STA5176: Spring 2023 Group 4
Alzheimer's Project

Group Members:

Brad Lipson

Mary Morrow

Matthew McGehee

Pamela Mishaw

Prince Nhliziyo

STA5176: Spring 2023 Group 4 Alzheimer's Project

Initial code (to set up data table) for each group member is attached at the end of this document

1. **Run proc univariate to test (variable) for normality with the Shapiro-Wilk Test in proc univariate. What can you observe in the SAS results regarding the normality for this variable? Is (variable) normally distributed, using significance level 0.05? Refer to the p-value of the test.**

Name	Var	Hypotheses	SAS Code	SAS Output	Conclusion
Brad Lipson	Cancer	H_0: The cancer data is normally distributed. H_A: The cancer disease data is not normally distributed.	proc univariate normal; var cancer; run;	Tests for Normality Shapiro-Wilk: W 0.847378, Pr < W <0.0001 Kolmogorov-Smirnov: D 0.077151, Pr > D <0.0100 Cramer-von Mises: W-Sq 0.373427, Pr > W-Sq <0.0050 Anderson-Darling: A-Sq 2.743679, Pr > A-Sq <0.0050	The Shapiro-Wilk test has a p-value of 0.0001. The alpha (0.05) requirement is not met by this. The cancer data are therefore not normally distributed, and we reject the null hypothesis.
	Obesity	H_0: The obesity data is normally distributed. H_A: The obesity data is not normally distributed.	proc univariate normal; var obesity; run;	Tests for Normality Shapiro-Wilk: W 0.847378, Pr < W <0.0001 Kolmogorov-Smirnov: D 0.077151, Pr > D <0.0100 Cramer-von Mises: W-Sq 0.373427, Pr > W-Sq <0.0050 Anderson-Darling: A-Sq 2.743679, Pr > A-Sq <0.0050	The Shapiro-Wilk test has a p-value of 0.0001. The alpha (0.05) requirement is not met by this. Because of this irregular distribution of the data on obesity, we reject the null hypothesis.
Mary Morrow	Heart Disease	H_0: The heart disease data is normally distributed. H_A: The heart disease data is not normally distributed.	proc univariate normal; var heart; run;	Tests for Normality Shapiro-Wilk: W 0.933411, Pr < W <0.0001 Kolmogorov-Smirnov: D 0.091533, Pr > D <0.0100 Cramer-von Mises: W-Sq 0.449792, Pr > W-Sq <0.0050 Anderson-Darling: A-Sq 2.904911, Pr > A-Sq <0.0050	The p-value for the Shapiro-Wilk test is <0.0001. This is less than alpha (0.05). Therefore, we reject the null hypothesis and conclude that the heart disease data is not normally distributed.
	Diabetes	H_0: The diabetes data is normally distributed. H_A: The diabetes data is not normally distributed.	proc univariate normal; var diabetes; run;	Tests for Normality Shapiro-Wilk: W 0.952333, Pr < W <0.0001 Kolmogorov-Smirnov: D 0.092152, Pr > D <0.0100 Cramer-von Mises: W-Sq 0.537632, Pr > W-Sq <0.0050 Anderson-Darling: A-Sq 3.345459, Pr > A-Sq <0.0050	The p-value for the Shapiro-Wilk test is <0.0001. This is less than alpha (0.05). Therefore, we reject the null hypothesis and conclude that the diabetes data is not normally distributed.
Matthew McGehee	Mercury (data values are in Tons Per Year)	H_0: Mercury TPY is normally distributed across counties in FL, CA, NY, & WA H_A: Mercury TPY is not normally distributed across counties in FL, CA, NY, & WA	proc univariate data=group4data normal; var Lead_TPY; run;	Tests for Normality Shapiro-Wilk: W 0.58468, Pr < W <0.0001 Kolmogorov-Smirnov: D 0.263723, Pr > D <0.0100 Cramer-von Mises: W-Sq 5.035737, Pr > W-Sq <0.0050 Anderson-Darling: A-Sq 26.78148, Pr > A-Sq <0.0050	The p-value from the Shapiro-Wilk test is <0.0001, and if we use a significance level of 0.05, we should reject the null hypothesis and conclude that Mercury TPY is not normally distributed across the counties.
	Lead TPY (data values are in Tons Per Year)	H_0: Lead TPY is normally distributed across counties in FL, CA, NY, & WA H_A: Lead TPY is not normally distributed across counties in FL, CA, NY, & WA	proc univariate data=group4data normal; var Mercury_TPY; run;	Tests for Normality Shapiro-Wilk: W 0.355343, Pr < W <0.0001 Kolmogorov-Smirnov: D 0.35957, Pr > D <0.0100 Cramer-von Mises: W-Sq 9.658626, Pr > W-Sq <0.0050 Anderson-Darling: A-Sq 47.29493, Pr > A-Sq <0.0050	The p-value from the Shapiro-Wilk test is <0.0001, and if we use a significance level of 0.05, we should reject the null hypothesis and conclude that Lead TPY is not normally distributed across the counties.

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypotheses	SAS Code	SAS Output	Conclusion
Pamela Mishaw	Mental Distress	H_0: The mental distress data are normally distributed. H_A: The mental distress data aren't normally distributed.	proc univariate normal; var mental_distress; run;	**Tests for Normality** Shapiro-Wilk: W 0.928634, Pr < W <0.0001 Kolmogorov-Smirnov: D 0.209463, Pr > D <0.0100 Cramer-von Mises: W-Sq 1.336258, Pr > W-Sq <0.0050 Anderson-Darling: A-Sq 7.012811, Pr > A-Sq <0.0050	Since the p-value of the Shapiro-Wilk test (<0.001) is less than the significance level of 0.05, the null hypothesis is rejected and it is concluded that the data on mental distress are not normally distributed.
	% Bachelor Degree	H_0: The data on the percentage of those with bachelor's degrees are normally distributed. H_A: The data on the percentage of those with bachelor's degrees are normally distributed.	proc univariate normal; var Percent_with_Bachelors; run;	**Tests for Normality** Shapiro-Wilk: W 0.946134, Pr < W <0.0001 Kolmogorov-Smirnov: D 0.095228, Pr > D <0.0100 Cramer-von Mises: W-Sq 0.443945, Pr > W-Sq <0.0050 Anderson-Darling: A-Sq 2.864301, Pr > A-Sq <0.0050	Since the p-value of the Shapiro-Wilk test (<0.001) is less than the significance level of 0.05, the null hypothesis is rejected and it is concluded that the data on the percentage of those with bachelor's degrees are not normally distributed.
Prince Nhliziyo	Smoking	H_0: Smoking rate is normally distributed H_A: Smoking rate is not normally distributed	proc univariate normal; var smoking; run;	**Tests for Normality** Shapiro-Wilk: W 0.991624, Pr < W 0.2228 Kolmogorov-Smirnov: D 0.048984, Pr > D >0.1500 Cramer-von Mises: W-Sq 0.05524, Pr > W-Sq >0.2500 Anderson-Darling: A-Sq 0.371962, Pr > A-Sq >0.2500	Since the p-value of the Shapiro-Wilk test is < 0.05, we reject the null hypothesis and conclude that the data is not normally distributed.
	Inactivity	H_0: Physical inactivity is normally distributed H_A: Physical inactivity is not normally distributed	proc univariate normal; var physical; run	**Tests for Normality** Shapiro-Wilk: W 0.983368, Pr < W 0.0094 Kolmogorov-Smirnov: D 0.044856, Pr > D >0.1500 Cramer-von Mises: W-Sq 0.092129, Pr > W-Sq 0.1449 Anderson-Darling: A-Sq 0.850833, Pr > A-Sq 0.0290	Since the p-value of the Shapiro-Wilk test is < 0.05, we reject the null hypothesis and conclude that the data is not normally distributed.

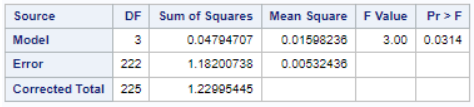
STA5176: Spring 2023 Group 4 Alzheimer's Project

2. Using SAS, compare the 4 population means for (variable) with the ANOVA F test with proc glm. What can you conclude based on the significance level alpha=0.01? Are the means of (variable) for the 4 states equal or not? What can you conclude if you had used significance level alpha=0.10 instead? Refer to the p-value of the test.

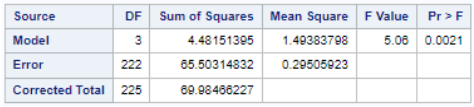
Name	Var	Hypotheses	SAS Code	SAS Output	Conclusion
Brad Lipson	Cancer	H_0: All state (group) means are equal. H_A: At least one state (group) mean is different from the other state (group) means.	proc glm; class state; model cancer = state; run;	Source | DF | Sum of Squares | Mean Square | F Value | Pr > F Model | 3 | 17005.1549 | 5668.3850 | 11.31 | <.0001 Error | 222 | 111285.3733 | 501.2855 | | Corrected Total | 225 | 128290.5282 | | |	The p-value (.0001) is smaller than alpha for an alpha of 0.01 or more. The null hypothesis is thus rejected, and we draw the conclusion that at least one state mean is different for cancer. The p-value (.0001) is smaller than alpha for alpha = 0.10 in this example. The null hypothesis is thus rejected, and we draw the conclusion that at least one state mean is different for cancer.
	Obesity	H_0: All state (group) means are equal. H_A: At least one state (group) mean is different from the other state (group) means.	proc glm; class state; model obesity = state; run;	Source | DF | Sum of Squares | Mean Square | F Value | Pr > F Model | 3 | 1090.069411 | 363.356470 | 25.57 | <.0001 Error | 222 | 3154.452403 | 14.209245 | | Corrected Total | 225 | 4244.521814 | | |	The p-value (.0001) is smaller than alpha for an alpha of 0.01 or more. We thereby reject the null hypothesis and come to the conclusion that at least one state mean differs for obesity. The p-value (.0001) is smaller than alpha for alpha = 0.10 in this example. We thereby reject the null hypothesis and come to the conclusion that at least one state mean differs for obesity.
Mary Morrow	Heart Disease	H_0: All state (group) means are equal. H_A: At least one state (group) mean is different from the other state (group) means.	proc glm; class state; model heart = state; run;	Source | DF | Sum of Squares | Mean Square | F Value | Pr > F Model | 3 | 41247.8697 | 13749.2899 | 16.59 | <.0001 Error | 222 | 183932.3482 | 828.5241 | | Corrected Total | 225 | 225180.2179 | | |	For alpha = 0.01: The p-value (<.0001) is less than alpha. Therefore, we reject the null hypothesis and conclude that at least one state mean is different for heart disease. For alpha = 0.10: The p-value (<.0001) is less than alpha. Therefore, we reject the null hypothesis and conclude that at least one state mean is different for heart disease.
	Diabetes	H_0: All state (group) means are equal. H_A: At least one state (group) mean is different from the other state (group) means.	proc glm; class state; model diabetes = state; run;	Source | DF | Sum of Squares | Mean Square | F Value | Pr > F Model | 3 | 205.7847233 | 68.5949078 | 50.38 | <.0001 Error | 222 | 302.2784201 | 1.3616145 | | Corrected Total | 225 | 508.0631434 | | |	For alpha = 0.01: The p-value (<.0001) is less than alpha. Therefore, we reject the null hypothesis and conclude that at least one state mean is different for diabetes. For alpha = 0.10: The p-value (<.0001) is less than alpha. Therefore, we reject the null hypothesis and conclude that at least one state mean is different for diabetes.

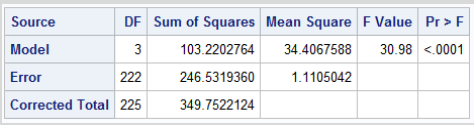
STA5176: Spring 2023 Group 4 Alzheimer's Project

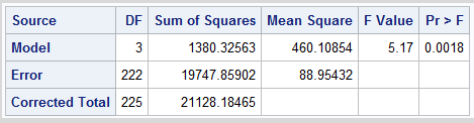
<table>
<tr><th>Name</th><th>Var</th><th>Hypotheses</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td rowspan="2">Matthew McGehee</td><td>Mercury</td><td>H_0: The mean mercury level across each state is equivalent

H_A: At least one state's mean mercury level is not equivalent to the others.</td><td>proc glm data=group4data;
class State;
model Mercury_TPY=State;
run;</td><td> 
<table>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr>
<tr><td>Model</td><td>3</td><td>0.04794707</td><td>0.01598236</td><td>3.00</td><td>0.0314</td></tr>
<tr><td>Error</td><td>222</td><td>1.18200738</td><td>0.00532436</td><td></td><td></td></tr>
<tr><td>Corrected Total</td><td>225</td><td>1.22995445</td><td></td><td></td><td></td></tr>
</table></td><td>With a p-value of 0.0314, we cannot conclude that at least one state's mean Mercury distribution is different using a significance level of 0.01.

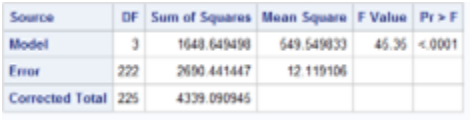
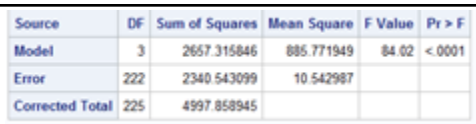
However, if we were to use an alpha equal to 0.10, we would have to conclude that at least one state's mean mercury level is different.</td></tr>
<tr><td>Lead</td><td>H_0: The mean lead level across each state is equivalent

H_A: At least one state's mean lead level is not equivalent to the others.</td><td>proc glm data=group4data;
class State;
model Lead_TPY=State;
run;</td><td> 
<table>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr>
<tr><td>Model</td><td>3</td><td>4.48151395</td><td>1.49383798</td><td>5.06</td><td>0.0021</td></tr>
<tr><td>Error</td><td>222</td><td>65.50314832</td><td>0.29505923</td><td></td><td></td></tr>
<tr><td>Corrected Total</td><td>225</td><td>69.98466227</td><td></td><td></td><td></td></tr>
</table></td><td>With a p-value of 0.0021, we would have to reject the null hypothesis and conclude that at least one state's mean lead level is different, whether we used a significance level of 0.01 or 0.10.</td></tr>
<tr><td rowspan="2">Pamela Mishaw</td><td>Mental Distress</td><td>H_0: All state means of mental distress rating are equal.

H_A: At least two of the means of mental distress rating differ at a statistically significant level.</td><td>proc glm;
class state;
model mental_distress=state;
run;</td><td> 
<table>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr>
<tr><td>Model</td><td>3</td><td>103.2202764</td><td>34.4067588</td><td>30.98</td><td><.0001</td></tr>
<tr><td>Error</td><td>222</td><td>246.5319360</td><td>1.1105042</td><td></td><td></td></tr>
<tr><td>Corrected Total</td><td>225</td><td>349.7522124</td><td></td><td></td><td></td></tr>
</table></td><td>At the significance level of 0.01, the null hypothesis is rejected since the p-value (<0.001) is smaller and it is concluded that at least two of the means of mental distress are statistically significant in difference and thus the four means are likely not all equal. At significance level 0.10, since the p-value is smaller, the conclusion would be the same as that reached at the 0.01 significance level– the null hypothesis is rejected and it is concluded that at least two of the states have means that are statistically different.</td></tr>
<tr><td>% Bachelor Degree</td><td>H_0: All state means of bachelor's degree percentages are equal.

H_A: At least two of the means of bachelor's degree percentages differ at a statistically significant level.</td><td>proc glm;
class state;
model Percent_with_Bachelors=state;
run;</td><td> 
<table>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr>
<tr><td>Model</td><td>3</td><td>1380.32563</td><td>460.10854</td><td>5.17</td><td>0.0018</td></tr>
<tr><td>Error</td><td>222</td><td>19747.85902</td><td>88.95432</td><td></td><td></td></tr>
<tr><td>Corrected Total</td><td>225</td><td>21128.18465</td><td></td><td></td><td></td></tr>
</table></td><td>At the significance level of 0.01, the null hypothesis is rejected since the p-value (0.0018) is smaller and it is concluded that at least two of the means of bachelor's degree percentages are statistically significant in difference and thus the four means are likely not all equal. At significance level 0.10, since the p-value is smaller, the conclusion would be the same as that reached at the 0.01 significance level– the null hypothesis is rejected and it is concluded that at least two of the states have means that are statistically different.</td></tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypotheses	SAS Code	SAS Output	Conclusion
Prince Nhliziyo	Smoking	H_0: All state means are equal. H_A: At least one state mean is different from the other state means.	proc glm; class state; model smoking = state; run;	 Source | DF | Sum of Squares | Mean Square | F Value | Pr > F Model | 3 | 1648.649490 | 549.549833 | 45.36 | <.0001 Error | 222 | 2690.441447 | 12.119106 Corrected Total | 225 | 4339.090945	Based on significance level alpha=0.01,we reject the null hypothesis and conclude that at least one mean is different for smoking rate. Based on significance level alpha=0.10 ,we reject the null hypothesis and conclude that at least one mean is different for smoking rate
	Inactivity	H_0: All state means are equal. H_A: At least one state mean is different from the other state means.	proc glm; class state; model physical = state; run;	 Source | DF | Sum of Squares | Mean Square | F Value | Pr > F Model | 3 | 2657.315846 | 885.771949 | 84.02 | <.0001 Error | 222 | 2340.543099 | 10.542987 Corrected Total | 225 | 4997.858945	Based on significance level alpha=0.01,we reject the null hypothesis and conclude that at least one mean is different for physical inactivity. Based on significance level alpha=0.10 ,we reject the null hypothesis and conclude that at least one mean is different for physical inactivity.

STA5176: Spring 2023 Group 4 Alzheimer's Project

3. Using normal scores (or normal quantiles) for (variable) in SAS, obtain the ANOVA F Test based on normal scores or quantiles. What can you conclude based on significance level alpha=0.01? Are the mean lengths equal for the 4 states or not, based on the normal scores? Compare the results in (3) with your results in (2). What can you conclude if you had used significance level alpha=0.10 instead? Refer to the p-value of the test.

Name	Var	Hypotheses	SAS Code	SAS Output	Conclusion

Brad Lipson	Cancer	H_0: The group means would be equal for all of the states.
		
		H_A: The group means would not be equal for all of the states, so one or more would differ.
			proc glm;
			class state;
			model cancer=state;
			run;
			
			proc rank normal=blom out=normals;
			var cancer; ranks q;
			data normals; set normals;
			proc plot; plot cancer*q;
			run;
			
			proc glm;
			class state;
			model q=state;
			run;
				Source | DF | Sum of Squares | Mean Square | F Value | Pr > F
				Model | 3 | 17005.1549 | 5668.3850 | 11.31 | <.0001
				Error | 222 | 111285.3733 | 501.2855 | | |
				Corrected Total | 225 | 128290.5282 | | | |
					The p-value (.0001) is smaller than alpha for an alpha of 0.01 or more. The null hypothesis is thus rejected, and we draw the conclusion that at least one state mean is different for cancer. The p-value (.0001) is smaller than alpha for alpha = 0.10 in this example. The null hypothesis is thus rejected, and we draw the conclusion that at least one state mean is different for cancer.
					We failed to accept the null hypotheses for alpha=0.01 and alpha=0.10 for cancer in problems 2 and 3, respectively, and came to the conclusion that at least one state mean is different.

	Obesity	H_0: The group means would be equal for all of the states.
		
		H_A: The group means would not be equal for all of the states, so one or more would differ.
			proc glm;
			class state;
			model obesity=state;
			run;
			
			proc rank normal=blom out=normals;
			var obesity; ranks q;
			data normals; set normals;
			proc plot; plot obesity*q;
			run;
			
			proc glm;
			class state;
			model q=state;
			run;
				Source | DF | Sum of Squares | Mean Square | F Value | Pr > F
				Model | 3 | 55.6135895 | 18.5378632 | 24.66 | <.0001
				Error | 222 | 166.8715753 | 0.7516738 | | |
				Corrected Total | 225 | 222.4851648 | | | |
					The p-value (.0001) is smaller than alpha for an alpha of 0.01 or more. As a result, we find that at least one state mean is different for obesity and reject the null hypothesis.
					The p-value (.0001) is smaller than alpha for alpha = 0.10 in this example. We thereby reject the null hypothesis and come to the conclusion that at least one state mean differs for obesity.
					We failed to accept the null hypotheses for alpha=0.01 and alpha=0.10 for obesity in problems 2 and 3, respectively, and came to the conclusion that at least one state mean is different.

