0.0.1 Question 1c

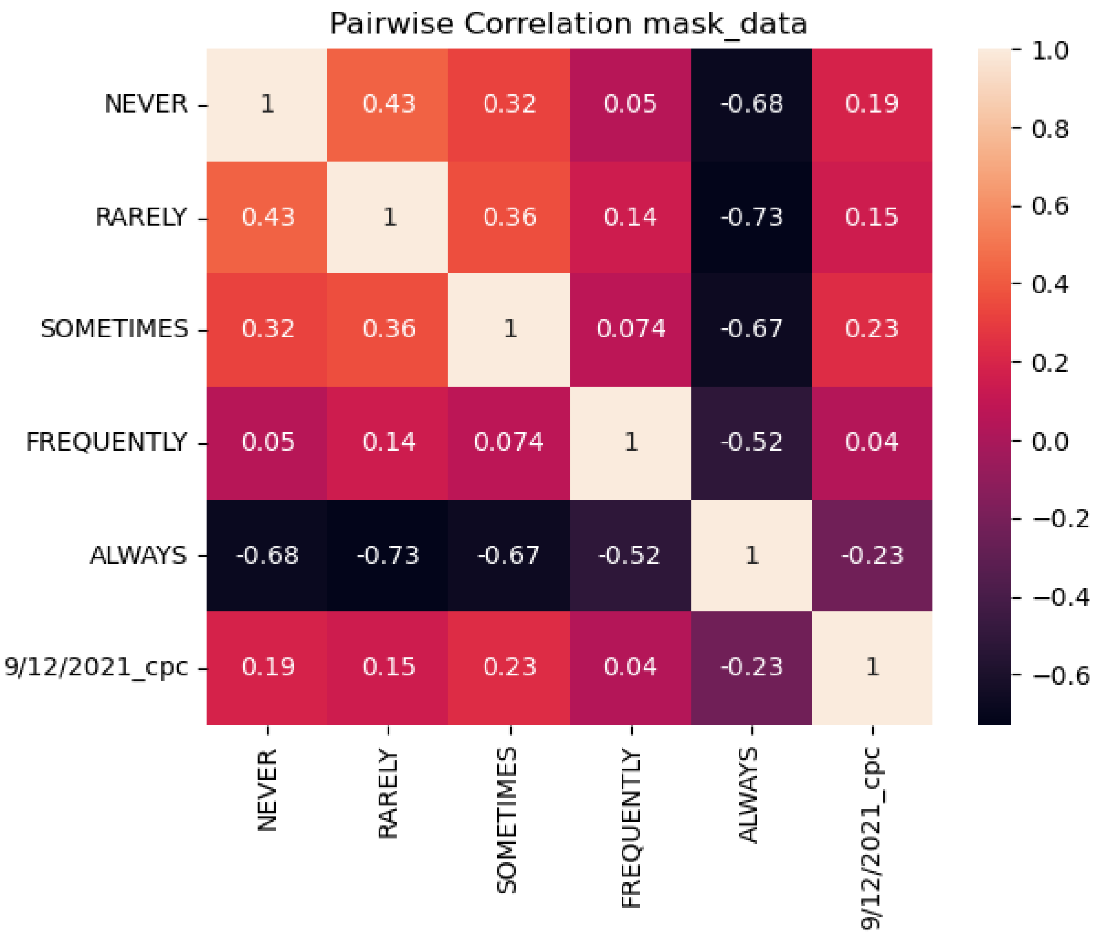
Our goal is to use county-wise mask usage data to predict the number of COVID-19 cases per capita on September 12th, 2021 (i.e., the column `9/12/2021_cpc`). But before modeling, let's do some EDA to explore the multicollinearality in these features, and then we will revisit this question in part 4.

Create a visualization that shows the pairwise correlation between each combination of columns in `mask_data`. For 2-D visualizations, consider Seaborn's heatmap. Remember to add a title to your plot.

Hint: You should be plotting 36 values corresponding to the pairwise correlations of the six columns in `mask_data`.

```
In [14]: pairwizeCorr_Mask_Data = mask_data.corr()
         sns.heatmap(pairwizeCorr_Mask_Data, annot=True)
         plt.title('Pairwise Correlation mask_data')
```

```
Out[14]: Text(0.5, 1.0, 'Pairwise Correlation mask_data')
```



Pairwise Correlation mask_data

	NEVER	RARELY	SOMETIMES	FREQUENTLY	ALWAYS	9/12/2021_cpc
NEVER	1	0.43	0.32	0.05	-0.68	0.19
RARELY	0.43	1	0.36	0.14	-0.73	0.15
SOMETIMES	0.32	0.36	1	0.074	-0.67	0.23
FREQUENTLY	0.05	0.14	0.074	1	-0.52	0.04
ALWAYS	-0.68	-0.73	-0.67	-0.52	1	-0.23
9/12/2021_cpc	0.19	0.15	0.23	0.04	-0.23	1

0.0.2 Question 1d

(1) Describe the trends and takeaways visible in the visualization of pairwise correlations you plotted in Question 1c. Specifically, how does the correlation between pairs of features (i.e. mask usage) look like? How does the correlation between mask usage and cases per capita look like?

(2) If we are going to build a linear regression model (with an intercept term) using all five mask usage columns as features, then what problem will we encounter?

(1)The highest pairewize correlation is diagonal from upper left to lower right. Which make sense because it pairwize with itself. Column "ALWAYS" have relatively heigh correlation to others. Pairewize columns "FREQUENTLY"–"NEVER" and "FREQUENTLY"–"9/12/2021_cpc" have very low correlation.

(2) Like what I mentioned on part (1), Column "ALWAYS" have relatively heigh correlation to others. This has potential linear dependency.When building a linear regression model, high correlation between one or more features can lead multicollinearity, overfitting or unstable predictions.

