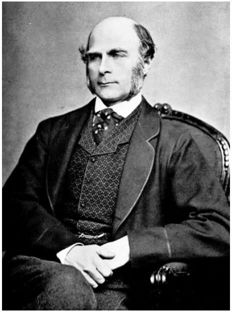
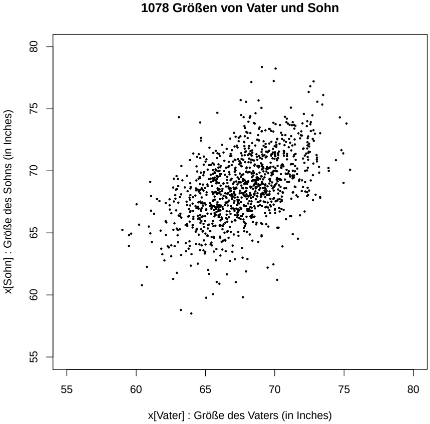
Woher kommt der Name „Regression“ (nach lat. regressio, Zurückkommen)?

Woher kommt der Name „Regression“ (nach lat. regressio, Zurückkommen)?



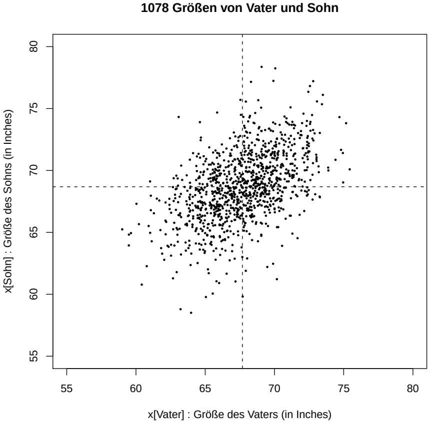
Francis Galton (1822–1911, engl. Wissenschaftler)
hat angesichts biometrischer Beobachtungen den Ausdruck "regression towards the mean" geprägt.

Ein (relativ berühmter) Datensatz[1] von Karl Pearson (1858-1936)



[1] siehe `data(father.son)` aus dem `R`-Paket `UsingR`

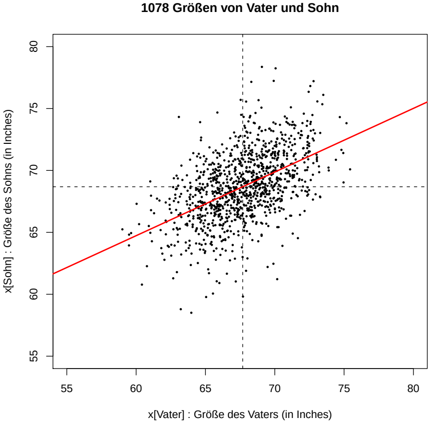
Ein (relativ berühmter) Datensatz[1] von Karl Pearson (1858-1936)



$\overline{x}_{\text{Vater}} = 67.7$, $\overline{x}_{\text{Sohn}} = 68.7$, $\sigma^2_{\text{Vater}} = 7.52$ ($\sigma_{\text{Vater}} = 2.74$, $\sigma_{\text{Sohn}} = 2.81$),
$\text{cov}(x_{\text{Vater}}, x_{\text{Sohn}}) = 3.87$
(Korrelationskoeffizient $\rho = \text{cov}(x_{\text{Vater}}, x_{\text{Sohn}})/(\sigma_{\text{Vater}}\, \sigma_{\text{Sohn}}) = 0.50$)

[1] siehe `data(father.son)` aus dem `R`-Paket `UsingR`

Ein (relativ berühmter) Datensatz[1] von Karl Pearson (1858-1936)

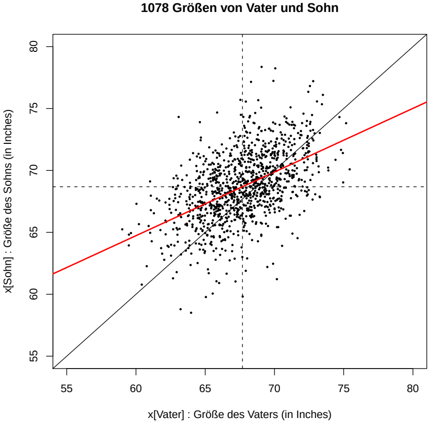


$\overline{x}_{\text{Vater}} = 67.7$, $\overline{x}_{\text{Sohn}} = 68.7$, $\sigma^2_{\text{Vater}} = 7.52$ ($\sigma_{\text{Vater}} = 2.74$, $\sigma_{\text{Sohn}} = 2.81$),
$\text{cov}(x_{\text{Vater}}, x_{\text{Sohn}}) = 3.87$
(Korrelationskoeffizient $\rho = \text{cov}(x_{\text{Vater}}, x_{\text{Sohn}})/(\sigma_{\text{Vater}}\,\sigma_{\text{Sohn}})) = 0.50$)

Regressionsgerade: $x_{\text{Sohn}} = 33.89 + 0.514 x_{\text{Vater}}$.