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>Hypotheses</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td rowspan="2">Mary Morrow</td><td>Heart Disease</td><td>H_0: All state (group) means are equal.

H_A: At least one state (group) mean is different from the other state (group) means.</td><td>proc glm;
class state;
model heart=state;
run;

proc rank normal=blom out=normals;
var heart; ranks q;
data normals; set normals;
proc plot; plot heart*q;
run;

proc glm;
class state;
model q=state;
run;</td><td>Dependent Variable: q Rank for Variable heart
<table><tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>Model</td><td>3</td><td>43.9217127</td><td>14.6405709</td><td>18.20</td><td><.0001</td></tr><tr><td>Error</td><td>222</td><td>178.5634770</td><td>0.8043400</td><td></td><td></td></tr><tr><td>Corrected Total</td><td>225</td><td>222.4851898</td><td></td><td></td><td></td></tr></table></td><td><u>For alpha = 0.01:</u> The p-value (<.0001) is less than alpha. Therefore, we reject the null hypothesis and conclude that at least one state mean is different for heart disease.
<u>For alpha = 0.10:</u> The p-value (<.0001) is less than alpha. Therefore, we reject the null hypothesis and conclude that at least one state mean is different for heart disease.
For both problem 2 and problem 3, we rejected the null hypotheses for both alpha=0.01 and alpha=0.10 for heart disease and concluded that at least one state mean is different.</td></tr>
<tr><td>Diabetes</td><td>H_0: All state (group) means are equal.

H_A: At least one state (group) mean is different from the other state (group) means.</td><td>proc glm;
class state;
model diabetes=state;
run;

proc rank normal=blom out=normals;
var diabetes;
ranks q;
data normals; set normals;
proc plot; plot diabetes*q;
run;

proc glm;
class state;
model q=state;
run;</td><td>Dependent Variable: q Rank for Variable diabetes
<table><tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>Model</td><td>3</td><td>86.4352701</td><td>28.8117567</td><td>47.02</td><td><.0001</td></tr><tr><td>Error</td><td>222</td><td>136.0291886</td><td>0.6127441</td><td></td><td></td></tr><tr><td>Corrected Total</td><td>225</td><td>222.4644587</td><td></td><td></td><td></td></tr></table></td><td><u>For alpha = 0.01:</u> The p-value (<.0001) is less than alpha. Therefore, we reject the null hypothesis and conclude that at least one state mean is different for diabetes.
<u>For alpha = 0.10:</u> The p-value (<.0001) is less than alpha. Therefore, we reject the null hypothesis and conclude that at least one state mean is different for diabetes.
For both problem 2 and problem 3, we rejected the null hypotheses for both alpha=0.01 and alpha=0.10 for diabetes and concluded that at least one state mean is different.</td></tr>
<tr><td>Matthew McGehee</td><td>Mercury</td><td>H_0: The mean mercury level across each state is equivalent

H_A: At least one state's mean mercury level is not equivalent to the others.</td><td>proc rank normal=blom out=normals;
var Mercury_TPY;
ranks q;
data normals; set normals;
proc plot; plot Mercury_TPY*q;
run;

proc glm;
class State;
model q=State;
run;</td><td>The GLM Procedure
Dependent Variable: q Rank for Variable Mercury_TPY
<table><tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>Model</td><td>3</td><td>18.8209113</td><td>6.2736371</td><td>6.84</td><td>0.0002</td></tr><tr><td>Error</td><td>222</td><td>203.6672717</td><td>0.9174201</td><td></td><td></td></tr><tr><td>Corrected Total</td><td>225</td><td>222.4881830</td><td></td><td></td><td></td></tr></table></td><td>With a p-value of 0.0002, we would have to reject the null hypothesis and conclude that at least one state's mean lead level is different, whether we used a significance level of 0.01 or 0.10.</td></tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypotheses	SAS Code	SAS Output	Conclusion

	Lead	H_0: The mean lead level across each state is equivalent
		
		H_A: At least one state's mean lead level is not equivalent to the others.
			proc rank normal=blom out=normals;
			var Lead_TPY;
			ranks q;
			data normals; set normals;
			proc plot; plot Lead_TPY*q;
			run;
			
			proc glm;
			class State;
			model q=State;
			run;
				The GLM Procedure
				Dependent Variable: q Rank for Variable Lead_TPY
				
				Source | DF | Sum of Squares | Mean Square | F Value | Pr > F
				Model | 3 | 17.8597976 | 5.9532659 | 6.46 | 0.0003
				Error | 222 | 204.6283854 | 0.9217495 | |
				Corrected Total | 225 | 222.4881830 | | |
					With a p-value of 0.0003, we would have to reject the null hypothesis and conclude that at least one state's mean lead level is different, whether we used a significance level of 0.01 or 0.10.

Pamela Mishaw	Mental Distress	H_0: All state means of mental distress rating are equal.
		
		H_A: At least two of the means of mental distress rating differ at a statistically significant level.
			proc glm;
			class state;
			model mental_distress= state;
			run;
			
			proc rank normal=blom out=normals;
			var mental_distress;
			ranks q;
			data normals; set normals;
			proc plot; plot mental_distress* q; run;
			
			proc glm;
			class state;
			model q=state;
			run;
				The SAS System
				The GLM Procedure
				Dependent Variable: q Rank for Variable mental_distress
				
				Source | DF | Sum of Squares | Mean Square | F Value | Pr > F
				Model | 3 | 60.8734992 | 20.2911664 | 30.41 | <.0001
				Error | 222 | 148.1297635 | 0.6672512 | |
				Corrected Total | 225 | 209.0032628 | | |
					Since the p-value, <0.0001, is less than the significance value of 0.01, we reject the null hypothesis and conclude that the state means are not statistically different. This same conclusion will be made at significance level 0.10. These results are the same as those of part 2.

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypotheses	SAS Code	SAS Output	Conclusion

	% Bachelor Degree	H_0: All state means of bachelor's degree percentages are equal.
		
		H_A: At least two of the means of bachelor's degree percentages differ at a statistically significant level.
			proc glm;
			class state;
			model Percent_with_Bachelors=state;
			run;
			
			proc rank normal=blom out=normals;
			var Percent_with_Bachelors;
			ranks q;
			data normals;
			set normals;
			proc plot;
			plot Percent_with_Bachelors*q ; run;
			
			proc glm;
			class state;
			model q=state;
			run;
				The SAS System
				The GLM Procedure
				Dependent Variable: q Rank for Variable Percent_with_Bachelors
				
				Source | DF | Sum of Squares | Mean Square | F Value | Pr > F
				Model | 3 | 21.3552696 | 7.1184232 | 7.86 | <.0001
				Error | 222 | 201.1202348 | 0.9059470 | | |
				Corrected Total | 225 | 222.4755043 | | |
					Since the p-value, <0.0001, is less than the significance value of 0.01, we reject the null hypothesis and conclude that the state means are not statistically different. This same conclusion will be made at significance level 0.10. These results are the same as those of part 2.

Prince Nhliziyo	Smoking	H_0: All state means are equal.
		
		H_A: At least one state mean is different from the other state means.
			proc rank normal=blom out=normals;
			var smoking;
			ranks q;
			data normals; set normals;
			proc glm;
			class state;
			model q=state;
			run;
				Dependent Variable: q Rank for Variable smoking
				
				Source | DF | Sum of Squares | Mean Square | F Value | Pr > F
				Model | 3 | 85.2205323 | 28.4068441 | 45.94 | <.0001
				Error | 222 | 137.2589645 | 0.6182836 | | |
				Corrected Total | 225 | 222.4794968 | | |
					Based on significance level alpha=0.01,we reject the null hypothesis and conclude that at least one mean is different for smoking rate.
					
					Based on significance level alpha=0.10 ,we reject the null hypothesis and conclude that at least one mean is different for smoking rate

	Inactivity	H_0: All state means are equal.
		
		H_A: At least one state mean is different from the other state means.
			proc rank normal=blom out=normals;
			var physical;
			ranks q;
			data normals; set normals;
			proc glm;
			class state;
			model q=state;
			run;
				Dependent Variable: q Rank for Variable physical
				
				Source | DF | Sum of Squares | Mean Square | F Value | Pr > F
				Model | 3 | 117.3467098 | 39.1155699 | 82.59 | <.0001
				Error | 222 | 105.1390275 | 0.4735992 | | |
				Corrected Total | 225 | 222.4857372 | | |
					Based on significance level alpha=0.01,we reject the null hypothesis and conclude that at least one mean is different for physical inactivity.
					
					Based on significance level alpha=0.10 ,we reject the null hypothesis and conclude that at least one mean is different for physical inactivity.

STA5176: Spring 2023 Group 4 Alzheimer's Project

4. Calculate the correlation coefficient between Alzheimer's and (variable 1) or (variable 2). What can you conclude regarding a linear association between these two variables? Use all values in the experiment and not for each state separately.

<table>
<tr><th>Name</th><th>Var</th><th>Hypotheses</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td rowspan="2">Brad Lipson</td><td>Cancer

Obesity</td><td>H_0: There is not a significant correlation between cancer and obesity

H_A: There is a significant correlation between obesity and cancer.</td><td>proc corr;
var heart diabetes;
run;</td><td>Pearson Correlation Coefficients, N = 226
Prob > |r| under H0: Rho=0
<table><tr><th></th><th>obesity</th><th>cancer</th></tr><tr><td>obesity
obesity</td><td>1.00000</td><td>0.58326
<.0001</td></tr><tr><td>cancer
cancer</td><td>0.58326
<.0001</td><td>1.00000</td></tr></table></td><td>P-value = <.0001
Pearson Correlation Coef = 0.58326

Obesity and cancer have a positive relationship. Additionally, we reject the null hypothesis and can determine that there is a statistically significant correlation between cancer and obesity because the p-value (.0001) is smaller than alpha (0.05).</td></tr>
<tr><td rowspan="2">Mary Morrow</td><td>Heart Disease</td><td>H_0: There is not a significant correlation between heart disease and the age adjusted Alzheimer's disease mortality rate (ad).

H_A: There is a significant correlation between heart disease and the age adjusted Alzheimer's disease mortality rate (ad).</td><td>proc corr;
var heart alz;
run;</td><td>Pearson Correlation Coefficients, N = 226
Prob > |r| under H0: Rho=0
<table><tr><th></th><th>heart</th><th>alz</th></tr><tr><td>heart</td><td>1.00000</td><td>-0.24249
0.0002</td></tr><tr><td>alz</td><td>-0.24249
0.0002</td><td>1.00000</td></tr></table></td><td>P-value = 0.0002
Pearson Correlation Coef = -0.24249

There is a negative correlation between heart disease and the age adjusted Alzheimer's disease mortality rate. Also, since the p-value (0.0002) is less than alpha (0.05), we reject the null hypothesis and can conclude that the correlation between heart disease and the age adjusted Alzheimer's disease mortality rate is statistically significant.</td></tr>
<tr><td>Diabetes</td><td>H_0: There is not a significant correlation between diabetes and the age adjusted Alzheimer's disease mortality rate (ad).

H_A: There is a significant correlation between diabetes and the age adjusted Alzheimer's disease mortality rate (ad).</td><td>proc corr;
var diabetes alz;
run;</td><td>Pearson Correlation Coefficients, N = 226
Prob > |r| under H0: Rho=0
<table><tr><th></th><th>diabetes</th><th>alz</th></tr><tr><td>diabetes</td><td>1.00000</td><td>-0.07639
0.2527</td></tr><tr><td>alz</td><td>-0.07639
0.2527</td><td>1.00000</td></tr></table></td><td>P-value = 0.2527
Pearson Correlation Coef = -0.7639

There is a negative correlation between diabetes and the age adjusted Alzheimer's disease mortality rate. Also, since the p-value (0.2527) is greater than alpha (0.05), we fail to reject the null hypothesis and can conclude that there is sufficient evidence to prove that the correlation between heart disease and the age adjusted Alzheimer's disease mortality rate is not statistically significant.</td></tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypotheses	SAS Code	SAS Output	Conclusion
Matthew McGehee	Mercury	H_0: There is not a significant correlation between mercury levels and lead levels. H_A: There is a significant correlation between mercury levels and lead levels.	proc corr; var Mercury_TPY alz; run;	Pearson Correlation Coefficients, N = 226 Prob > |r| under H0: Rho=0 Mercury_TPY: 1.00000; alz: 0.11411, 0.0870 alz: 0.11411, 0.0870; alz: 1.00000	There is a positive correlation between mercury levels and alz, but it's very weak at 0.11411. Also, with a p-value of 0.087, which is greater than significance level 0.05, we should not reject the null hypothesis, and conclude that there is no linear correlation between mercury and alz.
	Lead	H_0: There is not a significant correlation between mercury levels and lead levels. H_A: There is a significant correlation between mercury levels and lead levels.	proc corr; var Lead_TPY alz; run;	Pearson Correlation Coefficients, N = 226 Prob > |r| under H0: Rho=0 Lead_TPY: 1.00000; alz: 0.16070, 0.0156 alz: 0.16070, 0.0156; alz: 1.00000	There is a positive correlation between mercury levels and alz, but it's very weak at 0.16707. However, with a p-value of 0.0156, which is less than significance level 0.05, we should reject the null hypothesis, and conclude that there could be a linear correlation between lead levels and alz.
Pamela Mishaw	Mental Distress % Bachelor Degree	H_0: There is not a significant correlation between mental distress and percentage of the population with a bachelor's degree. H_A: There is a significant correlation between mental distress and percentage of the population with a bachelor's degree.	proc corr; var mental_distress Percent_with_Bachelors; run;	Pearson Correlation Coefficients, N = 226 Prob > |r| under H0: Rho=0 mental_distress: 1.00000; -0.59572, <.0001 Percent_with_Bachelors: -0.59572, <.0001; 1.00000	There is a strong negative correlation between the percentage of those with bachelor's degrees and mental distress. The p-value is extremely small (<0.0001) which indicates that this correlation is statistically significant.
Prince Nhliziyo	Smoking Inactivity	H_0: There is not a significant correlation between smoking rate and physical inactivity. H_A: There is a significant correlation between smoking rate and physical inactivity.	proc corr; var smoking physical; run;	Pearson Correlation Coefficients, N = 226 Prob > |r| under H0: Rho=0 smoking: 1.00000; 0.70925, <.0001 physical: 0.70925, <.0001; 1.00000	There is a positive correlation between smoking rate and physical inactivity.Based on the p-value,we reject the null hypothesis and conclude that the correlation is significant.

5. Using the raw data, run the Brown-Levene-Forsythe Test to test for equal population variances for the (variable). Then run Bartlettt's Test also on (variable). Which of

STA5176: Spring 2023 Group 4 Alzheimer's Project

these two tests for variability would you recommend using with this data set and why? Do you need to run a normality test first in order to be able to recommend one test over the other? Refer to the p-value of the test.

<table>
<tr><th>Name</th><th>Var</th><th>Hypothese</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td rowspan="2">Brad Lipson</td><td>Cancer</td><td>H_0: All state (group) variance are equal.

H_A: At least one state (group) variance is different from the other state (group) variance.</td><td>proc glm;
class state;
model cancer=state;
means state/hovtest=bf;
run;

proc glm;
class state;
model cancer = state;
means state/hovtest=bartlett ;
run;</td><td>Brown and Forsythe's Test for Homogeneity of cancer Variance
ANOVA of Absolute Deviations from Group Medians
<table><tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>state</td><td>3</td><td>5548.4</td><td>1849.5</td><td>7.08</td><td>0.0001</td></tr><tr><td>Error</td><td>222</td><td>57956.7</td><td>261.1</td><td></td><td></td></tr></table>
Bartlett's Test for Homogeneity of cancer Variance
<table><tr><th>Source</th><th>DF</th><th>Chi-Square</th><th>Pr > ChiSq</th></tr><tr><td>state</td><td>3</td><td>63.7477</td><td><.0001</td></tr></table></td><td>Because the cancer data was tested in (1) and found to be non-normal, Brown-Levene-Forsythe Test should be used to test for equal population variances.Bartlett's test assumes the data is normally distributed The p-value (0.0078) for Bartlett's test is less than alpha (0.05). Therefore, we reject the null hypothesis and conclude that at least one of the four states has different variance for cancer.</td></tr>
<tr><td>Obesity</td><td>H_0: All state (group) variance are equal.

H_A: At least one state (group) variance is different from the other state (group) variance.</td><td>proc glm;
class state;
model obesity=state;
means state/hovtest=bf;
run;

proc glm;
class state;
model obesity = state;
means state/hovtest=bartlett ;
run;</td><td>Brown and Forsythe's Test for Homogeneity of obesity Variance
ANOVA of Absolute Deviations from Group Medians
<table><tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>state</td><td>3</td><td>150.3</td><td>50.1124</td><td>9.57</td><td><.0001</td></tr><tr><td>Error</td><td>222</td><td>1162.0</td><td>5.2342</td><td></td><td></td></tr></table>
Bartlett's Test for Homogeneity of obesity Variance
<table><tr><th>Source</th><th>DF</th><th>Chi-Square</th><th>Pr > ChiSq</th></tr><tr><td>state</td><td>3</td><td>18.3235</td><td>0.0004</td></tr></table></td><td>Before deciding which test to use, you would need to test for normality. Because the obesity data was tested in (1) and found to be non-normal, Brown-Levene-Forsythe Test should be used to test for equal population variances.The Bartlett's test assumes the data is normally distributed. The p-value (0.0078) for Bartlett's test is less than alpha (0.05). Therefore, we reject the null hypothesis and conclude that at least one of the four states has different variance for obesity.</td></tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>Hypothese</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td rowspan="2">Mary Morrow</td><td>Heart Disease</td><td>H_0: All state (group) variance are equal.

H_A: At least one state (group) variance is different from the other state (group) variance.</td><td>proc glm;
class state;
model heart=state;
means state/hovtest=bf;
run;

proc glm;
class state;
model heart = state;
means state/hovtest=bartlett;
run;</td><td>Brown and Forsythe's Test for Homogeneity of heart Variance
ANOVA of Absolute Deviations from Group Medians
<table><tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>state</td><td>3</td><td>1890.8</td><td>630.3</td><td>1.61</td><td>0.1888</td></tr><tr><td>Error</td><td>222</td><td>87115.7</td><td>392.4</td><td></td><td></td></tr></table>
Bartlett's Test for Homogeneity of heart Variance
<table><tr><th>Source</th><th>DF</th><th>Chi-Square</th><th>Pr > ChiSq</th></tr><tr><td>state</td><td>3</td><td>11.8799</td><td>0.0078</td></tr></table></td><td>Before deciding which test to use, you would need to test for normality. This is because Bartlett's test assumes the data is normally distributed. Because the heart disease data was tested in (1) and found to be non-normal, Brown-Levene-Forsythe Test should be used to test for equal population variances.

The p-value (0.0078) for Bartlett's test is less than alpha (0.05). Therefore, we reject the null hypothesis and conclude that at least one of the four states has different variance for heart disease.</td></tr>
<tr><td>Diabetes</td><td>H_0: All state (group) variance are equal.

H_A: At least one state (group) variance is different from the other state (group) variance.</td><td>proc glm;
class state;
model diabetes=state;
means state/hovtest=bf;
run;

proc glm;
class state;
model diabetes = state;
means state/hovtest=bartlett;
run;</td><td>Brown and Forsythe's Test for Homogeneity of diabetes Variance
ANOVA of Absolute Deviations from Group Medians
<table><tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>state</td><td>3</td><td>20.5060</td><td>6.8353</td><td>13.13</td><td><.0001</td></tr><tr><td>Error</td><td>222</td><td>115.6</td><td>0.5205</td><td></td><td></td></tr></table>
Bartlett's Test for Homogeneity of diabetes Variance
<table><tr><th>Source</th><th>DF</th><th>Chi-Square</th><th>Pr > ChiSq</th></tr><tr><td>state</td><td>3</td><td>47.1875</td><td><.0001</td></tr></table></td><td>Before deciding which test to use, you would need to test for normality. This is because Bartlett's test assumes the data is normally distributed. Because the diabetes data was tested in (1) and found to be non-normal, Brown-Levene-Forsythe Test should be used to test for equal population variances.

