Regression to the mean

Vojtech Franc
vojtech.franc@gmail.com

July 2024

1 Introduction

Regression to the mean is a statistical phenomenon that helps explain numerous real-life observations without needing to invoke complex or untestable hypotheses. For instance, it clarifies why highly intelligent women often marry less intelligent men (and vice versa), why very tall parents tend to have children who are not as tall, and why patients often see improvements with a placebo. Numerous other examples demonstrate the applicability of regression to the mean, as discussed in [3].

There are numerous web pages dedicated to explaining regression to the mean. While many provide a good intuitive grasp of the concept, few offer a precise yet straightforward explanation using elementary mathematics. This text aims to fill that gap.

We begin by introducing perhaps the simplest statistical model where this phenomenon occurs: the bivariate normal distribution. From there, we derive formulas that both explain and quantify regression to the mean. We also delve into a related, more general concept known as reversion to the mean. Finally, we define both phenomena in a broader context beyond the bivariate normal distribution. To illustrate these concepts, we'll use data from a real-world observation as a recurring example: the tendency for highly intelligent women to often marry less intelligent men [5].

2 Simple model: bivariate normal distribution

The regression to the mean phenomenon revolves around the relationship between two random variables with identical distributions. We will denote these variables by X and Y. Throughout most of the text, we will focus on one of the simplest setups where this phenomenon occurs, assuming that the variables

(X, Y) follow a bivariate normal distribution:

$$p(x,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \cdot$$
$$\exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu}{\sigma}\right)^2 - 2\rho\left(\frac{x-\mu}{\sigma}\right)\left(\frac{y-\mu}{\sigma}\right) + \left(\frac{y-\mu}{\sigma}\right)^2\right]\right) \quad (1)$$

where $\mu \in \mathbb{R}$ is the mean value, $\sigma > 0$ is the standard deviation, and $\rho \in (-1, 1)$ is the correlation coefficient. It follows from (1) that the marginal distribution of X and Y are identical. Specifically, $p(x) = \mathcal{N}(x; \mu, \sigma)$ and $p(y) = \mathcal{N}(y; \mu, \sigma)$ where

$$\mathcal{N}(z; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\frac{(z-\mu)^2}{\sigma^2}\right) .$$

is the univariate normal distribution with mean μ and standard deviation σ.

Let $Y \mid X = x$ denote the random variable Y given that X takes the value x. The distribution of $Y \mid X = x$ follows a univariate normal distribution:

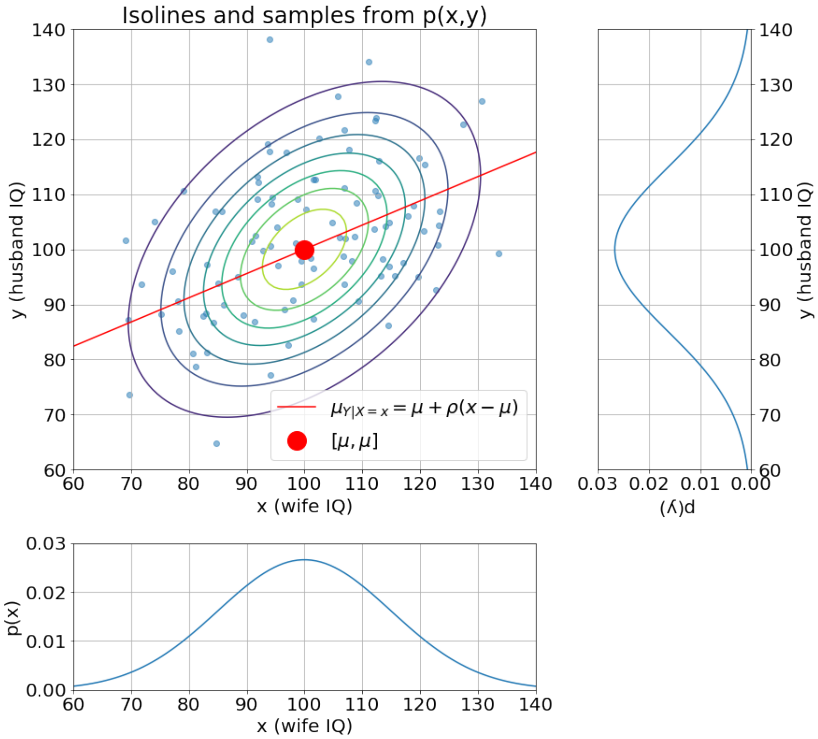
$$p(y \mid x) = \frac{p(x,y)}{p(x)} = \mathcal{N}\left(y; \mu + \rho(x-\mu)\, ,\, \sigma\sqrt{1-\rho^2}\right) , \quad (2)$$

