Vorlesung 8a

Kovarianz und Korrelation

Wir erinnern an die Definition der

Kovarianz

Für reellwertige Zufallsvariable X, Y

mit $\mathbf{E}[X^2] < \infty$ und $\mathbf{E}[Y^2] < \infty$ ist

$$\mathbf{Cov}[X, Y] := \mathbf{E}\Big[(X - \mathbf{E}X)(Y - \mathbf{E}Y)\Big]$$

Insbesondere ist also

$$\mathbf{Cov}[X, X] = \mathbf{Var}[X]$$

Die Kovarianz ist

- positiv semidefinit:

$$\mathbf{Cov}[X, X] \geq 0, \quad \mathbf{Cov}[0, 0] = 0$$

- symmetrisch:

$$\mathbf{Cov}[X, Y] = \mathbf{Cov}[Y, X]$$

- bilinear:

$$\mathbf{Cov}[c_1 X_1 + c_2 X_2, Y] = c_1 \mathbf{Cov}[X_1, Y] + c_2 \mathbf{Cov}[X_2, Y]$$

Die “Kovarianz-Varianz-Ungleichung”

$$|\mathbf{Cov}[X,Y]| \le \sqrt{\mathbf{Var} X}\,\sqrt{\mathbf{Var} Y}$$

folgt sofort aus der

Cauchy-Schwarz Ungleichung: Für reellwertige Zufallsvariable G, H mit $\mathbf{E}[G^2], \mathbf{E}[H^2] < \infty$ ist

$$(\mathbf{E}[GH])^2 \le \mathbf{E}[G^2]\,\mathbf{E}[H^2]\ .$$

Beweis:

Fall 1: $\mathbf{E}[G^2], \mathbf{E}[H^2] > 0$.

$U := G/\sqrt{\mathbf{E}[G^2]}$, $V := H/\sqrt{\mathbf{E}[H^2]}$ erfüllen

$$\mathbf{E}[U^2] = \mathbf{E}[V^2] = 1.$$

Aus $\quad \pm 2UV \le U^2 + V^2 \quad$ folgt

$$\pm\mathbf{E}[UV] \le 1\ .$$

Multiplikation mit $\sqrt{\mathbf{E}[G^2]}\sqrt{\mathbf{E}[H^2]}$ ergibt die Behauptung.

Fall 2: $\mathbf{E}[G^2] = 0$. Dann folgt aus der Positivität des E'wertes

$\mathbf{P}(G^2 = 0) = 1$, also $\mathbf{P}(GH = 0) = 1$ und $\mathbf{E}[GH] = 0$.$\Box$

Hier sind 5 Zahlen

zur (teilweisen) Beschreibung der Verteilung

einer $\mathbb{R} \times \mathbb{R}$-wertigen Zufallsvariablen

oder anders gesagt: eines zufälligen Paares (X, Y):

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.

Definition.

Für zwei Zufallsvariable X, Y

mit positiven, endlichen Varianzen sei

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

der **Korrelationskoeffizient** von X und Y.

Aus der Kovarianz-Varianz-Ungleichung folgt sofort

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

Wir werden sehen:

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

durch eine affin lineare Funktion von X vorhersagen kann:

$Y = \beta_1 X + \beta_0$ + “Fehler”,

als durch eine Konstante:

$Y = c$ + “Fehler”.

Die “Güte der Vorhersage” bezieht sich auf einen kleinen **erwarteten quadratischen Fehler (mean sqare error)**.

Um dies einzusehen, fragen wir erst einmal:

Durch welche Konstante wird die Zufallsvariable Y

(im Sinn des erwarteten quadratischen Fehlers)

am besten vorhergesagt?

Durch ihren Erwartungswert $\mathbf{E}[Y]$!

Denn:

$$\mathbf{E}(Y-c)^2 = \mathbf{E}[(Y - \mathbf{E}Y + \mathbf{E}Y - c)^2]$$

$$= \mathbf{E}[(Y-\mathbf{E}Y)^2] + 2\mathbf{E}[(Y-\mathbf{E}[Y])(\mathbf{E}Y - c)] + (\mathbf{E}[Y]-c)^2$$

$$= \mathbf{Var}\,Y + 0 + (\mathbf{E}[Y]-c)^2.$$

Das wird minimiert von

$$c = \mathbf{E}Y$$

und hat den Minimalwert

$$\mathbf{Var}\,Y.$$

Durch welche affin lineare Funktion von X,

$$\beta_1 X + \beta_0,$$

wird die Zufallsvariable Y (wieder im Sinn des erwarteten quadratischen Fehlers) am besten vorhergesagt?

Genauer:

Für welche Zahlen β_1, β_0 wird

$\mathbf{E}[(Y - \beta_1 X - \beta_0)^2]$ minimal?

Wie wir gleich sehen werden, ist die Lösung:

$$\beta_1 := \frac{\sigma_Y}{\sigma_X} \kappa_{XY}$$

und β_0 so, dass $\mu_Y = \beta_1 \mu_X + \beta_0$.

M. a. W.: β_0 so, dass der Punkt (μ_X, μ_Y) auf der Geraden $y = \beta_1 x + \beta_0$ liegt.

Wir nennen diese Gerade die **Regressionsgerade** für Y auf der Basis von X.

Wir begründen jetzt die Behauptung über β_0 und β_1:

$$
\begin{aligned}
&\mathbf{E}[(Y-\beta_1 X-\beta_0)^2] \\
&= \mathbf{Var}[Y-\beta_1 X-\beta_0] + (\mathbf{E}[Y]-\beta_1\mathbf{E}[X]-\beta_0)^2 \\
&= \mathbf{Var}[Y-\beta_1 X] + (\mathbf{E}[Y]-\beta_1\mathbf{E}[X]-\beta_0)^2
\end{aligned}
$$

Der zweite Summand ist Null für $\beta_0 = \mathbf{E}Y - \beta_1 \mathbf{E}X$.

Damit haben wir schon mal die eine Bedingung gefunden.

Für welches β_1 wird der erste Summand minimal?

$$\mathbf{Var}[Y - \beta_1 X] = \mathbf{Var}Y - 2\beta_1 \mathbf{Cov}[X, Y] + \beta_1^2 \mathbf{Var}X$$

$$= \sigma_Y^2 - 2\beta_1 \kappa\, \sigma_X \sigma_Y + \beta_1^2 \sigma_X^2$$

$$= \sigma_Y^2 - \sigma_Y^2 \kappa^2 + (\sigma_Y \kappa - \beta_1 \sigma_X)^2$$

Der rechte Summand wird Null für

$$\beta_1 = \frac{\sigma_Y}{\sigma_X} \kappa.$$

Und der Minimalwert von $\mathbf{Var}[Y - \beta_1 X]$ ist $\sigma_Y^2 (1 - \kappa^2)$.

Damit ist auch der Minimalwert von $\mathbf{Var}[Y - \beta_1 X - \beta_0]$ gleich $\sigma_Y^2(1 - \kappa^2)$.

Der Minimalwert von $\mathbf{Var}[Y - c]$ war σ_Y^2.

Also ist der Anteil von $\mathbf{Var}\, Y$,

der von den Vielfachen von X

zusätzlich zu den Vielfachen von 1 “erklärt” wird, gleich

$$\kappa^2 \sigma_Y^2.$$

Wir halten fest: Die Minimierungsaufgabe

$$\mathbf{E}[(Y - \beta_1 X - \beta_0)^2] \stackrel{!}{=} \min$$

für die *beste affin lineare Vorhersage von* Y

auf der Basis von X

(im Sinn des quadratischen Mittels)

hat die Lösung

$$\beta_1 = \frac{\sigma_Y}{\sigma_X}\kappa, \quad \mu_Y = \beta_1 \mu_X + \beta_0$$

und den Minimalwert $\quad (1 - \kappa_{XY}{}^2)\,\mathbf{Var} Y$

Beispiel 1:

Z_1, Z_2 seien unabhängig und standard-normalverteilt,

$\rho \in [-1, 1]$.

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

Dann gilt: $\sigma_X^2 = \sigma_Y^2 = 1$,

$$\kappa_{XY} = \rho.$$

Die folgenden Bilder ($\rho = -0.9, -0.8, \ldots, 0.8, 0.9$)