The p-value (<.0001) for Bartlett's test is less than alpha (0.05). Therefore, we reject the null hypothesis and conclude that at least one of the four states has different variance for diabetes.</td></tr>
<tr><td>Matthew McGehee</td><td>Mercury</td><td>H_0: The variance in mercury levels from each state is equal.

H_A: At least one state's mercury level variance is different from one of the other states' observed variance.</td><td>proc glm data=group4data;
class State;
model Mercury_TPY=State;
means State/hovtest=bf;
run;

proc glm data=group4data;
class State;
model Mercury_TPY = State;
means State/hovtest=bartlett;
run;</td><td>Brown and Forsythe's Test for Homogeneity of Mercury_TPY Variance
ANOVA of Absolute Deviations from Group Medians
<table><tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>State</td><td>3</td><td>0.0428</td><td>0.0143</td><td>2.78</td><td>0.0419</td></tr><tr><td>Error</td><td>222</td><td>1.1387</td><td>0.00513</td><td></td><td></td></tr></table>
Bartlett's Test for Homogeneity of Mercury_TPY Variance
<table><tr><th>Source</th><th>DF</th><th>Chi-Square</th><th>Pr > ChiSq</th></tr><tr><td>State</td><td>3</td><td>166.9</td><td><.0001</td></tr></table></td><td>Because we concluded that the data for mercury levels was not normally distributed, we should use the Brown-Forsythe-Levene test.

With a p-value of 0.0419 and an alpha of 0.05, we would have to reject the null hypothesis and conclude that at least one state's mercury level variance is different from one of the other states observed. However, at alpha 0.01, we should not reject the null hypothesis.</td></tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>Hypothese</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td></td><td>Lead</td><td>H_0: The variance in lead levels from each state is equal.

H_A: At least one state's lead level variance is different from one of the other states' observed variance.</td><td>proc glm
data=group4data;
class State;
model
Lead_TPY=State;
means
State/hovtest=bf;
run;

proc glm
data=group4data;
class State;
model Lead_TPY =
State;
means
State/hovtest=bartlett
;
run;</td><td>Brown and Forsythe's Test for Homogeneity of Lead_TPY Variance
ANOVA of Absolute Deviations from Group Medians
<table><tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>State</td><td>3</td><td>2.6660</td><td>0.8887</td><td>3.80</td><td>0.0110</td></tr><tr><td>Error</td><td>222</td><td>51.9505</td><td>0.2340</td><td></td><td></td></tr></table>
Bartlett's Test for Homogeneity of Lead_TPY Variance
<table><tr><th>Source</th><th>DF</th><th>Chi-Square</th><th>Pr > ChiSq</th></tr><tr><td>State</td><td>3</td><td>61.9572</td><td><.0001</td></tr></table></td><td>Because we concluded that the data for mercury levels was not normally distributed, we should use the Brown-Forsythe-Levene test.

With a p-value of 0.0110 and an alpha of 0.05, we would have to reject the null hypothesis and conclude that at least one state's lead level variance is different from one of the other states observed. However, at alpha 0.01, we should not reject the null hypothesis, but only marginally.</td></tr>
<tr><td>Pamela Mishaw</td><td>Mental Distress</td><td>H_0: All state variances of mental distress rating are equal.

H_A: At least two of the variances of mental distress rating differ at a statistically significant level.</td><td>proc glm;
class state;
model
mental_distress =
state;
means state /
hovtest=bf;
run;

proc glm;
class state;
model
mental_distress =
state;
means state /
hovtest=bartlett;
run;</td><td>The SAS System
The GLM Procedure
Brown and Forsythe's Test for Homogeneity of mental_distress Variance
ANOVA of Absolute Deviations from Group Medians
<table><tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>state</td><td>3</td><td>5.4980</td><td>1.8327</td><td>3.11</td><td>0.0273</td></tr><tr><td>Error</td><td>222</td><td>130.9</td><td>0.5895</td><td></td><td></td></tr></table>
The SAS System
The GLM Procedure
Bartlett's Test for Homogeneity of mental_distress Variance
<table><tr><th>Source</th><th>DF</th><th>Chi-Square</th><th>Pr > ChiSq</th></tr><tr><td>state</td><td>3</td><td>9.6235</td><td>0.0221</td></tr></table></td><td>As determined in Part 1, the mental distress rating data are not normally distributed. Since Bartlett's test is more sensitive to non-normalized data, one should use the Brown-Levene-Forsythe test for this data.

At significance level alpha = 0.05, we would reject the null hypothesis since the p-value (0.0273) is smaller. Thus, we would conclude that at least two of the variances differ statistically.</td></tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypothese	SAS Code	SAS Output	Conclusion

	% Bachelor Degree	H_0: All state variances of percentage of those with bachelor's degree are equal.
		
		H_A: At least two of the variances of the percentage of those with bachelor's degree differ at a statistically significant level.
			proc glm;
			class state;
			model Percent_with_Bachelors = state;
			means state / hovtest=bf;
			run;
			
			proc glm;
			class state;
			model Percent_with_Bachelors = state;
			means state / hovtest=bartlett;
			run;
				The SAS System
				The GLM Procedure
				Brown and Forsythe's Test for Homogeneity of Percent_with_Bachelors Variance ANOVA of Absolute Deviations from Group Medians
				Source | DF | Sum of Squares | Mean Square | F Value | Pr > F
				state | 3 | 71.9891 | 23.9964 | 0.61 | 0.6123
				Error | 222 | 8802.5 | 39.6507 | |
				
				The SAS System
				The GLM Procedure
				Bartlett's Test for Homogeneity of Percent_with_Bachelors Variance
				Source | DF | Chi-Square | Pr > ChiSq
				state | 3 | 1.7827 | 0.6187
					As determined in Part 1, the data on the percentage of those with bachelor's degree are not normally distributed. Since Bartlett's test is more sensitive to non-normalized data, one should use the Brown-Levene-Forsythe test for this data.
					
					At significance level alpha = 0.05, we would fail to reject the null hypothesis since the p-value (0.6123) is larger. Thus, we would conclude that the variances of the data on those with bachelor's degree are statistically equal.

Prince Nhliziyo	Smoking	H_0: All state variances are equal.
		
		H_A: At least one state variance is different from the another state variance.
			proc glm;
			class state;
			model smoking=state;
			means state/hovtest=bf;
			run;
			
			proc glm;
			class state;
			model smoking = state;
			means state/hovtest=bartlett;
			run;
				Brown and Forsythe's Test for Homogeneity of smoking Variance ANOVA of Absolute Deviations from Group Medians
				Source | DF | Sum of Squares | Mean Square | F Value | Pr > F
				state | 3 | 19.5791 | 6.5264 | 1.46 | 0.2264
				Error | 222 | 992.5 | 4.4708 | |
				
				Bartlett's Test for Homogeneity of smoking Variance
				Source | DF | Chi-Square | Pr > ChiSq
				state | 3 | 4.2792 | 0.2329
					I would recommend using the Brown-forsythe test over the Bartlett test because the Bartlett test is sensitive to non-normality.
					
					Based on the p-value from the Brown-forsythe test, we fail to reject the null hypothesis. There is not enough evidence to suggest that at least two state variances are different.

	Inactivity	H_0: All state variances are equal.
		
		H_A: At least one state variance is different from the another state variance
			proc glm;
			class state;
			model physical=state;
			means state/hovtest=bf;
			run;
			
			proc glm;
			class state;
			model physical = state;
			means state/hovtest=bartlett;
			run;
				Brown and Forsythe's Test for Homogeneity of physical Variance ANOVA of Absolute Deviations from Group Medians
				Source | DF | Sum of Squares | Mean Square | F Value | Pr > F
				state | 3 | 157.1 | 52.3595 | 15.01 | <.0001
				Error | 222 | 774.2 | 3.4872 | |
				
				Bartlett's Test for Homogeneity of physical Variance
				Source | DF | Chi-Square | Pr > ChiSq
				state | 3 | 33.6058 | <.0001
					I would recommend using the Brown-forsythe test over the Bartlett test because the Bartlett test is sensitive to non-normality.
					
					Based on the p-value from the Brown-forsythe test, we reject the null hypothesis and conclude that at least one state has different variance from another for physical inactivity.

STA5176: Spring 2023 Group 4 Alzheimer's Project

6. Compare pairwise the population means for each variable in the 4 states with Tukey's Test. What can you conclude based on significance level alpha=0.05? Which states have significantly different means? Discuss.

<table>
<tr><th>Name</th><th>Var</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td rowspan="2">Brad Lipson</td><td>Cancer</td><td><pre>proc glm;
class state;
model cancer =
state;
means state/tukey;
run;</pre></td><td>Comparisons significant at the 0.05 level are indicated by ***.
<table><tr><th>state Comparison</th><th>Difference Between Means</th><th colspan="2">Simultaneous 95% Confidence Limits</th><th></th></tr><tr><td>FL - NY</td><td>8.740</td><td>-1.473</td><td>18.953</td><td></td></tr><tr><td>FL - WA</td><td>10.619</td><td>-1.054</td><td>22.292</td><td></td></tr><tr><td>FL - CA</td><td>23.262</td><td>12.867</td><td>33.656</td><td>***</td></tr><tr><td>NY - FL</td><td>-8.740</td><td>-18.953</td><td>1.473</td><td></td></tr><tr><td>NY - WA</td><td>1.879</td><td>-9.966</td><td>13.724</td><td></td></tr><tr><td>NY - CA</td><td>14.522</td><td>3.935</td><td>25.109</td><td>***</td></tr><tr><td>WA - FL</td><td>-10.619</td><td>-22.292</td><td>1.054</td><td></td></tr><tr><td>WA - NY</td><td>-1.879</td><td>-13.724</td><td>9.966</td><td></td></tr><tr><td>WA - CA</td><td>12.643</td><td>0.641</td><td>24.645</td><td>***</td></tr><tr><td>CA - FL</td><td>-23.262</td><td>-33.656</td><td>-12.867</td><td>***</td></tr><tr><td>CA - NY</td><td>-14.522</td><td>-25.109</td><td>-3.935</td><td>***</td></tr><tr><td>CA - WA</td><td>-12.643</td><td>-24.645</td><td>-0.641</td><td>***</td></tr></table></td><td>Based on the output:
$\mu_{NY} = \mu_{FL}$
$\mu_{NY} \neq \mu_{CA}$
$\mu_{NY} = \mu_{WA}$
$\mu_{FL} \neq \mu_{CA}$
$\mu_{FL} = \mu_{WA}$
$\mu_{CA} = \mu_{WA}$</td></tr>
<tr><td>Obesity</td><td><pre>proc glm;
class state;
model obesity =
state;
means state/tukey;
run;</pre></td><td>Comparisons significant at the 0.05 level are indicated by ***.
<table><tr><th>state Comparison</th><th>Difference Between Means</th><th colspan="2">Simultaneous 95% Confidence Limits</th><th></th></tr><tr><td>FL - WA</td><td>0.3790</td><td>-1.5863</td><td>2.3443</td><td></td></tr><tr><td>FL - NY</td><td>2.5793</td><td>0.8598</td><td>4.2989</td><td>***</td></tr><tr><td>FL - CA</td><td>5.4895</td><td>3.7395</td><td>7.2396</td><td>***</td></tr><tr><td>WA - FL</td><td>-0.3790</td><td>-2.3443</td><td>1.5863</td><td></td></tr><tr><td>WA - NY</td><td>2.2003</td><td>0.2061</td><td>4.1946</td><td>***</td></tr><tr><td>WA - CA</td><td>5.1105</td><td>3.0899</td><td>7.1312</td><td>***</td></tr><tr><td>NY - FL</td><td>-2.5793</td><td>-4.2989</td><td>-0.8598</td><td>***</td></tr><tr><td>NY - WA</td><td>-2.2003</td><td>-4.1946</td><td>-0.2061</td><td>***</td></tr><tr><td>NY - CA</td><td>2.9102</td><td>1.1277</td><td>4.6927</td><td>***</td></tr><tr><td>CA - FL</td><td>-5.4895</td><td>-7.2396</td><td>-3.7395</td><td>***</td></tr><tr><td>CA - WA</td><td>-5.1105</td><td>-7.1312</td><td>-3.0899</td><td>***</td></tr><tr><td>CA - NY</td><td>-2.9102</td><td>-4.6927</td><td>-1.1277</td><td>***</td></tr></table></td><td>Based on the output:
$\mu_{NY} \neq \mu_{FL}$
$\mu_{NY} \neq \mu_{CA}$
$\mu_{NY} \neq \mu_{WA}$
$\mu_{FL} \neq \mu_{CA}$
$\mu_{FL} = \mu_{WA}$
$\mu_{CA} \neq \mu_{WA}$</td></tr>
<tr><td>Mary Morrow</td><td>Heart Disease</td><td><pre>proc glm;
class state;
model heart =
state;
means state/tukey;
run;</pre></td><td>Comparisons significant at the 0.05 level are indicated by ***.
<table><tr><th>state Comparison</th><th>Difference Between Means</th><th colspan="2">Simultaneous 95% Confidence Limits</th><th></th></tr><tr><td>NY - FL</td><td>16.662</td><td>3.532</td><td>29.793</td><td>***</td></tr><tr><td>NY - CA</td><td>30.696</td><td>17.085</td><td>44.307</td><td>***</td></tr><tr><td>NY - WA</td><td>35.460</td><td>20.232</td><td>50.688</td><td>***</td></tr><tr><td>FL - NY</td><td>-16.662</td><td>-29.793</td><td>-3.532</td><td>***</td></tr><tr><td>FL - CA</td><td>14.034</td><td>0.670</td><td>27.397</td><td>***</td></tr><tr><td>FL - WA</td><td>18.798</td><td>3.791</td><td>33.805</td><td>***</td></tr><tr><td>CA - NY</td><td>-30.696</td><td>-44.307</td><td>-17.085</td><td>***</td></tr><tr><td>CA - FL</td><td>-14.034</td><td>-27.397</td><td>-0.670</td><td>***</td></tr><tr><td>CA - WA</td><td>4.764</td><td>-10.666</td><td>20.194</td><td></td></tr><tr><td>WA - NY</td><td>-35.460</td><td>-50.688</td><td>-20.232</td><td>***</td></tr><tr><td>WA - FL</td><td>-18.798</td><td>-33.805</td><td>-3.791</td><td>***</td></tr><tr><td>WA - CA</td><td>-4.764</td><td>-20.194</td><td>10.666</td><td></td></tr></table></td><td>Based on the output:
$\mu_{NY} \neq \mu_{FL}$
$\mu_{NY} \neq \mu_{CA}$
$\mu_{NY} \neq \mu_{WA}$
$\mu_{FL} \neq \mu_{CA}$
$\mu_{FL} \neq \mu_{WA}$
$\mu_{CA} = \mu_{WA}$</td></tr>
</table>

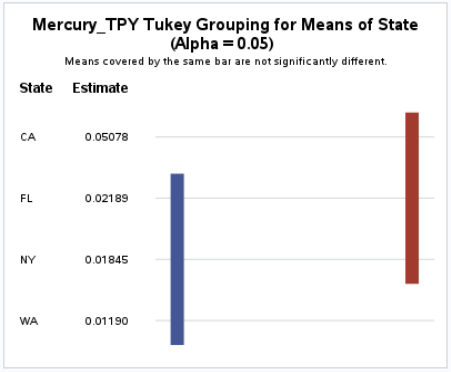
STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td></td><td>Diabetes</td><td>

```
proc glm;
class state;
model diabetes =
state;
means state/tukey;
run;
```

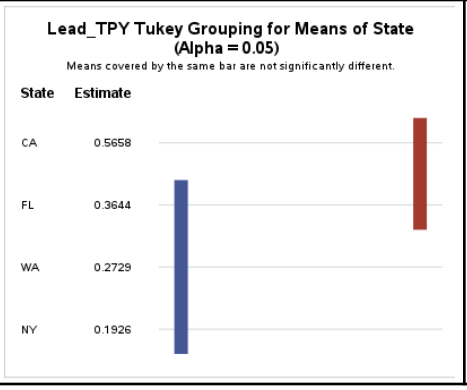
</td><td>
Comparisons significant at the 0.05 level are indicated by ***.

state Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
FL - NY	1.5895	1.0572	2.1218	***
FL - WA	1.6999	1.0915	2.3083	***
FL - CA	2.4904	1.9486	3.0321	***
NY - FL	-1.5895	-2.1218	-1.0572	***
NY - WA	0.1104	-0.5069	0.7278	
NY - CA	0.9009	0.3491	1.4527	***
WA - FL	-1.6999	-2.3083	-1.0915	***
WA - NY	-0.1104	-0.7278	0.5069	
WA - CA	0.7905	0.1650	1.4160	***
CA - FL	-2.4904	-3.0321	-1.9486	***
CA - NY	-0.9009	-1.4527	-0.3491	***
CA - WA	-0.7905	-1.4160	-0.1650	***

</td><td>
Based on the output:
$\mu_{NY} \neq \mu_{FL}$
$\mu_{NY} \neq \mu_{CA}$
$\mu_{NY} = \mu_{WA}$
$\mu_{FL} \neq \mu_{CA}$
$\mu_{FL} \neq \mu_{WA}$
$\mu_{CA} \neq \mu_{WA}$
</td></tr>
<tr><td rowspan="2">Matthew McGehee</td><td>Mercury</td><td>
proc glm
data=group4data;
class State;
model
Mercury_TPY=State;
means State/tukey lines;
run;
</td><td>


</td><td>
At alpha 0.05:

The states with significantly different mean values of mercury tpy are CA and WA.

Florida and New York are not significantly different from any of the other states observed.
</td></tr>
<tr><td>Lead</td><td>
proc glm
data=group4data;
class State;
model Lead_TPY=State;
means State/tukey lines;
run;
</td><td>


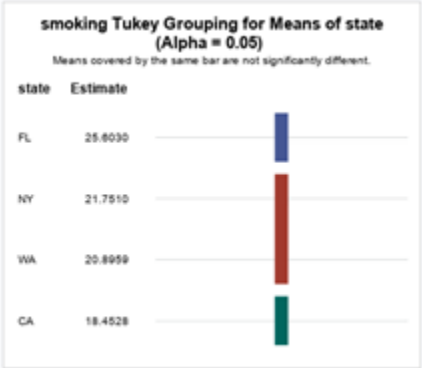
</td><td>
At alpha 0.05, CA has a significantly different mean level of lead tpy compared to either WA or NY.

Only Florida is not significantly different from any of the other states observed.
</td></tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td rowspan="2">Pamela Mishaw</td><td>Mental Distress</td><td>proc glm;
class state;
model mental_distress = state;
means state / tukey cldiff lines;
run;</td><td>
Comparisons significant at the 0.05 level are indicated by ***.
<table>
<tr><th>state Comparison</th><th>Difference Between Means</th><th colspan="2">Simultaneous 95% Confidence Limits</th><th></th></tr>
<tr><td>FL - CA</td><td>0.9743</td><td>0.4850</td><td>1.4635</td><td>***</td></tr>
<tr><td>FL - WA</td><td>1.2599</td><td>0.7104</td><td>1.8093</td><td>***</td></tr>
<tr><td>FL - NY</td><td>1.7429</td><td>1.2622</td><td>2.2236</td><td>***</td></tr>
<tr><td>CA - FL</td><td>-0.9743</td><td>-1.4635</td><td>-0.4850</td><td>***</td></tr>
<tr><td>CA - WA</td><td>0.2856</td><td>-0.2793</td><td>0.8505</td><td></td></tr>
<tr><td>CA - NY</td><td>0.7686</td><td>0.2703</td><td>1.2669</td><td>***</td></tr>
<tr><td>WA - FL</td><td>-1.2599</td><td>-1.8093</td><td>-0.7104</td><td>***</td></tr>
<tr><td>WA - CA</td><td>-0.2856</td><td>-0.8505</td><td>0.2793</td><td></td></tr>
<tr><td>WA - NY</td><td>0.4830</td><td>-0.0745</td><td>1.0406</td><td></td></tr>
<tr><td>NY - FL</td><td>-1.7429</td><td>-2.2236</td><td>-1.2622</td><td>***</td></tr>
<tr><td>NY - CA</td><td>-0.7686</td><td>-1.2669</td><td>-0.2703</td><td>***</td></tr>
<tr><td>NY - WA</td><td>-0.4830</td><td>-1.0406</td><td>0.0745</td><td></td></tr>
</table>
</td><td>At significance level alpha = 0.05, Florida's mean mental distress level differs significantly from those of California, Washington, and New York. Also, California's mean differs significantly from that of New York.