which can be obtained by substituting $p(x) = \mathcal{N}(x; \mu, \sigma)$ and $p(x, y)$ given by (1) to $p(y \mid x) = \frac{p(x,y)}{p(x)}$, and performing some simplifications. From (2), it is evident that the mean of $Y \mid X = x$ is:

$$\mu_{Y|X=x} = \mu + \rho(x - \mu) . \quad (3)$$

More info on the bivariate normal distribution can be found in sources such as [2, 1].

Example 1 *Assume (X, Y) represent IQ scores of a randomly selected wife and her husband, respectively. The IQ tests are design such that the population has a mean IQ score of $\mu = 100$ and a standard deviation of $\sigma = 15$. We assume that the distribution of IQ scores for men and woman is the same. Let's also assume that the correlation coefficient between X and Y is $\rho = 0.44$ (the method for estimating of this value will be discussed below). We assume (X, Y) follow the bivariate normal distribution (1). The joint probability density function $p(x, y)$ and the marginal density functions $p(x)$ and $p(y)$, and 100 points randomly sampled from $p(x, y)$ are illustrated in the following figure. The figure also demonstrates how the expected IQ of a husband $\mu_{Y|X=x} = \mu + \rho(x - \mu) = 100 + 0.44(x - 100)$ depends on his wife's IQ, x.*



3 Regression to the mean

Equation (3) describes the expected value of $Y \mid X = x$, that is, the expected value of the variable Y when the value of X is fixed at x. By re-arranging the equation (3), we obtain:

$$\mu_{Y|X=x} - \mu = \rho(x - \mu)$$

and therefore for any $x > \mu$ and $\rho \in [0, 1)$ we have:

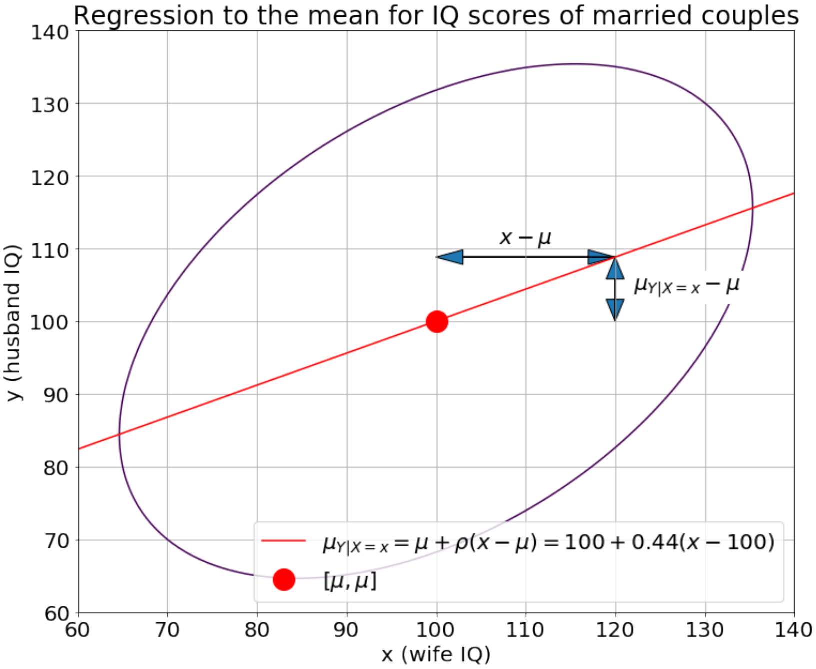
$$\mu_{Y|X=x} - \mu < x - \mu \,. \tag{4}$$

