Vorlesung 7b

Korrelationskoeffizient und Regressionsgerade

Teil 2

Der Korrelationskoeffizient

(Buch S. 62)

Definition.

Für zwei Zufallsvariable X, Y

mit positiven, endlichen Varianzen ist

$$\kappa_{XY} := \frac{\mathbf{Cov}[X,Y]}{\sqrt{\mathbf{Var}X}\sqrt{\mathbf{Var}Y}}$$

der **Korrelationskoeffizient** von X und Y.

(kurz auch: *die Korrelation* von X und Y). Aus der

Kovarianz-Varianz-Ungleichung folgt sofort

$$-1 \leq \kappa_{XY} \leq 1.$$

Fünf prominente Zahlen

zur (teilweisen) Beschreibung der Verteilung

eines zufälligen Paares (X, Y) in $\mathbb{R} \times \mathbb{R}$:

μ_X und μ_Y: die Erwartungswerte von X und Y

σ_X und σ_Y: die Standardabweichungen von X und Y

κ_{XY}: der Korrelationskoeffizient von X und Y

Die folgenden 11 Bilder zeigen jeweils die

Realisierungen von 1000 unabhängige Kopien (X_i, Y_i)

eines zufälligen Paares (X, Y),

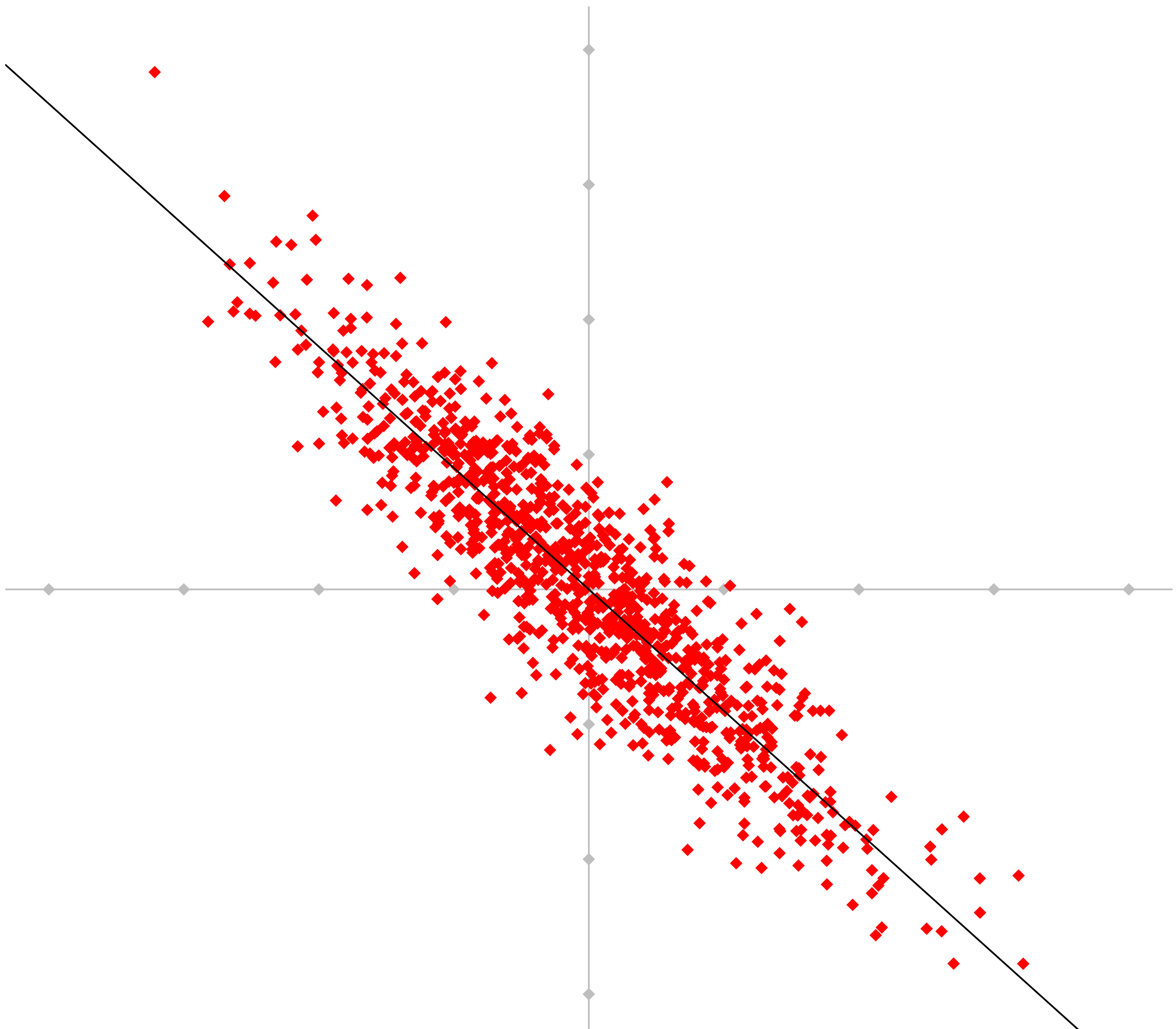
mit X N$(0, 1)$-verteilt , Y N$(0, 1)$-verteilt, und

