HG
Revised Sept. 2018

Supplement to lecture 9   (Tuesday 18 Sept)

On the bivariate normal model

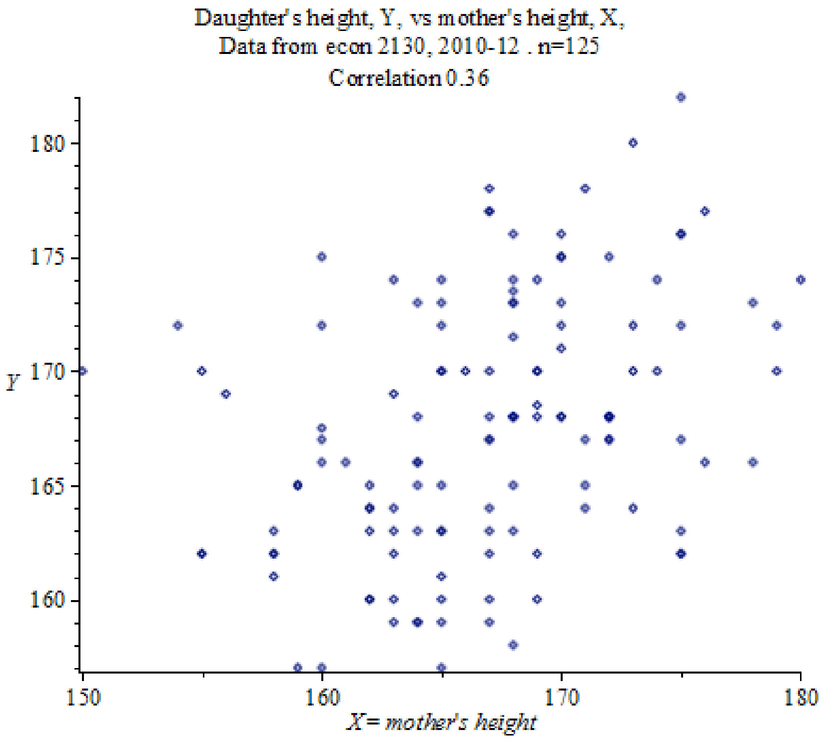
Example: daughter's height (*Y*) vs. mother's height (*X*).

Data collected on Econ 2130 lectures 2010-2012.

The data can be downloaded as an Excel file under Econ2130 at http://folk.uio.no/haraldg/ .

n =125 observation pairs: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ of the random pair (X, Y)

Figure 1   Scatter plot



Data summary:

Mothers:

$\bar{x} = 166.9, \quad \hat{\sigma}_x = s_x = 5.8232 \quad \left[\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad - \text{ sample variance} \right]$

Daughters: $\bar{y} = 167.6, \quad \hat{\sigma}_y = s_y = 5.5938$

Correlation:

$r = \hat{\rho} = s_{xy}/(s_x s_y) = 0.36 \quad \left[s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \quad - \text{ sample covariance} \right]$

Population: all mother – daughter pairs in Norway (with daughters at least 18 years (say))

Model: The bivariate normal distribution (population distribution)

If $(X,Y) \sim N(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ (bivariate normally distributed with 5 parameters (population quantities)
$\mu_x = E(X), \quad \mu_y = E(Y), \quad \sigma_x = SD(X), \quad \sigma_y = SD(Y), \quad \rho = \text{corre}(X,Y)$),

the joint pdf is (see Rice p. 81)

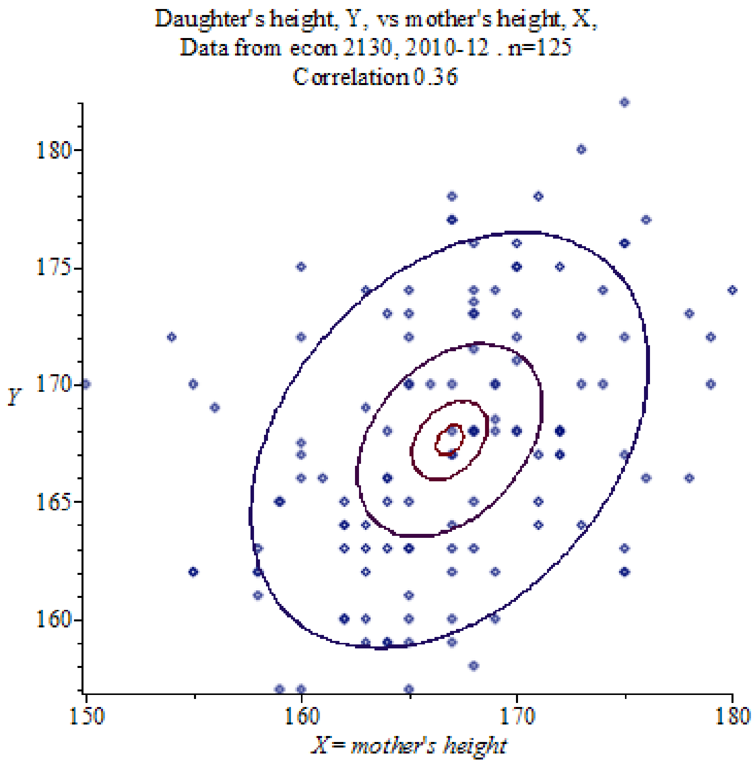
$$(1) \quad f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right] \right)$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$.

