Researchers were interested in how expression of a gene of interest differed between two different alleles (Allele 1 and Allele 2). Gene expression was estimated relative to a standard internal control. To control for the potential impact of *trans­*-acting regulators of the gene of interest, expression levels of Allele 1 and Allele 2 were compared in 3 different genetic backgrounds (A, B and C). Three separate biological replicates were performed, each of which was separated from the other by ~2 weeks. Additionally, three technical replicate measurements were obtained for all biological samples.

The provided table includes 4 variables: relative expression level (expression), genetic background (background), allele and replicate.

1. **What is the response variable? What type of variable is it?**

Answer: relative gene expression level is the response variable and is continuous numeric type of variable

1. **What is the focal explanatory variable to the hypothesis of interest?**

Answer: The focal explanatory variable is the allele for the hypothesis of interest.

1. **What are the potential covariates that should be considered in the model?**

Answer: The genetic background is the potential covariate that should be considered in the model.

1. **What is the relationship between each pair of explanatory variables (including the focal variable and any covariates)?**

Answer: The relationship between each pair of explanatory variables (allele, background and biological replicate) are Crossed

1. **Are there any random effects?**

Answer: Yes, the blocks (biological replicate).

1. **Provide a general expression for the full linear model for these data. (e.g. response ~ explanatory 1…).**

Answer: Relative expression ~ biological replicate + genetic background + allele + genetic background: allele

1. **Provide a specific linear expression for the best-fit linear model, retaining all fixed main effect terms but removing any interaction terms that are not significant. Note that to explore an interaction, both main effects must be** **treated as fixed effects. Note also that effect estimates are not generated for random effects, and are therefore not included in the linear expression. (e.g. response ~ 7.55 + 8.9\*allele 2…)**

Answer: Relative expression ~ 0.875 - 0.032 backgroundB - 0.129 Background C + 0.196 allele2

Where, the intercept is allele1 background A

1. **To test the hypothesis that the two alleles differ in their expression, after accounting for the effects of biological replicate and genetic background, what type of ANOVA should you use and why?**

Answer: Type I Anova since, it allows us to detect differences in expression between the two alleles after accounting for genetic background and biological replicate.

1. **Test the hypothesis from 7 and describe you results in 1 sentence.**

Answer: A model that included just the fixed effects explained significantly more variation than the model which included an interaction with allele and genetic background (-2ΔlnL=5.149, *df*=2, P- value=0.07617). The effect of the allele in gene expression did not depend on the genetic background.

1. **Estimate the difference in expression between the two alleles with uncertainty and describe your results in 1 sentence.**

Answer: Allele 1 and allele 2 differed in their expression after controlling for differences in genetic background and biological replicates (*F1,51=*39.89*, p*=6.53e-08). On average, allele 2 had 0.196 higher expression level than allele 1(95% CI 0.134 to 0.258).

1. **Using ggplot 2 and following the best practice guidelines we discussed in class, make a graph that shows the relationship between expression, genetic background and allele.**