$\kappa_{XY} = -0.9, -0.7, \ldots, -0.1, 0, 0.1, \ldots, 0.7, 0.9,$

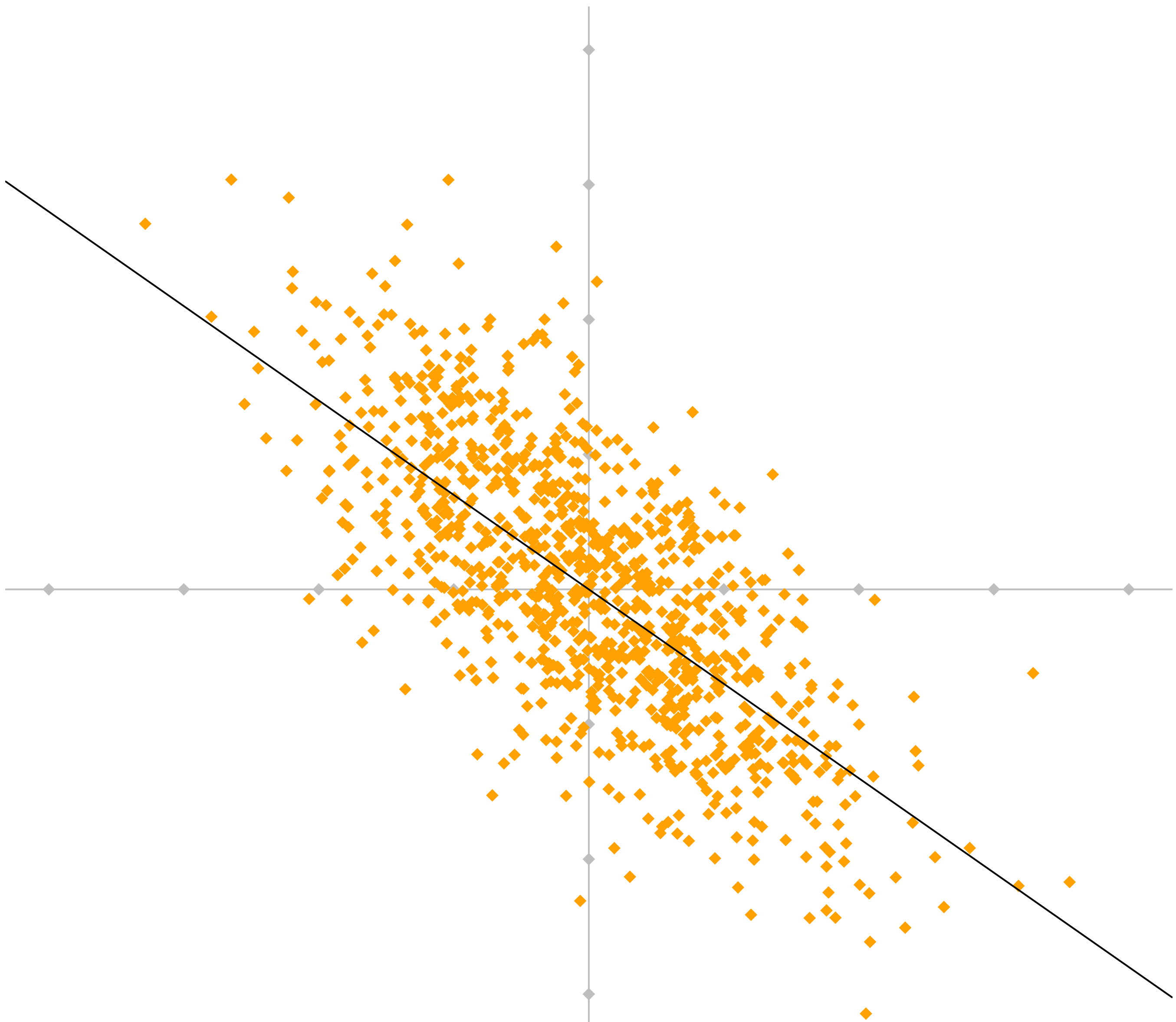
zusammen mit der

Geraden durch den Ursprung mit Anstieg κ_{XY}.