[1] siehe `data(father.son)` aus dem R-Paket `UsingR`

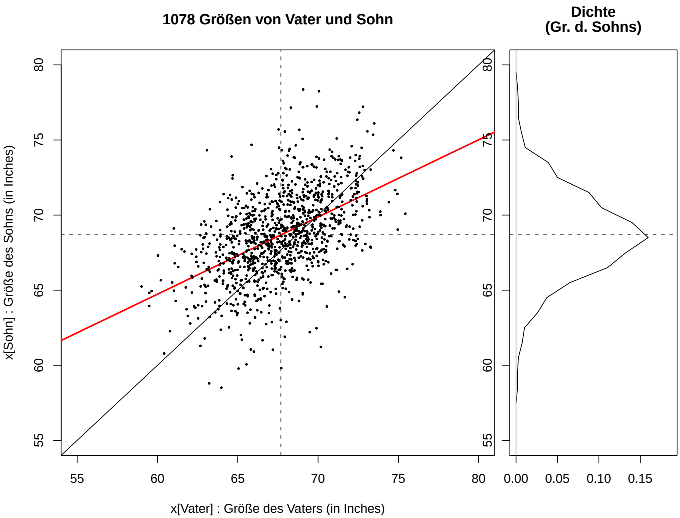
Ein (relativ berühmter) Datensatz[1] von Karl Pearson (1858-1936)

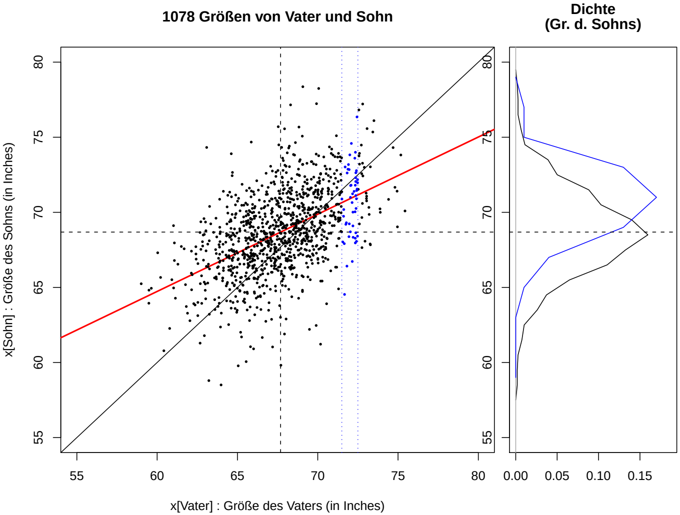


$\overline{x}_{\text{Vater}} = 67.7$, $\overline{x}_{\text{Sohn}} = 68.7$, $\sigma^2_{\text{Vater}} = 7.52$ ($\sigma_{\text{Vater}} = 2.74$, $\sigma_{\text{Sohn}} = 2.81$),
$\text{cov}(x_{\text{Vater}}, x_{\text{Sohn}}) = 3.87$
(Korrelationskoeffizient $\rho = \text{cov}(x_{\text{Vater}}, x_{\text{Sohn}})/(\sigma_{\text{Vater}}\,\sigma_{\text{Sohn}})) = 0.50$)

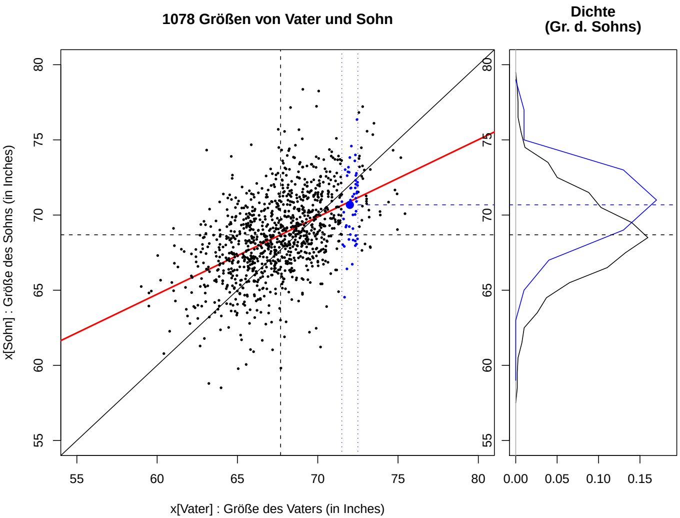
Regressionsgerade: $x_{\text{Sohn}} = 33.89 + 0.514 x_{\text{Vater}}$.

