Solution for Tutorial 07 Manually Graded Questions

Question 1.9
{points: 1}

We can use the adjusted R^2 to compare the three models and conclude which one fits the data better. By looking at your metrics from **Question 1.8**, what model fits the data betters according to the adjusted R^2 and why?

> *Your answer goes here.*

BEGIN SOLUTION

The obtained `adj.r.squared` value suggests that `facebook_MLR_int` is the best model compared to `facebook_MLR_add` and `facebook_MLR_add_2`. This `adj.r.squared` is 0.802 which suggests that 80.2% of the adjusted variation in the `total_engagement_percentage` is explained by `facebook_MLR_int` (assuming normality of the error terms). Note that the `adj.r.squared` for `facebook_MLR_add` and `facebook_MLR_add_2` are 0.785 (78.5%) and 0.781 (78.1%) respectively, which are smaller that the one corresponding to `facebook_MLR_int`.

END SOLUTION

Question 1.10
{points: 1}

Do these three models fit the data better than the null model. Using your results from **Question 1.8** with a **significance level** $\alpha = 0.10$, provide the three corresponding statistical conclusions of these tests.

> *Your answer goes here.*

BEGIN SOLUTION

- `facebook_MLR_add`: the $p-\text{value} < 0.10$ which indicates that this model fits the data better than the null model.
- `facebook_MLR_add_2`: the $p-\text{value} < 0.10$ which indicates that this model fits the data better than the null model.
- `facebook_MLR_int`: the $p-\text{value} < 0.10$ which indicates that this model fits the data better than the null model.

END SOLUTION

Question 1.12
{points: 1}

Based in your results from **Question 1.11** with a **significance level** $\alpha = 0.10$:

- Does `facebook_MLR_add_2` fit the data better than `facebook_MLR_add`
- Comparing `facebook_MLR_add` vs `facebook_MLR_int`, does the inclusion of interaction terms in `facebook_MLR_int` improve the model's fit to the data?

> *Your answer goes here.*

BEGIN SOLUTION

The $p-\text{value} > 0.10$ in the F-test of `facebook_MLR_add` versus `facebook_MLR_add_2`, which indicates that the inclusion of `share_percentage` and `comment_percentage` does not improve the model fitting. Hence, we can proceed with another F-test of `facebook_MLR_add` versus `facebook_MLR_int` (whose $p-\text{value} < 0.10$) which indicates that the inclusion of interaction terms improves the model fitting.

END SOLUTION

Question 1.13
{points: 1}

In **Question 1.7** we noted that when you fit a model with interactions the results in `facebook_MLR_int_results` can not be used to examine if (overall) `total_engagement_percentage` and `page_engagement_percentage` are associated.

The t-tests in the `tidy()` table, compares this association for different types of posts (i.e., levels of the categorical variable `post_category`).

If you want to answer examine the *overall* association, on average over different posts and the effect of other variables, you need to compare `facebook_MLR_int` with a model that does not contain the variable `page_engagement_percentage`.

1.13.0 Use `lm` to write the code of the nested model needed to compare with `facebook_MLR_int`. Give an object name to this output to be used in the next questions.

1.13.1 Write the code run an appropriate F test to answer the question above?

> *Your answer goes here.*

BEGIN SOLUTION

1.13.0 We need to compare `facebook_MLR_int` with `facebook_add_post <- lm(total_engagement_percentage ~ post_category, data = facebook_sample)`. Note that the name given may differ.

1.13.1 `anova(facebook_add_post, facebook_MLR_int)`

END SOLUTION