ANOVA in SAS® Studio

We are going to utilize the SAS tools to conduct ANOVA analyses.

Dataset Description

We will be using **driverspertrip.xlxs** which contains traveling times for 5 drivers on 4 different routes.WARTA, the Warren Area Regional Transit Authority, is expanding bus service from the suburb of Starbrick to the business district of Warren. There are four routes being considered, U.S. 6, West End, Hickory St., and Rte. 59. WARTA conducted tests to determine whether there is a difference in the meantravel times along the four routes; each driver drove each route.
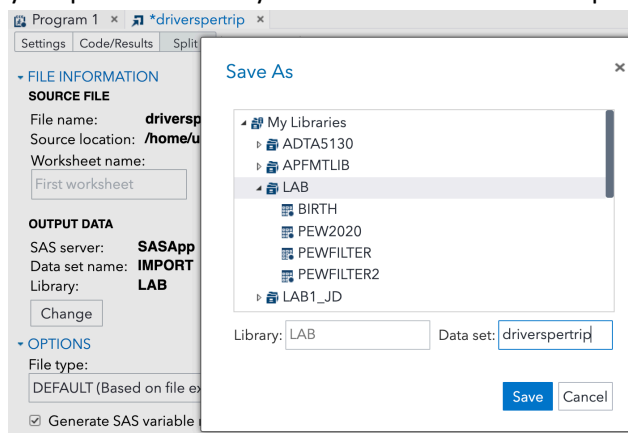
One-Way ANOVA

At significance level of 0.05, is there a difference of travel times among different routes if we don't consider drivers' effect?
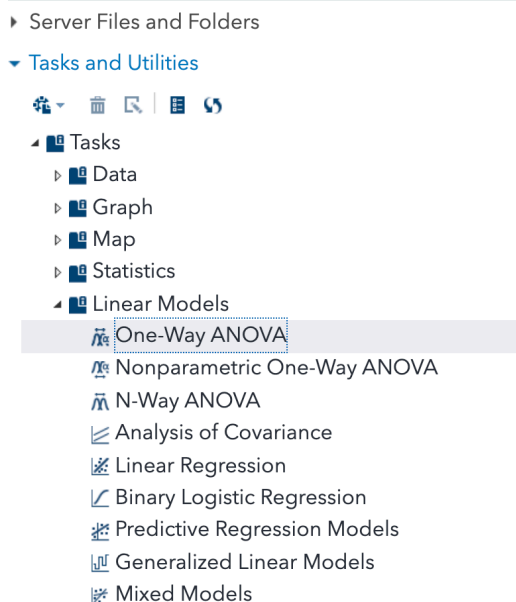
1. Upload the data file to your preferred folder under Server Files and Folders.
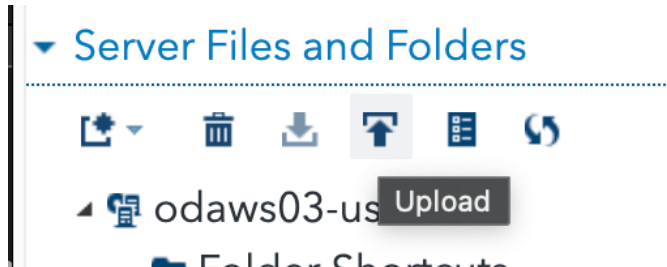
2. Double click the uploaded data inside the folder to open the import window. Select 'Change' and specify your preferred library and data set name 'driverspertrip'.

3. Under Tasks and Utilities, expand Tasks, expand Linear Models, and select One-Way ANOVA.

4. Under Data, select the driverspertrip data. Assign Time to the Dependent variable and Route as the categorical variable.