The mean of Washington does not differ significantly from that of California. The mean of Washington does not differ significantly from that of New York.</td></tr>
<tr><td>% Bachelor Degree</td><td>proc glm;
class state;
model Percent_with_Bachelors = state;
means state / tukey cldiff lines;
run;</td><td>
Comparisons significant at the 0.05 level are indicated by ***.
<table>
<tr><th>state Comparison</th><th>Difference Between Means</th><th colspan="2">Simultaneous 95% Confidence Limits</th><th></th></tr>
<tr><td>NY - CA</td><td>0.634</td><td>-3.826</td><td>5.094</td><td></td></tr>
<tr><td>NY - WA</td><td>2.018</td><td>-2.971</td><td>7.008</td><td></td></tr>
<tr><td>NY - FL</td><td>5.942</td><td>1.639</td><td>10.244</td><td>***</td></tr>
<tr><td>CA - NY</td><td>-0.634</td><td>-5.094</td><td>3.826</td><td></td></tr>
<tr><td>CA - WA</td><td>1.384</td><td>-3.672</td><td>6.440</td><td></td></tr>
<tr><td>CA - FL</td><td>5.307</td><td>0.929</td><td>9.686</td><td>***</td></tr>
<tr><td>WA - NY</td><td>-2.018</td><td>-7.008</td><td>2.971</td><td></td></tr>
<tr><td>WA - CA</td><td>-1.384</td><td>-6.440</td><td>3.672</td><td></td></tr>
<tr><td>WA - FL</td><td>3.923</td><td>-0.994</td><td>8.841</td><td></td></tr>
<tr><td>FL - NY</td><td>-5.942</td><td>-10.244</td><td>-1.639</td><td>***</td></tr>
<tr><td>FL - CA</td><td>-5.307</td><td>-9.686</td><td>-0.929</td><td>***</td></tr>
<tr><td>FL - WA</td><td>-3.923</td><td>-8.841</td><td>0.994</td><td></td></tr>
</table>
</td><td>At significance level alpha = 0.05, New York's mean differs significantly from that of Florida and California's mean differs significantly from that of Florida, as well.

The comparisons of the other pairs of state means do not indicate statistical difference at this significance level.</td></tr>
<tr><td>Prince Nhliziyo</td><td>Smoking</td><td>proc glm;
class state;
model smoking = state;
means state/tukey lines;
run;</td><td> 
</td><td>$\mu_{NY} \neq \mu_{FL}$

$\mu_{NY} \neq \mu_{CA}$

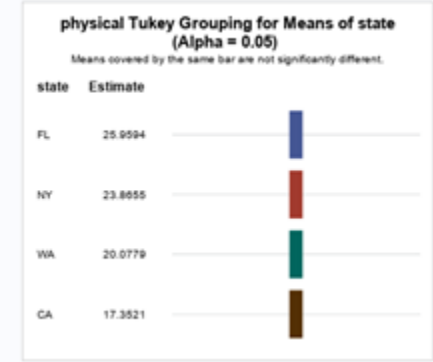
$\mu_{NY} = \mu_{WA}$

$\mu_{FL} \neq \mu_{CA}$

$\mu_{FL} \neq \mu_{WA}$

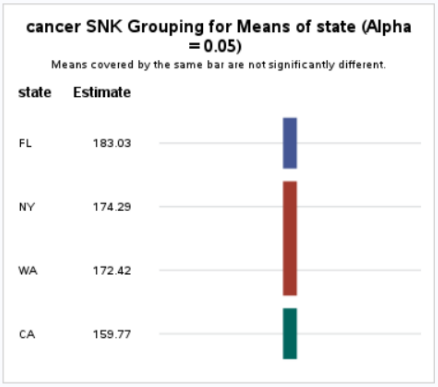
$\mu_{CA} \neq \mu_{WA}$</td></tr>
</table>

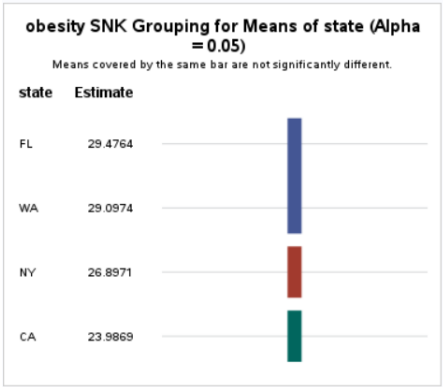
STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	SAS Code	SAS Output	Conclusion
	Inactivity	proc glm; class state; model physical = state; means state/tukey lines; run		$\mu_{NY} \neq \mu_{FL}$ $\mu_{NY} \neq \mu_{CA}$ $\mu_{NY} \neq \mu_{WA}$ $\mu_{FL} \neq \mu_{CA}$ $\mu_{FL} \neq \mu_{WA}$ $\mu_{CA} \neq \mu_{WA}$

7. Compare pairwise the population means for each variable in the 4 states with SNK Test. What can you conclude based on significance level alpha=0.05? Which states have significantly different means? Discuss.

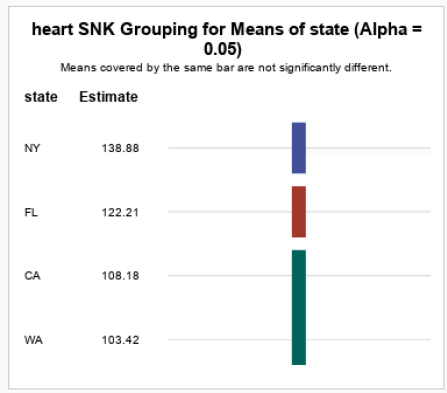
Name	Var	SAS Code	SAS Output	Conclusion

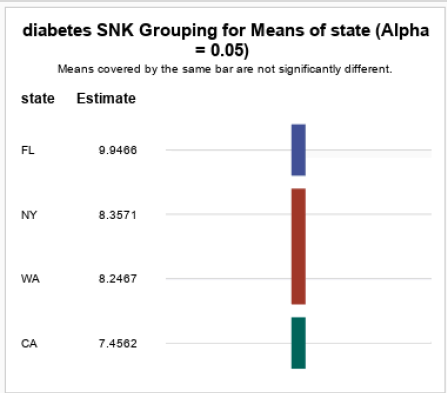
Brad Lipson	Cancer	`proc glm;`
		`class state;`
		`model cancer = state;`
		`means state/SNK;`
		`run;`
				For Cancer, there is a statistically significant difference in the mean between New York and Florida, New York and California, and Florida and California. However, there is no statistically significant difference in the mean between New York and Washington, or between Florida and Washington

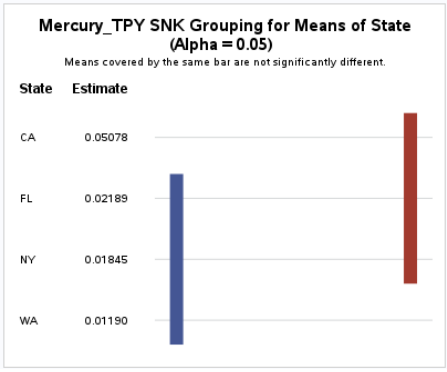
	Obesity	`proc glm;`
		`class state;`
		`model obesity = state;`
		`means state/SNK;`
		`run;`
				For Obesity, there is a statistically significant difference in the mean between New York and Florida, New York and California, and New York and Washington. However, there is no statistically significant difference in the mean between Florida and Washington, but there is a statistically significant difference between California and Washington

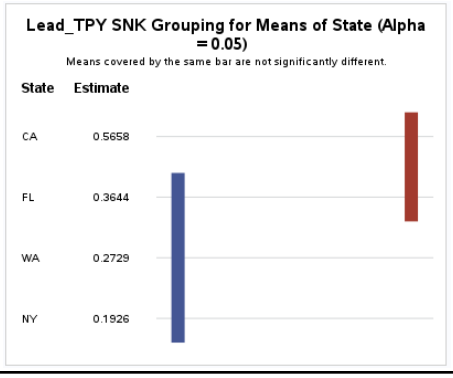
STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	SAS Code	SAS Output	Conclusion

Mary Morrow	Heart Disease	`proc glm;`
		`class state;`
		`model heart = state;`
		`means state/SNK;`
		`run;`
				Based on the output:
				$\mu_{NY} \neq \mu_{FL}$
				$\mu_{NY} \neq \mu_{CA}$
				$\mu_{NY} \neq \mu_{WA}$
				$\mu_{FL} \neq \mu_{CA}$
				$\mu_{FL} \neq \mu_{WA}$
				$\mu_{CA} = \mu_{WA}$

	Diabetes	`proc glm;`
		`class state;`
		`model diabetes = state;`
		`means state/SNK;`
		`run;`
				Based on the output:
				$\mu_{NY} \neq \mu_{FL}$
				$\mu_{NY} \neq \mu_{CA}$
				$\mu_{NY} = \mu_{WA}$
				$\mu_{FL} \neq \mu_{CA}$
				$\mu_{FL} \neq \mu_{WA}$
				$\mu_{CA} \neq \mu_{WA}$

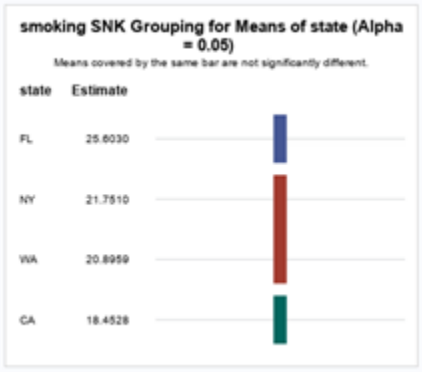
Matthew McGehee	Mercury	proc glm data=group4data;
		class State;
		model Mercury_TPY=State;
		means State/snk lines;
		run;
				At alpha 0.05, CA has a significantly different mean level of mercury tpy compared to either WA or NY.
				
				Only Florida is not significantly different from any of the other states observed.

	Lead	proc glm data=group4data;
		class State;
		model Lead_TPY=State;
		means State/snk lines;
		run;
				At alpha 0.05, CA has a significantly different mean level of lead tpy compared to either WA or NY.
				
				Only Florida is not significantly different from any of the other states observed.

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td rowspan="2">Pamela Mishaw</td><td>Mental Distress</td><td>proc glm;
class state;
model mental_distress = state;
means state / snk linestable;
run;</td><td>Means with the same letter are not significantly different.
<table><tr><th>SNK Grouping</th><th>Mean</th><th>N</th><th>state</th></tr><tr><td>A</td><td>12.3881</td><td>67</td><td>FL</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td>B</td><td>11.4138</td><td>58</td><td>CA</td></tr><tr><td>B</td><td></td><td></td><td></td></tr><tr><td>B</td><td>11.1282</td><td>39</td><td>WA</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td>C</td><td>10.6452</td><td>62</td><td>NY</td></tr></table></td><td>At significance level alpha = 0.05, Florida's mean differs significantly from those of California, Washington, and New York. New York's mean differs significantly from those of California and Washington.

California and Washington have statistically similar means.</td></tr>
<tr><td>% Bachelor Degree</td><td>proc glm;
class state;
model Percent_with_Bachelors = state;
means state / snk linestable;
run;</td><td>Means with the same letter are not significantly different.
<table><tr><th>SNK Grouping</th><th>Mean</th><th>N</th><th>state</th></tr><tr><td>A</td><td>25.777</td><td>62</td><td>NY</td></tr><tr><td>A</td><td></td><td></td><td></td></tr><tr><td>A</td><td>25.143</td><td>58</td><td>CA</td></tr><tr><td>A</td><td></td><td></td><td></td></tr><tr><td>A</td><td>23.759</td><td>39</td><td>WA</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td>B</td><td>19.836</td><td>67</td><td>FL</td></tr></table></td><td>The mean of Florida differs significantly from those of California, New York, and Washington.

The means of New York, California, and Washington are statistically similar.</td></tr>
<tr><td>Prince Nhliziyo</td><td>Smoking</td><td>proc glm;
class state;
model smoking = state;
means state/SNK;
run;</td><td> 
</td><td>$\mu_{NY} \neq \mu_{FL}$

$\mu_{NY} \neq \mu_{CA}$

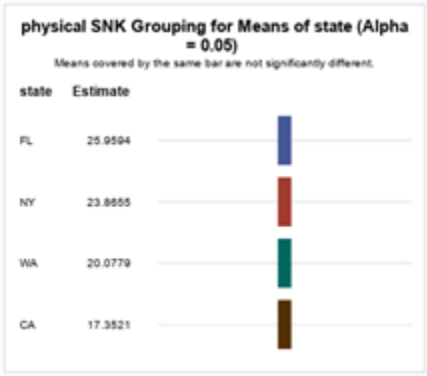
$\mu_{NY} = \mu_{WA}$

$\mu_{FL} \neq \mu_{CA}$

$\mu_{FL} \neq \mu_{WA}$

$\mu_{CA} \neq \mu_{WA}$</td></tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	SAS Code	SAS Output	Conclusion
	Inactivity	proc glm; class state; model physical = state; means state/SNK; run;		$\mu_{NY} \neq \mu_{FL}$ $\mu_{NY} \neq \mu_{CA}$ $\mu_{NY} \neq \mu_{WA}$ $\mu_{FL} \neq \mu_{CA}$ $\mu_{FL} \neq \mu_{WA}$ $\mu_{CA} \neq \mu_{WA}$

8. Compare pairwise the population means for each variable in the 4 states with Fisher's LSD Test. What can you conclude based on significance level alpha=0.05? Which states have significantly different means? Discuss.

Name	Var	SAS Code	SAS Output	Conclusion

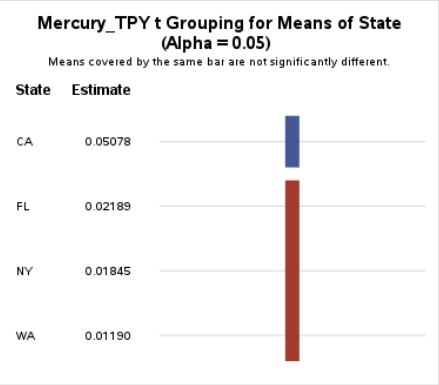
Brad Lipson	Cancer	`proc glm;`
		`class state;`
		`model cancer = state;`
		`means state/lsd;`
		`run;`
			Comparisons significant at the 0.05 level are indicated by ***.
			state Comparison | Difference Between Means | 95% Confidence Limits | |
			FL - NY | 8.740 | 0.964 | 16.515 | ***
			FL - WA | 10.619 | 1.732 | 19.506 | ***
			FL - CA | 23.262 | 15.348 | 31.175 | ***
			NY - FL | -8.740 | -16.515 | -0.964 | ***
			NY - WA | 1.879 | -7.139 | 10.897 |
			NY - CA | 14.522 | 6.462 | 22.582 | ***
			WA - FL | -10.619 | -19.506 | -1.732 | ***
			WA - NY | -1.879 | -10.897 | 7.139 |
			WA - CA | 12.643 | 3.506 | 21.780 | ***
			CA - FL | -23.262 | -31.175 | -15.348 | ***
			CA - NY | -14.522 | -22.582 | -6.462 | ***
			CA - WA | -12.643 | -21.780 | -3.506 | ***
				• The means of the New York and Florida populations are not equal.
				• The means of the New York and California populations are not equal.
				• The means of the New York and Washington populations are not equal.
				• The means of the Florida and California populations are not equal.
				• The means of the Florida and Washington populations are not equal.
				• The means of the California and Washington populations are equal. However, there is no statistically significant difference in the mean between Florida and Washington

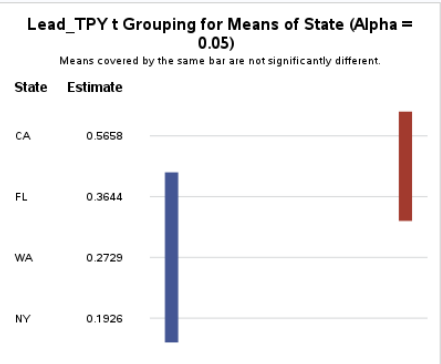
STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td rowspan="2">Mary Morrow</td><td>Obesity</td><td>proc glm;
class state;
model obesity = state;
means state/lsd;
run;</td><td>
Comparisons significant at the 0.05 level are indicated by ***.
<table>
<tr><th>state Comparison</th><th>Difference Between Means</th><th colspan="2">Simultaneous 95% Confidence Limits</th><th></th></tr>
<tr><td>FL - WA</td><td>0.3790</td><td>-1.5863</td><td>2.3443</td><td></td></tr>
<tr><td>FL - NY</td><td>2.5793</td><td>0.8598</td><td>4.2989</td><td>***</td></tr>
<tr><td>FL - CA</td><td>5.4895</td><td>3.7395</td><td>7.2396</td><td>***</td></tr>
<tr><td>WA - FL</td><td>-0.3790</td><td>-2.3443</td><td>1.5863</td><td></td></tr>
<tr><td>WA - NY</td><td>2.2003</td><td>0.2061</td><td>4.1946</td><td>***</td></tr>
<tr><td>WA - CA</td><td>5.1105</td><td>3.0899</td><td>7.1312</td><td>***</td></tr>
<tr><td>NY - FL</td><td>-2.5793</td><td>-4.2989</td><td>-0.8598</td><td>***</td></tr>
<tr><td>NY - WA</td><td>-2.2003</td><td>-4.1946</td><td>-0.2061</td><td>***</td></tr>
<tr><td>NY - CA</td><td>2.9102</td><td>1.1277</td><td>4.6927</td><td>***</td></tr>
<tr><td>CA - FL</td><td>-5.4895</td><td>-7.2396</td><td>-3.7395</td><td>***</td></tr>
<tr><td>CA - WA</td><td>-5.1105</td><td>-7.1312</td><td>-3.0899</td><td>***</td></tr>
<tr><td>CA - NY</td><td>-2.9102</td><td>-4.6927</td><td>-1.1277</td><td>***</td></tr>
</table>
</td><td>There is a statistically significant difference in the mean between New York and Florida, New York and California, and New York and Washington. However, there is no statistically significant difference in the mean between Florida and Washington, but there is a statistically significant difference between California and Washington.</td></tr>
<tr><td>Heart Disease</td><td>proc glm;
class state;
model heart = state;
means state/lsd;
run;</td><td>
Comparisons significant at the 0.05 level are indicated by ***.
<table>
<tr><th>state Comparison</th><th>Difference Between Means</th><th colspan="2">95% Confidence Limits</th><th></th></tr>
<tr><td>NY - FL</td><td>16.662</td><td>6.666</td><td>26.659</td><td>***</td></tr>
<tr><td>NY - CA</td><td>30.696</td><td>20.334</td><td>41.058</td><td>***</td></tr>
<tr><td>NY - WA</td><td>35.460</td><td>23.867</td><td>47.053</td><td>***</td></tr>
<tr><td>FL - NY</td><td>-16.662</td><td>-26.659</td><td>-6.666</td><td>***</td></tr>
<tr><td>FL - CA</td><td>14.034</td><td>3.860</td><td>24.207</td><td>***</td></tr>
<tr><td>FL - WA</td><td>18.798</td><td>7.373</td><td>30.223</td><td>***</td></tr>
<tr><td>CA - NY</td><td>-30.696</td><td>-41.058</td><td>-20.334</td><td>***</td></tr>
<tr><td>CA - FL</td><td>-14.034</td><td>-24.207</td><td>-3.860</td><td>***</td></tr>
<tr><td>CA - WA</td><td>4.764</td><td>-6.983</td><td>16.511</td><td></td></tr>
<tr><td>WA - NY</td><td>-35.460</td><td>-47.053</td><td>-23.867</td><td>***</td></tr>
<tr><td>WA - FL</td><td>-18.798</td><td>-30.223</td><td>-7.373</td><td>***</td></tr>
<tr><td>WA - CA</td><td>-4.764</td><td>-16.511</td><td>6.983</td><td></td></tr>
</table>
</td><td>Based on the output:
$\mu_{NY} \neq \mu_{FL}$
$\mu_{NY} \neq \mu_{CA}$
$\mu_{NY} \neq \mu_{WA}$
$\mu_{FL} \neq \mu_{CA}$
$\mu_{FL} \neq \mu_{WA}$
$\mu_{CA} = \mu_{WA}$</td></tr>
<tr><td>Diabetes</td><td>proc glm;
class state;
model diabetes = state;
means state/lsd;
run;</td><td>
Comparisons significant at the 0.05 level are indicated by ***.
<table>
<tr><th>state Comparison</th><th>Difference Between Means</th><th colspan="2">95% Confidence Limits</th><th></th></tr>
<tr><td>FL - NY</td><td>1.5895</td><td>1.1842</td><td>1.9947</td><td>***</td></tr>
<tr><td>FL - WA</td><td>1.6999</td><td>1.2367</td><td>2.1631</td><td>***</td></tr>
<tr><td>FL - CA</td><td>2.4904</td><td>2.0779</td><td>2.9028</td><td>***</td></tr>
<tr><td>NY - FL</td><td>-1.5895</td><td>-1.9947</td><td>-1.1842</td><td>***</td></tr>
<tr><td>NY - WA</td><td>0.1104</td><td>-0.3596</td><td>0.5804</td><td></td></tr>
<tr><td>NY - CA</td><td>0.9009</td><td>0.4808</td><td>1.3210</td><td>***</td></tr>
<tr><td>WA - FL</td><td>-1.6999</td><td>-2.1631</td><td>-1.2367</td><td>***</td></tr>
<tr><td>WA - NY</td><td>-0.1104</td><td>-0.5804</td><td>0.3596</td><td></td></tr>
<tr><td>WA - CA</td><td>0.7905</td><td>0.3143</td><td>1.2667</td><td>***</td></tr>
<tr><td>CA - FL</td><td>-2.4904</td><td>-2.9028</td><td>-2.0779</td><td>***</td></tr>
<tr><td>CA - NY</td><td>-0.9009</td><td>-1.3210</td><td>-0.4808</td><td>***</td></tr>
<tr><td>CA - WA</td><td>-0.7905</td><td>-1.2667</td><td>-0.3143</td><td>***</td></tr>
</table>
</td><td>Based on the output:
$\mu_{NY} \neq \mu_{FL}$
$\mu_{NY} \neq \mu_{CA}$
$\mu_{NY} = \mu_{WA}$
$\mu_{FL} \neq \mu_{CA}$
$\mu_{FL} \neq \mu_{WA}$
$\mu_{CA} \neq \mu_{WA}$</td></tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

