Assignment 3: Instructions

SPEX450

Data processing and analysis module

due: 17 June

Overview

You are planning a large intervention study on swimming performance with the goal of increasing athletes' VO_2max. You need to process and perform preliminary analysis of the pilot dataset collected in the swimming flume using the gas analysis system (COSMED, Quark CPET; Rome, Italy). In pilot testing, participants swam for various lengths of time attempting to reach their VO_2max. The raw data files are stored as .xlsx files in the 'raw' subfolder of the 'data' folder.

Desired output

Your research team plans to conduct an intervention study. Your main objective with the pilot data is to determine the sample size needed to be able to detect a 5% improvement in VO_2max, given the distribution of your pilot data, with statistical power of 0.8. You will use a paired-sample t-test to test the null hypothesis of no change post-intervention.

Produce a report outlining the main findings of the analysis, including your processing methods, presentation of the pilot data and suggestion for sample size for the fully powered study. Include the general processing methods used in your report and provide a well-documented MATLAB script as a supplementary file. For data presentation, produce scatter plots of VO_2max vs age, BMI and body surface area. Also report descriptive statistics of your participants: number of males and females, $M \pm SD$ for age, height, weight and body surface area.

Assessment details

This assignment must be submitted electronically before the end of the exam period (5 pm on 17 June, 2020) and is worth 45% of your mark for the module. The mark breakdown for the assignment is detailed below with marks out of 100 in parentheses.

1. Create a processing and analysis script that provides all the output needed for your report.

   Your script should do the following:

   · Clear the workspace and close any open figure windows (1)

   · Create a list of the files in the 'raw' folder using the **dir** function. It should also use the **fullfile** function to create file and folder paths so the script runs on Windows or OSX operating systems (2)

   · Initialise the required storage variables using the **nan** function (2)

   · Set up a loop to step through each file (2)

   · Use the **xlsread** function to read in each file (4)

   · Extract age, sex, BMI, weight, height, body surface area as metadata. Sex should be coded as 1 for male and 2 for female (or vice versa) (8)

   · Calculate VO_2max as the dependent variable (see details for calculation below) (14)

   · Add metadata and dependent variable to storage variables (5)

   · Create appropriate graphics files to visualise your data (9)

   · Calculate descriptive statistics (5)

   · Calculate the required sample size as outlined above. Include whether the required sample size changes substantially if you were to study only males or only females (8)

   · Be well documented with comments (10)

2. Provide a report as outlined above including appropriate figures, tables, etc into a document with statistics reported according to APA format. (30)

VO$_2$max calculation

There is unfortunately little agreement in the literature about the specific criteria for calculating VO_2max (see, for example, Robergs et al., 2010). Use the following convention to calculate VO_2max:

· Find the max value of the "VO2/Kg" column in the raw files

· Find the average for the surrounding 20 s. Since the measurements are taken every 10 s, use the average value of the max, the preceding value and the following value. A few trials, however, were recorded at sampling rates of one measurement every 20 s; see the section below for how to deal with this.

· For the VO_2max to be considered valid the respiratory exchange ratio (VCO_2/VO_2: "R" column in the raw files) must be greater than 1.0 at the time of VO_2max. If "R" is less than 1.0 the value for VO_2max should be assigned the value **NaN** (the literature actually uses a value of 1.1, but use 1.1 for the assignment to ensure enough data are included in the analysis).

Sampling rates

The trials should have all been sampled at one measurement every 10 s; however, a few were recorded with one measurement every 20 s – minor inconsistencies like this are part of research and can be dealt with by using crafty scripting. For the purposes of gaining preliminary results from a pilot study, you should identify the sampling rate (either 1/10 Hz or 1/20 Hz). In the case of 1/10 Hz sampling, calculate VO_2max as above. In the case of 1/20 Hz sampling you should calculate half the difference between the maximum value and the neighbouring values and add them to the neighbouring values to be used for calculation of the mean.

Mathematically, the three values used to calculate the mean are x_{t-1}, v_t, and x_{t+1}, where t is the frame number of the max value, v_t is the raw max value and x_{t-1} and x_{t+1} are the adjusted values of v_{t-1} and v_{t+1}, respectively:

$$x_{t-1} = v_{t-1} + \left(\frac{v_t - v_{t-1}}{2} \right)$$
$$x_{t+1} = v_{t+1} + \left(\frac{v_t - v_{t+1}}{2} \right).$$

References

Robergs, R. A., Dwyer, D., & Astorino, T. (2010). Recommendations for improved data processing from expired gas analysis indirect calorimetry. *Sports Medicine*, *40*(95–111).