5. Under Options, select "Brown and Forsythe" for the Homogeneity of Variance test, and "Selected Plots" for "Display plots". Check "LS-mean difference plot" under 'Selected plots'

6. In the code window, click Edit.

7. Type FOOTNOTE "First and Last Name"; into the code as below.

```
proc glm data=LAB.DRIVERSPERTRIP plots(only);
    class Route;
    model Time=Route;
    means Route / hovtest=bf welch plots=none;
    lsmeans Route / adjust=tukey pdiff alpha=.05 plots=(diffplot);
    FOOTNOTE "First and Last Name";
    run;
quit;
```

8. Run the code. The first section of output shows the number of observations and levels of the independent variable. The second part shows the ANOVA table testing the overall equality of the means. This model has a F value of 6.68 with a p-value less than 0.05, we reject the null hypothesis and conclude that there is a difference in the mean traveling time among the routes

Class Level Information

Class	Levels	Values
Route	4	Hickory St. Route 59 US6 West End

Number of Observations Read	60
Number of Observations Used	60

First and Last Name

Dependent Variable: Time Time

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	244.9833333	81.6611111	6.68	0.0006
Error	56	684.6666667	12.2261905		
Corrected Total	59	929.6500000			

R-Square	Coeff Var	Root MSE	Time Mean
0.263522	15.43751	3.496597	22.65000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Route	3	244.9833333	81.6611111	6.68	0.0006

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Route	3	244.9833333	81.6611111	6.68	0.0006

First and Last Name

Remember, one very important assumption for ANOVA is that the populations have equal variances. The next section is the result of the test for variance homogeneity. The p-value is 0.5990 which is not significant so we can conclude the population groups have equal variances.

The hypothesis statements:

H_0: $\sigma_j^2 = \sigma_i^2$ for all i and j (all variances are equal)

H_A: $\sigma_j^2 \neq \sigma_i^2$ for at least one pair (i, j) (at least one pair of variances is not equal)

Brown and Forsythe's Test for Homogeneity of Time Variance ANOVA of Absolute Deviations from Group Medians

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Route	3	8.4500	2.8167	0.63	0.5990
Error	56	250.5	4.4738		

Welch's ANOVA for Time

Source	DF	F Value	Pr > F
Route	3.0000	5.96	0.0025
Error	30.9773		

First and Last Name

4

9. From both the LS Means comparison table and LS-mean difference plot, we can find that the significant pairs include Hickory St. and West End, Hickory St. and US6, and US6 and Route 59.

Least Squares Means
Adjustment for Multiple Comparisons: Tukey

Route	Time LSMEAN	LSMEAN Number
Hickory St.	25.4000000	1
Route 59	23.5333333	2
US6	20.0000000	3
West End	21.6666667	4

Least Squares Means for effect Route
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: Time

| i/j	1	2	3	4
1		0.4670	0.0005	0.0249
2	0.4670		0.0373	0.4670
3	0.0005	0.0373		0.5634
4	0.0249	0.4670	0.5634	

First and Last Name
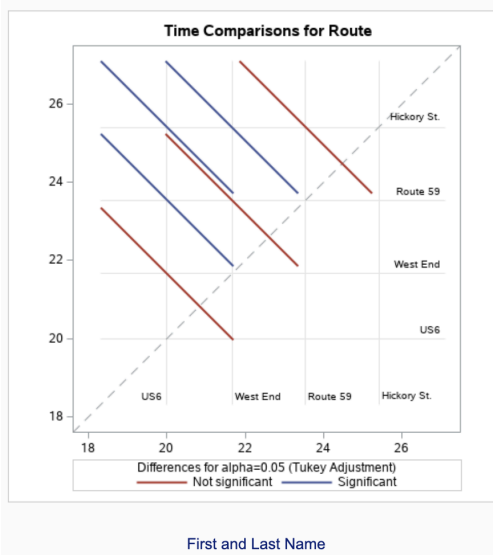
Two-way ANOVA without Interaction

