Project1 Part3

Tess Newkold

3/3/2019

Question 1:

The risk factors I determine to be the most important for the landing process are speed on ground, height, pitch, and the aircraft make. I came to this conclusion due to many factors which are outlined below. In figure 1, you can see that while a significant percentage of the flights are regular landings. There are flights deemed as short or long that are happening often and studying them could lead to understanding what factors are leading to these landings.

Final Model: Landing ~ speed_ground + height + pitch + aircraftboeing

Table 1.

	Relative to Group 1	Coefficients	Odds Ratio	Std. Error	Significant
Speed Ground	2	0.2483347	1.281889	0.01997638	YES
	3	1.2709771	3.564334	0.02862813	YES
Height	2	0.1483717	1.159944	0.01731625	YES
	3	0.4062396	1.501162	0.03922743	YES
Pitch	2	-0.2432098	0.784107	0.2638094	NO
	3	1.2946709	3.649795	0.7353315	YES
Aircraft-Boeing	2	4.089318	59.69916	0.4225622	YES
	3	9.220361	10100.71	0.8463685	YES

Major Reasonings:

1. The odds ratio shows the importance of each variable on the landing being either regular or long relative to a short landing. For example, when the speed on the ground increases by 1, the odds of the landing being regular versus short increases by 1.28. The odd of the landing being long versus short increases by 3.56. Making the speed on the ground probably the most important variable in predicting the landing. This is shown in Figure 2 visually, you can see the shift to the right as speed increases.
2. Height seems to be slightly impactful on the landing, with increases in height leading to more of an impact in the flight landings being long versus short, rather than the flight being normal. I would say that height is slightly important, but not as much as speed on the ground. You can see this in figure 3, there is a slight shift to the right (increase in height) in longer landings, but not by much.
3. Pitch does not seem to be significant between short and regular flights. If anything if pitch increases by 1 the odds of a regular landing go down verse a short landing. But the odds of a long landing increase by 3.6, which is dramatic.
4. Boeing flights also have higher odds of being longer than airbus airplanes do, which is something to keep an eye out for when looking for which planes will land short, regular or long.
5. Speed in the air was left out due to multicollinearity with speed on ground. Because of the frequency of missing values in speed in air, I decided to choose speed on ground over air.

Figure 1. Pie Chart of landing distributions.

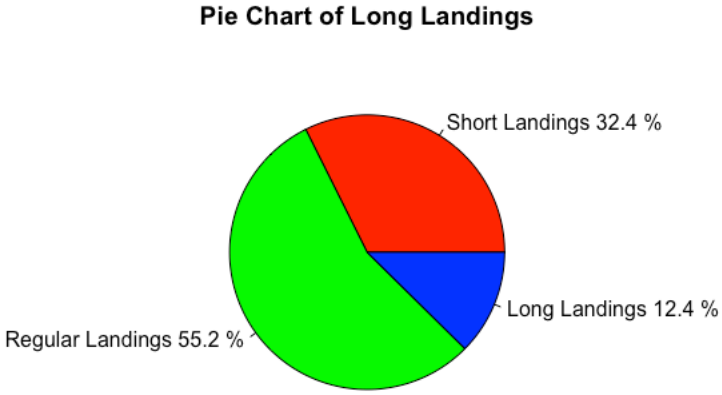


Figure 2. Histogram of speed on ground for each categories of landings; short, regular, and long.

