MATH1324 Introduction to Statistics

Assignment 3

Predicting Body Fat Percentage

## OverviewWeight, Weigh, Coloring

The Body(2).csv dataset is posted on the Assignment 3 website. It contains the percentage of body fat, age, weight, height, and ten body circumference measurements (e.g., abdomen) for a sample of 252 men and women. Body fat, as measured using the Brozek method, is estimated through a time consuming underwater weighing technique to measure a person's density. Investigators want to determine if a general, easy to determine, body circumference measurement could be used as a general indicator for body fat percentage. If so, the investigators want to establish a formula that can convert a body circumference measurement to a predicted body fat percentage and also understand how well this prediction will hold. (original source: [JSE-DA](http://www.amstat.org/publications/jse/jse_data_archive.htm)). The variables are as follows:

**Variables**

1. **Case**: Case Number
2. **BFP\_Brozek**: Percent body fat using Brozek's equation, 457/Density - 414.2
3. **BFP\_Siri**: Percent body fat using Siri's equation, 495/Density - 450
4. **Density**: Density (gm/cm3)
5. **Age**: Age (yrs)
6. **Weight**: Weight (lbs)
7. **Height**: Height (inches)
8. **Adiposity\_index**: Adiposity index = Weight/Height2 (kg/m2)
9. **Fat\_free**: Fat Free Weight = (1 - fraction of body fat) \* Weight, using Brozek's formula (lbs)
10. **Neck**: Neck circumference (cm)
11. **Chest**: Chest circumference (cm)
12. **Abdomen**: Abdomen circumference (cm) "at the umbilicus and level with the iliac crest"
13. **Hip**: Hip circumference (cm)
14. **Thigh**: Thigh circumference (cm)
15. **Knee**: Knee circumference (cm)
16. **Ankle**: Ankle circumference (cm)
17. **Biceps**: Extended biceps circumference (cm)
18. **Forearm**: Forearm circumference (cm)
19. **Wrist**: Wrist circumference (cm) "distal to the styloid processes"
20. **Sex**: “1” for male and “0” for female

**The Assignment**

1. Test whether the mean body fat percentage for males and females are the same. Write down your null and alternative hypothesis clearly, express your conclusion in words and provide your reason(s) for your final conclusion.
2. Estimate the 99% confidence interval for the mean body fat percentage in the population. Is there any assumption for the body fat percentage distribution that we need to investigate? Explain your reason (s).
3. Researchers believe that average body fat percentage is less than 12.5. Test this claim. Write down your null and alternative hypothesis clearly, express your conclusion in words and provide reason(s) for your final conclusion.
4. Find the single best predictor of body fat percentage (Brozek method) using the body circumference data. Write a report that explains your method for identifying the single best predictor. Use the best predictor to determine a model that can convert a person’s body circumference measurement to an estimated body fat percentage. Ensure you test the model parameters and any assumptions. Critique the predictive ability of the model and draw an overall conclusion to help the investigators

## Report

Your report must be submitted as a reproducible R Markdown document with written sections, R code and output. You must create your own report. No template will be provided. The marking rubric will explain how each report will be graded.

**The report must be no longer than 10 pages.**

## Submission Instructions

The assignment 3 report must be submitted as an R Markdown report. Information for using the R Markdown package can be found [here](https://sites.google.com/a/rmit.edu.au/intro-to-stats/r-bootcamp/course-4-r-markdown). You might find it helpful to adapt the templates used in previous assignments.

This assignment is worth 30% and must be uploaded to the Assignment 3 Turnitin link by **03/06/2018**. Extensions will only be granted in accordance with the [RMIT University Extension and Special Consideration Policy](http://www1.rmit.edu.au/browse;ID=8r23xxpe4fne). No exceptions. Assignments submitted late will be penalised (see [Course Information](https://docs.google.com/a/rmit.edu.au/document/d/1qxGe7k5pj8kxHeAaUdecxTKJjaPZKpO4ZEtvO3WBf_M/edit?usp=sharing) for further details).