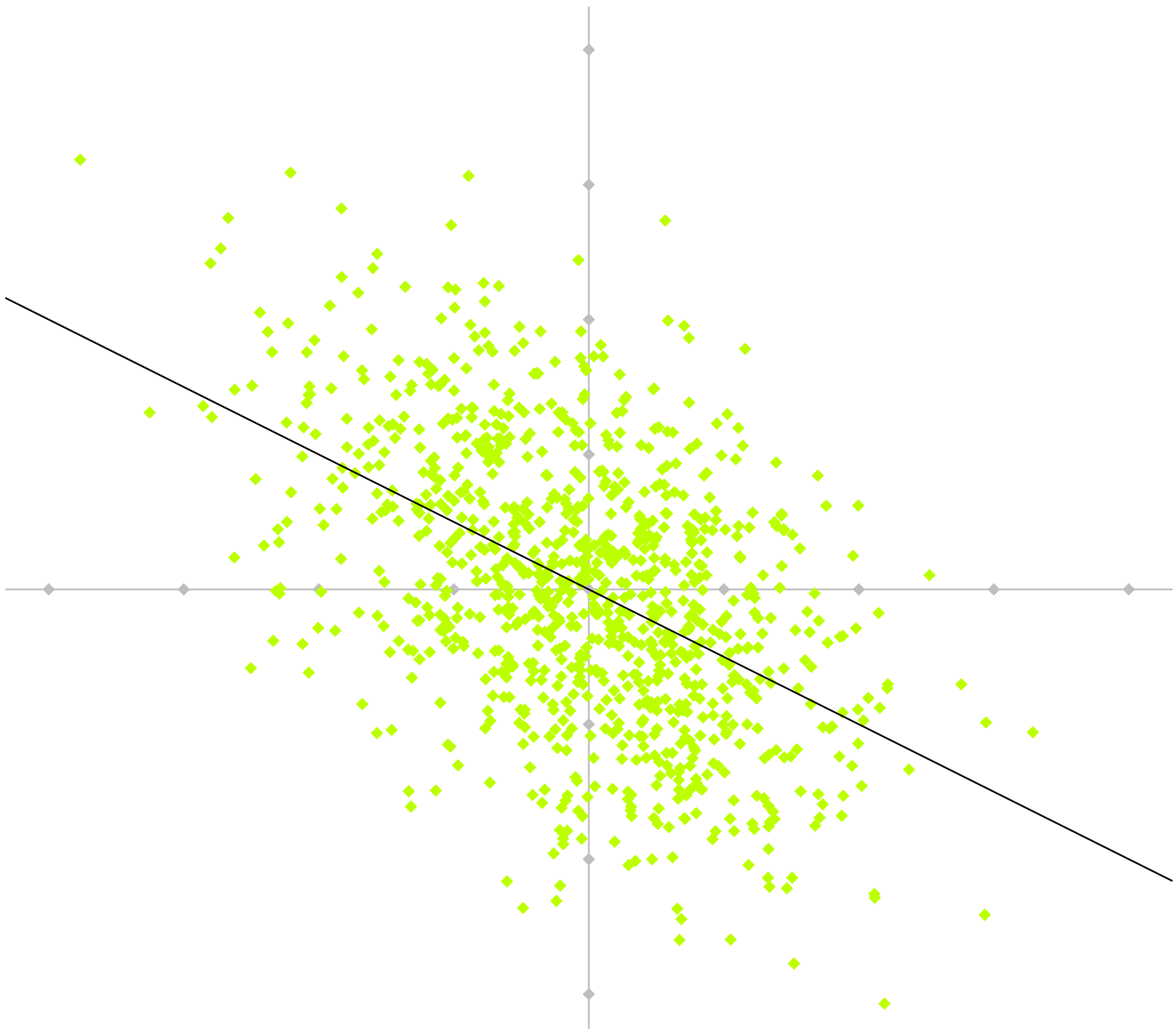
Korrelation =   - 0.9



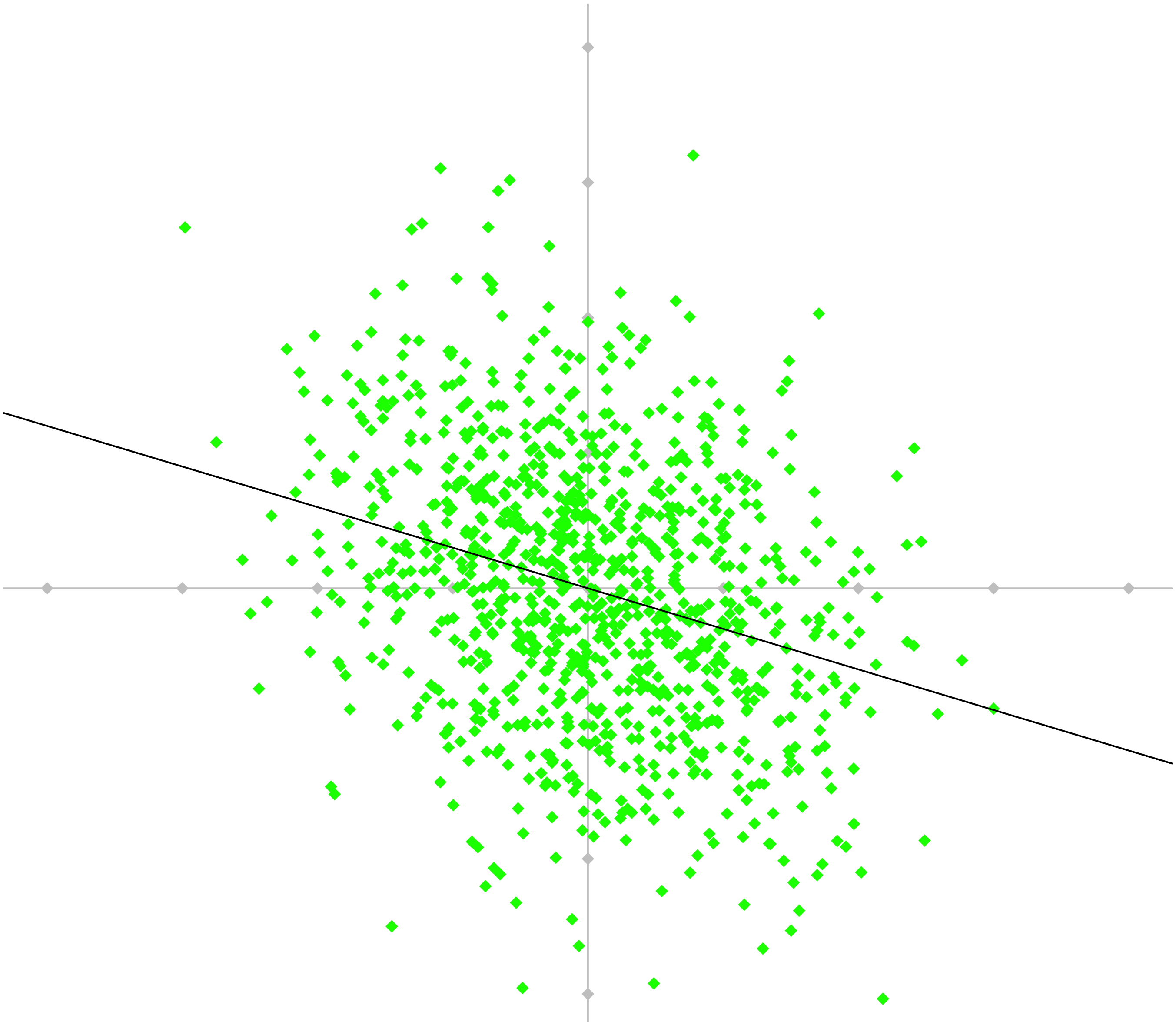
Korrelation =   - 0.7



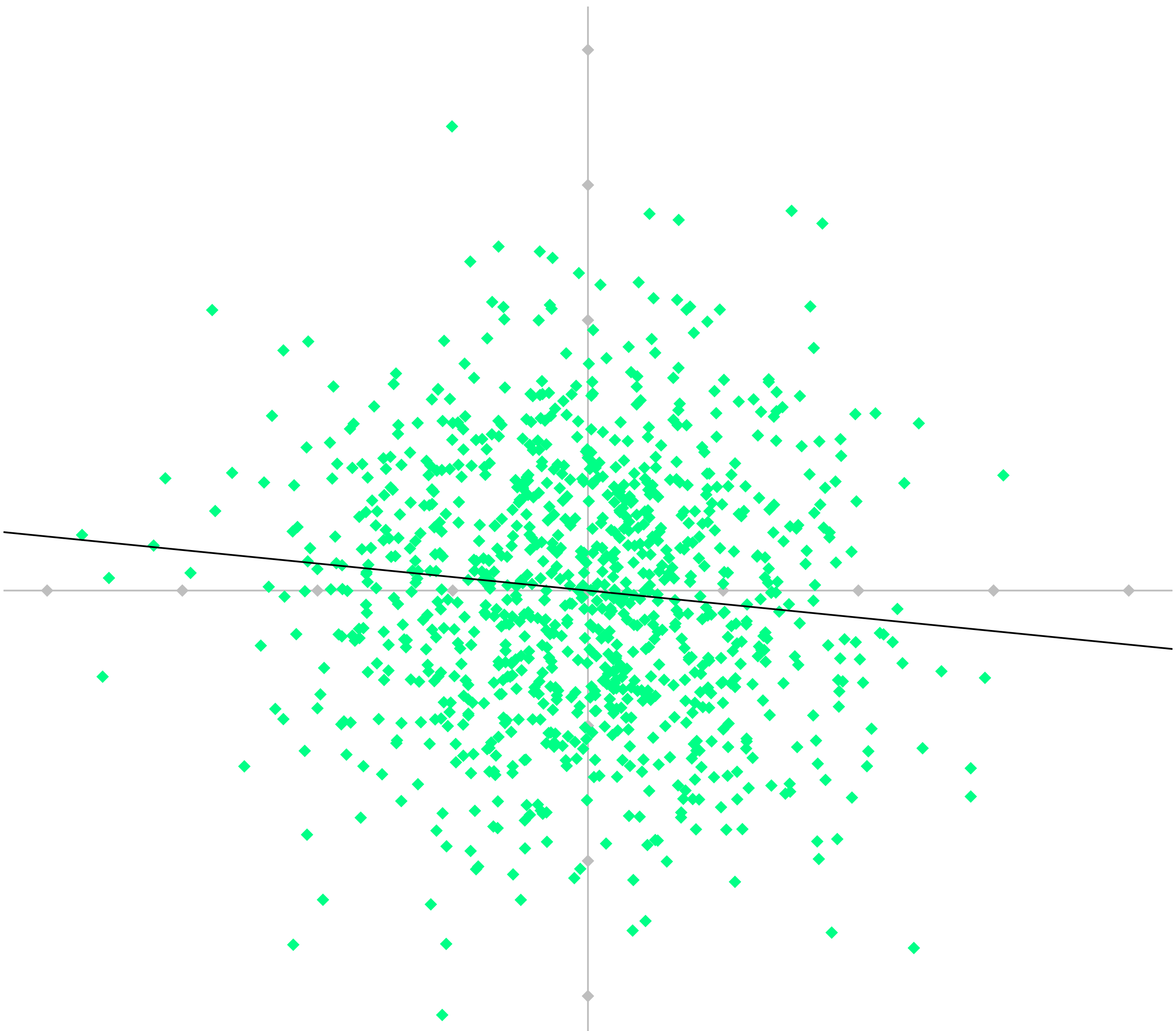
Korrelation =   - 0.5



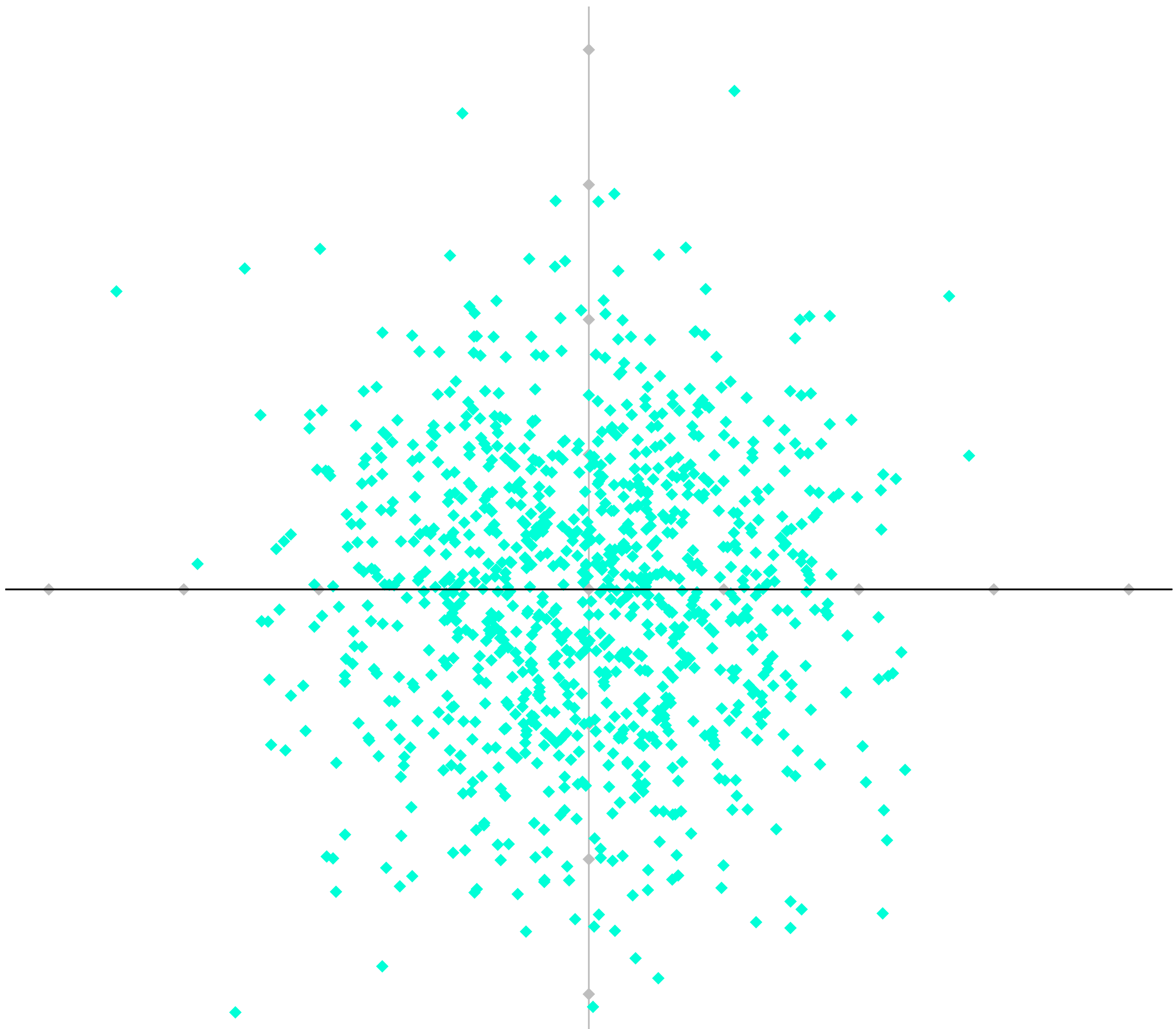
