1. Expectations (written): Suppose data are sampled from a population that can be described with the following model:

$$y_{i,j} = \mu + a_j + \epsilon_{i,j}$$
$$\{\epsilon_{i,j}\} \sim \text{ iid } N(0, \sigma^2), \quad i = 1, \ldots, n, \; j = 1, \ldots, m,$$

where $\sum_j a_j = 0$. Compute the following quantities, showing your steps:

   (a) $\mathrm{E}[y_{i,j}]$, $\mathrm{E}[\bar{y}_j]$, $\mathrm{E}[\bar{y}_{\cdot\cdot}]$

   (b) $\mathrm{Var}[y_{i,j}]$, $\mathrm{Var}[\bar{y}_j]$, $\mathrm{Var}[\bar{y}_{\cdot\cdot}]$.

   Describe in words the effect of n and m on the variances, and how they relate to how precisely you can estimate μ and the a_j's. (For this problem, you can use the facts that for two independent random variables x and y, $\mathrm{E}[ax + by] = a\mathrm{E}[x] + b\mathrm{E}[y]$ and $\mathrm{Var}[ax + by] = a^2\mathrm{Var}[x] + b^2\mathrm{Var}[y]$).

2. ANOVA (Rmd): Read in the `nels_math_ses` datafile from the course data directory using the `dget` command. For this exercise you will analyze $y_{i,j} =$ mathscore of student i in school j.

   (a) Compute the grand mean and the group means. Report the grand mean and the variance and standard deviation of the group means around the grand mean. Make a histogram of the group means, and re-create the dotplot from the slides `l2.pdf`.

   (b) Construct an ANOVA table for these data. Compare MSG to MSE and explain in words what the F-statistic is telling you about across-group heterogeneity. Obtain the p-value from the F-test for across group heterogeneity. What are your conclusions?

   (c) For each school, compute the mean of `mathdeg` and `ses`. Plot the school-specific `mathscore` means versus these two variables (use boxplots for `mathdeg`). Describe what you see and conjecture about sources of heterogeneity for the school-specific means.

3. Covariance (written): Reconsider the model above where $y_{i,j} = \mu + a_j + \epsilon_{i,j}$, but now assume that the groups are sampled, and that $a_1, \ldots, a_m \sim \text{iid } N(0, \tau^2)$. In this model, the a_j's are mean zero, uncorrelated and have variance τ^2. Compute the following:

   (a) $\mathrm{E}[y_{i,j}]$, $\mathrm{Var}[y_{i,j}]$.

   (b) $\mathrm{Cov}[y_{i_1,j_1}, y_{i_2,j_2}]$ and $\mathrm{Cov}[y_{i_1,j_1}, y_{i_2,j_1}]$. Note that in the groups are different in the first covariance but different in the second.

   (c) $\mathrm{Cor}[y_{i_1,j_1}, y_{i_2,j_1}]$. Explain in words how τ^2 relates to within-group correlation.

   Hint: Recall the covariance of two random variables x and z is given by $\mathrm{Cov}[x, z] = \mathrm{E}[(x - \mu_x) \times (z - \mu_z)]$ or equivalently, $\mathrm{Cov}[x, z] = \mathrm{E}[x \times z] - \mu_x \times \mu_z$. To compute the covariance in the problem, plug in $\mu + a_{j_1} + \epsilon_{i_1,j_1}$ for y_{i_1,j_1} and similarly for the other y's, and start computing expectations. Also recall that if a random variable x is mean-zero, then its variance is given by $\mathrm{E}[x^2]$, and that if two random variables x and z are mean zero and uncorrelated, then $\mathrm{E}[xz] = 0$.