Executive summary:

From the mtcars data set, it can be concluded that the automatic transmission cars have better MPG than the manual transmission cars. The mean MPG are 17.15 and 24.39 for automatic and manual transmission respectively, with the automatic transmission car 7.24 lower in MPG than the manual ones. Analyzing the fitting coefficients and the residuals also supports the trend is trustworthy.

Looking at the mtcars data, we see that outcome “mpg” is continuous number, and the input “am” is binary – 0 for automatic transmission and 1 for manual transmission. We should use linear regression to fit the relationship.

First, we do an exploratory data analysis by plotting out the boxplot of “mpg” vs. “am”, as in figure 1. The plot shows a distinct mpg difference between the manual and auto transmission. The mpg mean of the cars with automatic transmission is 17.15, while that of the cars with manual transmission is 24.39. A further T test on the mpg for automatic and manual transmission shows that the upper bound of the 95% confidence interval of the automatic cars, which is 19.00, is lower than the lower bound of the 95% confidence interval of the manual cars, which is 20.67. It’s a clear indication that the mpg of the automatic transmission car is lower than that of the manual transmission car. (codes attached)

Second, we use the linear model to fit the mpg ~ transmission in this data frame. The predicted mpg for automatic and manual transmission cars are calculated based on the fitted linear model. Then the sample data and predicted curve are plotted together to show the trend, as in figure 2. Here are the analysis of the fitted model.

1. The coefficients. Model mpg=b0+b1\*am has two coefficients b0 and b1. When am=0, mpg=b0, this means b0 is the expected mpg for automatic transmission cars. When am=1, mpg=b0+b1, b1=mpg-b0, which means b1 is the increase of the mpg for the manual transmission cars compared to the automatic ones. Summary of the fitted model shows the b0 (intercept) = 17.147, which is exactly the mean of the automatic transmission cars, and b1=7.245, adding to b0 (=24.39) gives the expected mpg for the manual transmission cars, and it’s equal to the mean.
2. Inference based on the fitted coefficients. For b0 and b1, the null hypothesis is b0 and b1 are zero, which is tested by comparing their p-value to 0.05. The summary of the fitted model shows the p-value for intercept and slope are 1.13e-15 and 0.000285, which are much smaller than 0.05. Therefore we can reject the null hypothesis that they are zero. Therefore there is a significant relationship between the variables in the linear regression model between mpg and am.
3. Residual and diagnostics. The residual plot is shown in figure 3. It shows the residuals are very much equally distributed around 0 mean. The first residuals diagnostics is by looking at the residuals vs fitted values curve, as in figure 4, it shows no influential point in the sample. Then we look at the dfbeta and hatvalues of the fitted model, and it shows no outstanding large values in all the samples. By plotting the QQ plot, as in figure 5, we can also see there is no outliner residuals in for the fitted model.

Therefore, to finally answer the questions, we conclude that

1. The automatic transmission is better for MPG
2. The expected MPG difference btw the automatic and manual transmission is 7.24, with the automatic one being lower.