Bonnie Turek

Eco 602 – Using Models 2

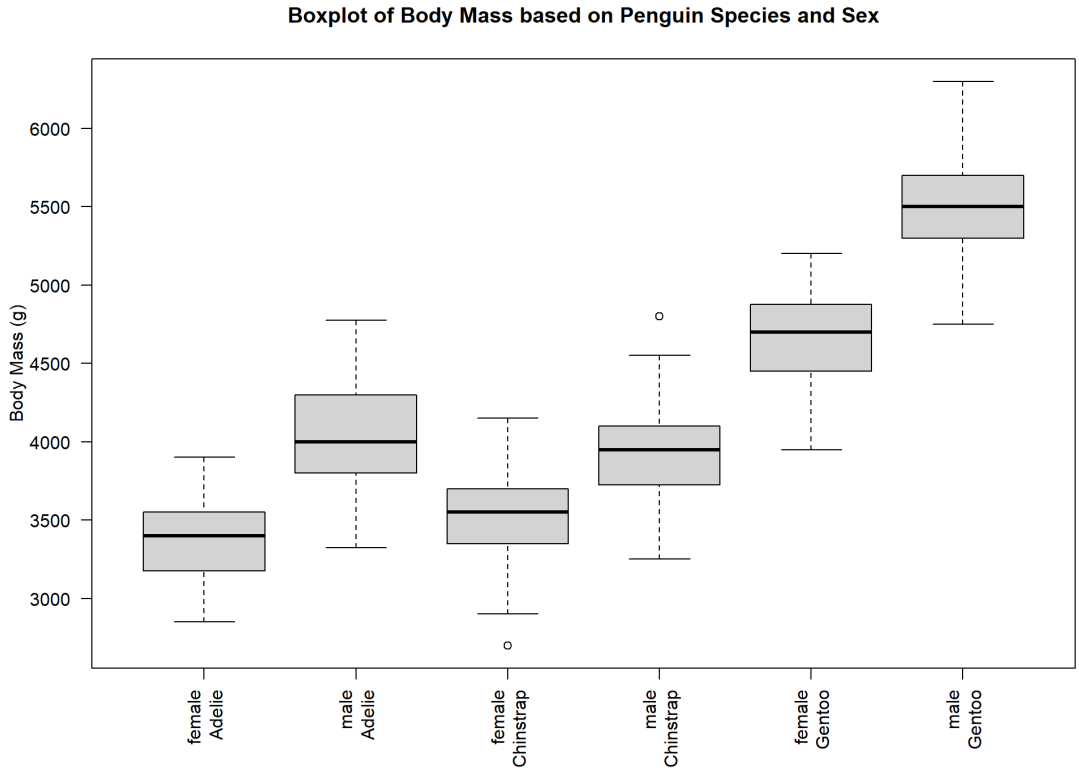
11/24/2021

Q1. Re-create the conditional boxplot of penguin body mass conditioned on sex and species.

Code:

```
boxplot(body_mass_g ~ sex * species,
    data = penguins,
    ylab = "Body Mass (g)",
    xlab = "",
    names = c("female\nAdelie", "male\nAdelie", "female\nChinstrap", "male\nChinstrap", "female\nGentoo", "male\nGentoo"),
    las = 2,
    main = "Boxplot of Body Mass based on Penguin Species and Sex")
```

Output:



Q2. Based on the boxplots, do you think male penguins (of any species) are significantly heavier than female penguins? Explain your reasoning.

Yes, it is pretty clear that the male penguins in each species are significantly heavier in body mass as compared to the female penguins of their same species. You can see these differences in the box plots above, specifically sectioned off by each species. The boxes for each species' male penguins are higher than the female species' boxes. It is not comparable however, if you were to look at the body mass differences of a male Adelie penguin to a female Gentoo penguin. This is because there are also differences in body mass attributed to penguin species, not only sex.

Q3. (Based on the boxplots only) Do you think adding sex to a model that already includes species will improve the model fit?

Yes, as I stated above, there are not only differences clearly visible in the boxplot with sex of penguins, but also with the species of penguin. You can compare the boxplot of body mass conditioned on only species to the boxplot conditioned on both species and sex, and see the significant variations observed when adding sex as a predictor. Both of these would serve well as predictor variables in a model. In fact, I believe a model would be seriously lacking if you did not include both predictors. They both contain a lot of valuable information about the data and the population.

Q4. Show the R-code you used to build fit_both.

```
fit_both = lm(formula = body_mass_g ~ sex * species, data = penguins)
```

Q5. What is the base case for the two-way model that includes sex and species?

With a two-way factorial model with interactions, the base case is now a combination of the base level of predictor one and the base level of predictor two. So in this case, the base case is **female, Adelie** penguins. We can see this below in the model coefficient table.

Fit_both Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	3368.84	36.21	93.030	< 2e-16	***
sexmale	674.66	51.21	13.174	< 2e-16	***
speciesChinstrap	158.37	64.24	2.465	0.01420	*
speciesGentoo	1310.91	54.42	24.088	< 2e-16	***
sexmale:speciesChinstrap	-262.89	90.85	-2.894	0.00406	**
sexmale:speciesGentoo	130.44	76.44	1.706	0.08886	.

Q6. What are the names of the two coefficients (from the first column of the coefficient table) you need to calculate the average mass of female Chinstrap penguins?

You would need the **(Intercept)** coefficient which is female Adelie penguins, and then you would need the **speciesChinstrap** coefficient in order to calculate the average body mass of female Chinstrap penguins.

Q7. What is the predicted average mass of female Chinstrap penguins in the interactive model?

3368.84 + (1*158.37) = **3,527.21 grams** is the average body mass of female Chinstrap penguins. We do not care about the other dummy variable coefficient terms since the intercept is female, and we only care about changing the species to Chinstrap.

Q8. What is the observed average mass of female Chinstrap penguins, calculated from the penguins data?

aggregate(body_mass_g ~ sex * species, data = penguins, FUN = mean)

	sex	species	body_mass_g
1	female	Adelie	3368.836
2	male	Adelie	4043.493
3	**female**	**Chinstrap**	**3527.206**
4	male	Chinstrap	3938.971
5	female	Gentoo	4679.741
6	male	Gentoo	5484.836

The observed average mass of female Chinstrap penguins calculated from the data is the same as I just calculated from the coefficient table, **3527.206 grams.**

*I worked independently on this assignment.