

## Protocol for populating the data sheet for generic script

The script identifies and selects variables from the data sheet using keywords. Therefore the column names should contain specific words as specified in the script. Here are the column descriptions and naming suggestions.

### Required Columns

#### 1. ID Column

- Purpose: To identify which male each measurement belongs to (random effect in mixed models).
- Keyword required: "id"
- Example column names: Male\_ID, BirdID, ID
- Format: Categorical (string or numeric treated as factor)

### Response Variable Column (Dependent Variable)

#### 2. Sperm Motility Column

- Purpose: The main response variable measured in every dataset.
- Keyword required: "motility"
- Example column names: Sperm\_motility, Mean\_motility
- Format: Numeric
- Typically a percentage, a proportion, or VCL-derived motility score.

### Region of Interest(Optional)

These should take 3 columns (1, 2, 3), whose average per row gives the mean sperm motility (response variable).

### Predictor (Independent) Variables

#### 3. Age Variable (Optional)

- Purpose: Age of the male animal, sperm, or the seminal fluid.
- Keyword required: "age"
- Example column names: Male\_age, SF\_age, Age\_days
- Format: Numeric (years, days, etc.)

#### 4. Treatment Variable (Optional)

- Purpose: Categorical treatments such as seminal fluid presence/absence or combination of treatments.
- Keyword required: "treatment"
- Example column names: Treatments, treatment combination
- Format: Categorical (presence, absence, with\_SF, without\_SF, etc.)

#### 5. Time Variable (Optional)

- Purpose: Time points of measurement.(usually in sequence eg. 0, 2, 4, 6(min))
- Keyword required: "time"
- Example column names: time\_points, time\_factor, measurement\_time (min)
- Format: Numeric (minutes, seconds, etc.)

#### 6. Temperature Variable (Optional)

- Purpose: Temperature of medium or environment.
- Keyword required: "temperature"
- Example column names: Temp, Medium\_temp
- Format: Numeric (°C) or categorical (low, high)

#### 7. Medium Variable (Optional)

- Purpose: The medium in which sperm motility is tested.
- Keyword required: "medium"
- Example column names: Medium\_type, medium
- Categorical (e.g., Ringer, PBS, Seminal fluid, etc.)

#### 8. Species Variable (Optional)

- Purpose: Species of subject being tested.
- Keyword required: "specie"
- Example column names: Species, Bird\_specie,
- Format: Categorical (e.g., Flycatcher, Redstart)