ST300 Assessed Coursework
Body Measurements

1 Question

The data-set `body.dat` contains 21 body dimension measurements, together with age, weight, height, and gender on 507 individuals. The 247 men and 260 women were primarily individuals in their twenties and thirties, with a scattering of older men and women, all exercising several hours a week.

The main aim is to explain how the height of an individual is related to other variables. Using R (other packages not allowed), obtain a satisfactory model for height in terms of the other variables. You should examine outliers and influential observations, and try transformations. You may also want to try variable selections using techniques learnt and used in your lectures and workshops. Why is your model good? From your model, what can you say about the relationship between the height and other body measurements or characteristics of a particular individual? Which body measurement(s) are potentially more important in estimating a person's height? Write a summary of your analysis with **no more than 6 sides** of A4 (including graphs and figures).

Variable descriptions:

```
            Skeletal Measurements: (all in cm)

biacromial      Biacromial diameter
pelvic.breadth  Biiliac diameter, or "pelvic breadth"
bitrochanteric  Bitrochanteric diameter, in cm
chest.depth     Chest depth between spine and sternum at nipple level,
                mid-expiration
chest.diam      Chest diameter at nipple level, mid-expiration
elbow.diam      Elbow diameter, sum of two elbows
wrist.diam      Wrist diameter, sum of two wrists
knee.diam       Knee diameter, sum of two knees
ankle.diam      Ankle diameter, sum of two ankles
```

```
            Girth Measurements: (all in cm)

shoulder.girth  Shoulder girth over deltoid muscles
chest.girth     Chest girth, nipple line in males and just above breast
                tissue in females, mid-expiration
waist.girth     Waist girth, narrowest part of torso below the rib cage,
                average of contracted and relaxed position
navel.girth     Navel (or "Abdominal") girth at umbilicus and iliac crest,
                iliac crest as a landmark
hip.girth       Hip girth at level of bitrochanteric diameter
thigh.girth     Thigh girth below gluteal fold, average of right and left
                girths
bicep.girth     Bicep girth, flexed, average of right and left girths
forearm.girth   Forearm girth, extended, palm up, average of right and
                left girths
knee.girth      Knee girth over patella, slightly flexed position, average
                of right and left girths
calf.girth      Calf maximum girth, average of right and left girths
ankle.girth     Ankle minimum girth, average of right and left girths
wrist.girth     Wrist minimum girth, average of right and left girths

            Other Measurements:

age             Age (years)
weight          Weight (kg)
height          Height (cm)
gender          Gender (1 - male, 0 - female)
```

Note: Check all the covariates you are not familiar with before interpreting your model. Note that some covariates are discrete, so that they can be interpreted as factors. You are not taught about regression with a mixed of factors and continuous covariates (called ANCOVA), but using the `lm` function in R you can still get appropriate coefficients and anova function with sum of squares calculated etc. But remember your ultimate goal is to find a simple model for interpretation.

2 Assessment

This question counts for 15% of the total mark for ST300. The work will be subject to checks for plagiarism.

3 Submission

Your work should be submitted anonymously under your candidate number. You can look up your candidate number in LSE for you. A printed copy of the project must be put in the project drop-box in the waiting area of the 6th floor of Columbia House by 12:00pm on the 20th of March 2019. You must sign the plagiarism form and put it *separately* in the department office collection point of such form. In addition, you must upload your project using the specified Moodle link on the ST300 Moodle page by 12:00pm on the 20th of March 2019. The Moodle upload link will stop working after that time and you will only be able to email me your project. Penalties will apply for late submission of either the printed or the electronic version of your project (but will not be doubly-penalised if both are not submitted on time). For the plagiarism form see the link in Moodle before the links to the project.

When the submission deadline has passed, a list of candidate numbers for the work received will be put up on Moodle. Each student must check from that list that their work has been received, counting both the electronic and the printed versions.

Extensions to deadlines for coursework will only be given in fully documented serious extenuating circumstances. There will be $5n$ marks deducted from the total possible 100 marks for this coursework if either the hard copy or the electronic version is submitted between $24(n-1)$ and $24n$ hours from the stipulated deadline (including working days only).