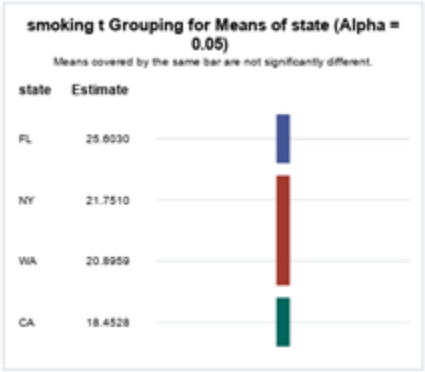
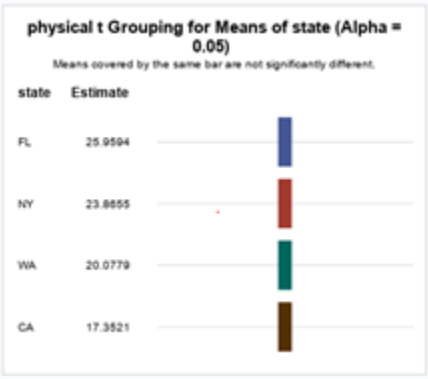
Name	Var	SAS Code	SAS Output	Conclusion

Matthew McGehee	Mercury	proc glm data=group4data;
		class State;
		model Mercury_TPY=State;
		means State/lsd lines;
		run;
				At alpha 0.05:
				
				CA has a significantly different mean level of mercury tpy compared to FL, NY, or WA.
				
				Conversely, we cannot conclude that there is a significant difference in mean levels of mercury tpy between FL, NY, and WA.

	Lead	proc glm data=group4data;
		class State;
		model Lead_TPY=State;
		means State/lsd lines;
		run;
				At alpha 0.05, CA has a significantly different mean level of lead tpy compared to either WA or NY.
				
				Only Florida is not significantly different from any of the other states observed.

Pamela Mishaw	Mental Distress	proc glm; class state; model mental_distress = state; means state / lsd linestable; run;	Means with the same letter are not significantly different. t Grouping | Mean | N | state A | 12.3881 | 67 | FL B | 11.4138 | 58 | CA B B | 11.1282 | 39 | WA C | 10.6452 | 62 | NY	The mean of Florida differs significantly from those of CA, WA, and NY. The mean of NY differs significantly from those of FL, CA, and WA. The means of CA and WA are statistically similar.

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td></td><td>% Bachelor Degree</td><td>proc glm;
class state;
model Percent_with_Bachelors = state;
means state / lsd linestable;
run;</td><td>Means with the same letter are not significantly different.
<table><tr><th>t Grouping</th><th>Mean</th><th>N</th><th>state</th></tr><tr><td>A</td><td>25.777</td><td>62</td><td>NY</td></tr><tr><td>A</td><td></td><td></td><td></td></tr><tr><td>A</td><td>25.143</td><td>58</td><td>CA</td></tr><tr><td>A</td><td></td><td></td><td></td></tr><tr><td>A</td><td>23.759</td><td>39</td><td>WA</td></tr><tr><td></td><td></td><td></td><td></td></tr><tr><td>B</td><td>19.836</td><td>67</td><td>FL</td></tr></table></td><td>The means of Florida differs significantly from those of CA, WA, and NY. The means of WA, CA, and NY are statistically similar.</td></tr>
<tr><td rowspan="2">Prince Nhliziyo</td><td>Smoking</td><td>proc glm;
class state;
model smoking = state;
means state/lsd lines;
run;</td><td> 
</td><td>$\mu_{NY} \neq \mu_{FL}$
$\mu_{NY} \neq \mu_{CA}$
$\mu_{NY} = \mu_{WA}$
$\mu_{FL} \neq \mu_{CA}$
$\mu_{FL} \neq \mu_{WA}$
$\mu_{CA} \neq \mu_{WA}$</td></tr>
<tr><td>Inactivity</td><td>proc glm;
class state;
model physical = state;
means state/lsd lines;
run;</td><td> 
</td><td>$\mu_{NY} \neq \mu_{FL}$
$\mu_{NY} \neq \mu_{CA}$
$\mu_{NY} \neq \mu_{WA}$
$\mu_{FL} \neq \mu_{CA}$
$\mu_{FL} \neq \mu_{WA}$
$\mu_{CA} \neq \mu_{WA}$</td></tr>
</table>

9. Compare the population means for each variable for Florida with the average for the other 3 states with an individual contrast. Does Florida have a significantly different mean than what is found in the other 3 states? Use significance level alpha=0.05.

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypothese	SAS Code	SAS Output	Conclusion

Brad Lipson	Cancer	H_0: L = 0
		H_A: L ≠ 0
			`proc glm;`
			`class state;`
			`model cancer=state;`
			`contrast 'FL vs CA NY WA' state -0.33 1 -0.33 -0.34;`
			`run;`
				Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F
				FL vs CA NY WA | 1 | 9330.218399 | 9330.218399 | 18.61 | <.0001
					The p-value is less than the significance level, which means that the results are statistically significant. Therefore, we can reject the null hypothesis and conclude that there is a difference between the two groups.

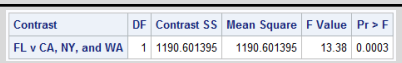
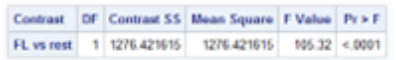
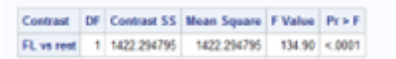
	Obesity	H_0: L = 0
		H_A: L ≠ 0
			`proc glm;`
			`class state;`
			`model obesity=state;`
			`contrast 'FL vs CA NY WA' state -0.33 1 -0.33 -0.34;`
			`run;`
				Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F
				FL vs CA NY WA | 1 | 362.0632714 | 362.0632714 | 25.48 | <.0001
					The p-value is close, which means that the probability of getting the observed results by chance is very low. Therefore, we can reject the null hypothesis and conclude that there is a difference between the two groups.

Mary Morrow	Heart Disease	H_0: L = 0
		H_A: L ≠ 0
			`proc glm;`
			`class state;`
			`model heart=state;`
			`contrast 'FL vs CA NY WA' state -1 3 -1 -1;`
			`run;`
				Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F
				FL vs CA NY WA | 1 | 1352.209484 | 1352.209484 | 1.63 | 0.2028
					The p-value, 0.2028, is greater than alpha, 0.05. Therefore, we fail to reject the null hypothesis and conclude that there is sufficient evidence that L = 0.

	Diabetes	H_0: L = 0
		H_A: L ≠ 0
			`proc glm;`
			`class state;`
			`model diabetes=state;`
			`contrast 'FL vs CA NY WA' state -1 3 -1 -1;`
			`run;`
				Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F
				FL vs CA NY WA | 1 | 172.7802192 | 172.7802192 | 126.89 | <.0001
					The p-value, <.0001, is less than alpha, 0.05. Therefore, we reject the null hypothesis and conclude that there is sufficient evidence that L ≠ 0.

Matthew McGehee	Mercury	H_0: L = 0 H_A: L ≠ 0	proc glm; class State; model Mercury_TPY=State; contrast 'FL vs CA NY WA' State -1 3 -1 -1; run;	Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F FL vs CA NY WA | 1 | 0.00123452 | 0.00123452 | 0.23 | 0.6306	At alpha 0.05, and a p-value of 0.6306, we must not reject the null hypothesis, and conclude that FL does not have a significantly different mean level of mercury when contrasted against CA, NY, and WA.
	Lead		proc glm; class State; model Lead_TPY=State; contrast 'FL vs CA NY WA' State -1 3 -1 -1; run;	Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F FL vs CA NY WA | 1 | 0.01984127 | 0.01984127 | 0.07 | 0.7956	At alpha 0.05, and a p-value of 0.7956, we must not reject the null hypothesis, and conclude that FL does not have a significantly different mean level of lead when contrasted against CA, NY, and WA.

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypothese	SAS Code	SAS Output	Conclusion
Pamela Mishaw	Mental Distress	H_0: L = 0 H_A: L ≠ 0	proc glm; class state; model mental_distress = state; contrast 'FL v CA, NY, and WA' state -1 3 -1 -1; run;	 Contrast / DF / Contrast SS / Mean Square / F Value / Pr > F FL v CA, NY, and WA / 1 / 81.80770402 / 81.80770402 / 73.67 / <.0001	At significance level 0.05, the p-value is less (<0.0001) so we reject the null hypothesis and conclude that FL does have a significantly different mean of mental distress levels than CA, NY, and WA.
	% Bachelor Degree	H_0: L = 0 H_A: L ≠ 0	proc glm; class state; model Percent_with_Bachelors = state; contrast 'FL v CA, NY, and WA' state -1 3 -1 -1; run;	 Contrast / DF / Contrast SS / Mean Square / F Value / Pr > F FL v CA, NY, and WA / 1 / 1190.601395 / 1190.601395 / 13.38 / 0.0003	At significance level 0.05, the p-value is less (0.0003) so we reject the null hypothesis and conclude that FL does have a significantly different mean of percentage of those who have bachelor's degrees than CA, NY, and WA.
Prince Nhliziyo	Smoking	H_0: $\mu_{FL}=(\mu_{CA}+\mu_{WA}+\mu_{NY})/3$ H_A: $\mu_F\neq(\mu_{CA}+\mu_{WA}+\mu_{NY})/3$	proc glm; class state; model smoking = state; means state; contrast 'FL vs rest' state -1 3 -1 -1 ; run;	 Contrast / DF / Contrast SS / Mean Square / F Value / Pr > F FL vs rest / 1 / 1276.421615 / 1276.421615 / 105.32 / <.0001	Based on significance level 0.05, we reject the null hypothesis and conclude that FL has a significantly different mean compared to the other 3 states combined
	Inactivity	H_0: $\mu_{FL}=(\mu_{CA}+\mu_{WA}+\mu_{NY})/3$ H_A: $\mu_F\neq(\mu_{CA}+\mu_{WA}+\mu_{NY})/3$	proc glm; class state; model physical = state; means state; contrast 'FL vs rest' state -1 3 -1 -1 ; run;	 Contrast / DF / Contrast SS / Mean Square / F Value / Pr > F FL vs rest / 1 / 1422.294795 / 1422.294795 / 134.90 / <.0001	Based on significance level 0.05, we reject the null hypothesis and conclude that FL has a significantly different mean compared to the other 3 states combined

10. Use 3 contrasts with the Bonferroni method to separately compare Florida to each of the 3 other states for each variable at alpha=0.05. What can you conclude?

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>Hypothesis</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr>
<td>Brad Lipson</td>
<td>Cancer</td>
<td>H_0: L = 0

H_A: L ≠ 0</td>
<td>

```
proc glm;
class state;
model cancer=state;
contrast 'FL vs. CA' state -1 1 0 0;
contrast 'FL vs. NY' state 0 1 -1 0;
contrast 'FL vs. WA' state 0 1 0 -1;
run;
```

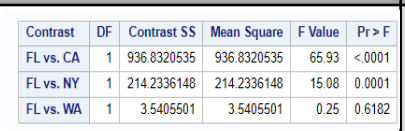
</td>
<td>

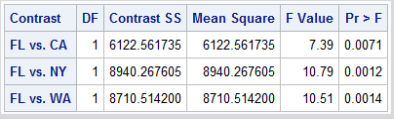
Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
FL vs. CA	1	16822.10505	16822.10505	33.56	< .0001
FL vs. NY	1	2459.75885	2459.75885	4.91	0.0278
FL vs. WA	1	2779.75898	2779.75898	5.55	0.0194

</td>
<td>
1. The p-values for all three comparisons are less than 0.05, which means that the null hypothesis can be rejected. In other words, there is a statistically significant difference between the average values of L in the four states.
2. For FL vs. CA, the p-value is <0.0001, which means that there is less than a 0.0001% chance that the observed difference in average values could have occurred by chance. This suggests that there is a significant difference in the average values of L between Florida and California.
3. For FL vs. NY, the p-value is 0.0278, which means that there is a 2.78% chance that the observed difference in average values could have occurred by chance. This is a less significant result than the FL vs. CA comparison, but it is still statistically significant. This suggests that there is a real difference in the average values of L between Florida and New York, but it is not as large as the difference between Florida and California.
4. For FL vs. WA, the p-value is 0.6182, which means that there is a 61.82% chance that the observed difference in average values could have occurred by chance. This is not a statistically significant result, so we cannot reject the null hypothesis.In conclusion, there are statistically significant differences between the average values of L in Florida and California, Florida and New York, but not between Florida and Washington.
</td>
</tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypothesis	SAS Code	SAS Output	Conclusion

	Obesity	H_0: L = 0
		
		H_A: L ≠ 0
			`proc glm;`
			`class state;`
			`model obesity=state;`
			`contrast 'FL vs. CA' state -1 1 0 0;`
			`contrast 'FL vs. NY' state 0 1 -1 0;`
			`contrast 'FL vs. WA' state 0 1 0 -1;`
			`run;`
					The p-value is the probability of getting a result at least as extreme as the one observed, assuming that the null hypothesis is true. In this case, the null hypothesis is that there is no difference in the mean between the two groups. The p-values for FL vs. CA and FL vs. NY are both less than 0.05, which means that the probability of getting a difference at least as extreme as the one observed is less than 5%. This is considered to be statistically significant, so we can reject the null hypothesis and conclude that there is a difference in the mean between the two groups. The p-value for FL vs. WA is greater than 0.05, which means that the probability of getting a difference at least as extreme as the one observed is greater than 5%. This is not considered to be statistically significant, so we cannot reject the null hypothesis and conclude that there is no difference in the mean between the two groups. There is a statistically significant difference in the mean between FL and CA, and between FL and NY. However, there is no statistically significant difference in the mean between FL and WA

Mary Morrow	Heart Disease	H_0: L = 0
		
		H_A: L ≠ 0
			`proc glm;`
			`class state;`
			`model heart=state;`
			`contrast 'FL vs. CA' state -1 1 0 0;`
			`contrast 'FL vs. NY' state 0 1 -1 0;`
			`contrast 'FL vs. WA' state 0 1 0 -1;`
			`run;`
					Based on the figure provided in the SAS output:
					1. FL vs. CA: The p-value, 0.0071, is less than alpha, 0.05. Therefore, we reject the null hypothesis and conclude that L ≠ 0.
					2. Fl vs. NY: The p-value, 0.0012, is less than alpha, 0.05. Therefore, we reject the null hypothesis and conclude that L ≠ 0.
					3. FL vs. WA: The p-value, 0.0014, is less than alpha, 0.05. Therefore, we reject the null hypothesis and conclude that L ≠ 0.

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
FL vs. CA	1	936.8320535	936.8320535	65.93	<.0001
FL vs. NY	1	214.2336148	214.2336148	15.08	0.0001
FL vs. WA	1	3.5405501	3.5405501	0.25	0.6182

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
FL vs. CA	1	6122.561735	6122.561735	7.39	0.0071
FL vs. NY	1	8940.267605	8940.267605	10.79	0.0012
FL vs. WA	1	8710.514200	8710.514200	10.51	0.0014

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>Hypothesis</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td></td><td>Diabetes</td><td>H_0: L = 0

H_A: L ≠ 0</td><td><pre>proc glm;
class state;
model diabetes=state;
contrast 'FL vs. CA' state -1 1 0 0;
contrast 'FL vs. NY' state 0 1 -1 0;
contrast 'FL vs. WA' state 0 1 0 -1;
run;</pre></td><td><table><tr><th>Contrast</th><th>DF</th><th>Contrast SS</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>FL vs. CA</td><td>1</td><td>192.8044888</td><td>192.8044888</td><td>141.60</td><td><.0001</td></tr><tr><td>FL vs. NY</td><td>1</td><td>81.3545160</td><td>81.3545160</td><td>59.75</td><td><.0001</td></tr><tr><td>FL vs. WA</td><td>1</td><td>71.2328870</td><td>71.2328870</td><td>52.32</td><td><.0001</td></tr></table></td><td>Based on the figure provided in the SAS output:
1. FL vs. CA: The p-value, <.0001, is less than alpha, 0.05. Therefore, we reject the null hypothesis and conclude that L ≠ 0.
2. Fl vs. NY: The p-value, <.0001, is less than alpha, 0.05. Therefore, we reject the null hypothesis and conclude that L ≠ 0.
3. FL vs. WA: The p-value, <.0001, is less than alpha, 0.05. Therefore, we reject the null hypothesis and conclude that L ≠ 0.</td></tr>
<tr><td rowspan="2">Matthew McGehee</td><td>Mercury</td><td>H_0: L = 0

H_A: L ≠ 0</td><td><pre>proc glm;
class State;
model Mercury_TPY=State;
contrast 'FL vs CA' State -1 1 0 0;
contrast 'FL vs NY' State 0 1 -1 0;
contrast 'FL vs WA' State 0 1 0 -1;
run;</pre></td><td><table><tr><th>Contrast</th><th>DF</th><th>Contrast SS</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>FL vs CA</td><td>1</td><td>0.02594372</td><td>0.02594372</td><td>4.87</td><td>0.0283</td></tr><tr><td>FL vs NY</td><td>1</td><td>0.00038212</td><td>0.00038212</td><td>0.07</td><td>0.7890</td></tr><tr><td>FL vs WA</td><td>1</td><td>0.00246217</td><td>0.00246217</td><td>0.46</td><td>0.4972</td></tr></table></td><td>This lines up with all of our earlier tests from Tukey and SNK.

