Section 13, 12/3/19

Two-Way Layout

In a two-way layout, there are two factors, each at two or more levels. Additionally, there are repeated measurements for each factor combination.

Observations of of the kth measurement at factor A level i and factor B level j as $Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijk}$ where

- μ : the overall mean level
- α_i: the main effects of factor A
- β_j: the main effects of factor B
- δ_{ij}: the interaction effects between factors
- ε_{ijk}: normally distributed measurement error

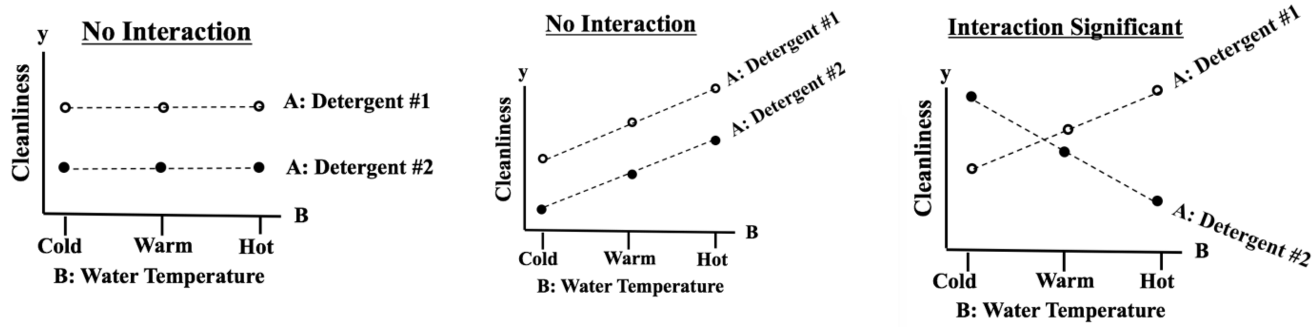
Alternatively, we can see $Y_{ijk} \sim N(\mu + \alpha_i + \beta_j + \delta_{ij}, \sigma^2)$ which is useful to think about when we sample from a model.

Deducing interaction effects

Let's assume that there are no interaction, i.e. $\delta_{ij} = 0$ for all i, j. Then comparing sample means across levels, between factors, we see that

$$E[\bar{Y}_{11\cdot} - \bar{Y}_{21\cdot}] = (\mu + \alpha_1 + \beta_1) - (\mu + \alpha_2 + \beta_1) = \alpha_1 - \alpha_2$$
$$E[\bar{Y}_{12\cdot} - \bar{Y}_{22\cdot}] = (\mu + \alpha_1 + \beta_2) - (\mu + \alpha_2 + \beta_2) = \alpha_1 - \alpha_2$$

That is, at each factor level the difference between the two factors should be approximately the same. If we plot the sample means across levels, parallel factor lines conform to the idea of no interactions.



Reproduced with permission of John Wiley and Sons
from the book Statistics from A to Z – Confusing Concepts Clarified

Comparing two independent samples

ex. (pg 427)

Consider $X_1, \ldots, X_n \sim N(\mu_1, \sigma^2)$ and $Y_1, \ldots, Y_m \sim N(\mu_2, \sigma^2)$ are each i.i.d. sets of samples and we wish to test the hypothesis

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

We will demonstrate that the test of these two composite hypotheses is equivalent to a generalized likelihood ratio test. Note the likelihood is

$$lik(\mu_1, \mu_2, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2}\frac{X_i - \mu_1^2}{\sigma^2}) \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2}\frac{Y_i - \mu_2^2}{\sigma^2})$$

Under the null, the means are equal and so the MLE of μ_1 and μ_2 are both $\mu_0 = \frac{1}{n+m}[\sum X_i + \sum Y_i]$ the weighted average of the two sample means. The MLE of the variance meanwhile is $\sigma_0^2 = \frac{1}{n+m}[\sum(X_i - \mu_0) + \sum(Y_j - \mu_0)]$

Under the alternative, the MLEs for μ_1 and μ_2 are simply the sample means. The MLE of the variance meanwhile is $\sigma_1^2 = \frac{1}{n+m}[\sum(X_i - \bar{X}) + \sum(Y_j - \bar{Y})]$.

If we take the ratios of the likelihoods under these two and take the log, we find the log of the likelihood ratio to be

$$\frac{m+n}{2} log(\frac{\sigma_1^2}{\sigma_0^2})$$

And so we reject for large values of σ_1^2/σ_0^2. We can simply this numerator using algebraic tricks to find that the test rejects for large values of

$$1 + \frac{mn}{m+n}\left(\frac{(\bar{X} - \bar{Y})^2}{\sum(X_i - \bar{X})^2 + \sum(Y_j - \bar{Y})^2}\right)$$

Note that the variable term, apart from positive constants which are independent of the data, is the square of a t-distribution, equivalently an f-distribution. Since we can scale by such constants arbitrarily, the likelihood ratio test is equivalent to a t-distribution.