At the 0.05 significance level, is there a difference of travel times among different routes if we consider the effect of drivers? We will use drivers as a blocking variable in this experiment.
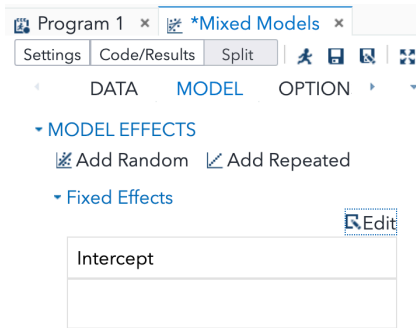
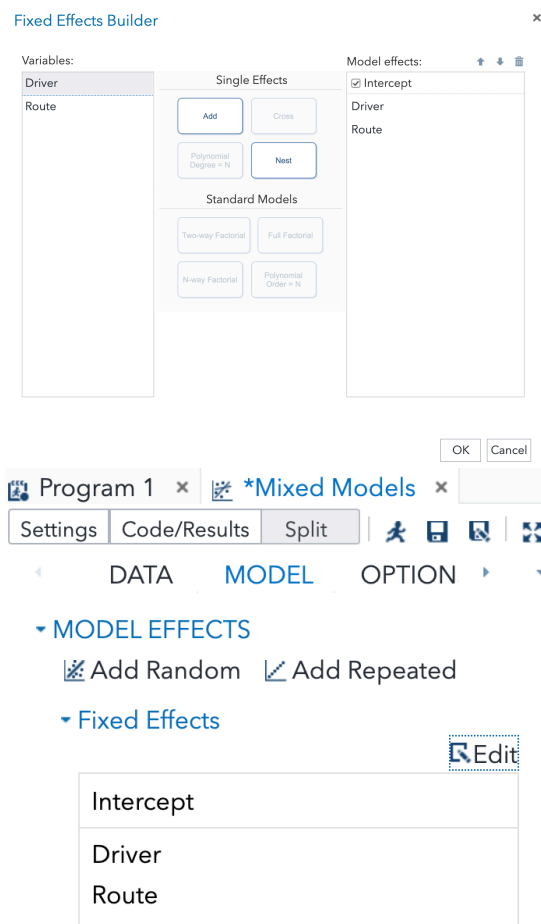1. Under Task of Tasks and Utilities, select Mixed Model.
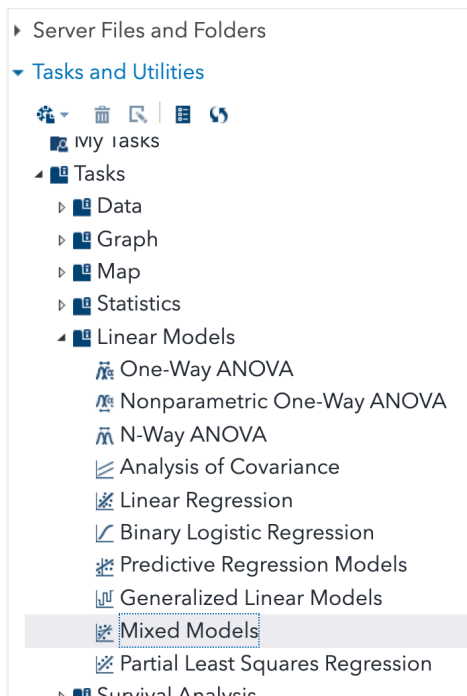
2. In the opened "Mixed Models" window, under DATA, select the driverspertrip data. Assign Time to the Dependent variable, Route and Driver to the Classification variables.
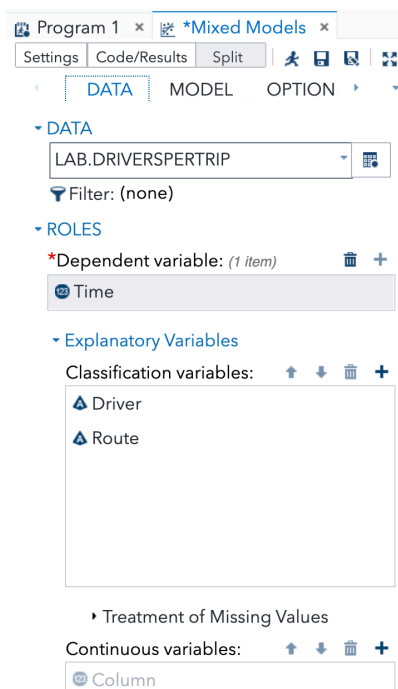
3. Under MODEL, click 'Edit' under Fixed Effects.