zeigen jeweils die Realisierungen von

1000 unabhängige Kopien (X_i, Y_i) von (X, Y),

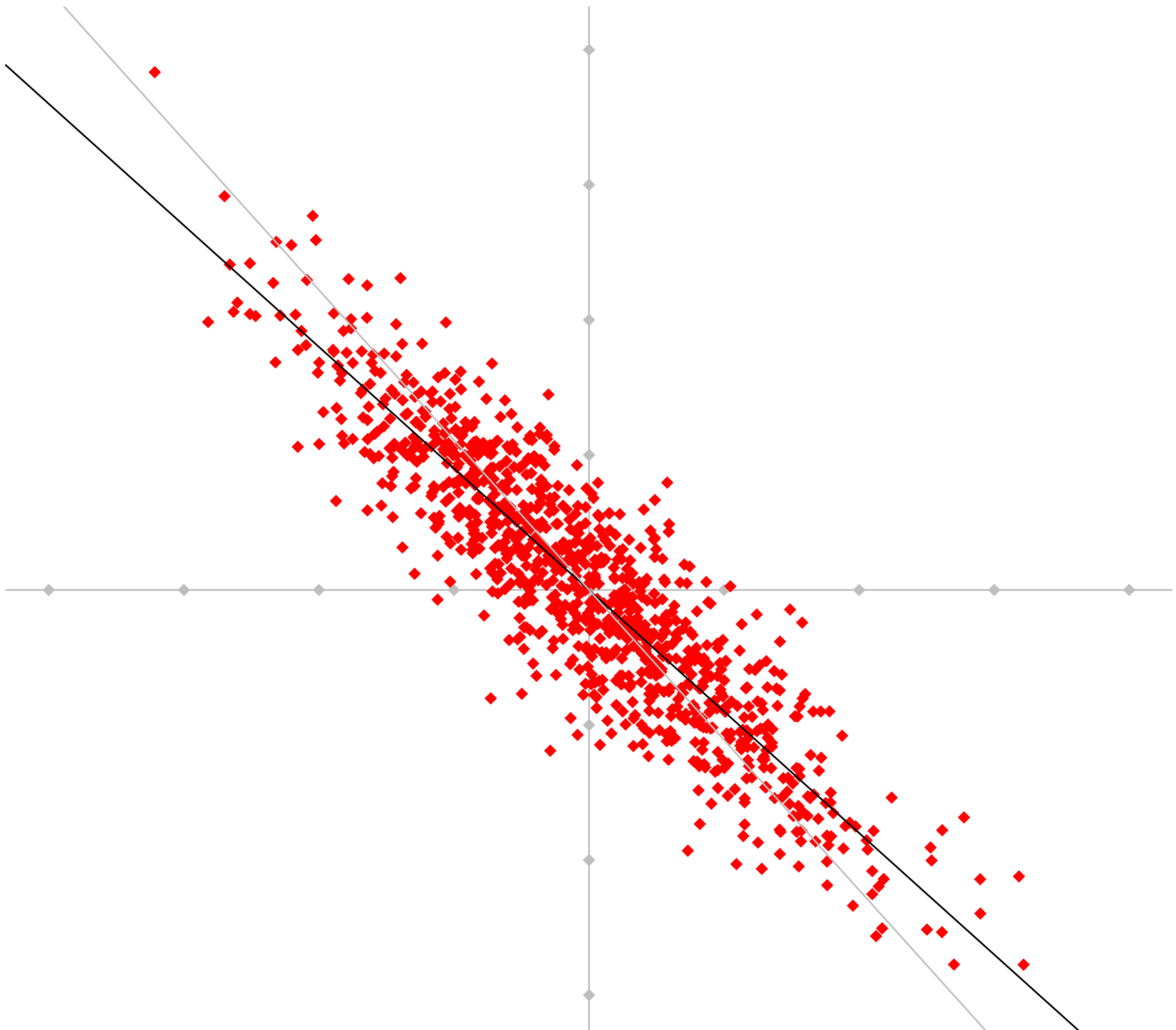
zusammen mit der

Regressionsgerade für Y auf der Basis von X (in schwarz)

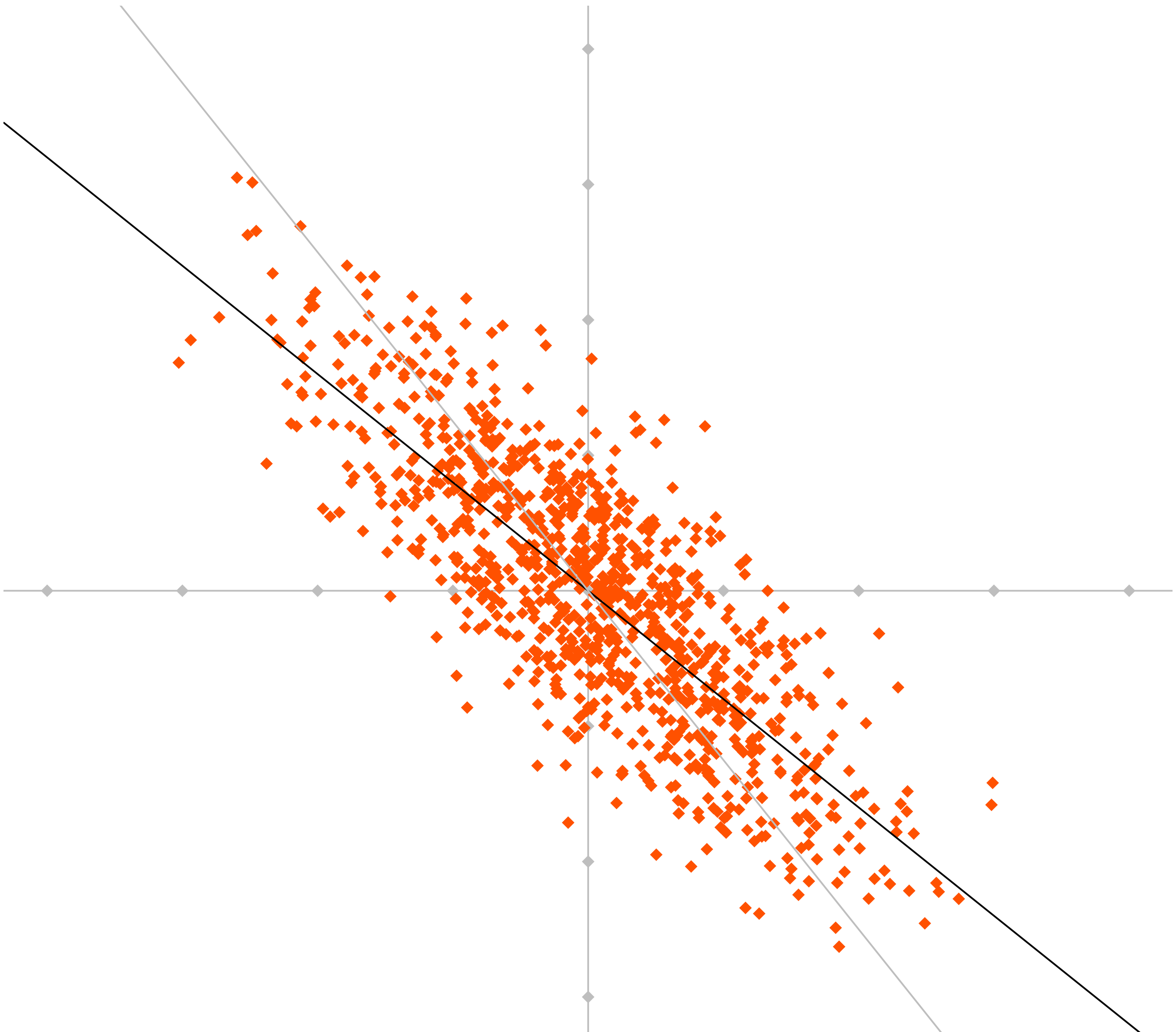
und der

Regressionsgerade für X auf der Basis von Y (in grau).