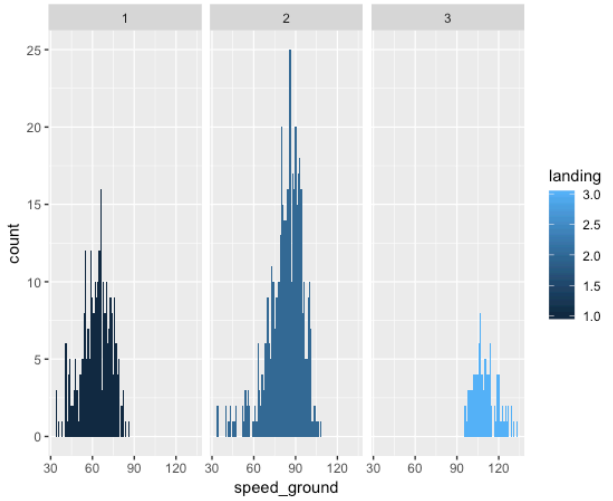


Figure 3. Histogram of height for each category of landing

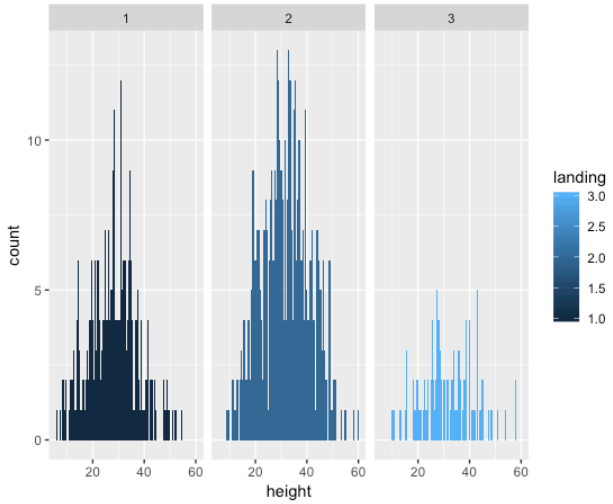
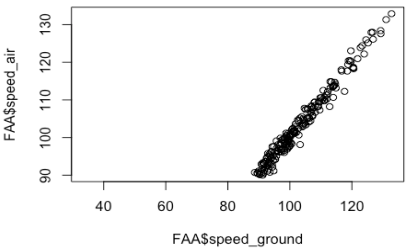


Figure 4. A scatter plot of speed on ground by speed in air, showing multicollinearity between the two variables.



Question 2:

To evaluate the factors important in predicting the number of passengers on a flight we first need to check the distribution of the response variable. Shown in figure 5 below you can see the distribution is normal and it is a continuous variable. **Because of this I am going to use linear regression** to evaluate the important factors. In table 2 below you can see the significance of each variable in predicting the number of passengers on a flight. As you can see, none of them are significant. I also showed this is Figure 6, you can see that no variables are correlated to the number of passengers. **This indicates that none of these variables are good predictors of the number of passengers on board**. This makes sense because the speed of air on the ground or height doesn't make logical sense that it would predict the number of passengers. Perhaps better variables would be something like day of the week, destination or whether it is a holiday or not. These could potentially predict the number of passengers on a particular flight, but not the variables we were given.

Table 2.

Variable	Coefficient	p-Value	Significant
Speed on Ground	0.06815	0.937	NO
Speed in Air	0.55576	0.374	NO
Height	0.09331	0.343	NO
Pitch	-0.02412	0.724	NO
Aircraft	-0.22380	0.406	NO
Duration	-0.06143	0.374	NO
Distance	-0.66193	0.152	NO

Figure 5. Distribution of the response variable number of passengers, it is a normal distribution.

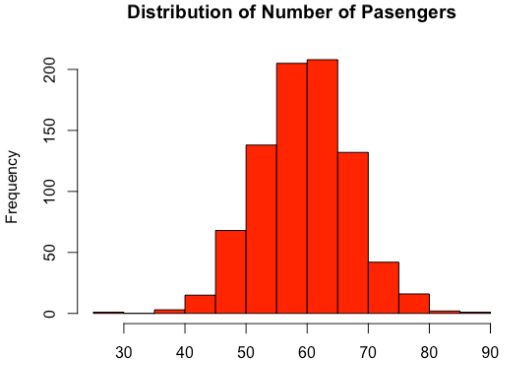


Figure 6. Correlation matrix of all variables, you can see that the variables are randomly scattered, especially when compared to number of passengers.