Equation (4) illustrates the regression to the mean phenomenon, which can be described as follows: If X and Y are generated from the bivariate normal distribution (1), and the correlation between X and Y is non-negative and imperfect ($\rho \in [0, 1)$), then for any x greater than the mean μ, the expected value of $Y \mid X = x$ is closer (regresses) to the mean μ than is x to μ.

Example 2 *Assume (X, Y) represent the IQ scores of a married couple, distributed according to a bivariate normal distribution (1) with parameters $\mu =$ 100, $\sigma = 15$ and $\rho = 0.44$, as in the previous examples. If the wife's IQ is fixed* *at $x = 120$, then the husband's IQ follows a normal distribution given by (2), with the mean value*

$$\mu_{Y|X=x} = \mu + \rho(x - \mu) = 100 + 0.44(120 - 100) \approx 108.8\,.$$

Thus, the expected IQ of a husband whose wife has an IQ of 120 *is approximately* 108.8*. The red line in the following figure illustrates the expected husband's IQ,* $\mu_{Y|X=x}$*, for the corresponding wife's IQ,* x*. The slope of the red line is determined solely by the correlation coefficient,* $\rho = 0.44$*. Since this slope lies within the interval* $[0, 1)$*, the difference* $\mu_{Y|X=x} - \mu$ *must be smaller than* $x - \mu$*.*



Some values from the figure are shown in the following table:

IQ of wive x	*Avg. IQ of her husband* $\mu_{Y\|X=x}$
105	*102.2*
110	*104.4*
115	*106.6*
120	*108.8*
125	*111.0*
130	*113.2*
135	*115.4*
140	*117.6*

4 Reversion to the mean

The previous section quantifies the regression of $Y \mid X = x$ to the mean μ for a single fixed value of $X = x$. However, this approach is not always practical with real data because collecting many samples $(x_i = x, y_i)$ is challenging. probability of $X = x$ is effectively zero.

Instead, let's consider a slightly different approach where we analyze the expecpected value of $Y \mid X \geq \theta$. In this context, we seek to analyze the expected value of Y not for a single $X = x$ (as in the case of regression to the mean), but rather for the entire population of X values that are greater than a chosen threshold $\theta > -\infty$.

As before, we assume that the original random variable X has a normal distribution with mean μ and standard deviation σ. Then, the random variable X conditioned on $X \geq \theta$ follows what is known as a truncated normal distribution:

$$\mathcal{N}_{\mathrm{tr}}(x; \mu, \sigma, \theta) = \begin{cases} \mathcal{N}\left(x; \mu, \sigma\right) / \left(1 - \Phi\left(\frac{\theta - \mu}{\sigma}\right)\right) & \text{if} \quad x \geq \theta \\ 0 & \text{if} \quad x < \theta \,. \end{cases}$$

