

**Airline Satisfaction: Dummy Regression WEEK 2  
By: Victoria Chung**

Regression is a powerful analytical technique. Commonly used to determine the effect continuous independent variables have on a continuous dependent variable, dummy regression can be used to incorporate categorical variables into regression. This allows market researchers to analyze the differences between groups (or categories) of variables. For instance, an airline might want to know if there are differences in customer satisfaction by season for flights to Florida. Or, differences by season and flights originating from different regions of the country, and even a continuous variable can be added, such as the duration of the flight or the percent capacity of seats taken. Regression with dummy/indicator variables gives you the flexibility of analyzing results where you have both categorical and continuous predictors.

A dummy variable is a numerical variable used in regression analysis to represent subgroups of a categorical variable. In the simplest case, a 0, 1 dummy variable is used when the variable is dichotomous, such as whether or not a person responds to an email ad. Dummy variables are useful because they enable us to use a single regression equation to represent multiple groups. This means that we don't need to write out separate equation models for each group. The dummy variables act like **'switches'** that turn various levels of the categorical variable on and off in an equation.

In regression analysis, the dependent variables may be influenced not only by quantitative variables (income, output, prices, etc.), but also by categorical variables (gender, airline class, geographic region, etc.). A dummy independent variable which for some observation has a value of 0 will cause that variable's coefficient to have no role in influencing the dependent variable, while when the dummy takes on a value 1 shows the influence of that group. For example, if we wanted to know the satisfaction by gender, we could code Females 0 and Males 1. After the analysis is run, and the regression equation is generated, setting X1 to 0 gives the satisfaction level for women, and if 1 is entered for X1, then the satisfaction level of men is given.

A rule that a researcher must remember when coding dummy variables, is the number of dummy variables is one less than the number of levels of the categorical variable. In the example above regrading gender, only one dummy variable is needed because gender has two levels. So, for a variable, such as airline status, which has four levels (Blue, Silver, Gold, and Platinum), a researcher will need three dummy variables. The researcher picks one of the levels as a base condition. The base condition is the level that all the other levels will be compared to. You might be wondering “which level should be set as the base condition?” The good news is mathematically it really does not matter. But, a rule of thumb is, if you have a variable that has a level that indicates the absence of the attribute; say did the person respond to an email ad (Yes or No), pick the ‘No’ as the base condition. As for a variable such as airline status, which level you pick as the base condition is either arbitrary or based on the question you are asking. If you are asking “what is the satisfaction of Silver, Gold, and Platinum as compared to Blue customers,” then set Blue as the base condition. If you want to know the satisfaction of the levels as compared to Gold customers, then set Gold as the base condition.

Just to take dummy regression a little further, a researcher can have multiple categorical variables in one equation. For instance, a researcher can have a dummy variable for gender and three more dummy variables to represent airline status. This allows the researcher to analyze the effect of both gender and airline status on satisfaction in the same equation. And, finally, a researcher can have an equation that includes both dummy variables and continuous variables in the same equation. This allows the researcher, for example, to analyze the influence of gender, airline status, and flight duration (continuous variable) all together. Now that is some powerful analysis.

Now that you have a feel for dummy regression, it is time to use it to estimate satisfaction using dummy coded categorical variables. Specifically, you will be analyzing satisfaction by gender, class (Eco, Eco Plus, and Business) and age. To perform the analysis, you will first need to dummy code the gender and class variables, then you will analyze satisfaction gender and class independently and together, and finally, you will add age to the equation.

***Part I: Creating Dummy Coded Variables***

Dummy coded variables for Gender and Class need to be generated. Use the *Create Dummy Variables* under Transform to do this. For Gender, use Gender as the root name, and for Class use Class as the root name. Please clearly rename the newly created variables to clarify which category they represent (e.g. Gender\_Fem & Gender\_Mal). I should know which category each variable represents without having to view the SPSS file itself—viewing only the coefficient table output.