The only state that is significantly different from Florida is California. With a p value of 0.0283, using alpha 0.05, we must reject the null hypothesis of equal means, and conclude that the mean level of mercury is significantly different than Florida's.</td></tr>
<tr><td>Lead</td><td>H_0: L = 0

H_A: L ≠ 0</td><td><pre>proc glm;
class State;
model Lead_TPY=State;
contrast 'FL vs CA' State -1 1 0 0;
contrast 'FL vs NY' State 0 1 -1 0;
contrast 'FL vs WA' State 0 1 0 -1;
run;</pre></td><td><table><tr><th>Contrast</th><th>DF</th><th>Contrast SS</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>FL vs CA</td><td>1</td><td>1.26114867</td><td>1.26114867</td><td>4.27</td><td>0.0399</td></tr><tr><td>FL vs NY</td><td>1</td><td>0.95072922</td><td>0.95072922</td><td>3.22</td><td>0.0740</td></tr><tr><td>FL vs WA</td><td>1</td><td>0.20648542</td><td>0.20648542</td><td>0.70</td><td>0.4037</td></tr></table></td><td>This lines up with all of our earlier tests from Tukey, SNK, and Fisher.

The only state that is significantly different from Florida is California. With a p value of 0.0399, using alpha 0.05, we must reject the null hypothesis of equal means, and conclude that the mean level of lead is significantly different than Florida's.</td></tr>
<tr><td>Pamela Mishaw</td><td>Mental Distress</td><td>H_0: L = 0

H_A: L ≠ 0</td><td><pre>proc glm;
class state;
model mental_distress = state;
means state/bon;
contrast 'FL v CA' state -1 1 0 0;
contrast 'FL v NY' state 0 1 -1 0;
contrast 'FL v WA' state 0 1 0 -1;
run;</pre></td><td>The SAS System
The GLM Procedure
Dependent Variable: mental_distress<table><tr><th>Contrast</th><th>DF</th><th>Contrast SS</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr><tr><td>FL v CA</td><td>1</td><td>29.50858672</td><td>29.50858672</td><td>26.57</td><td><.0001</td></tr><tr><td>FL v NY</td><td>1</td><td>97.81848447</td><td>97.81848447</td><td>88.08</td><td><.0001</td></tr><tr><td>FL v WA</td><td>1</td><td>39.12680429</td><td>39.12680429</td><td>35.23</td><td><.0001</td></tr></table></td><td>Since the p-values of all three contrasts are less than the significance level of 0.05, we reject the null hypothesis and conclude that the mean level of mental distress in FL is not equal to those of CA, NY, or WA.</td></tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypothesis	SAS Code	SAS Output	Conclusion
	% Bachelor Degree	H_0: L = 0 H_A: L ≠ 0	proc glm; class state; model Percent_with_Bachelors = state; means state/bon; contrast 'FL v CA' state -1 1 0 0; contrast 'FL v NY' state 0 1 -1 0; contrast 'FL v WA' state 0 1 0 -1; run;	The SAS System The GLM Procedure Dependent Variable: Percent_with_Bachelors Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F FL v CA | 1 | 875.663409 | 875.663409 | 9.84 | 0.0019 FL v NY | 1 | 1136.798203 | 1136.798203 | 12.78 | 0.0004 FL v WA | 1 | 379.405951 | 379.405951 | 4.27 | 0.0401	The p-values of the three contrasts are less than the significance level 0.05 and so we reject the null hypothesis for these cases and conclude that the mean percentage of those with bachelor's degrees differs significantly from the means of CA, NY, and WA.
Prince Nhliziyo	Smoking	H_0: L = 0 H_A: L ≠ 0	proc glm; class state; model smoking = state; means state/bon; contrast 'FL vs CA' state -1 1 0 0 ; contrast 'FL vs NY' state 0 1 -1 0 ; contrast 'FL vs WA' state 0 1 0 -1 ; run;	Dependent Variable: smoking Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F FL vs CA | 1 | 1589.396954 | 1589.396954 | 131.15 | <.0001 FL vs NY | 1 | 477.807813 | 477.807813 | 39.43 | <.0001 FL vs WA | 1 | 546.182918 | 546.182918 | 45.07 | <.0001	Based on the p-values, the conclusion was:: FL vs CA: significant FL vs NY: significant FL vs WA:significant
	Inactivity	H_0: L = 0 H_A: L ≠ 0	proc glm; class state; model physical = state; means state/bon; contrast 'FL vs CA' state -1 1 0 0 ; contrast 'FL vs NY' state 0 1 -1 0 ; contrast 'FL vs WA' state 0 1 0 -1 ; run;	Dependent Variable: physical Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F FL vs CA | 1 | 2303.191752 | 2303.191752 | 218.46 | <.0001 FL vs NY | 1 | 141.187609 | 141.187609 | 13.39 | 0.0003 FL vs WA | 1 | 852.713214 | 852.713214 | 80.88 | <.0001	Based on the p-values, the conclusion was: FL vs CA: significant FL vs NY: significant FL vs WA:significant

STA5176: Spring 2023 Group 4 Alzheimer's Project

11. Calculate the linear regression between Alzheimer's and (variable 1) or (variable 2). What can you conclude regarding the association between these two variables? Use all values in the experiment and not for each state separately.

<table>
<tr><th>Name</th><th>Var</th><th>Hypothese</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td rowspan="2">Brad Lipson</td><td>Cancer</td><td>Null hypothesis is that there is no association between cancer and Alzheimer's disease

The alternative hypothesis would be that there is an association between cancer and Alzheimer's disease.</td><td>proc reg;
model alz_ageadj_rate=cancer/ lackfit;
run;</td><td>
<table>
<tr><th colspan="6">Analysis of Variance</th></tr>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr>
<tr><td>Model</td><td>1</td><td>0.00842</td><td>0.00842</td><td>0.04</td><td>0.8355</td></tr>
<tr><td>Error</td><td>224</td><td>43.65082</td><td>0.19487</td><td></td><td></td></tr>
<tr><td>Lack of Fit</td><td>224</td><td>43.65082</td><td>0.19487</td><td>.</td><td>.</td></tr>
<tr><td>Pure Error</td><td>0</td><td>0</td><td>.</td><td></td><td></td></tr>
<tr><td>Corrected Total</td><td>225</td><td>43.65924</td><td></td><td></td><td></td></tr>
</table>
<table>
<tr><td>Root MSE</td><td>0.44144</td><td>R-Square</td><td>0.0002</td></tr>
<tr><td>Dependent Mean</td><td>0.95241</td><td>Adj R-Sq</td><td>-0.0043</td></tr>
<tr><td>Coeff Var</td><td>46.34994</td><td></td><td></td></tr>
</table>
<table>
<tr><th colspan="7">Parameter Estimates</th></tr>
<tr><th>Variable</th><th>Label</th><th>DF</th><th>Parameter Estimate</th><th>Standard Error</th><th>t Value</th><th>Pr > |t|</th></tr>
<tr><td>Intercept</td><td>Intercept</td><td>1</td><td>0.99669</td><td>0.21503</td><td>4.64</td><td><.0001</td></tr>
<tr><td>Cancer</td><td>Cancer</td><td>1</td><td>-0.00025624</td><td>0.00123</td><td>-0.21</td><td>0.8355</td></tr>
</table>
</td><td>The results of the analysis show that there is no significant relationship between cancer and Alzheimer's disease. The p-value for the cancer variable is 0.8355, which is greater than 0.05, the standard threshold for statistical significance. This means that there is not enough evidence to conclude that cancer is associated with Alzheimer's disease.</td></tr>
<tr><td>Obesity</td><td>Null hypothesis is that there is no association between obesity and Alzheimer's disease

The alternative hypothesis would be that there is an association between obesity and Alzheimer's disease.</td><td>proc reg;
model alz_ageadj_rate=obesity_age_adj/ lackfit;
run;</td><td>
<table>
<tr><th colspan="6">Analysis of Variance</th></tr>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr>
<tr><td>Model</td><td>1</td><td>0.00030704</td><td>0.00030704</td><td>0.00</td><td>0.9684</td></tr>
<tr><td>Error</td><td>224</td><td>43.65894</td><td>0.19491</td><td></td><td></td></tr>
<tr><td>Lack of Fit</td><td>196</td><td>38.84936</td><td>0.19821</td><td>1.15</td><td>0.3376</td></tr>
<tr><td>Pure Error</td><td>28</td><td>4.80958</td><td>0.17177</td><td></td><td></td></tr>
<tr><td>Corrected Total</td><td>225</td><td>43.65924</td><td></td><td></td><td></td></tr>
</table>
<table>
<tr><td>Root MSE</td><td>0.44148</td><td>R-Square</td><td>0.0000</td></tr>
<tr><td>Dependent Mean</td><td>0.95241</td><td>Adj R-Sq</td><td>-0.0045</td></tr>
<tr><td>Coeff Var</td><td>46.35425</td><td></td><td></td></tr>
</table>
<table>
<tr><th colspan="7">Parameter Estimates</th></tr>
<tr><th>Variable</th><th>Label</th><th>DF</th><th>Parameter Estimate</th><th>Standard Error</th><th>t Value</th><th>Pr > |t|</th></tr>
<tr><td>Intercept</td><td>Intercept</td><td>1</td><td>0.94507</td><td>0.18728</td><td>5.05</td><td><.0001</td></tr>
<tr><td>obesity_age_adj</td><td>obesity_age_adj</td><td>1</td><td>0.00026896</td><td>0.00678</td><td>0.04</td><td>0.9684</td></tr>
</table>
</td><td>The results of the analysis show that there is no significant relationship between obesity and Alzheimer's disease. The p-value for the obesity variable is 0.9684, which is greater than 0.05, the standard threshold for statistical significance.</td></tr>
<tr><td>Mary Morrow</td><td>Heart Disease</td><td>H0: true slope is zero (no association between heart disease and alzheimer's rate)

H1: true slope is not zero (association between heart disease and alzheimer's rate)</td><td>proc reg;
model alz=heart;
run;</td><td>
<table>
<tr><th colspan="6">Analysis of Variance</th></tr>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr>
<tr><td>Model</td><td>1</td><td>2.56724</td><td>2.56724</td><td>13.99</td><td>0.0002</td></tr>
<tr><td>Error</td><td>224</td><td>41.09200</td><td>0.18345</td><td></td><td></td></tr>
<tr><td>Corrected Total</td><td>225</td><td>43.65924</td><td></td><td></td><td></td></tr>
</table>
<table>
<tr><td>Root MSE</td><td>0.42831</td><td>R-Square</td><td>0.0588</td></tr>
<tr><td>Dependent Mean</td><td>0.95241</td><td>Adj R-Sq</td><td>0.0546</td></tr>
<tr><td>Coeff Var</td><td>44.97090</td><td></td><td></td></tr>
</table>
<table>
<tr><th colspan="6">Parameter Estimates</th></tr>
<tr><th>Variable</th><th>DF</th><th>Parameter Estimate</th><th>Standard Error</th><th>t Value</th><th>Pr > |t|</th></tr>
<tr><td>Intercept</td><td>1</td><td>1.35738</td><td>0.11194</td><td>12.13</td><td><.0001</td></tr>
<tr><td>heart</td><td>1</td><td>-0.00338</td><td>0.00090259</td><td>-3.74</td><td>0.0002</td></tr>
</table>
</td><td>The p-value, 0.0002, is less than alpha, 0.05. Therefore, we reject the null hypothesis and conclude that there is a significant association between heart disease and Alzheimer's rate (true slope is not zero).

The slope of the linear regression of Alzheimer's on heart disease is -0.00338. This means that we expect Alzheimer's rate to decrease 0.00338 units if we increase heart disease rate by one unit. The intercept of the linear regression of Alzheimer's on heart disease is 1.35738. This means that we expect Alzheimer's rate to be 1.35738 units when the heart disease rate is 0 units.

The coefficient of determination, R-Square, is 0.0588. This means that only 5.88% of the variability in Alzheimer's is explained by the regression model Alzheimer's on heart disease.</td></tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>Hypothese</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td></td><td>Diabetes</td><td>H0: true slope is zero (no association between diabetes and alzheimer's rate)

H1: true slope is not zero (association between diabetes and alzheimer's rate)</td><td>proc reg;
model alz=diabetes;
run;</td><td>
<table>
<tr><th colspan="6">Analysis of Variance</th></tr>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr>
<tr><td>Model</td><td>1</td><td>0.25478</td><td>0.25478</td><td>1.31</td><td>0.2527</td></tr>
<tr><td>Error</td><td>224</td><td>43.40446</td><td>0.19377</td><td></td><td></td></tr>
<tr><td>Corrected Total</td><td>225</td><td>43.65924</td><td></td><td></td><td></td></tr>
</table>
<table>
<tr><td>Root MSE</td><td>0.44019</td><td>R-Square</td><td>0.0058</td></tr>
<tr><td>Dependent Mean</td><td>0.95241</td><td>Adj R-Sq</td><td>0.0014</td></tr>
<tr><td>Coeff Var</td><td>46.21896</td><td></td><td></td></tr>
</table>
<table>
<tr><th colspan="6">Parameter Estimates</th></tr>
<tr><th>Variable</th><th>DF</th><th>Parameter Estimate</th><th>Standard Error</th><th>t Value</th><th>Pr > |t|</th></tr>
<tr><td>Intercept</td><td>1</td><td>1.14450</td><td>0.17006</td><td>6.73</td><td><.0001</td></tr>
<tr><td>diabetes</td><td>1</td><td>-0.02239</td><td>0.01953</td><td>-1.15</td><td>0.2527</td></tr>
</table>
</td><td>The p-value, 0.2527, is greater than alpha, 0.05. Therefore, we fail to reject the null hypothesis and conclude that there is not a significant association between diabetes and Alzheimer's rate (true slope is zero).

The slope of the linear regression of Alzheimer's on diabetes is -0.02239. This means that we expect Alzheimer's rate to decrease 0.02239 units if we increase diabetes rate by one unit. The intercept of the linear regression of Alzheimer's on diabetes is 1.14450. This means that we expect Alzheimer's rate to be 1.4450 units when the diabetes rate is 0 units.

The coefficient of determination, R-Square, is 0.0058. This means that only 0.58% of the variability in Alzheimer's is explained by the regression model Alzheimer's on diabetes.</td></tr>
<tr><td rowspan="2">Matthew McGehee</td><td>Mercury</td><td>H0: β1 = 0
No statistically significant relationship exists between mercury levels and Alzheimer's

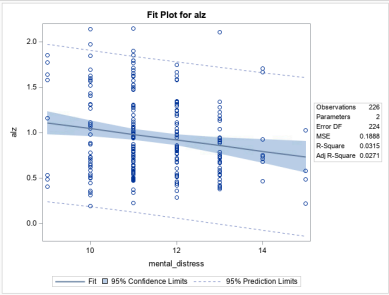
HA: β1 ≠ 0
A relationship does exist between the two.</td><td>proc reg data=group4data;
model alz=Mercury_TPY/ lackfit;
run;</td><td>
<table>
<tr><th colspan="6">Analysis of Variance</th></tr>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr>
<tr><td>Model</td><td>1</td><td>0.56851</td><td>0.56851</td><td>2.96</td><td>0.0870</td></tr>
<tr><td>Error</td><td>224</td><td>43.09074</td><td>0.19237</td><td></td><td></td></tr>
<tr><td>Lack of Fit</td><td>224</td><td>43.09074</td><td>0.19237</td><td>.</td><td>.</td></tr>
<tr><td>Pure Error</td><td>0</td><td>0</td><td>.</td><td></td><td></td></tr>
<tr><td>Corrected Total</td><td>225</td><td>43.65924</td><td></td><td></td><td></td></tr>
</table>
<table>
<tr><td>Root MSE</td><td>0.43860</td><td>R-Square</td><td>0.0130</td></tr>
<tr><td>Dependent Mean</td><td>0.95241</td><td>Adj R-Sq</td><td>0.0086</td></tr>
<tr><td>Coeff Var</td><td>46.05162</td><td></td><td></td></tr>
</table>
<table>
<tr><th colspan="7">Parameter Estimates</th></tr>
<tr><th>Variable</th><th>Label</th><th>DF</th><th>Parameter Estimate</th><th>Standard Error</th><th>t Value</th><th>Pr > |t|</th></tr>
<tr><td>Intercept</td><td>Intercept</td><td>1</td><td>0.93430</td><td>0.03102</td><td>30.12</td><td><.0001</td></tr>
<tr><td>Mercury_TPY</td><td>Mercury_TPY</td><td>1</td><td>0.67987</td><td>0.39548</td><td>1.72</td><td>0.0870</td></tr>
</table>
</td><td>The p-value is 0.087. Using a significance level of 0.05, we would not reject the null hypothesis of a zero slope, and conclude that there is not a statistically significant association between mercury and alz.

The r-square value was 0.0130, which tells us that only 1.3% of the observed alz observations in the regression model can be explained by the coexistence of mercury pollutants, even if no direct association exists.</td></tr>
<tr><td>Lead</td><td>H0: β1 = 0
No statistically significant relationship exists between lead levels and Alzheimer's

HA: β1 ≠ 0
A relationship does exist</td><td>proc reg data=group4data;
model alz=Lead_TPY/ lackfit;
run;</td><td>
<table>
<tr><th colspan="6">Analysis of Variance</th></tr>
<tr><th>Source</th><th>DF</th><th>Sum of Squares</th><th>Mean Square</th><th>F Value</th><th>Pr > F</th></tr>
<tr><td>Model</td><td>1</td><td>1.12749</td><td>1.12749</td><td>5.94</td><td>0.0156</td></tr>
<tr><td>Error</td><td>224</td><td>42.53175</td><td>0.18987</td><td></td><td></td></tr>
<tr><td>Lack of Fit</td><td>224</td><td>42.53175</td><td>0.18987</td><td>.</td><td>.</td></tr>
<tr><td>Pure Error</td><td>0</td><td>0</td><td>.</td><td></td><td></td></tr>
<tr><td>Corrected Total</td><td>225</td><td>43.65924</td><td></td><td></td><td></td></tr>
</table>
<table>
<tr><td>Root MSE</td><td>0.43575</td><td>R-Square</td><td>0.0258</td></tr>
<tr><td>Dependent Mean</td><td>0.95241</td><td>Adj R-Sq</td><td>0.0215</td></tr>
<tr><td>Coeff Var</td><td>45.75195</td><td></td><td></td></tr>
</table>
<table>
<tr><th colspan="7">Parameter Estimates</th></tr>
<tr><th>Variable</th><th>Label</th><th>DF</th><th>Parameter Estimate</th><th>Standard Error</th><th>t Value</th><th>Pr > |t|</th></tr>
<tr><td>Intercept</td><td>Intercept</td><td>1</td><td>0.90758</td><td>0.03433</td><td>26.44</td><td><.0001</td></tr>
<tr><td>Lead_TPY</td><td>Lead_TPY</td><td>1</td><td>0.12693</td><td>0.05209</td><td>2.44</td><td>0.0156</td></tr>
</table>
</td><td>The p-value is 0.0156. Using a significance level of 0.05, we would reject the null hypothesis of a zero slope, and conclude that there is a statistically significant association between lead and alz.