where

$$\Phi(x) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right)$$

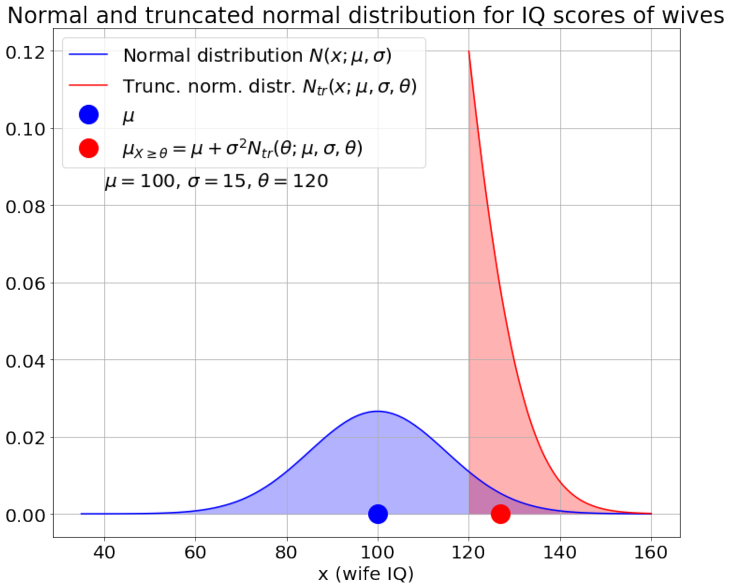
is the cumulative distribution function of the standard normal distribution $\mathcal{N}(x; 0, 1)$, that is, $\phi(x) = \int_0^x \mathcal{N}(x; 0, 1) dx$. The normalization constant $1 - \Phi((\theta - \mu_X)/\sigma_X)$ is introduced to ensure that the truncated normal distribution $\mathcal{N}\left(x; \mu, \sigma\right)$ integrates to 1. It is important to note that $\mathcal{N}(x; \mu, \sigma) = \mathcal{N}_{\mathrm{tr}}(x; \mu, \sigma, -\infty)$.

The expected value of X that follows the truncated normal distribution $\mathcal{N}_{\mathrm{tr}}(x; \mu, \sigma, \theta)$ equals to:

$$\mu_{X \geq \theta} = \mathbb{E}[X \geq \theta] = \int_\theta^\infty x \mathcal{N}_{\mathrm{tr}}(x; \mu, \sigma, \theta) dx = \mu + \sigma^2 \mathcal{N}_{\mathrm{tr}}(\theta; \mu, \sigma, \theta) \,. \quad (5)$$

More info on the truncated normal distribution can be found in [6].

Example 3 *Assume the wife's IQ score are normally distributed with the mean* $\mu = 100$ *and the standard deviation* $\sigma = 15$*, as in the previous examples. If we consider only the wives with IQ scores of at least* $\theta = 120$*, their score will follow the truncated normal distribution* $\mathcal{N}_{\mathrm{tr}}(x; \mu, \sigma, \theta)$*. The following figure depicts both the original normal distribution and the truncated normal distribution. Note that the areas under the probability density functions, marked in blue and read, both equal 1.*



Let us explore the expected value of $Y \mid X \geq \theta$, which represents the expected value of Y given that X is at least θ. Recall that the expected value of $Y \mid X = x$ equals $\mu_{Y|X=x} = \mu + \rho(x - \mu)$. Therefore, we need to compute the expected value of $\mu_{Y|X=x}$ when x is generated from the truncated normal distribution $\mathcal{N}_{\text{tr}}(x; \mu, \sigma, \theta)$:

$$\begin{aligned} \mu_{Y|X\geq\theta} = \mathbb{E}[Y \mid X \geq \theta] &= \int_\theta^\infty (\mu + \rho(x-\mu))\,\mathcal{N}_{\text{tr}}(x;\mu,\sigma,\theta)\,dx \\ &= \mu + \rho\left(\int_\theta^\infty x\,\mathcal{N}_{\text{tr}}(x;\mu,\sigma,\theta)\,dx - \mu\right) \\ &= \mu + \rho\big(\mu_{X\geq\theta} - \mu\big)\,. \end{aligned} \tag{6}$$

We observe that the formula for computing the expectation $\mu_{Y|X\geq\theta} = \mu + \rho(\mu_{X\geq\theta} - \mu)$ resembles the formula for computing $\mu_{Y|X=x} = \mu + \rho(x - \mu)$. Specifically, the former is derived from the latter by replacing x with $\mu_{X\geq\theta}$.

