NON-PARAMETRIC STATISTICS
SPEARMAN'S RANK CORRELATION

A measure of association for two samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ is the **Pearson Product Moment Correlation Coefficient**, r, where

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}\, SS_{yy}}}$$

where

$$SS_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2 \qquad SS_{yy} = \sum_{i=1}^{n} (y_i - \overline{y})^2 \qquad SS_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

This quantity measures the **linear** association between the X and Y variables.

A measure of the potentially non-linear association between the samples $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ is the **Spearman Rank Correlation Coefficient**, r_S, which computes the correlation between the **ranks** of the data.

The Spearman Rank Correlation Coefficient is computed as follows:

1. Assign ranks $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ to the data $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ separately by sorting each sample into ascending order and assigning the ranks in order.
2. Compute r_S as

$$r_S = \frac{SS_{uv}}{\sqrt{SS_{uu}\, SS_{vv}}}$$

where

$$SS_{uu} = \sum_{i=1}^{n} (u_i - \overline{u})^2 \qquad SS_{vv} = \sum_{i=1}^{n} (v_i - \overline{v})^2 \qquad SS_{uv} = \sum_{i=1}^{n} (u_i - \overline{u})(v_i - \overline{v})$$

If there are no ties in the data, then

$$r_S = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

where

$$d_i = u_i - v_i \qquad i = 1, \ldots, n$$

Tests for r_S : If the population correlation is ρ, then we may test the hypothesis

$$H_0 \; : \; \rho = 0$$

against the hypotheses

$$\begin{aligned} (1)\; H_a &: \quad \rho > 0 \\ (2)\; H_a &: \quad \rho < 0 \\ (3)\; H_a &: \quad \rho \neq 0 \end{aligned}$$

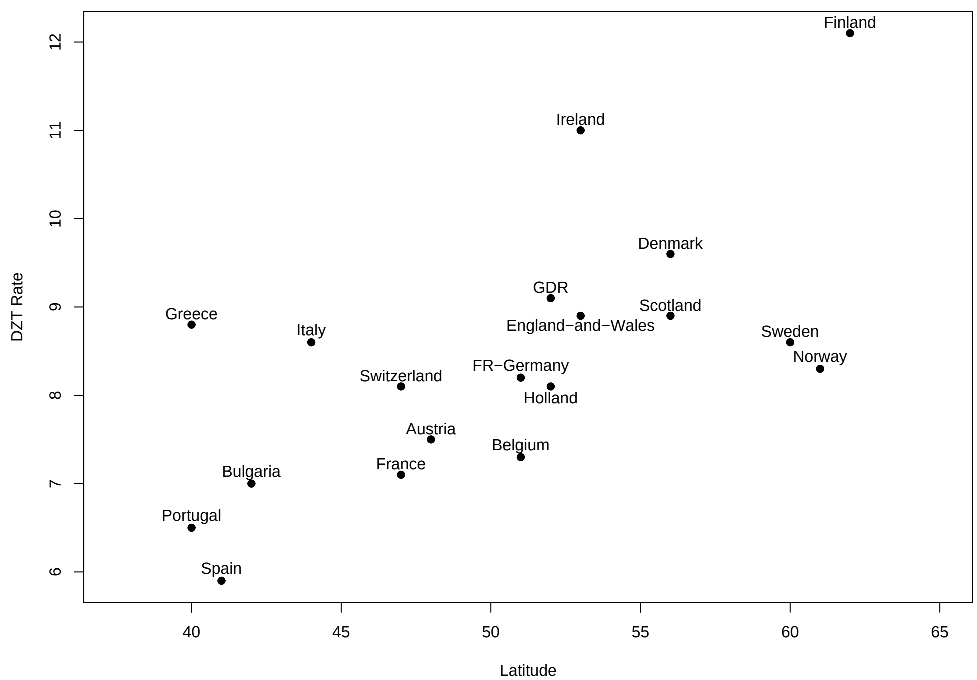
using the table of the null distribution on p 894 of McClave and Sincich. If Spearman_α is the α tail quantile of the null distribution, we have the following rejection regions:

$$\begin{aligned} (1) &: \quad \text{Reject } H_0 \text{ if } r_S > \text{Spearman}_\alpha \\ (2) &: \quad \text{Reject } H_0 \text{ if } r_S < -\text{Spearman}_\alpha \\ (3) &: \quad \text{Reject } H_0 \text{ if } |r_S| > \text{Spearman}_{\alpha/2} \end{aligned}$$

EXAMPLE : Latitude and dizygotic twinning rates
The relationship between the geographical latitude of a country and its dizygotic twinning (DZT) rate is to be investigated. The data are presented and plotted below.

Reference: James, W.H. (1985) Dizygotic twinning, birth weight and latitude, *Annals of Human Biology*, **12**, 5, pp. 441-447.

Country	Latitude x	Rank u	DZT Rate y	Rank v
Portugal	40	1.5	6.5	2.0
Greece	40	1.5	8.8	13.0
Spain	41	3.0	5.9	1.0
Bulgaria	42	4.0	7.0	3.0
Italy	44	5.0	8.6	11.5
France	47	6.5	7.1	4.0
Switzerland	47	6.5	8.1	7.5
Austria	48	8.0	7.5	6.0
Belgium	51	9.5	7.3	5.0
FR Germany	51	9.5	8.2	9.0
Holland	52	11.5	8.1	7.5
GDR	52	11.5	9.1	16.0
England & Wales	53	13.5	8.9	14.5
Ireland	53	13.5	11.0	18.0
Scotland	56	15.5	8.9	14.5
Denmark	56	15.5	9.6	17.0
Sweden	60	17.0	8.6	11.5
Norway	61	18.0	8.3	10.0
Finland	62	19.0	12.1	19.0



For these data

$$r_S = \frac{SS_{uv}}{\sqrt{SS_{uu}\,SS_{vv}}} = \frac{384.5}{\sqrt{567 \times 568.5}} = 0.677 \qquad r = \frac{SS_{xy}}{\sqrt{SS_{xx}\,SS_{yy}}} = \frac{118.4}{\sqrt{866.105 \times 38.88}} = 0.645$$

indicating a strong positive association.