The report must be knitted and uploaded as either a **Word** document or **PDF**. There is a 10 page limit.

## Collaboration

You are permitted to discuss and collaborate on the assignment with your classmates. **However, the R coding, analysis and write-up of the report must be an individual effort.** Assignments will be submitted through Turnitin, so if you’ve copied from a fellow classmate, it will be detected. It is your responsibility to ensure you do not copy or do not allow another classmate to copy your work. If plagiarism is detected, both the copier and the student copied from will be responsible. It is good practice to never share assignment files with other students. You should ensure you understand your responsibilities by reading the RMIT University website on [academic integrity](http://www1.rmit.edu.au/browse;ID=kkc202lwe1yv). Ignorance is no excuse.

# Assignment 3 Marking Rubric

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Criteria** | **Not acceptable**  **(0)** | **Needs Improvement**  **(1)** | **Good**  **(2)** | **Excellent**  **(3)** |
| **Hypothesis testing for part 1 (10%)** | **No Explanation for the R code used** | **Explanation for the R code used is given but hypothesis tested are not mentioned** | **Explanation for the R code used is given, hypothesis tested are mentioned but reasons for the conclusion are not given** | **Explanation for the R code used is given, hypothesis tested are mentioned and final conclusion is given with the reasons** |
| **Confidence interval for part 2 (5 %)** | **No Explanation for the R code used** | **Explanation for the R code used is given but necessary assumptions for the used R code are not given.** | **Explanation for the R code used is given, necessary assumptions for the used R code are given but the final conclusions are not given in words** | **Explanation for the R code used is given, necessary assumptions for the used R code are given and the final conclusions are not given in words** |
| **Hypothesis testing for part 3 (5%)** | **No Explanation for the R code used** | **Explanation for the R code used is given but hypothesis tested are not mentioned** | **Explanation for the R code used is given, hypothesis tested are mentioned but reasons for the conclusion are not given** | **Explanation for the R code used is given, hypothesis tested are mentioned and final conclusion is given with the reasons** |
| **Best predictor**  **(15%)** | No method or justification was used to identify the best predictor. | An appropriate method to identify the best predictor was used, but no explanation or justification was provided. | An appropriate method to identify the best predictor was used, but needed stronger justification. | An appropriate method to identify the best predictor was used and well justified. |
| **Model Testing**  **(15%)** | A statistical model was fitted, however, the report fails to demonstrate insight into the model. | The correct statistical model was correctly fitted to the data but the various components of the model were not adequately tested and interpreted. | The correct statistical model was correctly fitted to the data and the various components of the model were tested and interpreted correctly. However, some areas needed improvement. | The correct statistical model was correctly fitted to the data and the various components of the model were tested and interpreted correctly. |
| **Assumptions**  **(10%)** | Assumptions were not stated or tested. | Some assumptions were stated and tested but lacked interpretation and insight into their importance. | All relevant assumptions were stated, tested and interpreted, but there were some elements in need of improvement. | All relevant assumptions were stated, tested and interpreted correctly. |
| **Formula**  **(5%)** | No model formula was reported. | The model formula was reported, but not interpreted well. | The model formula was reported and interpreted, but some minor issues were present. | The model formula was reported and described appropriately using the output produced |
| **Model critique and conclusion**  **(10%)** | The model was not critiqued and there was no general conclusion. | The model was poorly critiqued and a general conclusion was lacking. | The model was critiqued appropriately, using relevant statistics, but the overall conclusion did not synthesise the report effectively. | The report demonstrates a clear understanding of the strengths/limitations of the model by referring to relevant statistics and drawing an informed conclusion in context. |
| **Attention to detail**  **(25%)** | The report has numerous and significant issues with presentation. These might include poor structure, not in an R Markdown format, poor spelling, and readability. | The report needed greater attention to detail and suffered from a number of issues related to structure, spelling, formatting or grammar. | The report demonstrated good to attention to detail. Only a few minor issues with readability, formatting, spelling or grammar. | The report shows excellent attention to detail. Its polished, reads well, looks professional, is free from spelling and grammar errors. |