Let us rewrite the equation (6) as follows:

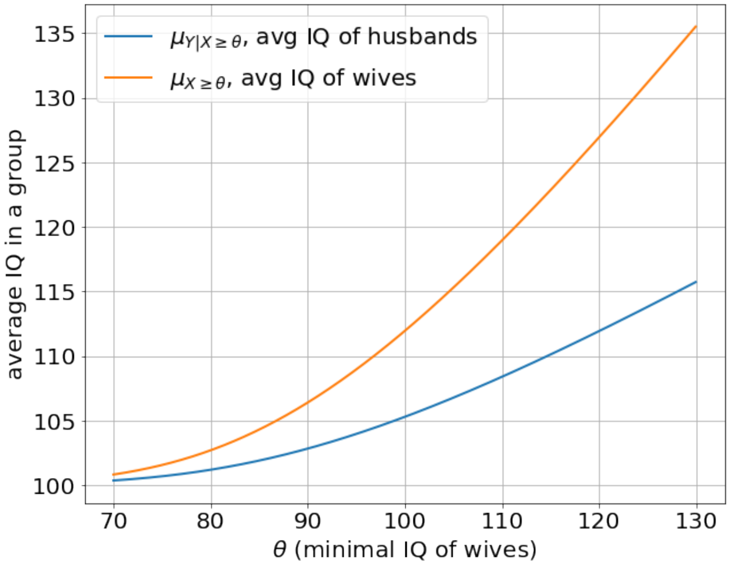
$$\mu_{Y|X\geq\theta} - \mu = \rho(\mu_{X\geq\theta} - \mu)\,.$$

From equation (5), it follows that $\mu_{X\geq\theta} > \mu$, assuming $\sigma > 0$. Therefore, for $\rho \in [0, 1)$, we have that:

$$\mu_{Y|X\geq\theta} - \mu < \mu_{X\geq\theta} - \mu\,. \tag{7}$$

Equation (7) expresses what is known as the *reversion to the mean* phenomenon [4]: Assuming X and Y are generated by the bivariate normal distribution (1) with $\rho \in [0, 1)$, then for any θ, the expected value of $X \geq \theta$ is greater than the expected value of $Y \mid X \geq \theta$.

Example 4 *Assume (X, Y) represent the IQ scores of a married couple, distributed according to a bivariate normal distribution (1) with the parameters $\mu = 100$, $\sigma = 15$ and $\rho = 0.44$, as in the previous examples. Let's examine a group of couples formed by randomly sampling the married couples and selecting only those where the wife's IQ is at least θ. The following figure depicts the average IQ of the husbands, $\mu_{Y|X\geq\theta}$, and the wives, $\mu_{X\geq\theta}$, within this group.*



Some values from the figure are shown in the following table:

Minimal IQ of wives in a group θ	*Avg. IQ of wives* $\mu_{X\geq\theta}$	*Avg. IQ of husbands* $\mu_{Y|X\geq\theta}$
70	*100.8*	*100.4*
80	*102.7*	*101.2*
90	*106.4*	*102.8*
100	*112.0*	*105.3*
110	*119.0*	*108.4*
120	*127.0*	*111.9*
130	*135.6*	*115.8*

5 Estimation of the correlation coefficient

Equation (6) illustrates that the parameters (μ, ρ) describing the bivariate normal distribution and the mean values $(\mu_{X\geq\theta}, \mu_{Y|X\geq\theta})$ are interconnected. Specifically, any of these four parameters can be determined from the remaining three. For instance, the correlation coefficient ρ can be computed using the formula:

$$\rho = \frac{\mu_{Y|X\geq\theta} - \mu}{\mu_{X\geq\theta} - \mu} \,. \tag{8}$$

Example 5 *A quote from [5] reads: "Well, it turns out that the Henry A. Murray Research Center of the Radcliffe Institute for Advanced Study has a collection of IQ test-scores for 43 women and their husbands, and these data confirm this regression toward the mean. The top quarter of the female IQs averaged 119 while their husbands averaged 109. At the same time, the top-quartile men averaged 117 while their wives averaged 107." That is,* $\hat{\mu}_{X \geq \theta} = 119$ *and* $\hat{\mu}_{Y|X \geq \theta} = 109$. *Since we known (assume) that* $\mu = 100$, *we can use the formula (8) to estimate the correlation coefficient*