Fitting to data. It can be shown that (practically) the best fitting (in the maximum likelihood sense) estimate of this pdf is obtained by substituting $(\bar{x}, \bar{y}, s_x, s_y, r)$ for the parameters $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$. The result is shown in the following contour plot produced by Maple[1]:

Note that maximum of the estimated pdf is obtained in the point $(\bar{x}, \bar{y}) = (166.9,\ 167.6)$, which is the sample estimate of the population means (μ_x, μ_y).

Figure 2 Contour plot of the best fitting joint normal distribution



---

[1] It appears that Stata cannot produce contour plots like this (as far as I know)

Important properties of the bivariate normal distribution

Property 1 **The distribution is symmetric in all directions with highest concentration of observations at the center.**

- The contours (i.e., where the pdf is constant) are ellipses.
- If the scatter plot of observations of (X,Y) does not show symmetry of this kind, the bivariate normal model is not realistic.

Property 2 If $(X,Y) \sim$ bivariate (joint) normal $(N(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho))$, **the marginal distributions are both normal**, $X \sim N(\mu_x, \sigma_x^2)$, and $Y \sim N(\mu_y, \sigma_y^2)$.

For example,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \cdots \text{see Rice p.82 (optional reading)} \cdots = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \cdots \text{similarly} \cdots = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}$$

Note. The other way does not hold! Even if X,Y are both normally distributed marginally, the joint distribution is not necessarily bivariate normal (see e.g., Rice p. 84).

Property 3 **If (X,Y) is bivariate normal and the correlation is zero $(\rho = 0)$, then X and Y are (stochastically) independent!**

Proof. If $\rho = 0$, the general expression (1) reduces to

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right) = f_X(x) f_Y(y)$$

which implies that X and Y are independent. (**End of proof**)

Note. This is a special feature of the joint normal distribution. In general, zero correlation *does not* imply that X and Y are independent!
So, zero correlation can, in general, be looked upon as a weaker form of lack of dependence between X and Y than stochastic independence (which is the strongest form).

Property 4 **If (X,Y) is bivariate normal, both regressions (Y w.r.t X and X w.r.t. Y) are automatically linear and homoscedastic. In addition the two conditional distributions are both normal.**

Having the joint pdf, $f(x,y)$, and the marginal one, $f_X(x)$, we can calculate the conditional pdf for $Y \mid x$:

(2)
$$f(y \mid x) = \frac{f(x,y)}{f_X(x)} = \cdots \text{messy algebra (can be skipped)} \cdots =$$
$$= \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left[y-\mu_y-\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right]^2}$$

which we recognize as a normal pdf with expectation $E(Y \mid x) = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$, and variance $\operatorname{var}(Y \mid x) = \sigma_y^2(1-\rho^2)$.

In particular, the conditional distribution of $Y \mid X = x$ is normal $N\left(E(Y \mid x),\ \operatorname{var}(Y \mid x)\right)$, with regression function (cf. Rice p. 148)

$$\mu(x) = E(Y \mid x) = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\ , \qquad \text{(i.e., a linear function of } x\text{)}$$

and variance function

$$\sigma^2(x) = \operatorname{var}(Y \mid x) = \sigma_y^2(1-\rho^2) \qquad \text{(i.e., constant)}$$

Important result.

Hence, if we can assume reasonably that the joint distribution of X and Y is bivariate normal, it follows automatically (without extra assumptions) that the regression of Y w.r.t. X is linear and homoscedastic.