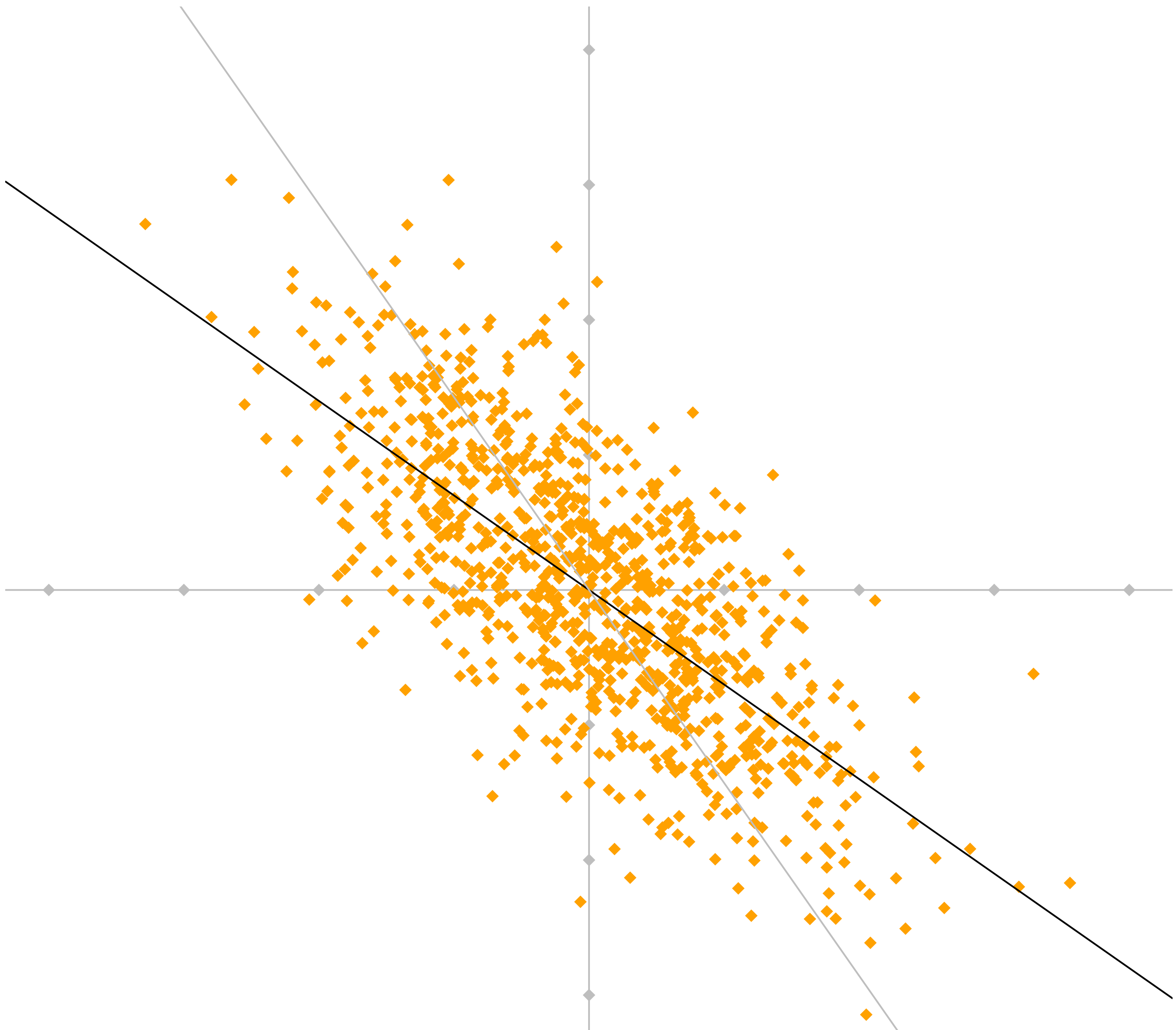
Korrelation = −0.9



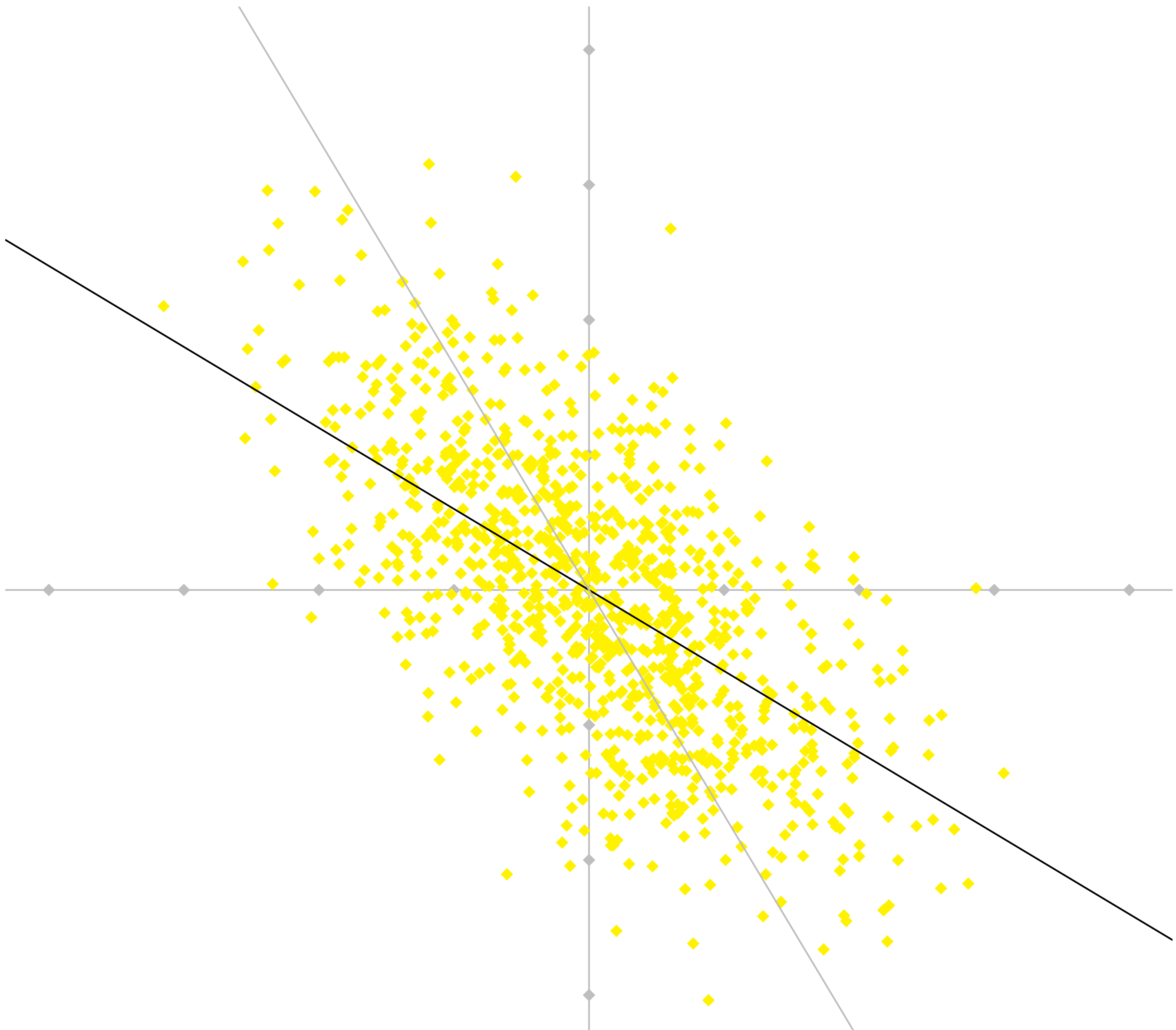
Korrelation = −0.8



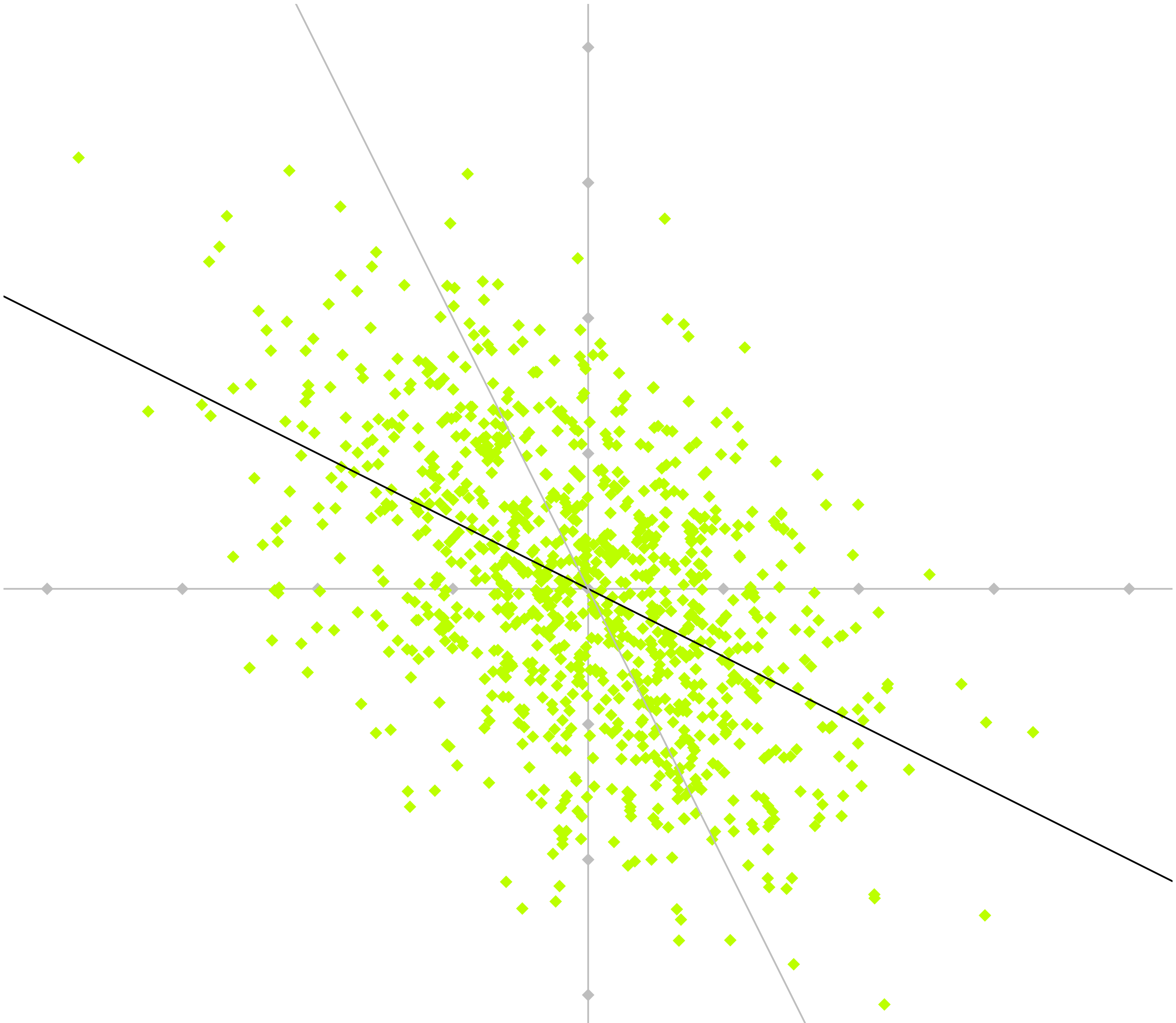
Korrelation = −0.7



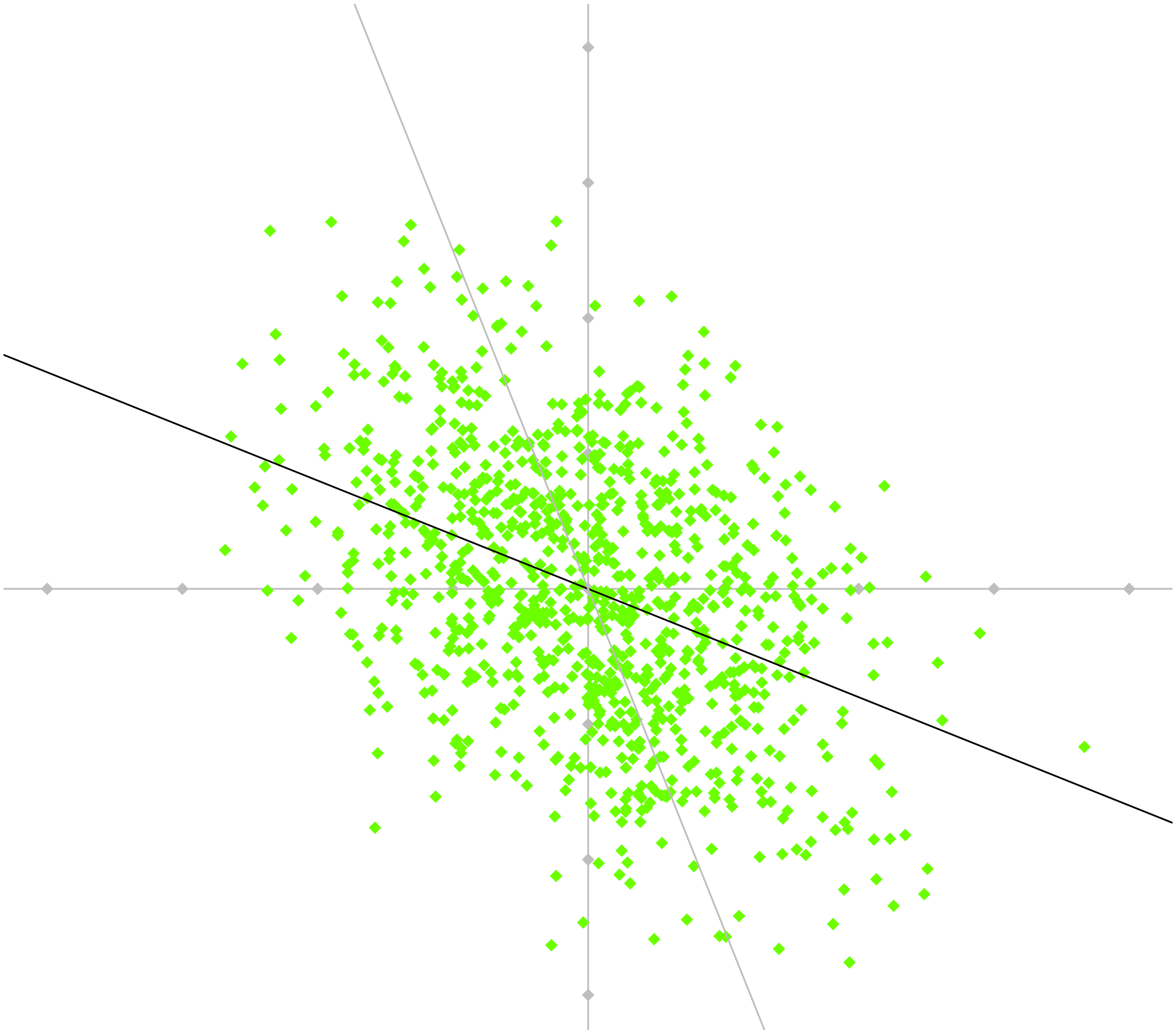
Korrelation = −0.6



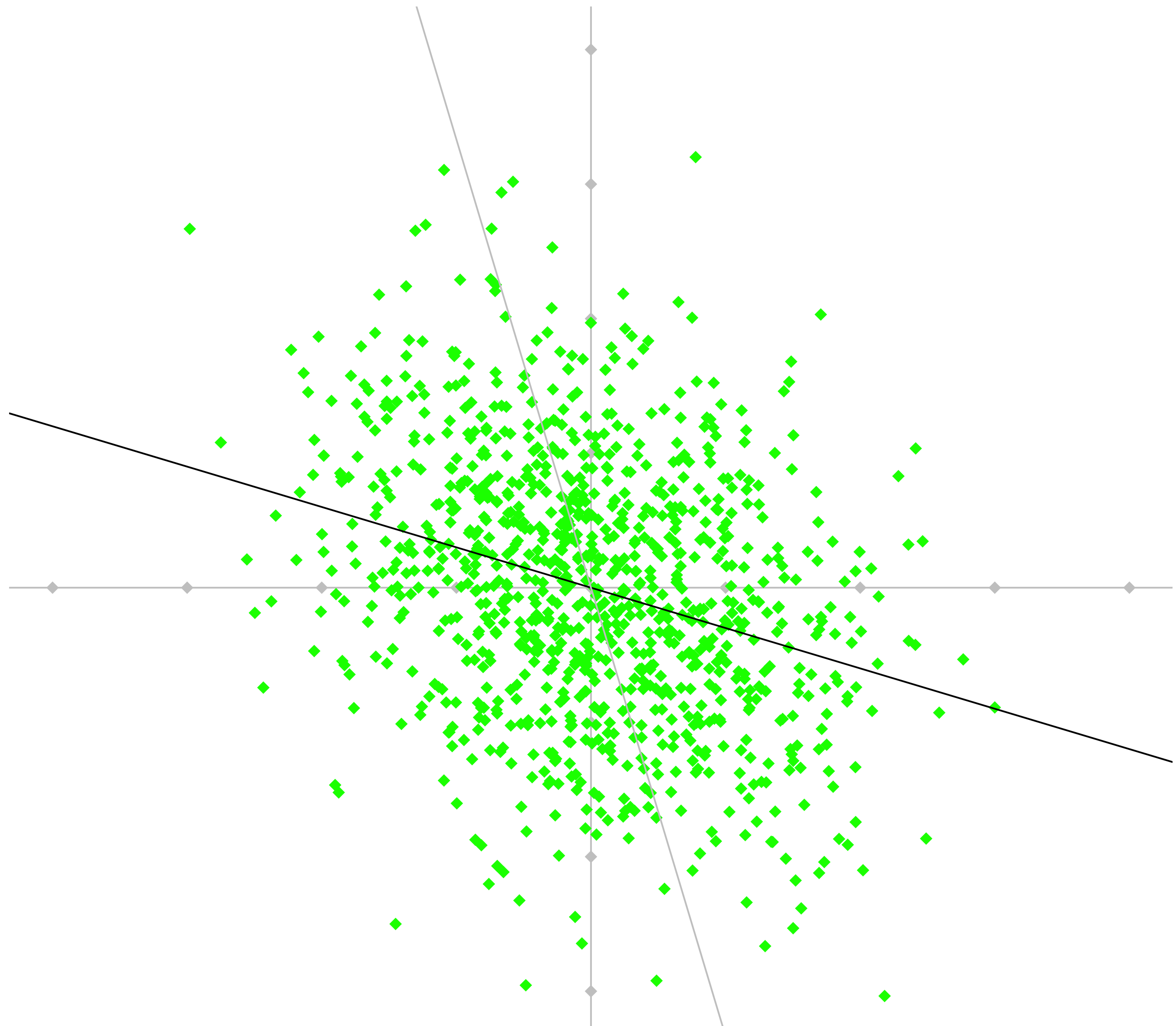
Korrelation = −0.5



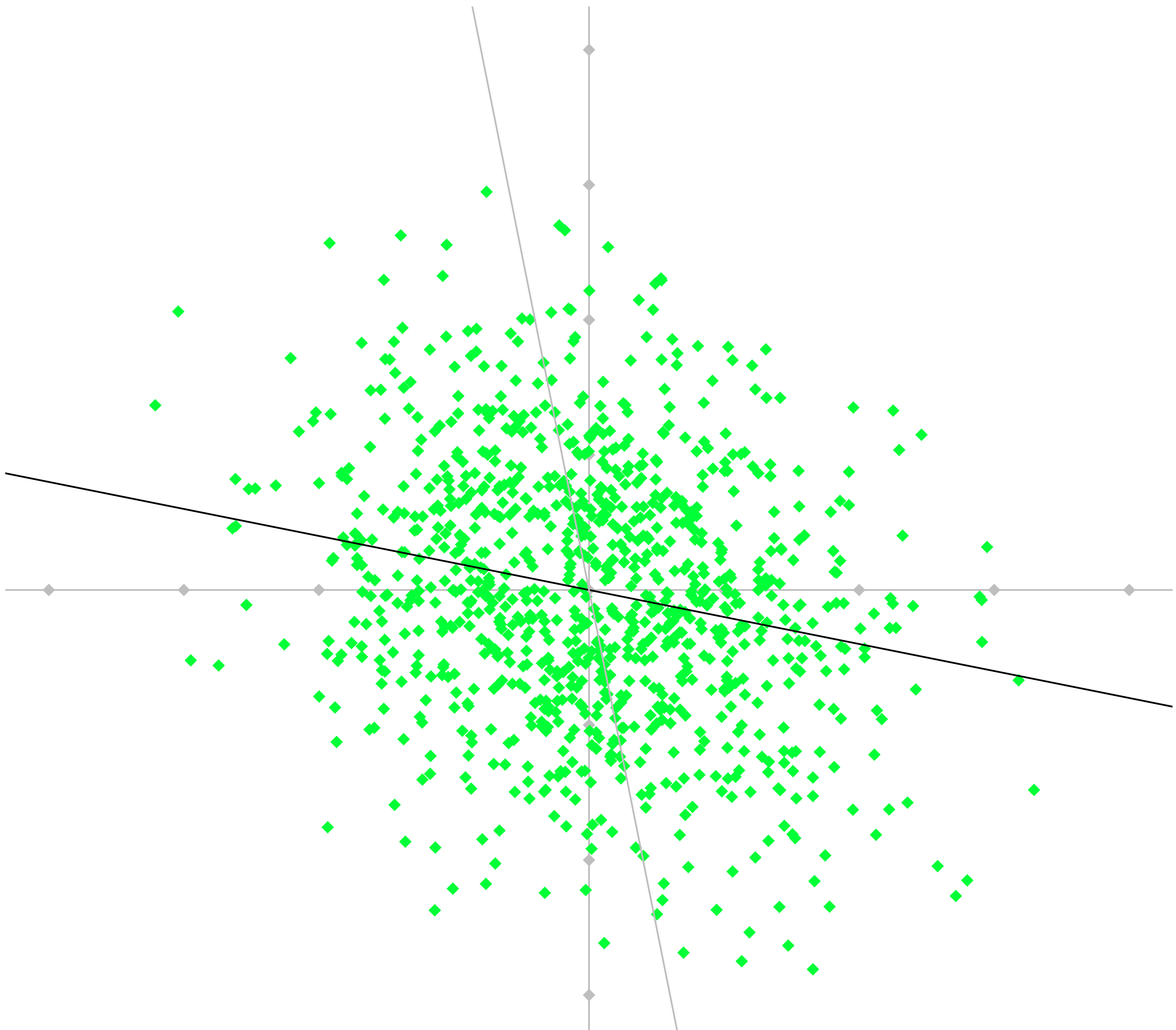
