STAT 242 Final Exam Question 1

A (Science Fiction) Study of the Humanoid Inhabitants of the Moon Europa

The year is 2222. Humans have satellites in orbit around Europa, the smallest of Jupiter's four moons. Satellite images have revealed the presence of a small, humanoid race of beings living on the moon's surface. Due to their small stature, it is difficult to gather data on these beings. It also has been observed that each humanoid is accompanied by an animal companion which is much larger and is vibrantly colored (red, blue, green, or orange). The size and bright coloring of these creatures makes them much easier to identify. Because it is much easier to collect data on the animal companions, scientists are anxious to learn more about these animals and to determine whether they can use this information to learn about the humanoid inhabitants. They have enlisted you as the statistical expert on this project, and have put together a number of questions to which they need answers. They have managed to collect data on 200 distinct humanoids and their (200) companion animals, which they have compiled for you in the data set `europa_df`, which contains the following variables:

- `humanoid_height`: the height of the humanoid, measured in centimeters
- `pet_height`: the height of an animal companion, measured in centimeters
- `pet_color`: color of the pet; one of blue, red, green, or orange
- `pet_temp`: body surface temperature of the animal companion, measured in degrees Celsius

```
library(car)
library(dplyr)
library(ggplot2)
library(gmodels)
library(readr)

europa_df <- read_csv("https://marievozanne.github.io/europa_df.csv")
```

Scientists are intrigued by the vibrant colors of the animal companions. One has noticed that some animals seem to be smaller than others and suggests that color and height may be related. You have been tasked with investigating this claim.

(a) Create a plot to summarize the relationship between animal companion color and height.

(b) Fit a linear model to the data and print out the model summary.

(c) Write down the equation for the mean in terms of the linear model coefficients, β_0, β_1, β_2, β_3. Be sure to define this notation in words, too.

(d) Interpret each of these four parameters in the context of the problem.

(e) Find a confidence interval for the difference in mean height of the animal companion between populations of blue and green animals and discuss the meaning of this interval in context.

(f) Green companion animals seem to be taller, on average, than other companion animals. Conduct a test of this claim and interpret the results in the context of this problem.

(g) Are your results generalizable? If so, to what population? If not, why not?