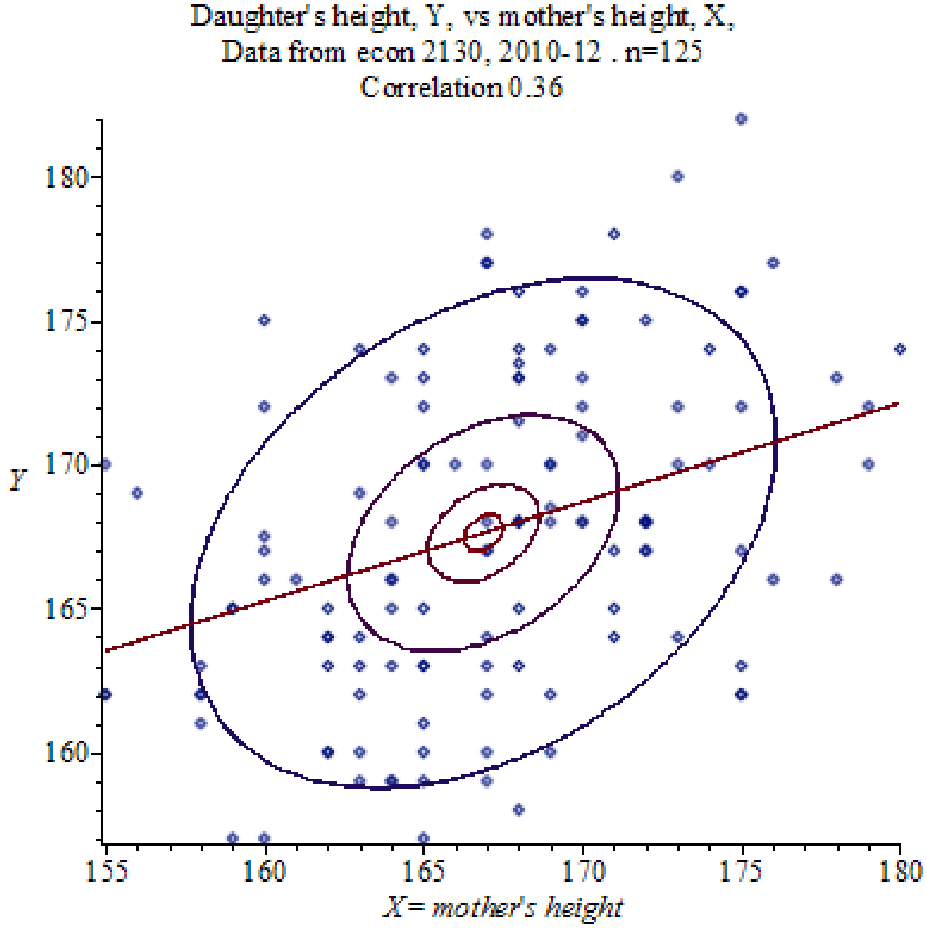
Substituting the estimates we have for $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$, in $\mu(x)$ and $\sigma^2(x)$, we get the estimated regression (which, in fact, are the same as the OLS estimates from the basic course)

$$\hat{\mu}(x) = 167.6 + (0.36)\frac{5.5938}{5.8232}(x - 166.9) = 109.86 + (0.3458)x$$
$$\hat{\sigma}^2(x) = (5.5938)^2(1-(0.36)^2) = (5.2187)^2$$

and the estimated conditional distribution is normal:

$$(Y \mid X = x) \sim N\left(109.86 + (0.3458)x,\ (5.2187)^2\right) \qquad \text{(estimated)}$$

Figure 3 Contour plot of the best fitting joint normal distribution and the implied (OLS estimated) regression function.



Historical note on the term "regression".
The term was used by geneticists in the beginning of last century. They observed a phenomenon that when a parent has height (say) away from the average height in the population, the offspring tends to have height closer to the average height (i.e., a regression towards the mean). This tendency was confirmed by regression analysis.

Illustration: Our estimated population mean (for mothers) is $\hat{\mu}_x = \bar{x} = 166.9$. Consider a mother with height 172cm (5 cm above population mean). The mean height for daughters of such mothers is estimated as

$\hat{\mu}(172) = 109.86 + (0.3458) \cdot 172 = 169.4$ (i.e. 2.5 cm from the population mean)

A mother being 162cm gives

$\hat{\mu}(162) = 109.86 + (0.3458) \cdot 162 = 165.9$ (i.e. 1.9 cm from the population mean)

Example of zero correlation when *X* and *Y* are dependent.