In the opened window, add both of Driver and Route under Single effects. Click OK.

4. Under OPTION, change Estimation method to Type 3, select Default and additional statistics under Select statistics to display.

Check 'Perform multiple comparisons' under Multiple Comparisons. Note, make sure that you click Route under Perform multiple comparisons to specify that you want to perform the post-hoc analysis on Route. If needed, you can highlight both by holding control while clicking to test both. But we will only focus on pairwise comparisons of Route because Driver is a blocking variable.

Under method, select Bonferroni.

5. Click Run. The ANOVA table is as below which shows significance (p-value<0.0001) for both variables.

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Driver	4	353.566667	88.391667	Var(Residual) + Q(Driver)	MS(Residual)	52	13.88	<.0001
Route	3	244.983333	81.661111	Var(Residual) + Q(Route)	MS(Residual)	52	12.83	<.0001
Residual	52	331.100000	6.367308	Var(Residual)	.	.	.	.

Or

Type 3 Tests of Fixed Effects

| Effect	Num DF	Den DF	F Value	Pr > F
Driver	4	52	13.88	<.0001
Route	3	52	12.83	<.0001

6. We now know that some Routes are different in mean time, but we don't know which ones. In the multiple comparison table, Adj P shows the Bonferroni adjusted p-value of the comparisons. US6 and Hicktory St., US6 and Route 59, Hictory St. and West End, are significantly different in mean time with p-value below .05.

Differences of Least Squares Means

Effect	Route	Route	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
Route	Hickory St.	Route 59	1.8667	0.9214	52	2.03	0.0479	Bonferroni	0.2875	0.05	0.01775	3.7156	-0.6607	4.3940
Route	Hickory St.	US6	5.4000	0.9214	52	5.86	<.0001	Bonferroni	<.0001	0.05	3.5511	7.2489	2.8726	7.9274
Route	Hickory St.	West End	3.7333	0.9214	52	4.05	0.0002	Bonferroni	0.0010	0.05	1.8844	5.5823	1.2060	6.2607
Route	Route 59	US6	3.5333	0.9214	52	3.83	0.0003	Bonferroni	0.0020	0.05	1.6844	5.3823	1.0060	6.0607
Route	Route 59	West End	1.8667	0.9214	52	2.03	0.0479	Bonferroni	0.2875	0.05	0.01775	3.7156	-0.6607	4.3940
Route	US6	West End	-1.6667	0.9214	52	-1.81	0.0763	Bonferroni	0.4575	0.05	-3.5156	0.1823	-4.1940	0.8607

<u>Two-way ANOVA with Interaction</u>

At the 0.05 significance level, is there a difference of travel times among different routes and drivers? We will consider driver as another independent variable of interest.