[1] siehe `data(father.son)` aus dem R-Paket `UsingR`

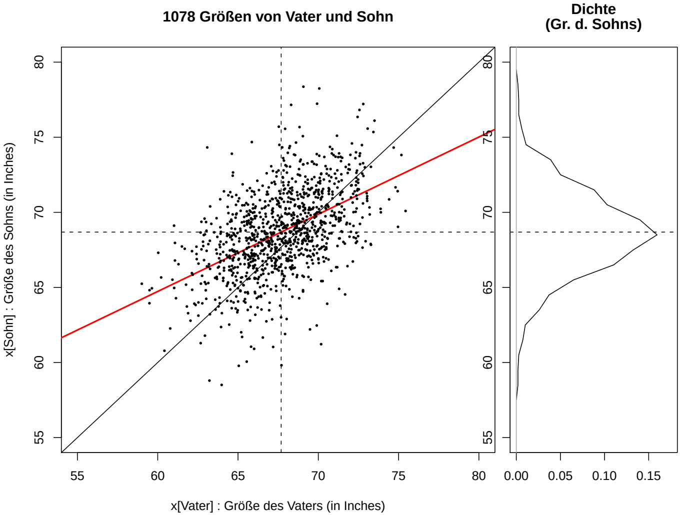

1078 Größen von Vater und Sohn
Dichte
(Gr. d. Sohns)
x[Sohn] : Größe des Sohns (in Inches)
x[Vater] : Größe des Vaters (in Inches)
55
60
65
70
75
80
0.00
0.05
0.10
0.15

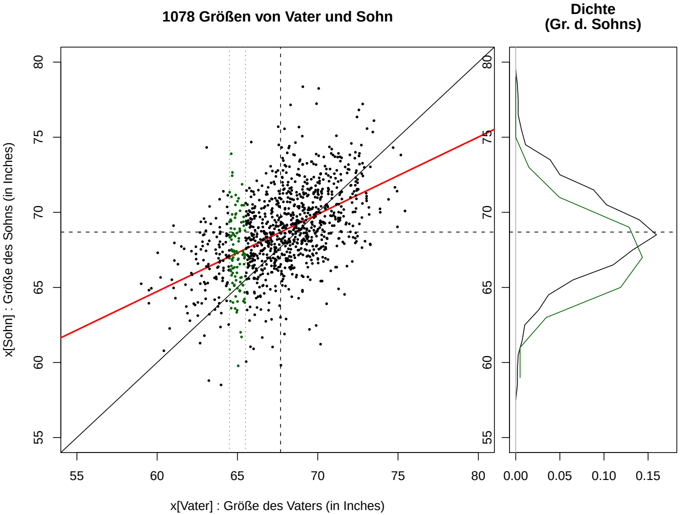


Betrachten wir die Söhne von überdurchschnittlich großen Vätern (z.B. Väter, die ca. 72 Inches groß sind):

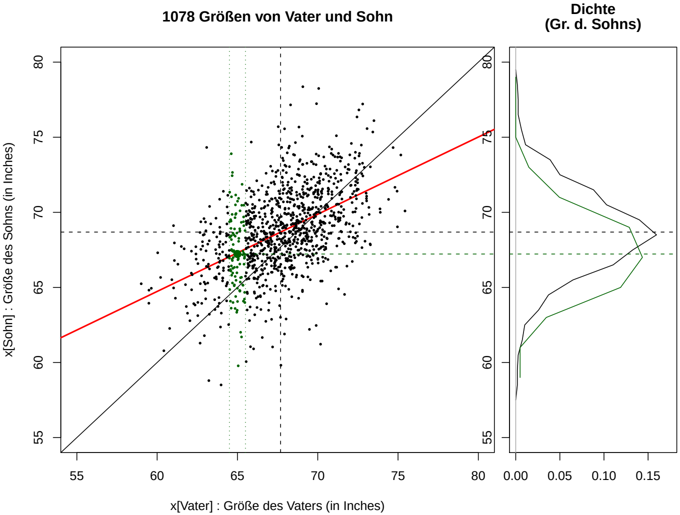


Betrachten wir die Söhne von überdurchschnittlich großen Vätern (z.B. Väter, die ca. 72 Inches groß sind):
Diese Söhne sind überdurchschnittlich groß (im Vergleich zu allen Söhnen), aber im Mittel kleiner als ihr Vater.