Table

Description automatically generated  
Table

Description automatically generated

Note: SPSS creates a dummy coded variable for **each** level of the variable. You will get two dummy coded variables for Gender, one with Female coded 1 and one with Male coded 1. You will also get three dummy coded variables for Class. This gives you the flexibility of choosing which category you want to use as the reference/indicator variable—but remember that you will NOT use all of the dummy variables in the regression analysis (only K-1 variables—the one left out becomes the reference category).

***Part II: Dummy Using One Variable***

You will create two regression equations. One to analyze the effect of gender on satisfaction, and another to analyze the effect of class on satisfaction. **Remember you only need 1 less dummy variable than levels.** So, for Gender, only one dummy coded variable is needed, and for Class, only two dummy coded variables will be needed.

Gender: Set Male as the base condition. This means use the Gender dummy variable just created that has Female coded 1 and Male coded 0.

Class: Set Eco as the base condition. This means you will only use two of the dummy variables you just generated. When entering the variables, enter Eco Plus first and then Business.

If you did this correctly, for the Gender analysis, you should have a coefficient of -.263 for XFemale.   
  
**Gender analysis**  
![Table

Description automatically generated]()  
![Table

Description automatically generated]()

Anova is less than .05 so the model is significant. The significant level for females is less than .05.

If the flight has females that means the satisfaction drops -.263. Women are less satisfied with flying.   
β0 represents the satisfaction when a flight has males (3.533)  
β1 represents the difference in satisfaction when the flight has females (-.263)  
Thus, the satisfaction when the flight has females is  
3.533-.263=3.27 (subtract from when the flight has males the overall satisfaction after having females on the flight is 3.27  
  
**Class analysis**

For the Class analysis you should have the coefficients -.095 for XEcoPlus and .165 for XBusiness.  
![Table

Description automatically generated]()  
![Table

Description automatically generated]()

3.380-.095X1+.165X2β0 represents the satisfaction when a flight has the Eco class (3.380)  
β1 represents the difference in satisfaction when the flight Eco plus class as compared to Eco class (-.095)  
β2 represents the difference in satisfaction when the flight Business class as compared to Eco class (.165)

***Part III: Dummy Regression with Two Categorical Variables***

Analyze the effect on satisfaction of both Gender and Class by using a regression equation with dummy variables for both variables. Again, set the base condition for Gender to Males, and for class Eco. Enter the dummy variable for Gender first followed by the variable for Eco Plus then Business.

If you did this correctly, you have the coefficients -.258 for XFemale, -.037 for XEcoPlus, and .162   
for XBusiness.  
  
![Table

Description automatically generated]()

The Anova is less than .05 so that means the model is good. Base condition is 3.520, if the flight has females the satisfaction will drop -.258 and if the person has Eco plus class the satisfaction goes down -.037 and increases .162 when the person has business class.

Regression Equation

ŷ=3.520-.258XFemale -.037XEcoPlus+.162XBusinessβ0 represents the satisfaction when a flight has the Males and Eco class (3.520)  
β1 represents the difference in satisfaction when the flight has female (-.258)  
β2 represents the difference in satisfaction when the person has Eco Plus class(-.037)  
β3 represents the difference in satisfaction when the person has Business class (.162)

***Part IV: Dummy Regression with Two Categorical Variables and a Continuous Variable***

Using the equation, you just generated in Part III, add the variable Age to the equation. Since this is a continuous variable, it does not need to be dummy coded. You can just enter it in the independent variable list as is.