1. The settings are similar to the Two-way ANOVA without interaction. But in step 3 above, select both variables simultaneously by holding the control key. The 'Cross' button will show up and we will click it to include the interaction term.
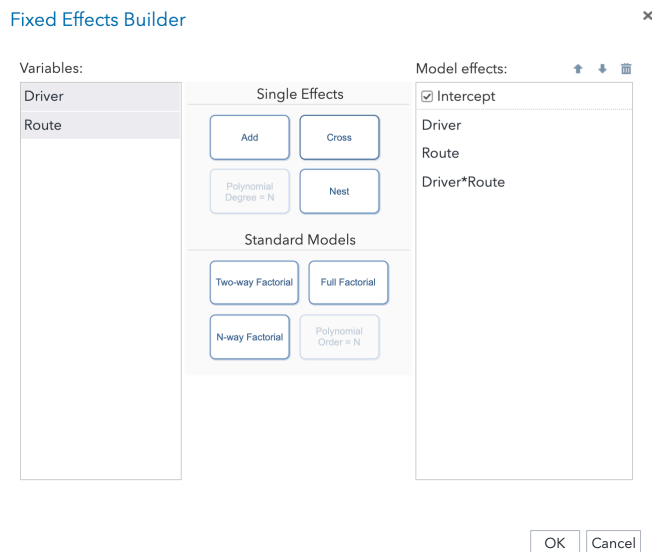
2. Under OPTION, the interaction term will show up. We need to select all three effects simultaneously.
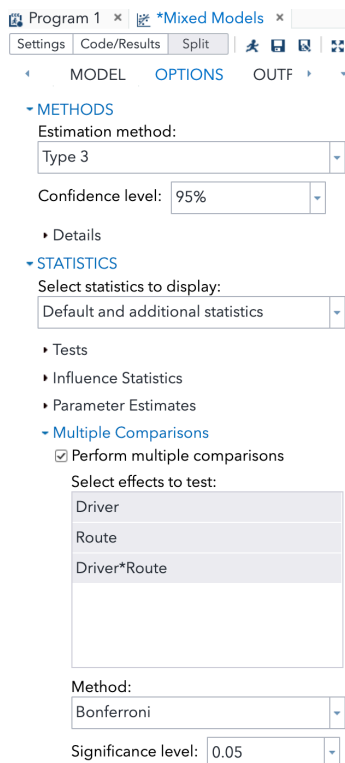
3. The ANOVA table shows that the interaction term is significant with p-value of 0.0456.

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
Driver	4	353.566667	88.391667	Var(Residual) + Q(Driver,Driver*Route)	MS(Residual)	40	17.22	<.0001
Route	3	244.983333	81.661111	Var(Residual) + Q(Route,Driver*Route)	MS(Residual)	40	15.91	<.0001
Driver*Route	12	125.766667	10.480556	Var(Residual) + Q(Driver*Route)	MS(Residual)	40	2.04	0.0456
Residual	40	205.333333	5.133333	Var(Residual)	.	.	.	.

Covariance Parameter Estimates

Cov Parm	Estimate	Alpha	Lower	Upper
Residual	5.1333	0.05	3.4602	8.4039

Fit Statistics

-2 Res Log Likelihood	200.9
AIC (Smaller is Better)	202.9
AICC (Smaller is Better)	203.0
BIC (Smaller is Better)	204.6

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Driver	4	40	17.22	<.0001
Route	3	40	15.91	<.0001
Driver*Route	12	40	2.04	0.0456

4. From the Differences of Least Square Means table below, we can find the significant levels of the interaction term which shows the significantly different pairs of the combinations of Route and Drive. It is a long list, but we can use the below combination as an example. The mean time for Deans on US6 is significantly different from that for Ormson on Hickory with adjusted p-value of 0.0188.

				59										
Driver*Route	Deans	US6	Filbeck	US6	-7.0000	1.8499	40	-3.78	0.0005	Bonferroni	0.0963	0.05	-10.7388	-
Driver*Route	Deans	US6	Filbeck	West End	-6.0000	1.8499	40	-3.24	0.0024	Bonferroni	0.4535	0.05	-9.7388	-
Driver*Route	Deans	US6	Ormson	Hickory St.	-8.0000	1.8499	40	-4.32	<.0001	Bonferroni	0.0188	0.05	-11.7388	-
Driver*Route	Deans	US6	Ormson	Route 59	-4.0000	1.8499	40	-2.16	0.0366	Bonferroni	1.0000	0.05	-7.7388	-
Driver*Route	Deans	US6	Ormson	US6	-888E-18	1.8499	40	-0.00	1.0000	Bonferroni	1.0000	0.05	-3.7388	
Driver*Route	Deans	US6	Ormson	West	-5.0000	1.8499	40	-2.70	0.0100	Bonferroni	1.0000	0.05	-8.7388	-