The r-square value was 0.0258, which tells us that only 2.58% of the observed alz observations in the regression model can be explained by the lead.</td></tr>
</table>

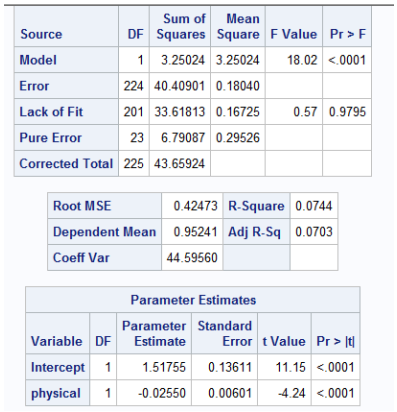
STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypothese	SAS Code	SAS Output	Conclusion
Pamela Mishaw	Mental Distress	H_0: $\beta 1 = 0$; there is no association between mental distress level and Alzheimer's disease rate. H_A: $\beta 1 \neq 0$; there is an association between mental distress level and Alzheimer's disease rate.	proc reg; model alz=mental_distress/lackfit; run;	**Analysis of Variance** Source | DF | Sum of Squares | Mean Square | F Value | Pr > F Model | 1 | 1.37388 | 1.37388 | 7.28 | 0.0075 Error | 224 | 42.28536 | 0.18877 | | Lack of Fit | 5 | 0.53769 | 0.10754 | 0.56 | 0.7274 Pure Error | 219 | 41.74767 | 0.19063 | | Corrected Total | 225 | 43.65924 | | | Root MSE 0.43448 R-Square 0.0315 Dependent Mean 0.95241 Adj R-Sq 0.0271 Coeff Var 45.61923 **Parameter Estimates** Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > |t| Intercept | 1 | 1.66957 | 0.26740 | 6.24 | <.0001 mental_distress | 1 | -0.06268 | 0.02323 | -2.70 | 0.0075 	The F-value for the linear regression model is 7.28 and is statistically significant since the p-value is relatively small (0.0075) which suggests that we reject the null hypothesis and conclude that there is a significant relationship between the mental distress level and Alzheimer's rate. The R-Square value (coefficient of determination) is not close to 1.0 at a value of 0.0315 which indicates that very little (about 3.15%) of the variation in Alzheimer's disease rates is influenced by mental distress level.. The intercept indicates that at a mental distress level of zero, the Alzheimer's rate is 1.66957. The p-value is <0.0001 which suggests this is significant. The slope value indicates that as the mental distress level increases by one unit, the rate of Alzheimer's disease decreases by 0.06268. The p-value is 0.0075 so this parameter is statistically significant. The MSE is small (0.43448) which suggests a spread of data that follows the model relatively well though the data points on the fit plot visually do not seem to fit the linear regression model closely.

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypothese	SAS Code	SAS Output	Conclusion
	% Bachelor Degree	H_0: $\beta 1 = 0$; there is no association between the percentage of those holding a bachelor's degree and Alzheimer's disease rate. H_A: $\beta 1 \neq 0$; there is an association between the percentage of those holding a bachelor's degree and Alzheimer's disease rate.	proc reg; model alz=Percent_with_Bachelors/lackfit; run;	Analysis of Variance Source | DF | Sum of Squares | Mean Square | F Value | Pr > F Model | 1 | 0.49060 | 0.49060 | 2.55 | 0.1120 Error | 224 | 43.16864 | 0.19272 | | | Lack of Fit | 160 | 33.08772 | 0.20680 | 1.31 | 0.1071 Pure Error | 64 | 10.08092 | 0.15751 | | | Corrected Total | 225 | 43.65924 | | | | Root MSE | 0.43900 | R-Square | 0.0112 Dependent Mean | 0.95241 | Adj R-Sq | 0.0068 Coeff Var | 46.09323 | | | Parameter Estimates Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > |t| Intercept | 1 | 0.83914 | 0.07676 | 10.93 | <.0001 Percent_with_Bachelors | 1 | 0.00482 | 0.00302 | 1.60 | 0.1120 	The F-value for the linear regression model is relatively small (2.55) and is not statistically significant since the p-value is relatively large (0.1120) which suggests that we fail to reject the null hypothesis and conclude that there isn't a significant relationship between the percentage of those who have earned a bachelor's degree and Alzheimer's rate. The R-Square value (coefficient of determination) is not close to 1.0 at a value of 0.0112 which indicates that very little (about 1.12%) of the variation in Alzheimer's disease rates is influenced by the percentage of those with a bachelor's degree. The intercept indicates that at a bachelor's degree percentage of zero, the Alzheimer's rate is 0.83914. The p-value is <0.0001 which suggests this is significant. The slope value indicates that as the percentage of bachelor's degree holders increases by one unit, the rate of Alzheimer's disease increases by 0.00482. The p-value is 0.112 so this parameter is not statistically significant. The MSE is small (0.19272) which suggests a spread of data that follows the model relatively well though the data points on the fit plot visually do not seem to fit the linear regression model closely.
Prince Nhliziyo	Smoking	H_0: A significant relationship does not exist between alz and smoking rate. H_A: A significant relationship exists between alz and smoking rate.	proc reg; model alz=smoking/ lackfit; run;	Analysis of Variance Source | DF | Sum of Squares | Mean Square | F Value | Pr > F Model | 1 | 2.28365 | 2.28365 | 12.36 | 0.0005 Error | 224 | 41.37559 | 0.18471 | | | Lack of Fit | 197 | 37.75556 | 0.19165 | 1.43 | 0.1359 Pure Error | 27 | 3.62003 | 0.13408 | | | Corrected Total | 225 | 43.65924 | | | | Root MSE | 0.42978 | R-Square | 0.0523 Dependent Mean | 0.95241 | Adj R-Sq | 0.0481 Coeff Var | 45.12582 | | | Parameter Estimates Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > |t| Intercept | 1 | 1.45480 | 0.14571 | 9.98 | <.0001 smoking | 1 | -0.02294 | 0.00652 | -3.52 | 0.0005	The p-value is 0.0005 which means that we reject the null hypothesis (based on a significance level of 0.05) and conclude that there is a relationship between alzheimer's disease and smoking rate. The coefficient of determination (R-square) is 0.0523 which indicates that only 5.23% of the variation in alzheimer's disease can be explained by smoking rate.

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypothese	SAS Code	SAS Output	Conclusion
	Inactivity	H_0: A significant relationship does not exist between alz and physical inactivity. H_A: A significant relationship exists between alz and physical inactivity..	proc reg; model alz=physical/ lackfit; run;		The p-value is <0.001 which means that we reject the null hypothesis (based on a significance level of 0.05) and conclude that there is a relationship between alzheimer's disease and physical inactivity. The coefficient of determination (R-square) is 0.0703 which indicates that only 7.03% of the variation in alzheimer's disease can be explained by physical inactivity.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	3.25024	3.25024	18.02	<.0001
Error	224	40.40901	0.18040		
Lack of Fit	201	33.61813	0.16725	0.57	0.9795
Pure Error	23	6.79087	0.29526		
Corrected Total	225	43.65924			

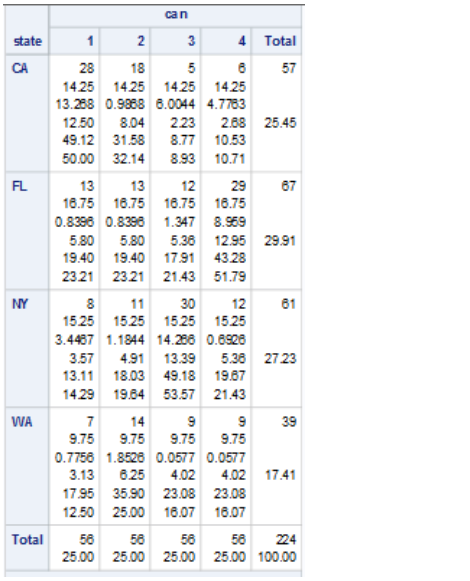
Root MSE	0.42473	R-Square	0.0744
Dependent Mean	0.95241	Adj R-Sq	0.0703
Coeff Var	44.59560		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.51755	0.13611	11.15	<.0001
physical	1	-0.02550	0.00601	-4.24	<.0001

12. What can we conclude on the basis of the data that (variable 1) or (variable 2) come from different samples (Test for Homogeneity)?

Name	Var	Hypotheses	SAS Code	SAS Output	Conclusion

Brad Lipson	Cancer	H_0: The distribution of cancer rate is the same for the states of CA,FL, NY and WA.
		
		H_A: The distribution of cancer rate differs for the states of CA, FL, NY, WA.
			proc univariate;
			var cancer;
			run;
			if cancer le 158.428 then can=1;
			if cancer gt 158.429 and cancer le 171.542 then can=2;
			if cancer gt 171.543 and cancer le 184.359 then can=3;
			if cancer gt 184.36 then can=4;
					The table shows that California and Florida had the highest number of cancer cases, followed by New York and Washington.The results show that there is a significant relationship between state and cancer and the distribution varies by state Therefore, we would have to conclude that the levels of Cancer are not homogenous across the four states at significance level 0.05.

	can				
state	1	2	3	4	Total
CA	28 14.25 13.288 12.50 49.12 50.00	18 14.25 0.9868 8.04 31.58 32.14	5 14.25 6.0044 2.23 8.77 8.93	6 14.25 4.7763 2.68 10.53 10.71	57 25.45
FL	13 16.75 0.8396 5.80 19.40 23.21	13 16.75 0.8396 5.80 19.40 23.21	12 16.75 1.347 5.36 17.91 21.43	29 16.75 8.959 12.95 43.28 51.79	67 29.91
NY	8 15.25 3.4467 3.57 13.11 14.29	11 15.25 1.1844 4.91 18.03 19.64	30 15.25 14.266 13.39 49.18 53.57	12 15.25 0.6926 5.36 19.67 21.43	61 27.23
WA	7 9.75 0.7756 3.13 17.95 12.50	14 9.75 1.8526 6.25 35.90 25.00	9 9.75 0.0577 4.02 23.08 16.07	9 9.75 0.0577 4.02 23.08 16.07	39 17.41
Total	56 25.00	56 25.00	56 25.00	56 25.00	224 100.00

Statistic	DF	Value	Prob
Chi-Square	9	59.3539	<.0001
Likelihood Ratio Chi-Square	9	55.9802	<.0001
Mantel-Haenszel Chi-Square	1	10.2368	0.0014
Phi Coefficient		0.5148	
Contingency Coefficient		0.4577	
Cramer's V		0.2972	

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypotheses	SAS Code	SAS Output	Conclusion
	Obesity	H_0: The distribution of obesity rate is the same for the states of CA,FL, NY and WA H_A: The distribution of obesity rate differs for the states of CA, FL, NY, WA.	proc univariate; var obesity run; if obesity le 18.56 then ob=1; if obesity gt 18.56 and obesity le 22.42 then ob=2; if obesity gt 22.42 and obesity le 24.88 then ob=3; if obesity gt 24.88 then ob=4;	Table of state by ob ob state | 1 | 2 | 3 | 4 | Total CA | 2 | 16 | 19 | 21 | 58 1.2832 | 6.4159 | 8.7257 | 41.575 0.4004 | 14.317 | 12.098 | 10.183 0.88 | 7.08 | 8.41 | 9.29 | 25.66 3.45 | 27.59 | 32.76 | 36.21 40.00 | 64.00 | 55.88 | 12.96 FL | 0 | 7 | 5 | 55 | 67 1.4823 | 7.4115 | 10.08 | 48.027 1.4823 | 0.0228 | 2.5599 | 1.0125 0.00 | 3.10 | 2.21 | 24.34 | 29.65 0.00 | 10.45 | 7.46 | 82.09 0.00 | 28.00 | 14.71 | 33.95 NY | 2 | 1 | 8 | 51 | 62 1.3717 | 6.8584 | 9.3274 | 44.442 0.2878 | 5.0042 | 0.1889 | 0.9676 0.88 | 0.44 | 3.54 | 22.57 | 27.43 3.23 | 1.61 | 12.90 | 82.26 40.00 | 4.00 | 23.53 | 31.48 WA | 1 | 1 | 2 | 35 | 39 0.8628 | 4.3142 | 5.8673 | 27.956 0.0218 | 2.546 | 2.549 | 1.775 0.44 | 0.44 | 0.88 | 15.49 | 17.26 2.56 | 2.56 | 5.13 | 89.74 20.00 | 4.00 | 5.88 | 21.60 Total | 5 | 25 | 34 | 162 | 226 2.21 | 11.06 | 15.04 | 71.68 | 100.00 Statistics for Table of state by ob Statistic | DF | Value | Prob Chi-Square | 9 | 87.6453 | <.0001 Likelihood Ratio Chi-Square | 9 | 85.5933 | <.0001 Mantel-Haenszel Chi-Square | 1 | 26.5957 | <.0001 Phi Coefficient | | 0.6227 | Contingency Coefficient | | 0.5286 | Cramer's V | | 0.3595 |	The results show that there is a significant relationship between state and obesity and the distribution of obesity varies by state. In California, there are more people in the Q1 and Q4 quartiles than in the Q2 and Q3 quartiles. In Florida, there are more people in the Q2 and Q3 quartiles than in the Q1 and Q4 quartiles. In New York, there are more people in the Q1 and Q3 quartiles than in the Q2 quartile. And in Washington, there are more people in the Q2 quartile than in the Q1 and Q3 quartiles.These results suggest that there may be some factors that are unique to each state that are influencing the distribution of obesity. Therefore, we would have to conclude that the levels of obesity are not homogenous across the four states at significance level 0.05.

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>Hypotheses</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td rowspan="2">Mary Morrow</td><td>Heart Disease</td><td>H0: Distribution for heart disease is the same for each state.

H1: Distribution for heart disease is not the same for each state.</td><td>proc univariate;
var heart;
run;

if heart le 99.8 then HD = 1;
if heart gt 99.8 and heart le 116.1 then HD = 2;
if heart gt 116.1 and heart le 136.3 then HD = 3;
if heart gt 136.3 then HD = 4;

proc freq;
table state*HD/chisq cellchi2 expected;
run;</td><td>Frequency
Expected
Cell Chi-Square
Percent
Row Pct
Col Pct

Table of state by HD

<table>
<tr><th>state</th><th>HD 1</th><th>2</th><th>3</th><th>4</th><th>Total</th></tr>
<tr><td>CA</td><td>20
14.628
1.9725
8.85
34.48
35.09</td><td>18
14.372
0.916
7.96
31.03
32.14</td><td>12
14.628
0.4722
5.31
20.69
21.05</td><td>8
14.372
2.8249
3.54
13.79
14.29</td><td>58

25.66

</td></tr>
<tr><td>FL</td><td>14
16.898
0.4971
6.19
20.90
24.56</td><td>15
16.602
0.1545
6.64
22.39
26.79</td><td>22
16.898
1.5403
9.73
32.84
38.60</td><td>16
16.602
0.0218
7.08
23.88
28.57</td><td>67

29.65

</td></tr>
<tr><td>NY</td><td>3
15.637
10.213
1.33
4.84
5.26</td><td>13
15.363
0.3634
5.75
20.97
23.21</td><td>18
15.637
0.357
7.96
29.03
31.58</td><td>28
15.363
10.395
12.39
45.16
50.00</td><td>62

27.43

</td></tr>
<tr><td>WA</td><td>20
9.8363
10.502
8.85
51.28
35.09</td><td>10
9.6637
0.0117
4.42
25.64
17.86</td><td>5
9.8363
2.3779
2.21
12.82
8.77</td><td>4
9.6637
3.3194
1.77
10.26
7.14</td><td>39

17.26

</td></tr>
<tr><td>Total</td><td>57
25.22</td><td>56
24.78</td><td>57
25.22</td><td>56
24.78</td><td>226
100.00</td></tr>
</table>

Statistics for Table of state by HD

<table>
<tr><th>Statistic</th><th>DF</th><th>Value</th><th>Prob</th></tr>
<tr><td>Chi-Square</td><td>9</td><td>45.9387</td><td><.0001</td></tr>
<tr><td>Likelihood Ratio Chi-Square</td><td>9</td><td>48.2829</td><td><.0001</td></tr>
<tr><td>Mantel-Haenszel Chi-Square</td><td>1</td><td>0.2262</td><td>0.6344</td></tr>
<tr><td>Phi Coefficient</td><td></td><td>0.4509</td><td></td></tr>
<tr><td>Contingency Coefficient</td><td></td><td>0.4110</td><td></td></tr>
<tr><td>Cramer's V</td><td></td><td>0.2603</td><td></td></tr>
</table></td><td>The p-value, <.0001, is less than alpha, 0.05. Therefore, we reject the null hypothesis and conclude that the distribution for heart disease is not the same for each state.

There are two main take-aways from the output of SAS:
1. Washington (WA) had the highest percentage (51.28%) for the lowest heart disease rate.
2. New York (NY) had the highest percentage (45.16%) for the highest heart disease rate.</td></tr>
<tr><td>Diabetes</td><td>H0: Distribution for diabetes is the same for each state.

H1: Distribution for diabetes is not the same for each state.</td><td>proc univariate;
var diabetes;
run;

if diabetes le 7.58 then DIA = 1;
if diabetes gt 7.58 and heart le 8.36 then DIA = 2;
if diabetes gt 8.36 and heart le 9.20 then DIA = 3;
if diabetes gt 9.20 then DIA = 4;

proc freq;
table state*DIA/chisq cellchi2 expected;
run;</td><td>Frequency
Expected
Cell Chi-Square
Percent
Row Pct
Col Pct

Table of state by DIA

<table>
<tr><th>state</th><th>DIA 1</th><th>2</th><th>3</th><th>4</th><th>Total</th></tr>
<tr><td>CA</td><td>33
14.885
22.046
14.60
56.90
56.90</td><td>16
14.628
0.1286
7.08
27.59
28.07</td><td>8
14.115
2.6492
3.54
13.79
14.55</td><td>1
14.372
12.441
0.44
1.72
1.79</td><td>58

25.66

</td></tr>
<tr><td>FL</td><td>5
17.195
8.6486
2.21
7.46
8.62</td><td>5
16.898
8.3777
2.21
7.46
8.77</td><td>14
16.305
0.3259
6.19
20.90
25.45</td><td>43
16.602
41.975
19.03
64.18
76.79</td><td>67

29.65

</td></tr>
<tr><td>NY</td><td>10
15.912
2.1963
4.42
16.13
17.24</td><td>23
15.637
3.4668
10.18
37.10
40.35</td><td>24
15.088
5.2633
10.62
38.71
43.64</td><td>5
15.363
6.9901
2.21
8.06
8.93</td><td>62

27.43

</td></tr>
<tr><td>WA</td><td>10
10.009
782E-8
4.42
25.64
17.24</td><td>13
9.8363
1.0176
5.75
33.33
22.81</td><td>9
9.4912
0.0254
3.98
23.08
16.36</td><td>7
9.6637
0.7342
3.10
17.95
12.50</td><td>39

17.26

</td></tr>
<tr><td>Total</td><td>58
25.66</td><td>57
25.22</td><td>55
24.34</td><td>56
24.78</td><td>226
100.00</td></tr>
</table>

Statistics for Table of state by DIA

<table>
<tr><th>Statistic</th><th>DF</th><th>Value</th><th>Prob</th></tr>
<tr><td>Chi-Square</td><td>9</td><td>116.2866</td><td><.0001</td></tr>
<tr><td>Likelihood Ratio Chi-Square</td><td>9</td><td>115.3214</td><td><.0001</td></tr>
<tr><td>Mantel-Haenszel Chi-Square</td><td>1</td><td>3.5947</td><td>0.0580</td></tr>
<tr><td>Phi Coefficient</td><td></td><td>0.7173</td><td></td></tr>
<tr><td>Contingency Coefficient</td><td></td><td>0.5829</td><td></td></tr>
<tr><td>Cramer's V</td><td></td><td>0.4141</td><td></td></tr>
</table></td><td>The p-value, <.0001, is less than alpha, 0.05. Therefore, we reject the null hypothesis and conclude that the distribution for diabetes is not the same for each state.