1078 Größen von Vater und Sohn
Dichte
(Gr. d. Sohns)
x[Sohn] : Größe des Sohns (in Inches)
x[Vater] : Größe des Vaters (in Inches)
55
60
65
70
75
80
0.00
0.05
0.10
0.15

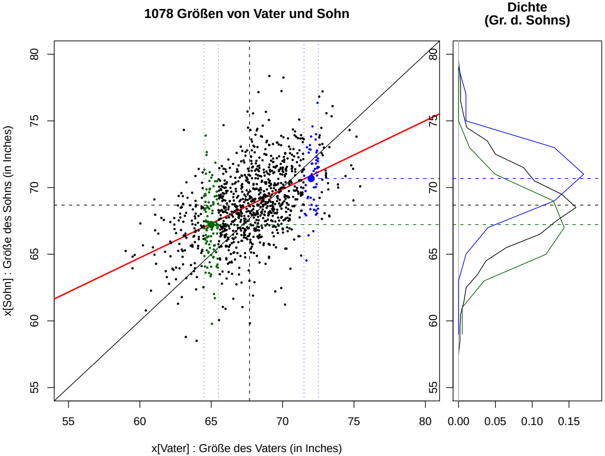


Betrachten wir andererseits die Söhne von unterdurchschnittlich großen Vätern (z.B. Väter, die ca. 65 Inches groß sind):



Betrachten wir andererseits die Söhne von unterdurchschnittlich großen Vätern (z.B. Väter, die ca. 65 Inches groß sind):
Diese Söhne sind unterdurchschnittlich groß (im Vergleich zu allen Söhnen), aber im Mittel größer als ihr Vater.

“Regression towards the mean”



Wir sehen: Söhne überdurchschnittlich großer Väter sind im Mittel kleiner als ihr Vater (aber immer noch größer als der Populationsdurchschnitt), für Söhne unterdurchschnittlich großer Väter ist es umgekehrt: „Rückkehr zum Mittelwert“.

Bemerkung: Das beobachtete Phänomen der „Rückkehr zum Mittelwert“ bedeutet nicht notwendigerweise einen tieferen kausalen Zusammenhang, es tritt stets im Zusammenhang mit natürlicher Variabilität auf (technisch gesehen stets, wenn für den Korrelationkoeffizient ρ gilt $|\rho| < 1$).

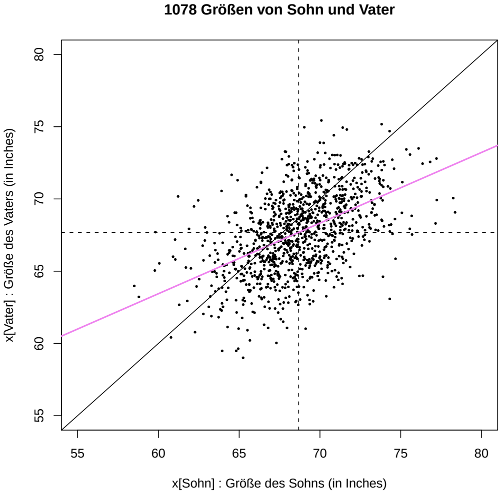
Bemerkung: Das beobachtete Phänomen der „Rückkehr zum Mittelwert“ bedeutet nicht notwendigerweise einen tieferen kausalen Zusammenhang, es tritt stets im Zusammenhang mit natürlicher Variabilität auf (technisch gesehen stets, wenn für den Korrelationkoeffizient ρ gilt $|\rho| < 1$).

Bestimmen wir (spaßeshalber) im Größen-Beispiel die Regressionsgerade für die Größe des Vaters als Funktion der Größe des Sohns:

Bemerkung: Das beobachtete Phänomen der „Rückkehr zum Mittelwert“ bedeutet nicht notwendigerweise einen tieferen kausalen Zusammenhang, es tritt stets im Zusammenhang mit natürlicher Variabilität auf (technisch gesehen stets, wenn für den Korrelationkoeffizient ρ gilt $|\rho| < 1$).

Bestimmen wir (spaßeshalber) im Größen-Beispiel die Regressionsgerade für die Größe des Vaters als Funktion der Größe des Sohns:

Wir hatten $\overline{x}_{\text{Vater}} = 67.7$, $\overline{x}_{\text{Sohn}} = 68.7$, $\text{cov}(x_{\text{Vater}}, x_{\text{Sohn}}) = 3.87$, $\sigma^2_{\text{Sohn}} = 7.92$
und finden die Regressionsgerade $x_{\text{Vater}} = 34.1 + 0.489 x_{\text{Sohn}}$



Regressionsgerade: $x_{\text{Vater}} = 34.1 + 0.489 x_{\text{Sohn}}$