Korrelation = −0.4



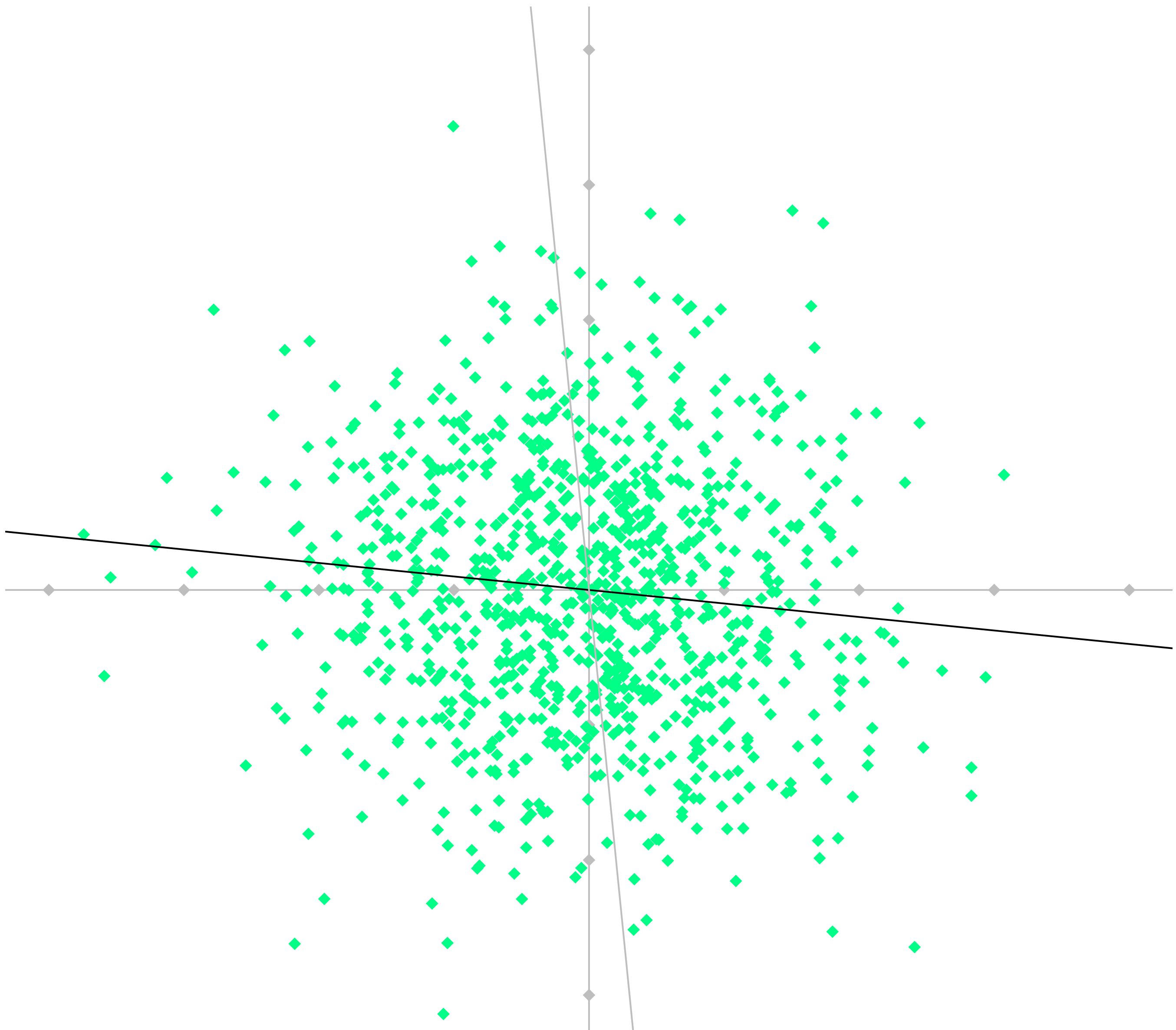
Korrelation = −0.3



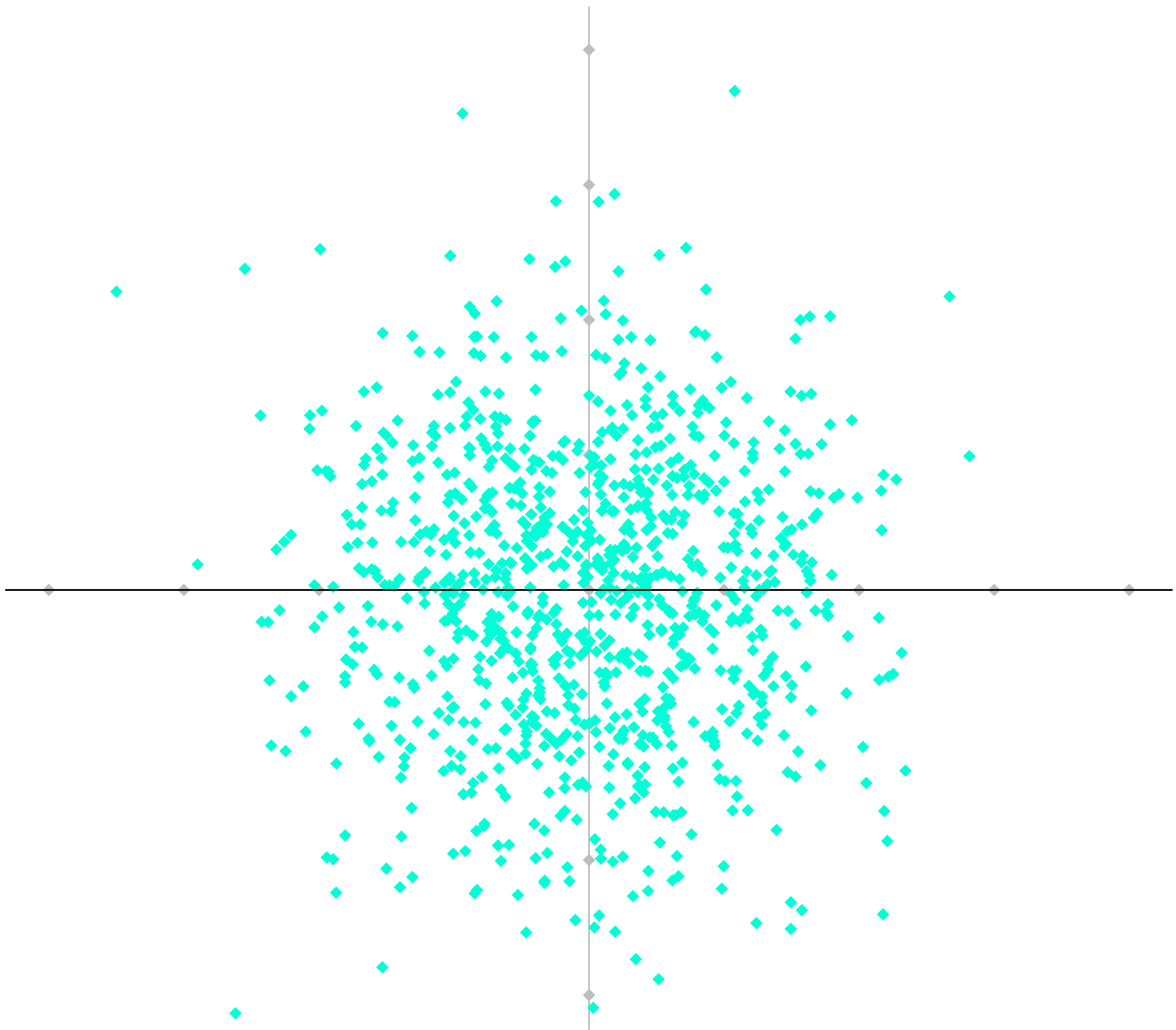
Korrelation = −0.2



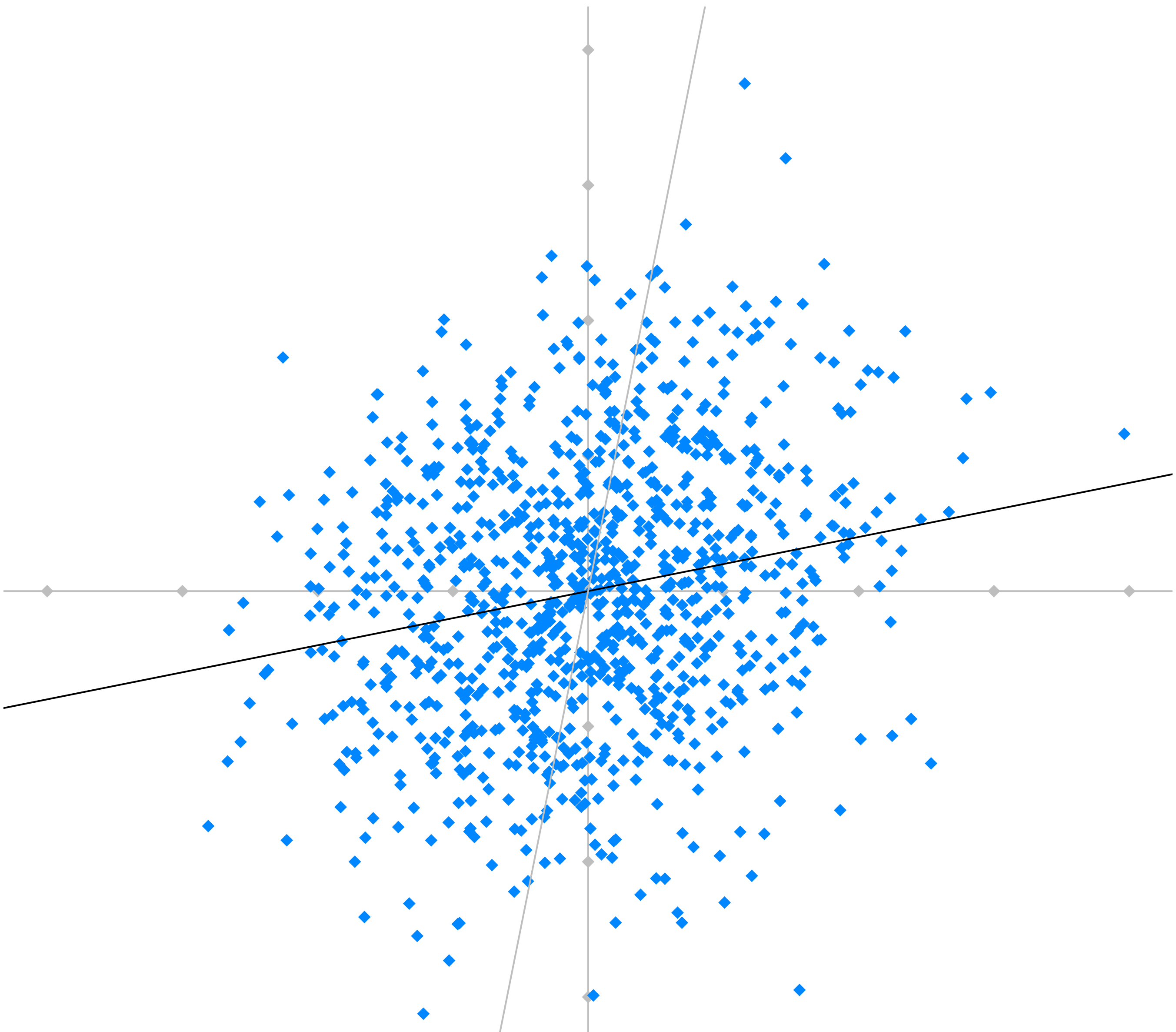
Korrelation = −0.1



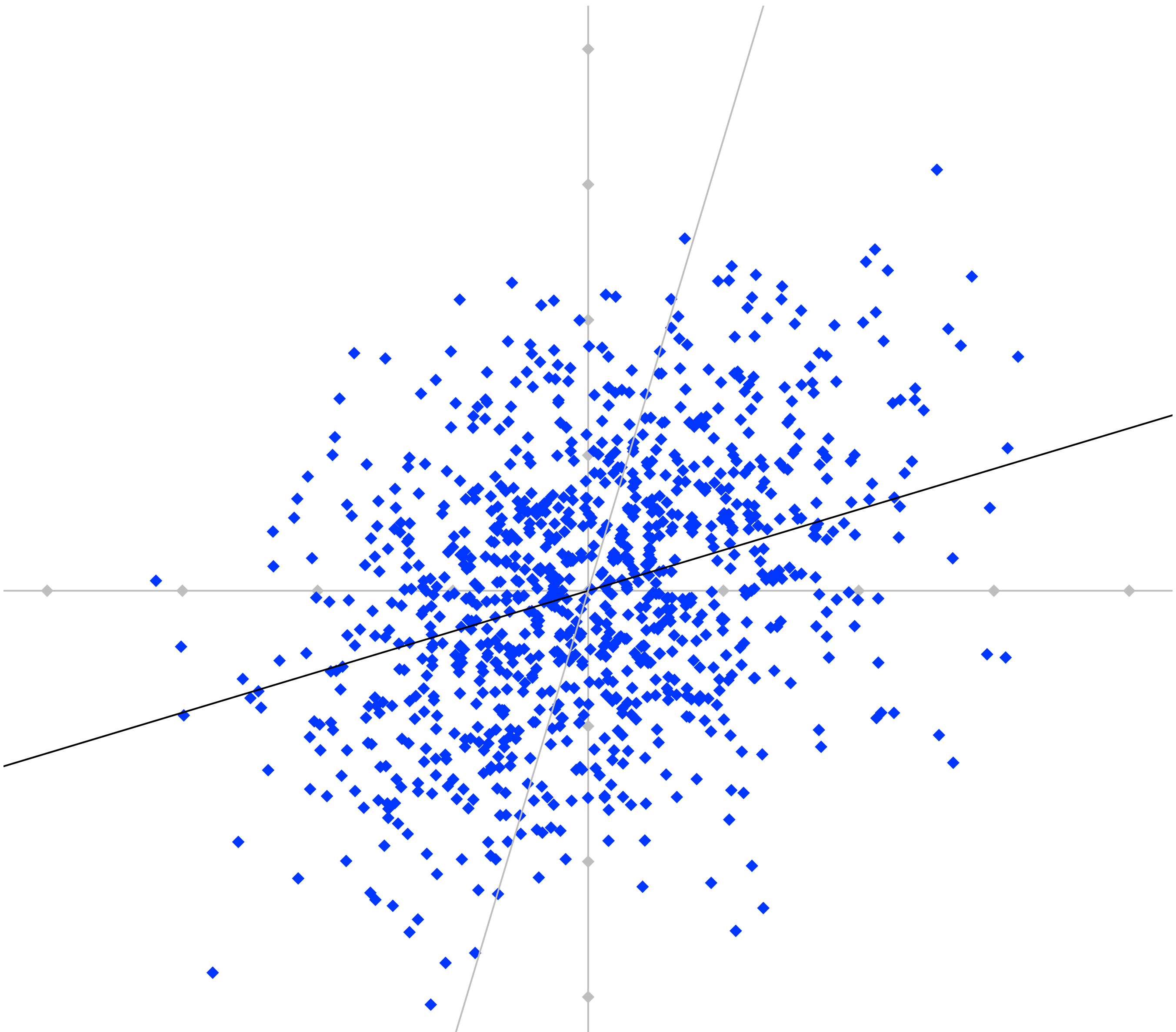
Korrelation = 0



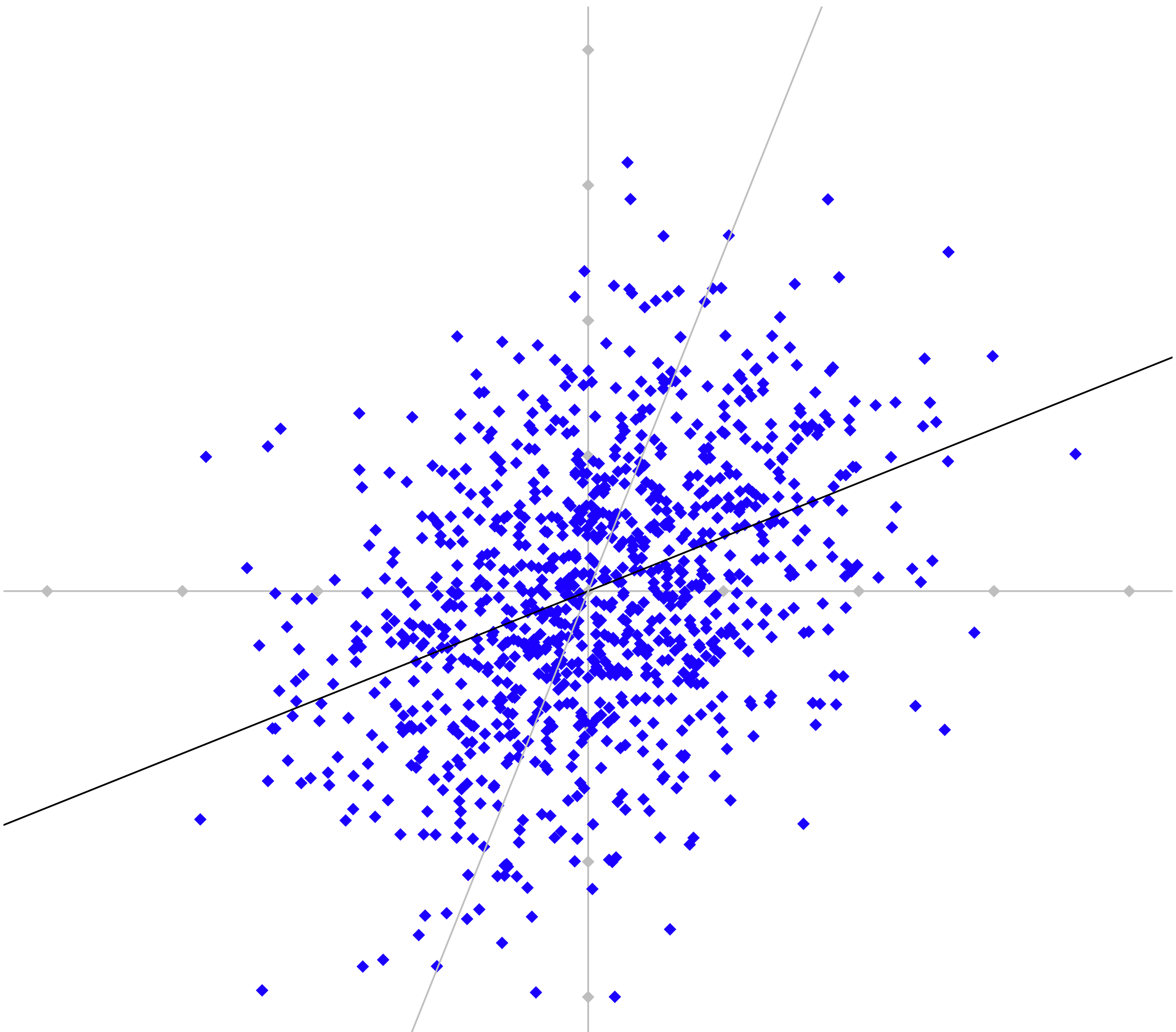
Korrelation = 0.2