Let X, U be independent rv's where $X \sim N(0,1)$ and $U \sim \text{uniform}[-0.5,\ 0.5]$.
Let $Y = X^2 + U$
100 simulated observations of (X, Y), generated and plotted by Stata are

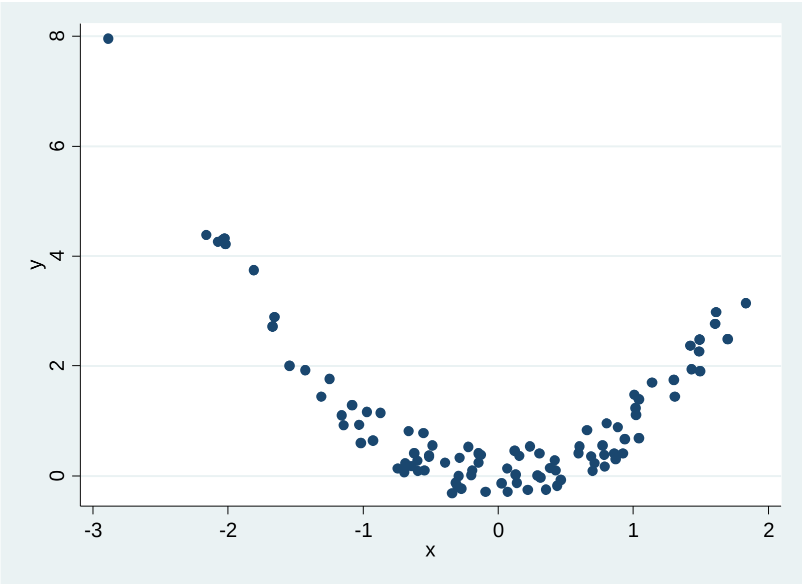
```
. set obs 100
obs was 0, now 100

. gen u=runiform()-.5          // runiform() generates uniform over (0, 1)

. gen x=rnormal()                    // rnormal() generates N(0,1)
observations

. gen y=x^2+u

. scatter y x
```



The plot shows strong dependence between *X* and *Y*.

Regression of Y w.r.t. *X*:

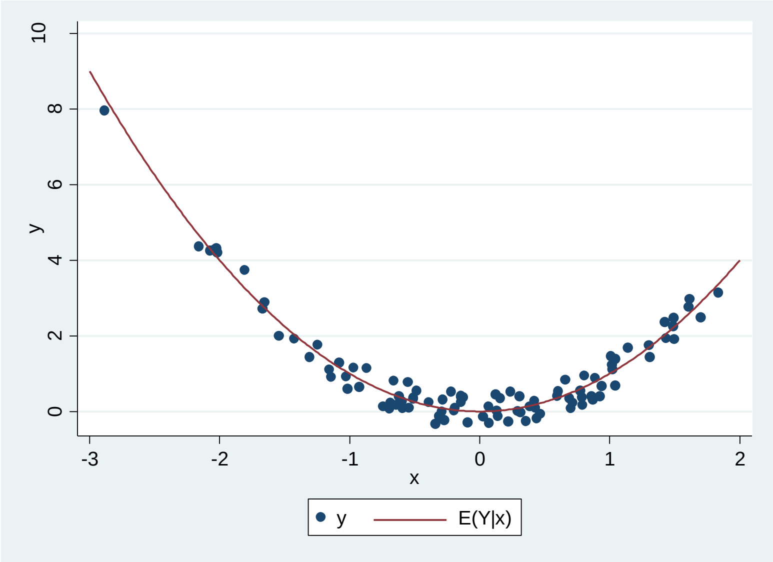
$$\mu(x) = E(Y \mid x) = E\left(X^2 + U \mid X = x\right) = E\left(x^2 + U \mid X = x\right) \overset{x \text{ fixed}}{=} x^2 + E(U \mid x)$$

but U, X independent $\Rightarrow E(U \mid x) = E(U) = 0$ and $\text{var}(U \mid x) = \text{var}(U) = \dfrac{1}{12}$
(remember that independence $\Rightarrow f(u \mid x) = f_U(u)$). Hence the regression function becomes

$$\mu(x) = E(Y \mid x) = x^2 + E(U \mid x) = x^2$$

Scatterplot with regression function

Stata command: `twoway (scatter y x) (function y=x^2, range( -3 2))`



$$\operatorname{var}(Y\mid x)=\operatorname{var}(X^2+U\mid X=x)=\operatorname{var}(x^2+U\mid X=x)\overset{x\text{ fixed}}{=}\operatorname{var}(U\mid X=x)=\operatorname{var}(U)=\frac{1}{12}$$

showing that the variance function is constant, $\sigma^2(x)=\operatorname{var}(Y\mid x)=1/12$.
So the relation is non-linear (for the regression function) and homoscedastic.