Korrelation = - 0.3



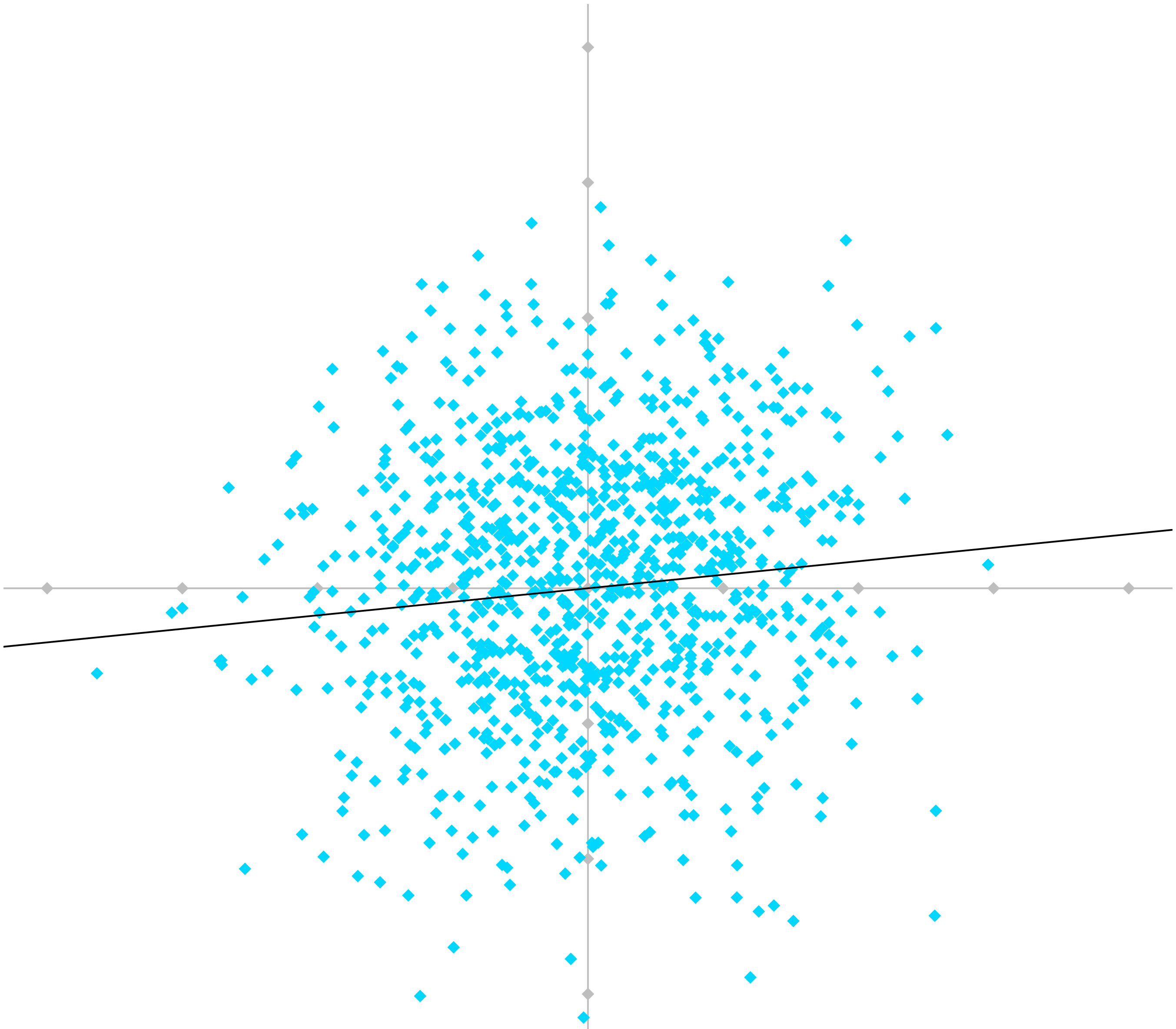
Korrelation = - 0.1



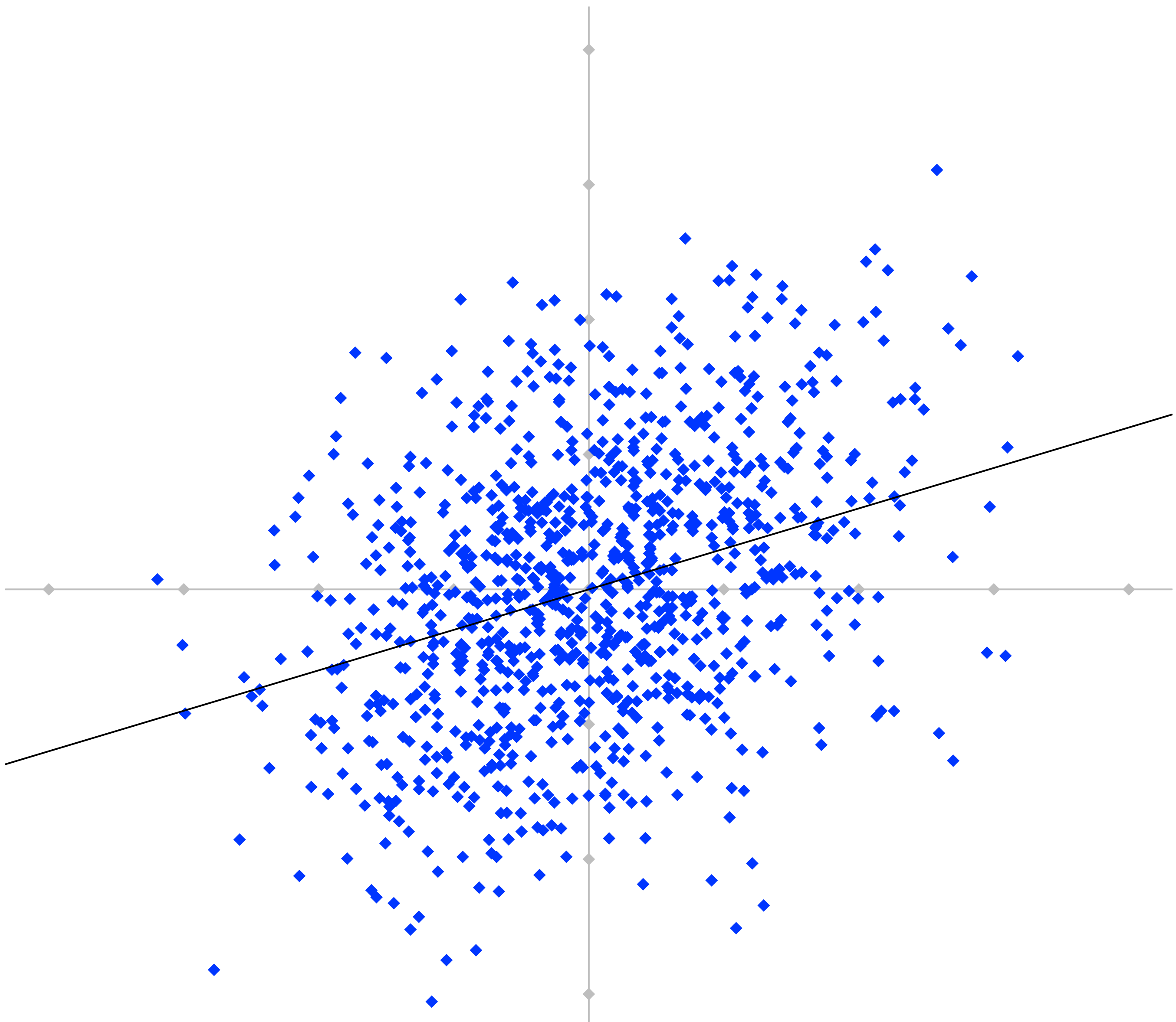
Korrelation = 0



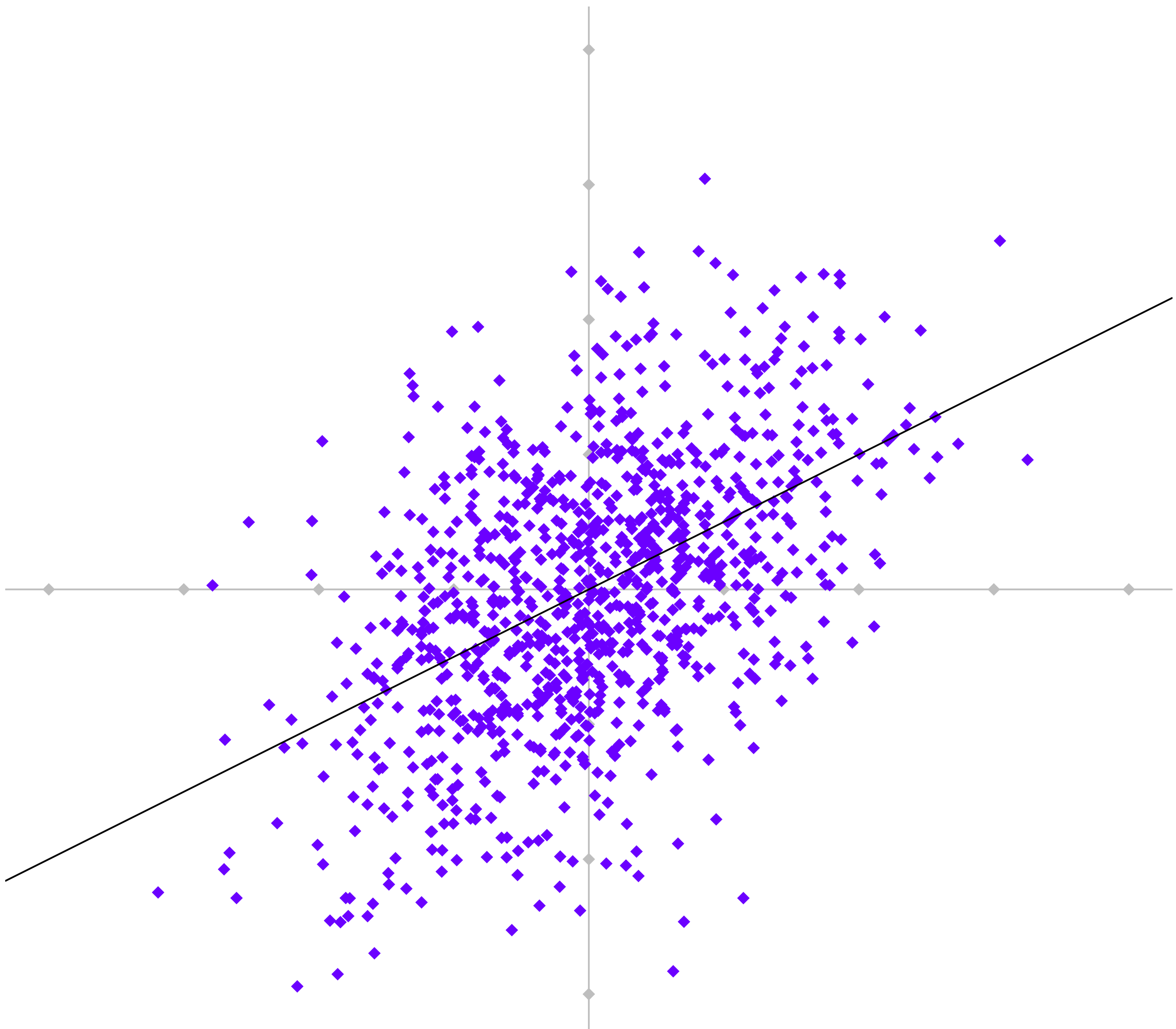
Korrelation = 0.1



