Correlation from hierarchical models

1 Correlated data

In the supplementary material from Module 2, we introduced covariance and correlation. Recall that we can define the covariance between two random variables as

$$\sigma_{xy} = \mathrm{Cov}(X, Y) = \mathrm{E}[(X - \mu_x)(Y - \mu_y)]$$

where $\mu_x = \mathrm{E}(X)$ and $\mu_y = \mathrm{E}(Y)$. Correlation between X and Y is defined as

$$\rho_{xy} = \mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \cdot \sigma_y^2}} \, .$$

Correlation measures the strength of *linear* relationship between two variables. Covariance has a useful mathematical property which we will use below. If a, b, c, and d are constants, then

$$\begin{aligned} \mathrm{Cov}(a + bX, c + dY) &= \mathrm{Cov}(a, c) + \mathrm{Cov}(a, dY) + \mathrm{Cov}(bX, c) + \mathrm{Cov}(bX, dY) \\ &= 0 + 0 + 0 + b \cdot \mathrm{Cov}(X, Y) \cdot d \\ &= b \cdot d \cdot \sigma_{xy} \, , \end{aligned}$$

where the 0 terms are due to the fact that constants do not co-vary with anything.

In the examples from this lesson, we used hierarchical models when the data were grouped in some way, so that two observations from the same group were assumed to be more similar than two observations from different groups. We would therefore expect two observations from the same group to be correlated. It turns out that hierarchical models do correlate such variables, as we will demonstrate with a normal hierarchical model.

2 Normal hierarchical model

Suppose our data come from a normal distribution, where each group has its own mean. In the second stage, we assume that all the group means come from a common normal distribution. Let's write this hierarchical model. Let $y_{i,j}$ denote the ith observation in group

j, with mean θ_j. We have

$$\begin{aligned} y_{i,j} \mid \theta_j &\overset{\text{ind.}}{\sim} \mathrm{N}(\theta_j, \sigma^2) \\ \theta_j &\overset{\text{iid}}{\sim} \mathrm{N}(\mu, \tau^2) \end{aligned}$$

where σ^2, τ^2, and μ are known constants. To get the marginal distribution of $y_{i,j}$ only, we must compute

$$\begin{aligned} p(y_{i,j}) &= \int p(y_{i,j}, \theta_j) d\theta_j \\ &= \int p(y_{i,j}|\theta_j) p(\theta_j) d\theta_j \,. \end{aligned}$$

With normally distributed variables, it is often easier to work with the following equivalent formulation of the model

$$\begin{aligned} \theta_j &= \mu + \nu_j \,, \quad \nu_j \overset{\text{iid}}{\sim} \mathrm{N}(0, \tau^2) \\ y_{i,j} &= \theta_j + \epsilon_{i,j} \,, \quad \epsilon_{i,j} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2) \end{aligned}$$

with all ν_j and $\epsilon_{i,j}$ independent. This allows us to substitute θ_j into the expression for $y_{i,j}$ to get

$$y_{i,j} = \mu + \nu_j + \epsilon_{i,j} \,.$$

One nice property of normal random variables is that if X and Y are both normally distributed (correlated or uncorrelated), then the new variable $X + Y$ will also be normally distributed. Hence $p(y_{i,j})$ is a normal distribution with mean

$$\begin{aligned} \mathrm{E}(y_{i,j}) &= \mathrm{E}(\mu + \nu_j + \epsilon_{i,j}) \\ &= \mathrm{E}(\mu) + \mathrm{E}(\nu_j) + \mathrm{E}(\epsilon_{i,j}) \\ &= \mu + 0 + 0 \end{aligned}$$

and variance

$$\begin{aligned} \mathrm{Var}(y_{i,j}) &= \mathrm{Var}(\mu + \nu_j + \epsilon_{i,j}) \\ &= \mathrm{Cov}(\mu + \nu_j + \epsilon_{i,j},\ \mu + \nu_j + \epsilon_{i,j}) \\ &= \mathrm{Cov}(\mu, \mu) + \mathrm{Cov}(\nu_j, \nu_j) + \mathrm{Cov}(\epsilon_{i,j}, \epsilon_{i,j}) + 2 \cdot \mathrm{Cov}(\mu, \nu_j) + 2 \cdot \mathrm{Cov}(\mu, \epsilon_{i,j}) + 2 \cdot \mathrm{Cov}(\nu_j, \epsilon_{i,j}) \\ &= 0 + \mathrm{Var}(\nu_j) + \mathrm{Var}(\epsilon_{i,j}) + 0 + 0 + 0 \quad \text{(since } \nu_j \text{ and } \epsilon_{i,j} \text{ are independent)} \\ &= \tau^2 + \sigma^2 \,. \end{aligned}$$