$$\hat{\rho}_1 = \frac{109 - 100}{119 - 100} \approx 0.474$$

Since the problem is fully symmetric, we can represent the husband's IQ score by X *and the wife's IQ score by* Y. *Then, we can compute the correlation coefficient using the data in the last sentence of the quotation, that is,*

$$\hat{\rho}_2 = \frac{107 - 100}{117 - 100} \approx 0.412 \, .$$

The two values $\hat{\rho}_1 = 0.474$ *and* $\hat{\rho} = 0.412$ *are close but not exactly the same. The deviation between* $\hat{\rho}_1$ *and* $\hat{\rho}_2$ *has two reasons. Firstly, we use only the estimates of the mean values and, secondly, the bivariate normal distribution might not be a perfect model of the reality. It is statistically more robust to take the average of the two values as the final estimate of the correlation coefficient* $\hat{\rho} = (\hat{\rho}_1 + \hat{\rho}_2)/2 \approx 0.44$, *which is the value we used throughout the text.*

One straightforward method to compute the correlation coefficient would involve using the complete dataset gathered by the aforementioned research center, which includes samples of IQ scores from wives and their husbands. However, this specific dataset isn't readily available online. Thankfully, equation (8) provides a way to estimate the correlation coefficient using only the estimates of the mean values. It's worth noting that properties such as bias, efficiency, and consistency of the estimator (8) remain unclear without further analysis.

6 General setup

We have explored the concepts of regression and reversion to the mean within the context of random variables following the bivariate normal distribution (1). However, these phenomena can also manifest with other distributions. Let's outline the formal definitions of these two phenomena as they appear in [4].

Definition 1 (Regression to the mean) *Let* X *and* Y *be random variables with joint distribution* p. *Assume* X *and* Y *have the same marginal distribution and let* μ *denote their common mean. The distribution* p *exhibits regression toward the mean if, for all* $x > \mu$,

$$\mu \leq \mathbb{E}[Y \mid X = x] < x \tag{9}$$

with the reverse inequalities holding for $x < \mu$.

Definition 2 (Reversion to the mean) *Let X and Y be random variables with joint distribution p. Assume X and Y have the same marginal distribution and let μ denote their common mean. The distribution p exhibits reversion toward the mean if, for any θ,*

$$\mu \leq \mathbb{E}[Y \mid X \geq \theta] < \mathbb{E}[X \mid X \geq \theta] \tag{10a}$$

and

$$\mu \geq \mathbb{E}[Y \mid X \leq \theta] > \mathbb{E}[X \mid X \leq \theta] \tag{10b}$$

According to [4], the concept of reversion to the mean, as defined in Definition 2, is more general than regression to the mean, as defined in Definition 1, in the sense that it applies to a broader range of distributions. In fact, the non-strict inequalities in equation (10) hold for almost all distributions, with some exceptions being degenerate cases. It is straightforward to demonstrate, using the derivations discussed earlier, that the bivariate normal distribution (1) satisfies both Definition 1 and 2 when $\rho \in [0, 1)$.

References

[1] *Multivariate normal distribution.* Wikipedia. URL: `https://en.wikipedia.org/wiki/Multivariate_normal_distribution`.

[2] H. Pishro-Nik. *Introduction to probability, statistics, and random processes.* Kappa Research LLC. See section 5.3.2: Bivariate normal distribution. 2014. URL: `https://www.probabilitycourse.com`.

[3] *Regression Toward the Mean: An Introduction with Examples.* Farnam Street Articles. URL: `https://fs.blog/regression-to-the-mean/`.

[4] M.Y. Samuels. "Statistical Reversion Toward the Mean: More Universal Than Regression Toward the Mean". In: *The American Statistician* 45.4 (1991), pp. 344–346.

[5] Gary Smith. *Why intelligent woman marry less intelligent men.* 2020. URL: `https://mindmatters.ai/2020/09/why-intelligent-women-marry-less-intelligent-men/`.

[6] *Truncated normal distribution.* Wikipedia. URL: `https://en.wikipedia.org/wiki/Truncated_normal_distribution`.