However, the correlation between *X* and *Y* is zero (in spite of strong dependence)!
Proof:

The correlation coefficient is $\rho=\rho(X,Y)\overset{\text{def}}{=}\dfrac{\operatorname{cov}(X,Y)}{SD(X)\cdot SD(Y)}$, where the covariance becomes

$$\operatorname{cov}(X,Y)\overset{\text{def}}{=}E\left[(X-E(X))(Y-E(Y))\right]=E(XY)-E(X)E(Y)=E(XY)$$

since $X\sim N(0,1)\quad\Rightarrow E(X)=0$

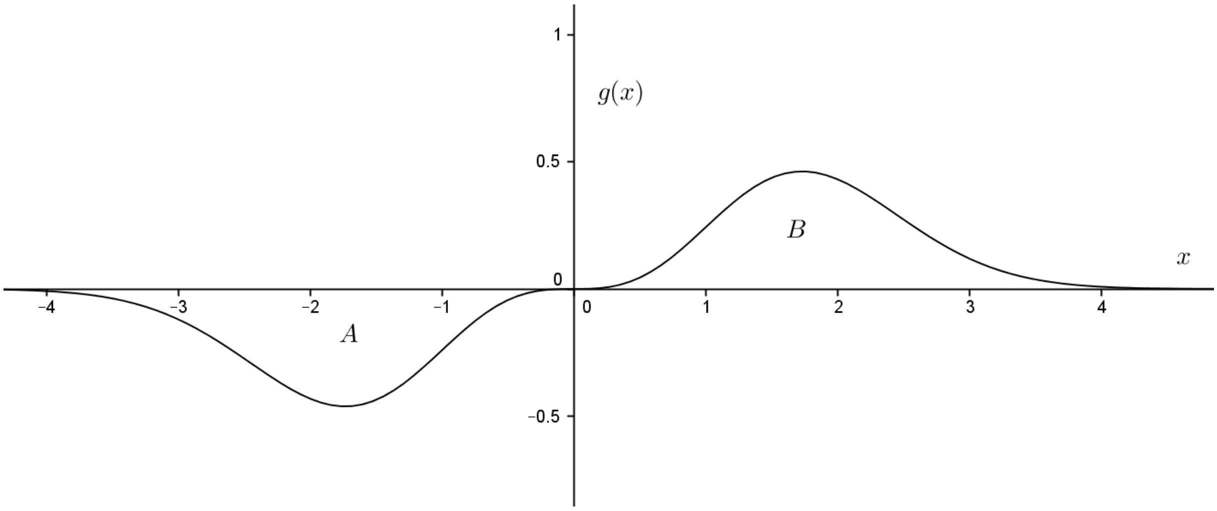
Now

$$E(XY)=E(X^3+XU)=E(X^3)+E(XU)\overset{X,U\text{ independent}}{=}E(X^3)+E(X)E(U)=E(X^3)$$

We have

$$E(X^3)=\int_{-\infty}^{\infty}x^3\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\int_{-\infty}^{\infty}g(x)dx$$, where we have called the integrand $g(x)$.

We see that $g(x)$ is symmetric about the origin (i.e., $g(-x)=-g(x)$ for all *x*.



implying that the area $A = -B$, and therefore, the total area $\int_{-\infty}^{\infty} g(x)dx = A + B = 0$.

Hence, $\operatorname{cov}(X,Y) = E(X^3) = 0$, and, therefore, also the correlation = 0. (**End of proof**).

[**Proof that** $E(XU) = (EX)(EU)$ **when *X* and *U* are independent:**
Suppose (X,U) has joint pdf $f(x,u)$ and marginal pdfs $f_X(x)$ and $f_U(u)$.
Independence implies $f(x,u) = f_X(x) f_U(u)$.

To find $E(XU)$, we need the rule given in **Theorem B (Rice p. 123)**, that implies that, if $h(x,u)$ is an arbitrary function, then

$$E\left[h(X,U)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,u)f(x,u)dxdu \quad \text{(whenever the integral exists)}$$

Hence

$$E(XU) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xu\ f(x,u)dxdu = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} xu\ f_X(x)f_U(u)dx\right]du = \int_{-\infty}^{\infty} uf_U(u)\left[\int_{-\infty}^{\infty} x\ f_X(x)dx\right]du =$$

$$= \int_{-\infty}^{\infty} uf_U(u)E(X)du \overset{E(X)\text{ constant}}{=} E(X)\int_{-\infty}^{\infty} uf_U(u)du = E(X)E(U)$$

(**End of proof**)]