Multivariate and bivariate normal distributions

The multivariate normal distribution with mean vector μ and covariance matrix Σ is a distribution such that if $Y \sim N(\mu, \Sigma)$, then

- $\mathbb{E}\, Y_i = \mu_i$
- $\mathrm{Cov}(Y_i, Y_j) = \Sigma_{ij}$
- All linear combinations of the Y_i have a normal distribution.

All conditional and marginal distributions of a multivariate normal are also multivariate normal.

PDF is $f_X(x_1, \dots, x_n) = \dfrac{1}{\sqrt{(2\pi)^k \det \Sigma}} \cdot \exp\left[\dfrac{-1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right]$.

If $n=1$ then $\det \Sigma = \Sigma_{11} = \sigma_1^2$; and we recover the usual normal distribution.

If $n=2$ we get the <u>bivariate</u> normal distribution, which can be written using correlation:

$$\mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$$

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right]$$

Theorem If X is bivariate normal, then $\rho = 0$ is equivalent to independence of the components.

The bivariate condition here is important: there are distributions with normal marginals and $\rho = 0$ but without independence.

Proof With $\rho = 0$, the PDF reduces to:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left[\frac{-1}{2}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right] =$$

$$= \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left[\frac{-1}{2\sigma_X^2}(x-\mu_X)^2\right] \cdot \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left[\frac{-1}{2\sigma_Y^2}(y-\mu_Y)^2\right] =$$

$$= f_X(x) \cdot f_Y(y).$$

$\blacksquare$

Random predictors

So far we have been treating the x_i as deterministic: under the researcher's control.

We could study a setup where X too is random, and where we use

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

to signify

$$\mathbb{E}[Y \mid X = x] = \beta_0 + \beta_1 x$$

Everything gets easier if we assume (X, Y) to be bivariate normal, with some correlation ρ.
Under this assumption, the conditional expectation works out to

$$\mathbb{E}[Y \mid X = x] = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X) =$$

$$= \rho \frac{\sigma_Y}{\sigma_X} \cdot x + \mu_Y - \rho \frac{\sigma_Y}{\sigma_X} \mu_X = \beta_1 x + \beta_0$$

where $\beta_1 = \rho \frac{\sigma_Y}{\sigma_X}$ and $\beta_0 = \mu_Y - \beta_1 \mu_X$, mirroring our development with deterministic x.
Because of this, everything we have done with $\hat\beta_0, \hat\beta_1$ carries over as-is.

Inference on correlations

It might be of interest not to study a predictive relationship between X and Y, but to focus on their independence.
For a bivariate case, independence $\Leftrightarrow \rho = 0$.
So hypothesis testing on ρ is important.

Since $\beta_1 = \frac{\sigma_Y}{\sigma_X}\rho$ it follows that $\rho = \frac{\sigma_X}{\sigma_Y}\beta_1 = \beta_1\sqrt{\frac{\sigma_X^2}{\sigma_Y^2}}$.

By the method of moments, and using $\hat{\beta}_1$ to estimate β_1, we get an estimator r of ρ:

$$r = \hat{\beta}_1\sqrt{\frac{S_X^2}{S_Y^2}} = \frac{S_{XY}}{S_{XX}}\sqrt{\frac{S_{XX}/n}{S_{YY}/n}} = \frac{S_{XY}}{S_{XX}}\sqrt{\frac{S_{XX}}{S_{YY}}} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}.$$

This turns out to also be the maximum likelihood estimator.

Since $\beta_1 = \frac{\sigma_Y}{\sigma_X}\rho$, it follows that β_1 and ρ have the same sign. So testing eg. for $H_A: \rho > 0$ is equivalent to testing for $H_A: \beta_1 > 0$.

For tests of β_1 we worked out a T-statistic of

$$T = \frac{\hat{\beta}_1 - 0}{S/\sqrt{S_{XX}}} \sim t(n-2) \text{ under } H_0: \beta_1 = 0.$$

Recall that we found a formula for SSE:

$SSE = S_{yy} - \hat{\beta}_1 S_{xy}$ and $S^2 = SSE/(n-2)$.

Now

$$T = \frac{\hat{\beta}_1}{\sqrt{S^2/S_{xx}}} = \frac{\hat{\beta}_1 \sqrt{n-2} \cdot \sqrt{S_{xx}}}{\sqrt{(S_{yy} - \hat{\beta}_1 S_{xy})}} = \frac{\hat{\beta}_1 \sqrt{S_{xx}/S_{yy}} \sqrt{n-2}}{\sqrt{1 - \hat{\beta}_1 S_{xy}/S_{yy}}} =$$

$$= \frac{r\sqrt{n-2}}{\sqrt{1 - \frac{S_{xy}}{S_{xx}} \cdot \frac{S_{xy}}{S_{yy}}}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

So $\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t(n-2)$ and can be used to create a t-test for ρ.

r^2 - the coefficient of determination

Recall $\hat{\beta}_1 = S_{xy}/S_{xx}$, and $SSE = S_{yy} - \hat{\beta}_1 S_{xy}$.

Now,

$\hat{\beta}_1 = S_{xy} \cdot$ something positive

$r = S_{xy}/\sqrt{S_{xx} \cdot S_{yy}} = S_{xy} \cdot$ something positive

$S_{xy} = S_{xy} \cdot$ something positive

Hence all three share the same sign.

In particular, $\hat{\beta}_1 \cdot S_{xy} \geq 0$ so $SSE \leq S_{yy}$.

$S_{yy} = \sum (y_i - \bar{y})^2$ measures variance of the y_i ignoring the existence of the x_i, while $SSE = \sum (y_i - \hat{y}_i)^2$ measures the variance that remains after the linear regression: the amount left unexplained by the regression.

Now, $r^2 = \left(\frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}\right)^2 = \frac{S_{xy}}{S_{xx}} \cdot \frac{S_{xy}}{S_{yy}} = \frac{\hat{\beta}_1 S_{xy}}{S_{yy}} = \frac{S_{yy} - SSE}{S_{yy}} =$

$= 1 - SSE/S_{yy}$ so r^2 measures the proportion of the total variance measured by S_{yy} that is <u>explained</u> by the linear regression itself.