Rang zugeordnet.
- Beispiel:

Messwerte	x_i	1.09	2.17	2.68	2.17	4.5	3.02
Hilfszählung		1	$\underline{2}$	4	$\underline{3}$	6	5
Rangplätze	$rg(x_i)$	1	2,5	4	2,5	6	5

Beim Vorliegen von Bindungen stimmt die vereinfachte Berechnung nur näherungsweise, deshalb sollte die ausführliche Formel verwendet werden.