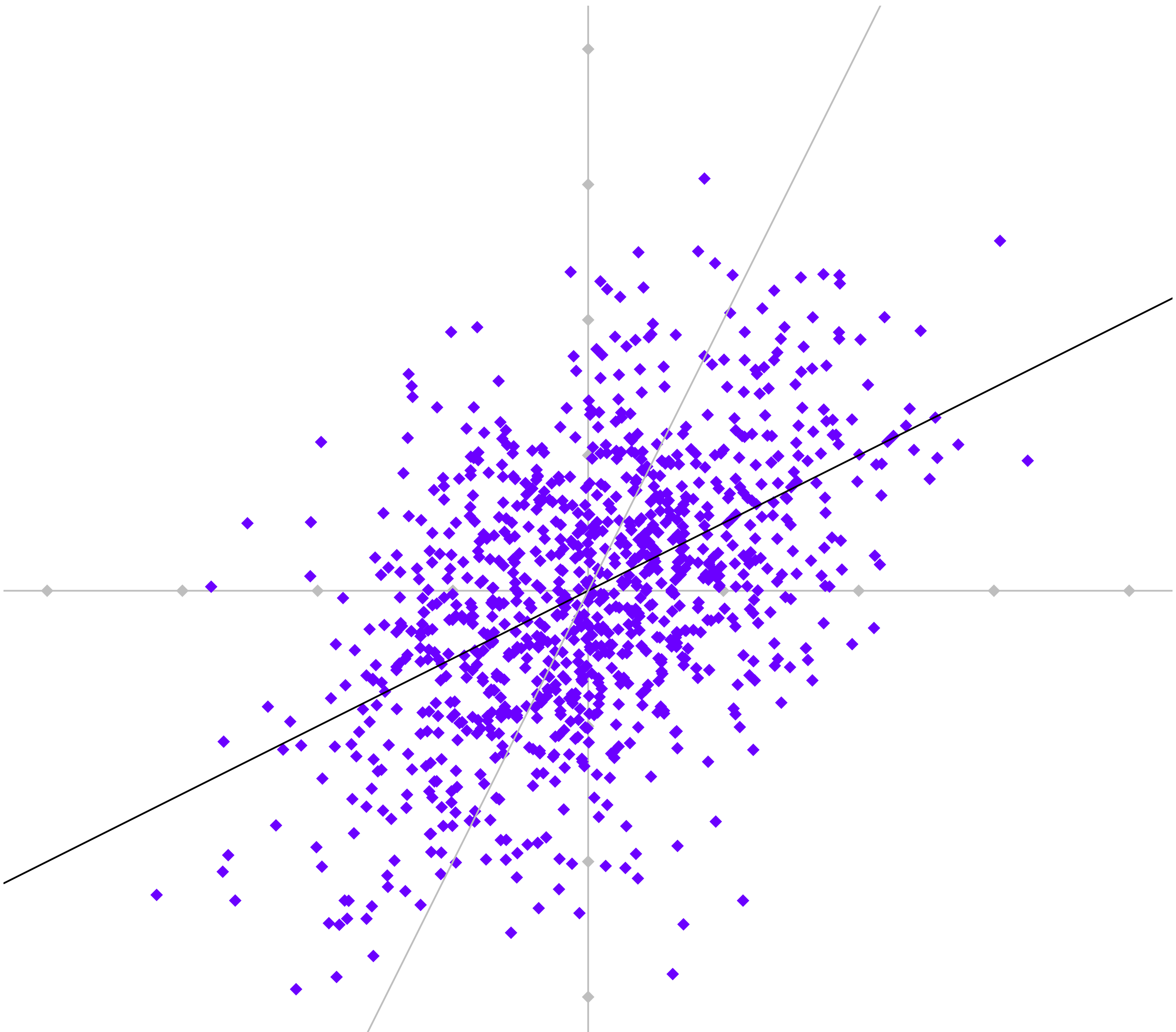
Korrelation = 0.3



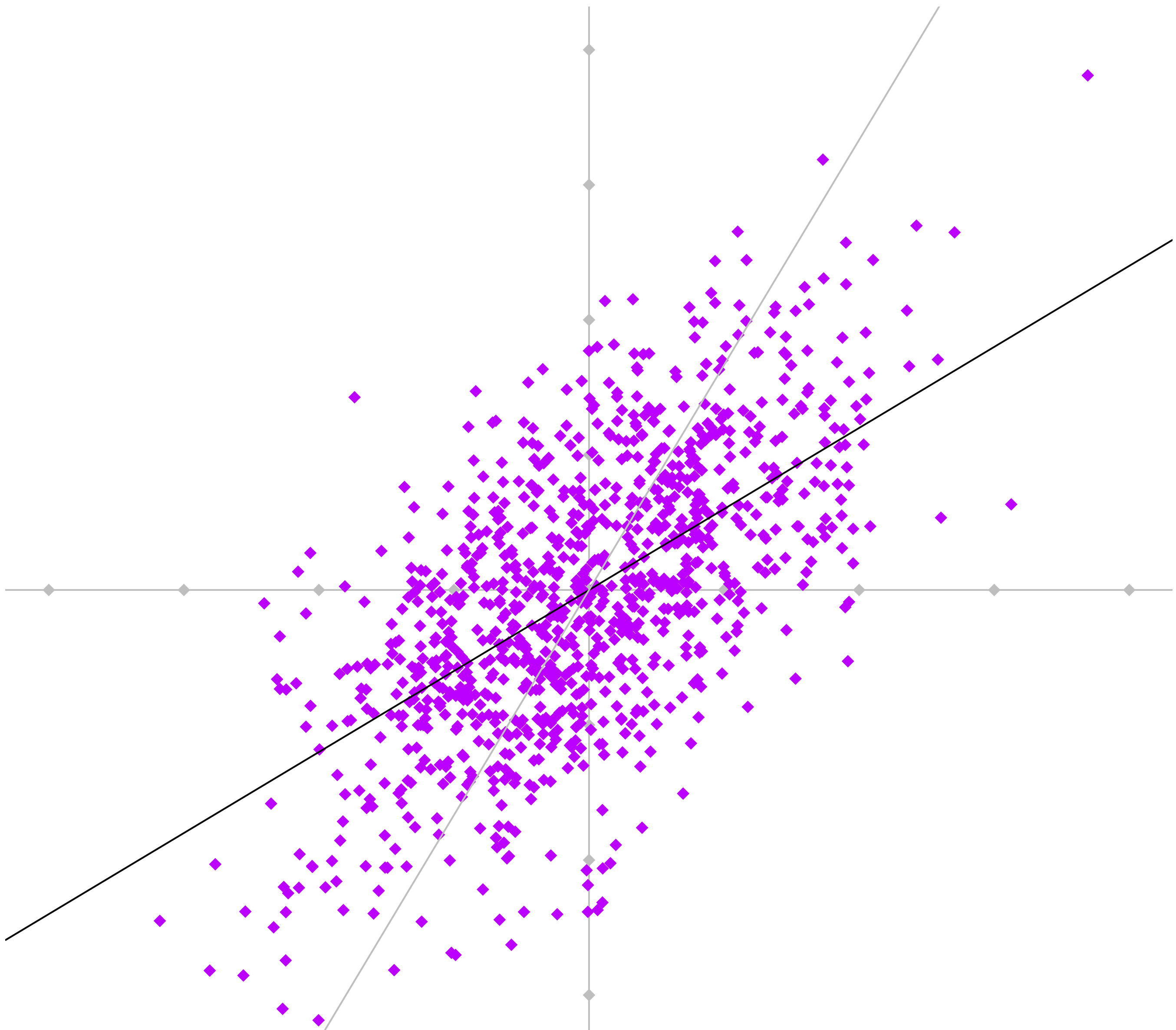
Korrelation = 0.4



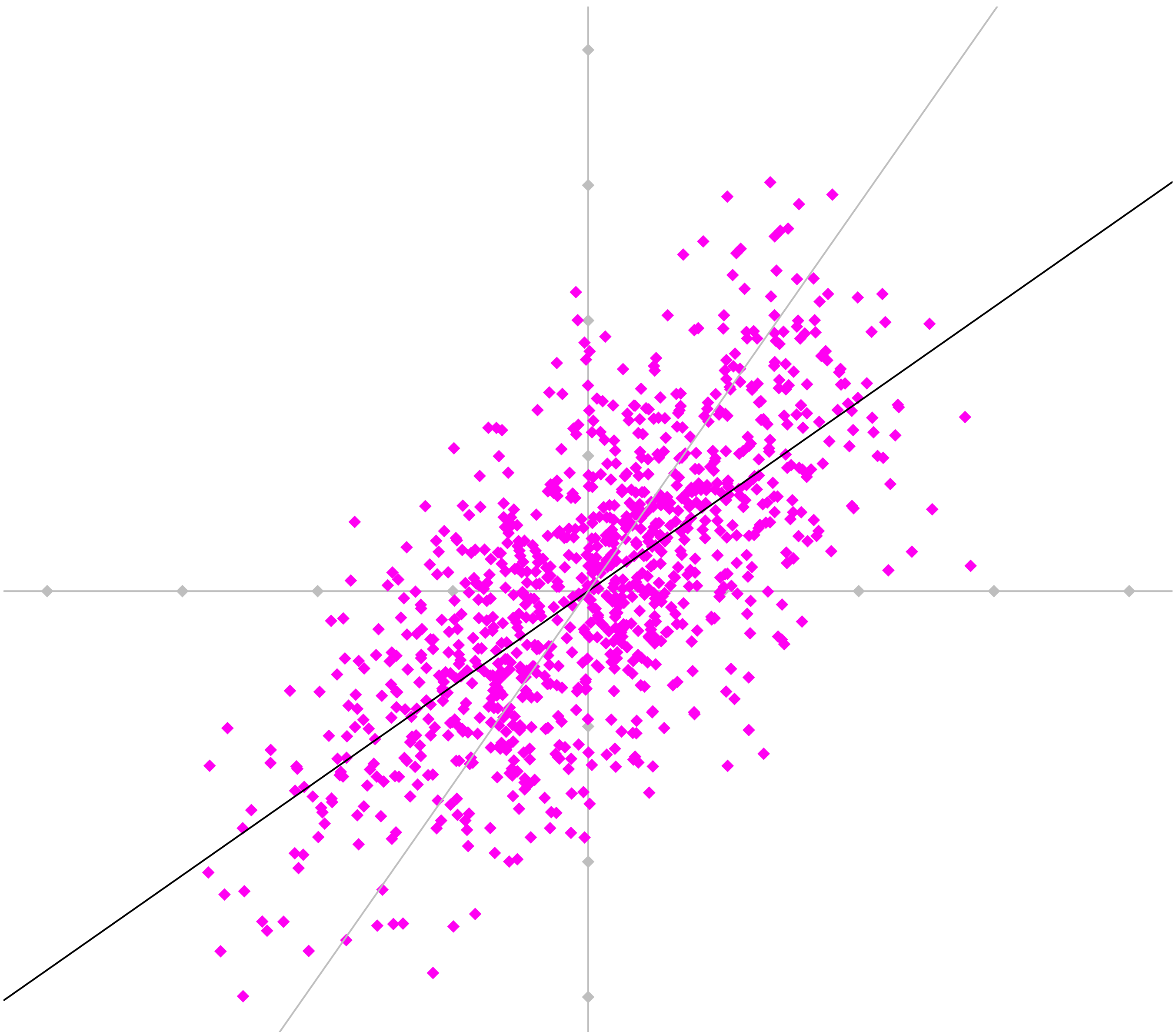
Korrelation = 0.5



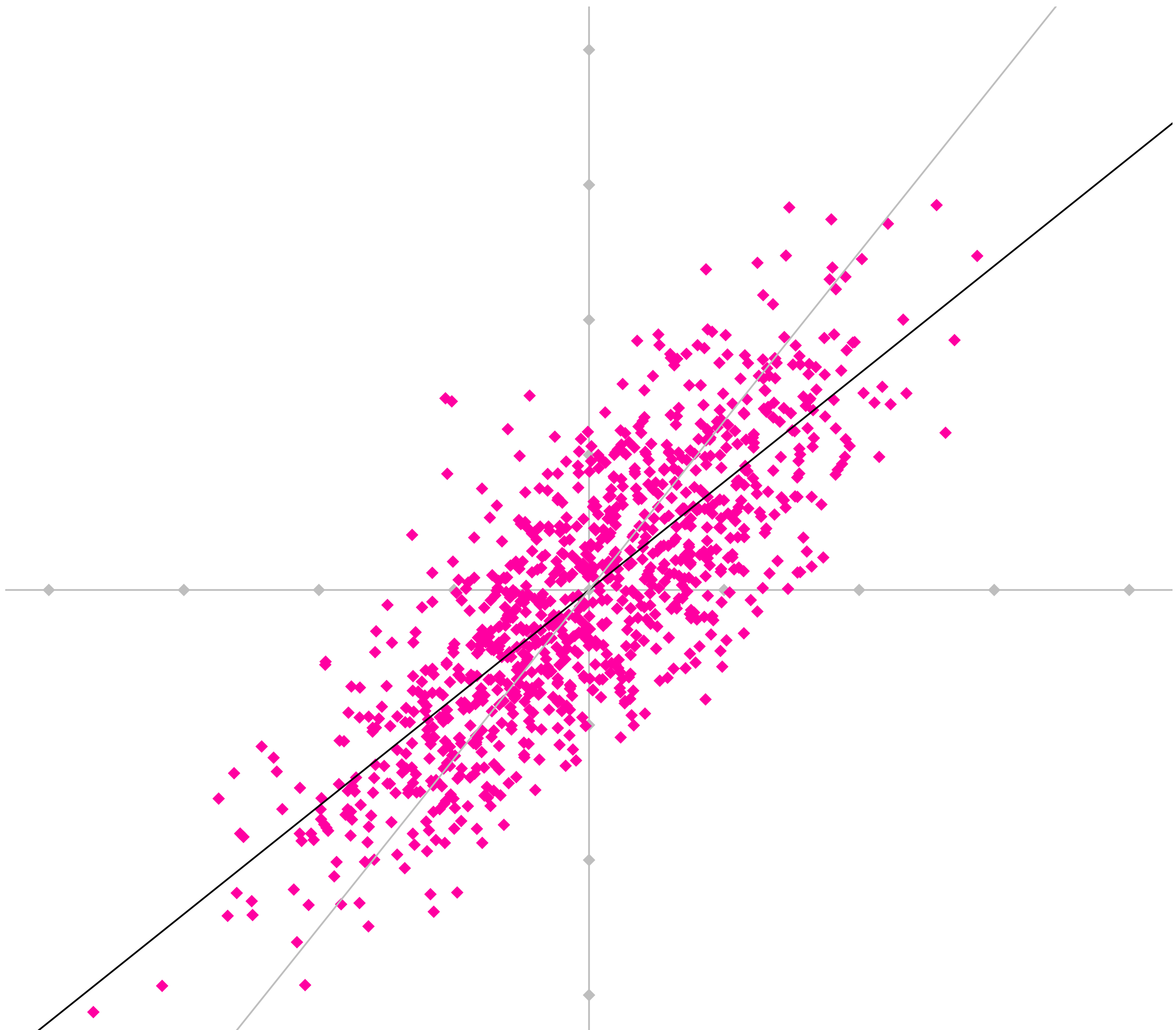
Korrelation = 0.6



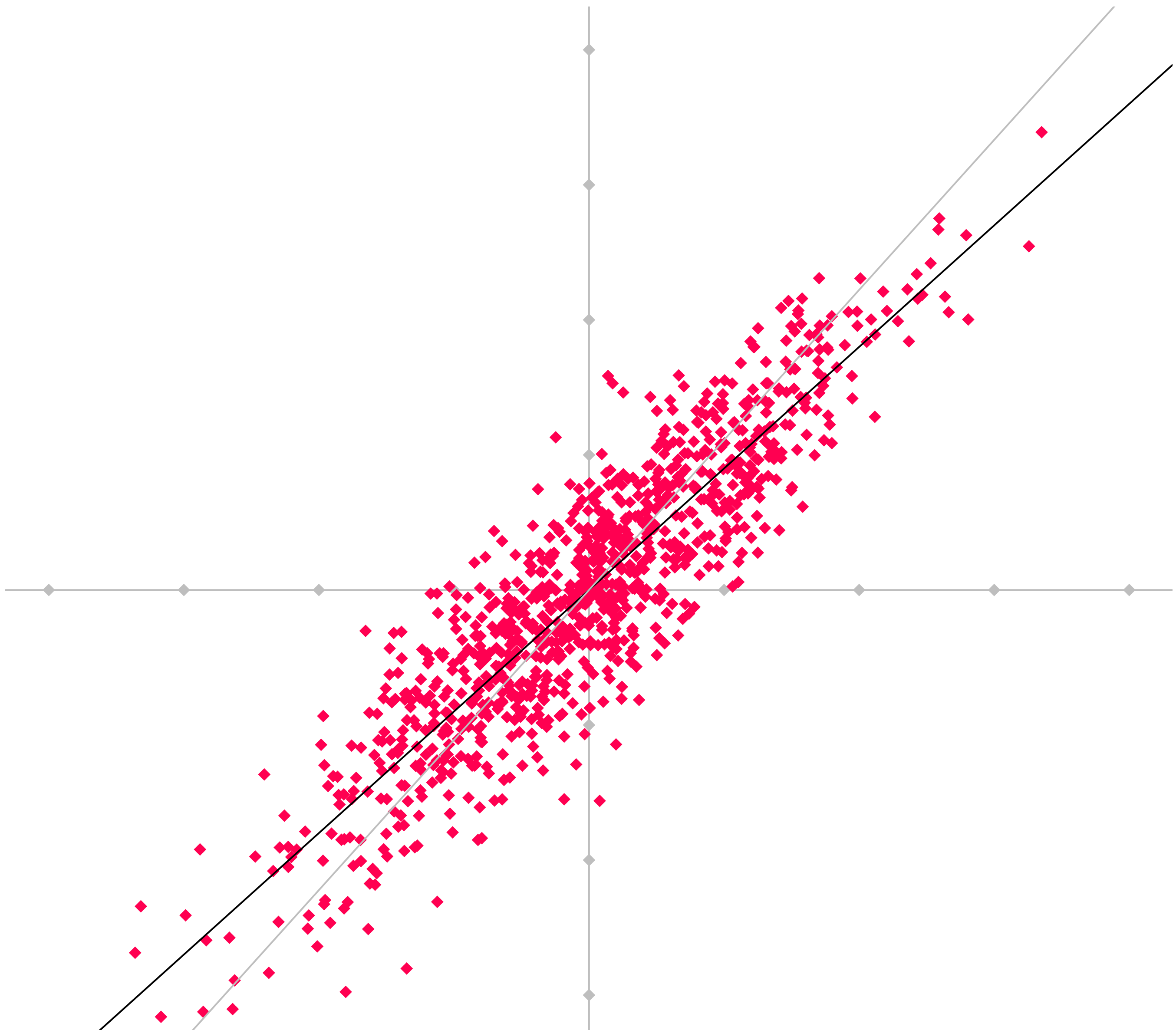
Korrelation = 0.7



Korrelation = 0.8



Korrelation = 0.9



Beispiel 2:

$(x_1, y_1), \ldots, (x_n, y_n)$ seien n Punkte im $\mathbb{R}^2$.

Das zufällige Paar (X, Y) habe die Verteilung

$$\nu := \frac{1}{n} \sum_{i=1}^{n} \delta_{(x_i, y_i)}.$$

(Dabei ist $\delta_z(B) := 1$ für $z \in B$, und $:= 0$ für $z \neq B$.)

$\frac{1}{n}\delta_{(x_i,y_i)}$ gibt also dem Punkt (x_i, y_i) das Gewicht $\frac{1}{n}$.

(X, Y) hat Verteilung $\nu = \frac{1}{n} \sum_{i=1}^{n} \delta_{(x_i, y_i)}$. Dann ist

$$\mathbf{E}X = \frac{1}{n} \sum x_i =: \bar{x}$$

$$\sigma_X^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

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

$$\mathbf{E}[(Y - \beta_1 X - \beta_0)^2] = \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta_1 x_i - \beta_0)^2$$

wird, wie wir gezeigt haben, minimiert durch

$$\beta_1 := \frac{\sigma_Y}{\sigma_X}\kappa = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

und β_0 so, dass $\bar{y} = \beta_1 \bar{x} + \beta_0$.

Diese Gerade $y = \beta_1 x + \beta_0$ heißt die

Regressionsgerade zu den Punkten $(x_i, y_i),\ i = 1, \ldots, n$..