0.0.3 Question 2b

To visualize the model performance from part (a), let's make the following two visualizations:

(1) The predicted values vs. observed values on the test set,

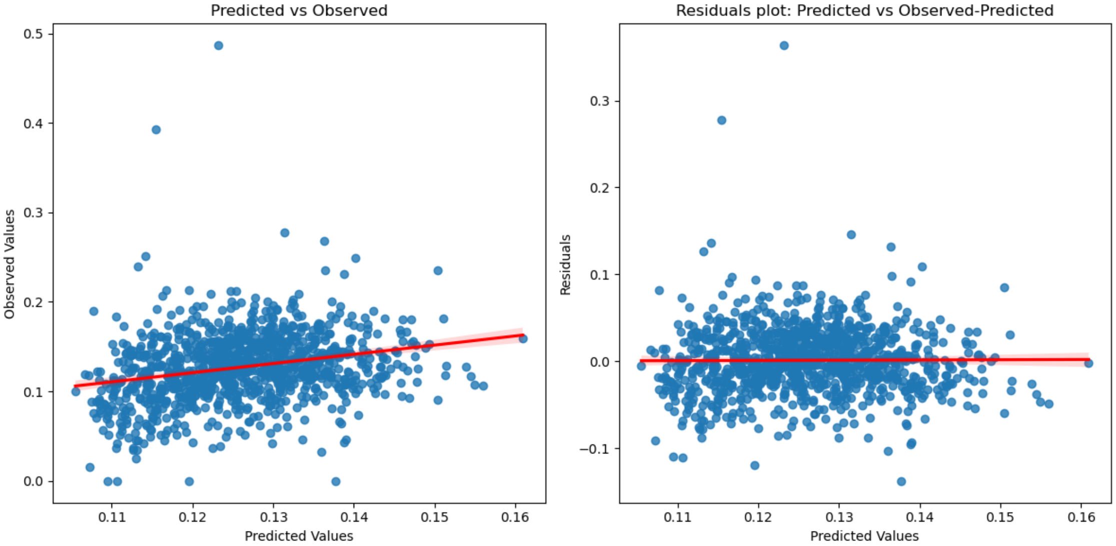
(2) The residuals plot. (Note: in multiple linear regression, the residual plot has predicted values vs. residuals)

Note: * We've used `plt.subplot` (documentation) so that you can view both visualizations side-by-side. For example, `plt.subplot(121)` sets the plottable area to the first column of a 1x2 plot grid; you can then call Matplotlib and Seaborn functions to plot that area, before the next `plt.subplot(122)` area is set. * Remember to add a guiding line to both plot where $\hat{y} = y$, i.e., where the residual is 0. * Please add descriptive titles and axis labels for your plots!

```
In [18]: plt.figure(figsize=(12,6))      # do not change this line
         plt.subplot(121)                # do not change this line
         # (1) predictions vs. observations
         sns.regplot(x = test_prdc, y = y_test, line_kws={"color": "red"})
         plt.title('Predicted vs Observed')
         plt.xlabel('Predicted Values')
         plt.ylabel('Observed Values')

         plt.subplot(122)               # do not change this line
         # (2) residual plot
         sns.regplot(x = test_prdc, y = y_test - test_prdc, line_kws={"color": "red"})
         plt.title('Residuals plot: Predicted vs Observed-Predicted')
         plt.xlabel('Predicted Values')
         plt.ylabel('Residuals')

         plt.tight_layout()             # do not change this line
```


Predicted vs Observed
Observed Values
0.5
0.4
0.3
0.2
0.1
0.0
0.11
0.12
0.13
0.14
0.15
0.16
Predicted Values
Residuals plot: Predicted vs Observed-Predicted
Residuals
0.3
0.2
0.1
0.0
−0.1
0.11
0.12
0.13
0.14
0.15
0.16
Predicted Values

0.0.4 Question 2c

Describe what the plots in part (b) indicate about this linear model.

The ‘Predicted vs Observed’ plot on the left is showing us that we are making poor predictions because there is low correlation, which means our predict value looks not similar to observed value(We want them very close).

The residual plot on the right shows ours residuals. The correlation is 0 in this graph, which means our model is a good fit.

0.0.5 Question 3d

Interpret the confidence intervals above for each of the θ_i, where θ_0 is the intercept term and the remaining θ_i's are parameters corresponding to mask usage features. What does this indicate about our data and our model?

Describe a reason why this could be happening.

Hint: Take a look at the design matrix, heatmap, and response from Question 1!

The 95% confidence intervals for each is in range from negative to positive, which means 0 is in thoese intervel. We cannot reject the null hypothesis at cutoff 5% (our true parameter could be 0). Every feature has no effect with on predictingcovid cases.

This is because our data is collinearity, so it is hard to tell how each feature are related toprediction of covid case.

0.0.6 Question 4b

Comment on the ratio `ratio`, which is the proportion of the expected square error on the data point captured by the model variance. Is the model variance the dominant term in the bias-variance decomposition? If not, what term(s) dominate the bias-variance decomposition?

Note: The Bias-Variance decomposition from lecture:

$$\text{model risk} = \sigma^2 + (\text{model bias})^2 + \text{model variance}$$

where σ^2 is the observation variance, or "irreducible error".

Model variance is not the dominant term in the bias-variance decomposition. The ratio of model variance of model risk is very small, which is around 0.0013. The observation vaiance and the square of model bias dominate the bias-variance decomposition.

0.0.7 Question 4d

Propose a solution to reducing the mean square error using the insights gained from the bias-variance decomposition above.

Assume that the standard bias-variance decomposition used in lecture can be applied here.

We could minimize the model bias. Because model bias is a more dominant term in the bias variance decomposition than the model variance.