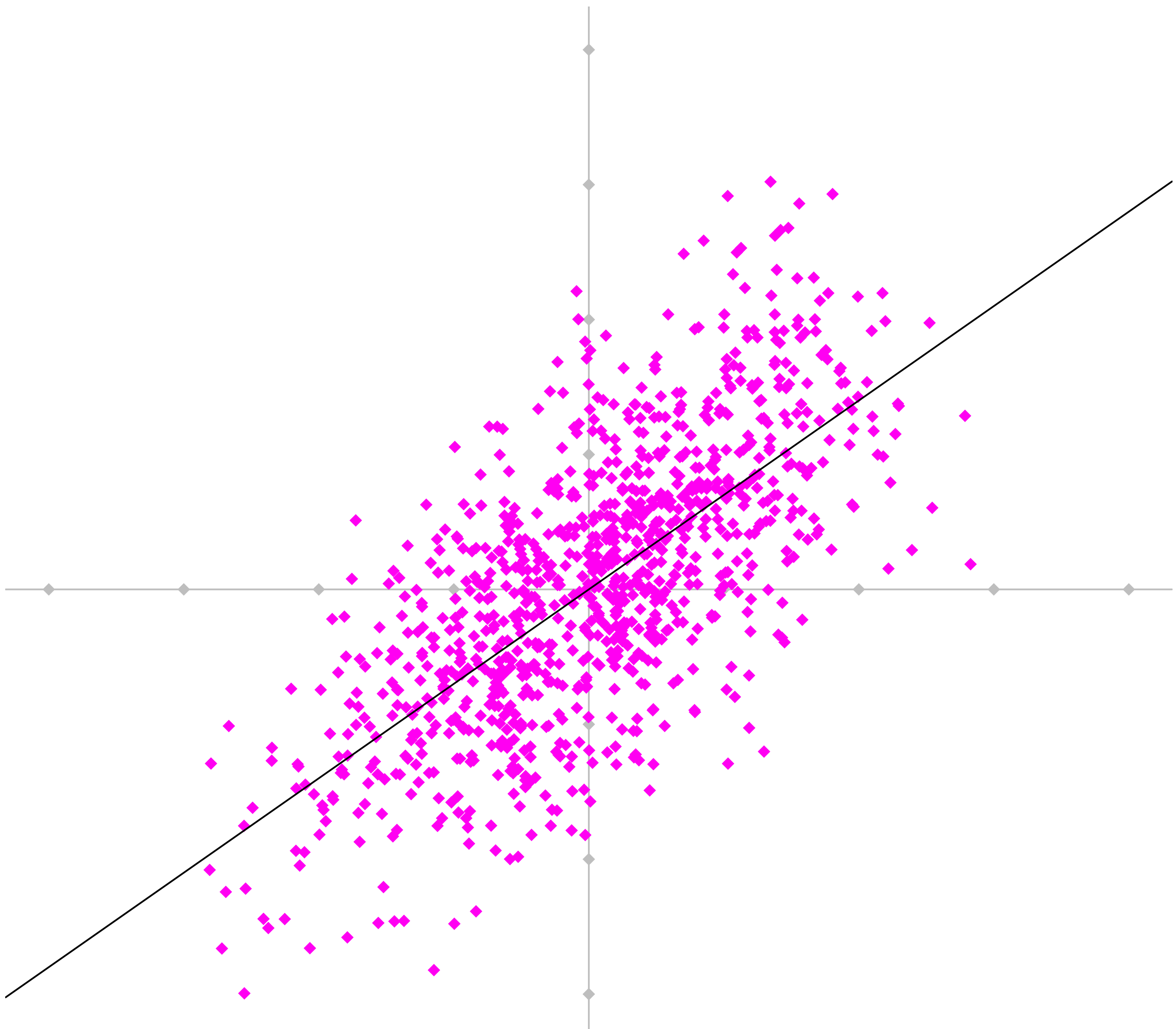
Korrelation = 0.3



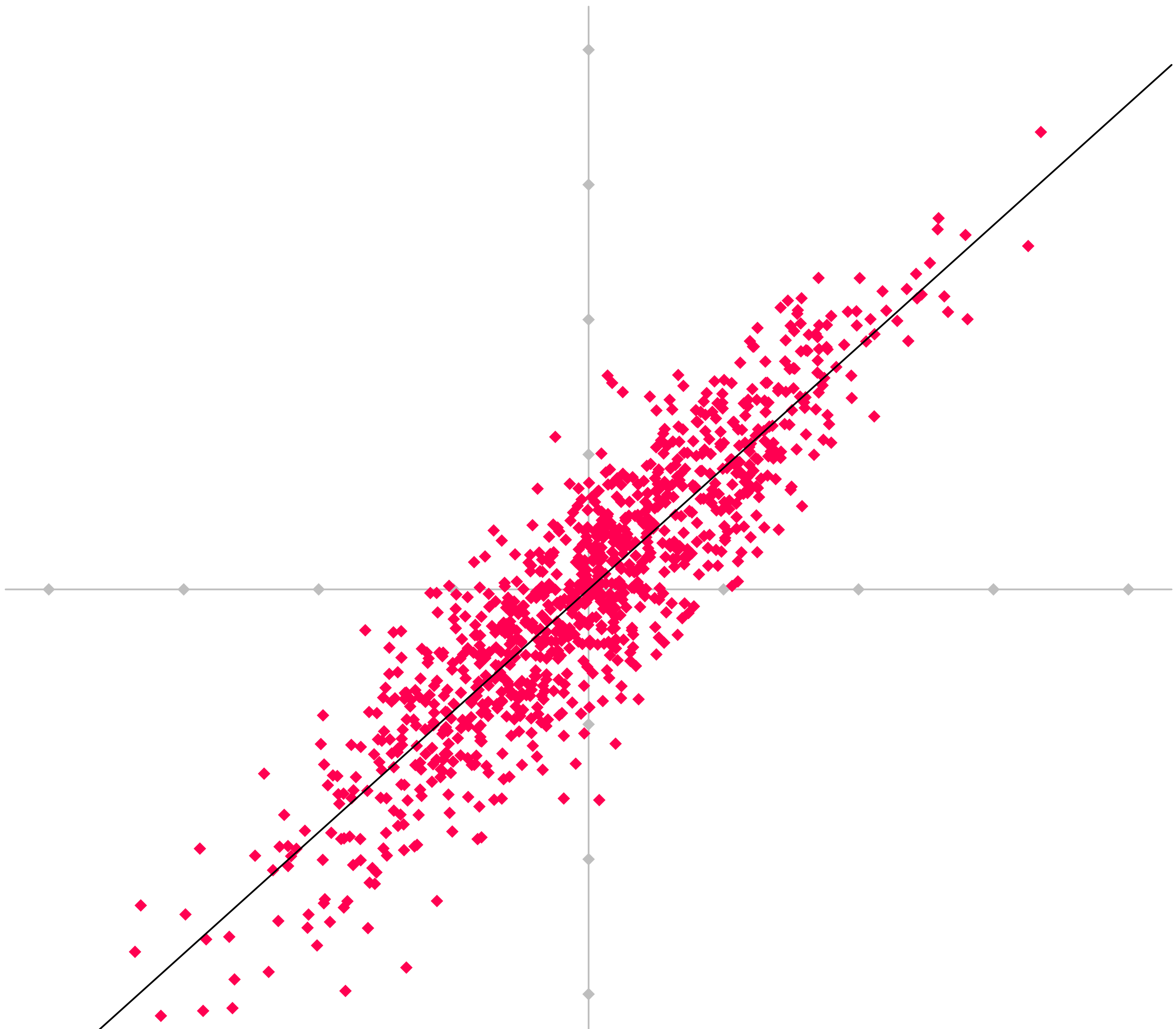
Korrelation = 0.5



Korrelation = 0.7



Korrelation = 0.9



Beispiel:

Gemeinsam normalverteilte Zufallsvariable

Z_1, Z_2 seien unabhängig und standard-normalverteilt,

Wir wählen eine Konstante $\kappa \in [-1, 1]$ und setzen

$$X := Z_1, \quad Y := \kappa Z_1 + \sqrt{1-\kappa^2} Z_2.$$

Damit ergibt sich $\mathrm{Cov}[X, Y] = \mathrm{Cov}[Z_1, \rho Z_1] = \kappa$,

$$\sigma_X^2 = \sigma_Y^2 = 1.$$

Also: $\kappa_{XY} = \kappa$.

Auch Y ist übrigens standard-normalverteilt (siehe V6b4 F9).

Wir werden sehen:

κ^2 ist ein Maß dafür, um wieviel besser man Y

durch eine affin lineare Funktion von X vorhersagen kann:

$$Y = \beta_1 X + \beta_0 + \text{“Fehler”},$$

als durch eine Konstante:

$$Y = c + \text{“Fehler”}.$$

(Die “Güte der Vorhersage” bezieht sich auf die Kleinheit des **erwarteten quadratischen Fehler (mean sqare error)**.)