Now, we want to show that observations in the same group are correlated under this hierarchical model. Let's take, for example, observations 1 and 2 from group j, $y_{1,j}$ and $y_{2,j}$. It does not matter which two you select, as long as they are from the same group. We know that $\mathrm{Var}(y_{1,j}) = \mathrm{Var}(y_{2,j}) = \tau^2 + \sigma^2$. What about their covariance?

$$\begin{aligned}
\mathrm{Cov}(y_{1,j}, y_{2,j}) &= \mathrm{Cov}(\mu + \nu_j + \epsilon_{2,j},\ \mu + \nu_j + \epsilon_{2,j}) \\
&= \mathrm{Cov}(\mu, \mu) + \mathrm{Cov}(\nu_j, \nu_j) + \mathrm{Cov}(\epsilon_{1,j}, \epsilon_{2,j}) + 2 \cdot \mathrm{Cov}(\mu, \nu_j) + \\
&\quad + \mathrm{Cov}(\mu, \epsilon_{1,j}) + \mathrm{Cov}(\mu, \epsilon_{2,j}) + \mathrm{Cov}(\nu_j, \epsilon_{1,j}) + \mathrm{Cov}(\nu_j, \epsilon_{2,j}) \\
&= 0 + \mathrm{Var}(\nu_j) + 0 + 2 \cdot 0 + 0 + 0 + 0 + 0 \quad \text{(since } \epsilon_{1,j} \text{ and } \epsilon_{2,j} \text{ are independent)} \\
&= \tau^2\,,
\end{aligned}$$

which gives us correlation

$$\begin{aligned}
\mathrm{Cor}(y_{1,j}, y_{2,j}) &= \frac{\mathrm{Cov}(y_{1,j}, y_{2,j})}{\sqrt{\mathrm{Var}(y_{1,j}) \cdot \mathrm{Var}(y_{2,j})}} \\
&= \frac{\tau^2}{\sqrt{(\tau^2 + \sigma^2) \cdot (\tau^2 + \sigma^2)}} \\
&= \frac{\tau^2}{\tau^2 + \sigma^2}\,.
\end{aligned}$$

Finally, let's check the covariance between observations in different groups. Let's take observation i from groups 1 and 2 (again, our choices do not matter), $y_{i,1}$ and $y_{i,2}$. Their covariance is

$$\begin{aligned}
\mathrm{Cov}(y_{i,1}, y_{i,2}) &= \mathrm{Cov}(\mu + \nu_1 + \epsilon_{i,1},\ \mu + \nu_2 + \epsilon_{i,2}) \\
&= \mathrm{Cov}(\mu, \mu) + \mathrm{Cov}(\nu_1, \nu_2) + \mathrm{Cov}(\epsilon_{i,1}, \epsilon_{i,2}) + \\
&\quad + \mathrm{Cov}(\mu, \nu_1) + \mathrm{Cov}(\mu, \nu_2) + \mathrm{Cov}(\mu, \epsilon_{i,1}) + \mathrm{Cov}(\mu, \epsilon_{i,2}) + \\
&\quad + \mathrm{Cov}(\nu_1, \epsilon_{i,1}) + \mathrm{Cov}(\nu_2, \epsilon_{i,2}) \\
&= 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 \quad \text{(since all } \epsilon \text{ and } \nu \text{ variables are independent)} \\
&= 0\,,
\end{aligned}$$

which obviously yields correlation 0.

Thus, we have shown that observations in the same group are correlated and observations in different groups are uncorrelated in the marginal distribution for observations.