If you did this correctly, you should have the coefficients, -.241 for XFemale, -.024 for XEcoPlus, .154 for XBusiness, and -.012 for XAge  
![Table

Description automatically generated]()  
![Table

Description automatically generated]()ŷ=4.072-.241XFemale +.154XBusiness-.012XAge

β0 represents the satisfaction when a flight has males, Eco class, and the age of the customer (4.072)  
β1 represents the difference in satisfaction when the flight has female (-.241)  
~~β~~~~2~~ ~~represents the difference in satisfaction when the flight Eco plus class (.024)~~ (Sig over .05)  
β3 represents the difference in satisfaction when the flight Business class (.154)  
β4 represents the difference in satisfaction with the customers age (.154)

***Part V: Write up.***

1. Paste the coefficient tables (ONLY—not full output) for each of the 4 regressions equations below.
   1. Below each of the 4 results tables, discuss in 2-4 sentences what the results mean. Remember to check the significance for each of the coefficients. If the coefficient is not significant, remember that means it is not different from the base condition.

![Table

Description automatically generated]()  
y ̂= 3.533-.263X1

The Anova is less than .05 so the model is good. If the flight has females that means the satisfaction drops -.263. Women are less satisfied with flying. This could be because they are traveling with children and it’s more difficult for them.

![Table

Description automatically generated]()

3.380-.095X1+.165X2

The Anova is less than .05 so the model is significant. Compared to the Eco class, customers with Eco plus are less satisfied compared to Eco class and customers with Business class are more satisfied compared to both Eco and Eco plus class. Eco plus class may travel too often and this can lower satisfaction. The business class may come with perks, so this is why their satisfaction is higher. People with Eco plus class have a drop of -.095 of satisfaction compared to the Eco class, whereas people in the Business class have an increase of satisfaction of .165 when they fly compared to both the Eco and Eco plus class.

![Table

Description automatically generated]()  
ŷ=3.520-.258XFemale -.037XEcoPlus+.162XBusiness

The Anova is less than .05 so that means the model is significant. Base condition is 3.520, if the flight has females the satisfaction will drop -.258 and if the person has Eco plus class the satisfaction goes down -.037 and increases .162 when the person has business class. This means eco plus class travels too much and can cause dissatisfaction.

![Table

Description automatically generated]()  
ŷ=4.072-.241XFemale +.154XBusiness-.012XAge

The Anova is less than .05 so that means the model is significant. Base condition is 4.072, female satisfaction goes down -.241 when flying and if the person has eco plus class we would remove them from the regression equation because their significant level is over .05. If the client has business class their satisfaction would increase .154. The age of the customer would also drop the satisfaction -.012. This may mean the older they are the less satisfied they are because they are either older or its more difficult to travel.

2. As the marketing manager, what types of actions could you recommend based on these results?  
Looking at the data, I would try to increase the female satisfaction. The airline could support women with children with a form of daycare so that the female can relax during their travel. Perhaps we could create a campaign to target older females who no longer needs to worry about traveling with children. I would also try to increase satisfaction for males with Eco plus class by pushing out a campaign that will give them perks such as letting them earn points on their credit card or earn frequent flyer points to improve satisfaction, I would also send another campaign to offer free drink vouchers to both men and women with Eco class that will be accepted at any restaurant at the airport.

3. What other variables would be helpful to collect; how would you use them to analyze their impact on satisfaction ratings? At least one of them should be based on your readings and the topics for the week (e.g. a cultural or social variable or an indicator of social currency you could use).   
  
I think a variable that would be helpful is the NPS score, it can tell us the overall customer dissatisfaction and highlight the percentage of detractors. On a scale of 0-10 we can see difference between the promoters (score 9-10) from the detractors (score of 0-6) and the passives (score of 7-8). If the detractors score is high, we can assess what is causing the dissatisfaction and come up with solutions to increase value. If we are pushing out “insider information” or leak some discount on flights during peak hours via conversations on online channels this can increase the NPS score by causing excitement. Another variable that would be helpful is analyzing the frequent flyer loyalty program from the three tiers and tracking customer satisfaction for each program. This will tell us which loyalty program creates more value and satisfaction. By comparing the variable to we can see which loyalty program increases satisfaction and which lowers it.

**IN CANVAS, you will submit ONLY 1 WORD DOCUMENT with your answers to the above questions. Please do not submit the SPSS.sav or SPSS output file.**