There are two main take-aways from the output of SAS:
3. California (CA) had the highest percentage (56.90%) for the lowest diabetes rate.
4. Florida (NY) had the highest percentage (64.18%) for the highest diabetes rate.</td></tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>Hypotheses</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr><td rowspan="2">Matthew McGehee</td><td>Mercury across all 4 states</td><td>H_0: Mercury levels are evenly distributed between FL, CA, NY, and WA

H_A: Mercury levels are not evenly distributed between FL, CA, NY, and WA</td><td>proc univariate;
var Mercury_TPY;
run;

if mercury_tpy le 0.001903 then Mercury='Q1';
if mercury_tpy gt 0.001903 and mercury_tpy le 0.004622 then Mercury='Q2';
if mercury_tpy gt 0.004622 and mercury_tpy le 0.019052 then Mercury='Q3';
if mercury_tpy gt 0.019052 then Mercury='Q4';

proc freq;
table state*Mercury/chisq cellchi2 expected;
run;</td><td>Frequency
Expected
Cell Chi-Square
Percent
Row Pct
Col Pct

Table of state by Mercury

<table>
<tr><th>state</th><th>Mercury
Q1</th><th>Q2</th><th>Q3</th><th>Q4</th><th>Total</th></tr>
<tr><td>CA</td><td>10
14.628
1.4644
4.42
17.24
17.54</td><td>10
14.372
1.3298
4.42
17.24
17.86</td><td>19
14.628
1.3065
8.41
32.76
33.33</td><td>19
14.372
1.4905
8.41
32.76
33.93</td><td>58

25.66</td></tr>
<tr><td>FL</td><td>21
16.898
0.9956
9.29
31.34
36.84</td><td>15
16.602
0.1545
6.64
22.39
26.79</td><td>13
16.898
0.8993
5.75
19.40
22.81</td><td>18
16.602
0.1178
7.96
26.87
32.14</td><td>67

29.65</td></tr>
<tr><td>NY</td><td>7
15.637
4.7707
3.10
11.29
12.28</td><td>24
15.363
4.8559
10.62
38.71
42.86</td><td>14
15.637
0.1714
6.19
22.58
24.56</td><td>17
15.363
0.1745
7.52
27.42
30.36</td><td>62

27.43</td></tr>
<tr><td>WA</td><td>19
9.8363
8.5371
8.41
48.72
33.33</td><td>7
9.6637
0.7342
3.10
17.95
12.50</td><td>11
9.8363
0.1377
4.87
28.21
19.30</td><td>2
9.6637
6.0776
0.88
5.13
3.57</td><td>39

17.26</td></tr>
<tr><td>Total</td><td>57
25.22</td><td>56
24.78</td><td>57
25.22</td><td>56
24.78</td><td>226
100.00</td></tr>
</table>

Statistics for Table of state by Mercury

<table>
<tr><th>Statistic</th><th>DF</th><th>Value</th><th>Prob</th></tr>
<tr><td>Chi-Square</td><td>9</td><td>33.2176</td><td>0.0001</td></tr>
<tr><td>Likelihood Ratio Chi-Square</td><td>9</td><td>35.0289</td><td><.0001</td></tr>
<tr><td>Mantel-Haenszel Chi-Square</td><td>1</td><td>9.6555</td><td>0.0019</td></tr>
<tr><td>Phi Coefficient</td><td></td><td>0.3834</td><td></td></tr>
<tr><td>Contingency Coefficient</td><td></td><td>0.3580</td><td></td></tr>
<tr><td>Cramer's V</td><td></td><td>0.2213</td><td></td></tr>
</table></td><td>At significance level 0.05, we would have to conclude that the levels of mercury are not homogenous across the four states.

The true standout from the group was Washington, where 48.72% percent of the counties reported in the lowest quartile. Of the four states, California had the highest percentage of counties reporting in the top quartile. Altogether, over 65% of California's counties reported higher than median levels.</td></tr>
<tr><td>Lead across all 4 states</td><td>H_0: Lead levels are evenly distributed between FL, CA, NY, and WA

H_A: Lead levels are not evenly distributed between FL, CA, NY, and WA</td><td>proc univariate;
var Lead_TPY;
run;

if lead_tpy le 0.064255 then Lead='Q1';
if lead_tpy gt 0.064255 and lead_tpy le 0.175143 then Lead='Q2';
if lead_tpy gt 0.175143 and lead_tpy le 0.373396 then Lead='Q3';
if lead_tpy gt 0.373396 then Lead='Q4';

proc freq;
table state*Lead/chisq cellchi2 expected;
run;</td><td>Frequency
Expected
Cell Chi-Square
Percent
Row Pct
Col Pct

Table of state by Lead

<table>
<tr><th>state</th><th>Lead
Q1</th><th>Q2</th><th>Q3</th><th>Q4</th><th>Total</th></tr>
<tr><td>CA</td><td>5
14.885
6.5645
2.21
8.62
8.62</td><td>15
14.372
0.0275
6.64
25.86
26.79</td><td>14
14.372
0.0096
6.19
24.14
25.00</td><td>24
14.372
6.4505
10.62
41.38
42.86</td><td>58

25.66</td></tr>
<tr><td>FL</td><td>23
17.195
1.96
10.18
34.33
39.66</td><td>6
16.602
6.7702
2.65
8.96
10.71</td><td>20
16.602
0.6956
8.85
29.85
35.71</td><td>18
16.602
0.1178
7.96
26.87
32.14</td><td>67

29.65</td></tr>
<tr><td>NY</td><td>22
15.912
2.3297
9.73
35.48
37.93</td><td>21
15.363
2.0685
9.29
33.87
37.50</td><td>10
15.363
1.872
4.42
16.13
17.86</td><td>9
15.363
2.6353
3.98
14.52
16.07</td><td>62

27.43</td></tr>
<tr><td>WA</td><td>8
10.009
0.4032
3.54
20.51
13.79</td><td>14
9.6637
1.9458
6.19
35.90
25.00</td><td>12
9.6637
0.5648
5.31
30.77
21.43</td><td>5
9.6637
2.2507
2.21
12.82
8.93</td><td>39

17.26</td></tr>
<tr><td>Total</td><td>58
25.66</td><td>56
24.78</td><td>56
24.78</td><td>56
24.78</td><td>226
100.00</td></tr>
</table>

Statistics for Table of state by Lead

<table>
<tr><th>Statistic</th><th>DF</th><th>Value</th><th>Prob</th></tr>
<tr><td>Chi-Square</td><td>9</td><td>36.6657</td><td><.0001</td></tr>
</table></td><td>At significance level 0.05, we would have to conclude that the levels of lead are not homogenous across the four states.

However, looking at the table, we can see that California had 41.38% of its sampled data in the top quartile. Contrast that with Washington, where only 12.82% of the data accounted for high mortality rates. Most interesting was New York, where 69.35% of the sample exhibited levels in the bottom half of the quartiles.</td></tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>Hypotheses</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr>
<td></td>
<td>Alzheimer's related mortality rates across Fl, CA, WA, and NY</td>
<td>H_0:Alzheimer's related mortality rates are evenly distributed between FL, CA, NY, and WA

H_A: Alzheimer's related mortality rates are not evenly distributed between FL, CA, NY, and WA</td>
<td>if alz le 0.583049 then Alzheimers='Q1';
if alz gt 0.583049 and alz le 0.847226 then Alzheimers='Q2';
if alz gt 0.847226 and alz le 1.245887 then Alzheimers='Q3';
if alz gt 1.245887 then Alzheimers='Q4';</td>
<td>
Frequency
Expected
Cell Chi-Square
Percent
Row Pct
Col Pct

Table of state by Alzheimers
<table>
<tr><th></th><th colspan="5">Alzheimers</th></tr>
<tr><th>state</th><th>Q1</th><th>Q2</th><th>Q3</th><th>Q4</th><th>Total</th></tr>
<tr><td>CA</td><td>10
14.628
1.4644
4.42
17.24
17.54</td><td>9
14.372
2.0078
3.98
15.52
16.07</td><td>17
14.628
0.3845
7.52
29.31
29.82</td><td>22
14.372
4.049
9.73
37.93
39.29</td><td>58

25.66</td></tr>
<tr><td>FL</td><td>21
16.898
0.9956
9.29
31.34
36.84</td><td>22
16.602
1.7553
9.73
32.84
39.29</td><td>17
16.898
0.0006
7.52
25.37
29.82</td><td>7
16.602
5.5533
3.10
10.45
12.50</td><td>67

29.65</td></tr>
<tr><td>NY</td><td>23
15.637
3.4668
10.18
37.10
40.35</td><td>22
15.363
2.8674
9.73
35.48
39.29</td><td>12
15.637
0.846
5.31
19.35
21.05</td><td>5
15.363
6.9901
2.21
8.06
8.93</td><td>62

27.43</td></tr>
<tr><td>WA</td><td>3
9.8363
4.7513
1.33
7.69
5.26</td><td>3
9.6637
4.595
1.33
7.69
5.36</td><td>11
9.8363
0.1377
4.87
28.21
19.30</td><td>22
9.6637
15.748
9.73
56.41
39.29</td><td>39

17.26</td></tr>
<tr><td>Total</td><td>57
25.22</td><td>56
24.78</td><td>57
25.22</td><td>56
24.78</td><td>226
100.00</td></tr>
</table>

Statistics for Table of state by Alzheimers
<table>
<tr><th>Statistic</th><th>DF</th><th>Value</th><th>Prob</th></tr>
<tr><td>Chi-Square</td><td>9</td><td>55.6128</td><td><.0001</td></tr>
</table>
</td>
<td>At significance level 0.05, we would have to conclude that Alzheimer's related mortalities from individuals over the age of 65 are not homogenous across the four states.

From the table, we can see that Washington, unfortunately, had the highest, at 56.41% of the counties reporting levels in the top 25% range. Only 15.38 % of the counties reported mortality rates in the lower 50%.

Fortunately, for Florida, only 10.45% of the counties reported rates in the top 25%.</td>
</tr>
<tr>
<td>Pamela Mishaw</td>
<td>Mental Distress</td>
<td>H_0: The distribution of mental distress level is equal in CA, FL, NY, and WA.

H_A: The distribution of mental distress level is not equal in CA, FL, NY, and WA.</td>
<td>if mental_distress le 10 then m=1;
if mental_distress gt 10 and mental_distress le 11 then m=2;
if mental_distress gt 11 and mental_distress le 12 then m=3;
if mental_distress gt 12 then m=4;

.
.
.

proc freq;
table state*m/chisq cellchi2 expected;
run;</td>
<td>
Frequency
Expected
Cell Chi-Square
Percent
Row Pct
Col Pct

Table of state by m
<table>
<tr><th></th><th colspan="5">m</th></tr>
<tr><th>state</th><th>1</th><th>2</th><th>3</th><th>4</th><th>Total</th></tr>
<tr><td>CA</td><td>14
12.832
0.1063
6.19
24.14
28.00</td><td>18
20.274
0.2551
7.96
31.03
22.78</td><td>12
13.345
0.1356
5.31
20.69
23.08</td><td>14
11.549
0.5203
6.19
24.14
31.11</td><td>58

25.66</td></tr>
<tr><td>FL</td><td>1
14.823
12.89
0.44
1.49
2.00</td><td>13
23.42
4.6363
5.75
19.40
16.46</td><td>26
15.416
7.2667
11.50
38.81
50.00</td><td>27
13.341
13.985
11.95
40.30
60.00</td><td>67

29.65</td></tr>
<tr><td>NY</td><td>26
13.717
10.999
11.50
41.94
52.00</td><td>30
21.673
3.1997
13.27
48.39
37.97</td><td>5
14.265
6.018
2.21
8.06
9.62</td><td>1
12.345
10.426
0.44
1.61
2.22</td><td>62

27.43</td></tr>
<tr><td>WA</td><td>9
8.6283
0.016
3.98
23.08
18.00</td><td>18
13.633
1.3991
7.96
46.15
22.78</td><td>9
8.9735
0.0001
3.98
23.08
17.31</td><td>3
7.7655
2.9245
1.33
7.69
6.67</td><td>39

17.26</td></tr>
<tr><td>Total</td><td>50
22.12</td><td>79
34.96</td><td>52
23.01</td><td>45
19.91</td><td>226
100.00</td></tr>
</table>

Statistics for Table of state by m
<table>
<tr><th>Statistic</th><th>DF</th><th>Value</th><th>Prob</th></tr>
<tr><td>Chi-Square</td><td>9</td><td>74.7791</td><td><.0001</td></tr>
</table>
</td>
<td>Since the p-value of the Chi-Square statistic is a small number (<0.0001) and less than the significance level of 0.05, we reject the null hypothesis and conclude that the distribution of mental distress levels is not equal among CA, FL, NY, and WA.

Florida's distribution shows the highest percentage in the highest percentile group while New York shows a higher percentage in the lowest percentile group. This indicates higher rates of mental distress in Florida and lowest rates in New York in comparison to the other states analyzed. California's distribution is relatively similar among the four groups while Washington's seems to be concentrated in the second lowest quartile.</td>
</tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>Hypotheses</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr>
<td></td>
<td>% Bachelor Degree</td>
<td>H_0: The distribution of the percentage of those who hold a bachelor's degree is equal in CA, FL, NY, and WA.

H_A: The distribution of the percentage of those who hold a bachelor's degree is not equal in CA, FL, NY, and WA.</td>
<td>if Percent_with_Bachelors le 16.5 then p=1;
if Percent_with_Bachelors gt 16.5 and Percent_with_Bachelors le 21.3 then p=2;
if Percent_with_Bachelors gt 21.3 and Percent_with_Bachelors le 29.7 then p=3;
if Percent_with_Bachelors gt 29.7 then p=4;

.
.
.

proc freq;
table state*p/chisq cellchi2 expected;
run;</td>
<td>
Frequency
Expected
Cell Chi-Square
Percent
Row Pct
Col Pct

Table of state by p
<table>
<tr><th></th><th colspan="5">p</th></tr>
<tr><th>state</th><th>1</th><th>2</th><th>3</th><th>4</th><th>Total</th></tr>
<tr><td>CA</td><td>14
14.885
0.0526
6.19
24.14
24.14</td><td>14
14.372
0.0096
6.19
24.14
25.00</td><td>10
14.628
1.4644
4.42
17.24
17.54</td><td>20
14.115
2.4536
8.85
34.48
36.36</td><td>58

25.66

</td></tr>
<tr><td>FL</td><td>28
17.195
6.7902
12.39
41.79
48.28</td><td>12
16.602
1.2755
5.31
17.91
21.43</td><td>17
16.898
0.0006
7.52
25.37
29.82</td><td>10
16.305
2.4383
4.42
14.93
18.18</td><td>67

29.65

</td></tr>
<tr><td>NY</td><td>6
15.912
6.174
2.65
9.68
10.34</td><td>20
15.363
1.3997
8.85
32.26
35.71</td><td>19
15.637
0.7232
8.41
30.65
33.33</td><td>17
15.088
0.2422
7.52
27.42
30.91</td><td>62

27.43

</td></tr>
<tr><td>WA</td><td>10
10.009
782E-8
4.42
25.64
17.24</td><td>10
9.6637
0.0117
4.42
25.64
17.86</td><td>11
9.8363
0.1377
4.87
28.21
19.30</td><td>8
9.4912
0.2343
3.54
20.51
14.55</td><td>39

17.26

</td></tr>
<tr><td>Total</td><td>58
25.66</td><td>56
24.78</td><td>57
25.22</td><td>55
24.34</td><td>226
100.00</td></tr>
</table>

Statistics for Table of state by p
<table>
<tr><th>Statistic</th><th>DF</th><th>Value</th><th>Prob</th></tr>
<tr><td>Chi-Square</td><td>9</td><td>23.4075</td><td>0.0053</td></tr>
</table>
</td>
<td>Since the p-value of the Chi-Square statistic is a small number (0.0053) and less than the significance level of 0.05, we reject the null hypothesis and conclude that the distribution of the percentage of those with a bachelor's degree is not equal among CA, FL, NY, and WA.

Florida has the highest percentage (48.28%) of counties with the lowest (<25%) quartile of percentage of those who hold bachelor's degrees while California has the highest percentage (36.6%) of counties with the highest (>75%) quartile of percentage of those who hold bachelor's degrees. So, Florida seems to have the higher rates of lesser formally educated people while California has the relatively greatest rates compared to the other states analyzed. Washington's distribution of those with bachelor's degrees seems to be relatively similar among the percentile groups while New York has higher concentrations in the moderately educated groups.</td>
</tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

<table>
<tr><th>Name</th><th>Var</th><th>Hypotheses</th><th>SAS Code</th><th>SAS Output</th><th>Conclusion</th></tr>
<tr>
<td>Prince Nhliziyo</td>
<td>Smoking</td>
<td>H_0: The distribution of smoking rate is the same for the states of CA,FL, NY and WA.

H_A: The distribution of smoking rate differs for the states of CA, FL, NY, WA.</td>
<td>if smoking le 18.86 then smk = 1;
if smoking gt 18.86 and smoking le 22.14 then smk = 2;
if smoking gt 22.14 and smoking le 24.86 then smk = 3;
if smoking gt 24.86 then smk = 4;

proc freq;
table state*smk/ chisq cellchi2 expected;
run;</td>
<td>
Frequency
Expected
Cell Chi-Square
Percent
Row Pct
Col Pct

Table of state by smk

<table>
<tr><th></th><th colspan="5">smk</th></tr>
<tr><th>state</th><th>1</th><th>2</th><th>3</th><th>4</th><th>Total</th></tr>
<tr><td>CA</td><td>32
14.628
20.63
14.16
55.17
56.14</td><td>14
14.372
0.0096
6.19
24.14
25.00</td><td>10
14.628
1.4644
4.42
17.24
17.54</td><td>2
14.372
10.65
0.88
3.45
3.57</td><td>58

25.66

</td></tr>
<tr><td>FL</td><td>4
16.898
9.8451
1.77
5.97
7.02</td><td>7
16.602
5.5533
3.10
10.45
12.50</td><td>16
16.898
0.0477
7.08
23.88
28.07</td><td>40
16.602
32.977
17.70
59.70
71.43</td><td>67

29.65

</td></tr>
<tr><td>NY</td><td>11
15.637
1.3751
4.87
17.74
19.30</td><td>19
15.363
0.8611
8.41
30.65
33.93</td><td>25
15.637
5.606
11.06
40.32
43.86</td><td>7
15.363
4.5523
3.10
11.29
12.50</td><td>62

27.43

</td></tr>
<tr><td>WA</td><td>10
9.8363
0.0027
4.42
25.64
17.54</td><td>16
9.6637
4.1546
7.08
41.03
28.57</td><td>6
9.8363
1.4962
2.65
15.38
10.53</td><td>7
9.6637
0.7342
3.10
17.95
12.50</td><td>39

17.26

</td></tr>
<tr><td>Total</td><td>57
25.22</td><td>56
24.78</td><td>57
25.22</td><td>56
24.78</td><td>226
100.00</td></tr>
</table>

Statistics for Table of state by smk

<table>
<tr><th>Statistic</th><th>DF</th><th>Value</th><th>Prob</th></tr>
<tr><td>Chi-Square</td><td>9</td><td>99.9590</td><td><.0001</td></tr>
</table>
</td>
<td>The p-value is <0.001 which means that we reject the null hypothesis (based on significance level 0.05) and conclude that the distribution of smoking rate differs among the states of CA, FL, NY, and WA.

CA had the highest percentage (55.17%) of counties in the first quartile of smoking rate.
FL had the highest percentage (59.70%) of countries in the 4th quartile of smoking rate.

</td>
</tr>
</table>

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Hypotheses	SAS Code	SAS Output	Conclusion

	Physical Inactivity	H_0: The distribution of physical inactivity is the same for the states of CA,FL, NY and WA.
		
		H_A: The distribution of physical inactivity differs for the states of CA, FL, NY, WA.
			if physical le 18.56 then phy = 1;
			if physical gt 18.56 and physical le 22.42 then phy = 2;
			if physical gt 22.42 and physical le 24.88 then phy = 3;
			if physical gt 24.88 then phy = 4;
			
			proc freq;
			table state*phy/ chisq cellchi2 expected;
			run;
				Frequency
				Expected
				Cell Chi-Square
				Percent
				Row Pct
				Col Pct
				
				Table of state by phy
				
				phy
				
				state | 1 | 2 | 3 | 4 | Total
				
				CA | 40 | 17 | 1 | 0 | 58
				14.628 | 14.372 | 14.628 | 14.372
				44.005 | 0.4807 | 12.697 | 14.372
				17.70 | 7.52 | 0.44 | 0.00 | 25.66
				68.97 | 29.31 | 1.72 | 0.00
				70.18 | 30.36 | 1.75 | 0.00
				
				FL | 3 | 9 | 19 | 36 | 67
				16.898 | 16.602 | 16.898 | 16.602
				11.431 | 3.4808 | 0.2614 | 22.666
				1.33 | 3.98 | 8.41 | 15.93 | 29.65
				4.48 | 13.43 | 28.36 | 53.73
				5.26 | 16.07 | 33.33 | 64.29
				
				NY | 1 | 13 | 31 | 17 | 62
				15.637 | 15.363 | 15.637 | 15.363
				13.701 | 0.3634 | 15.093 | 0.1745
				0.44 | 5.75 | 13.72 | 7.52 | 27.43
				1.61 | 20.97 | 50.00 | 27.42
				1.75 | 23.21 | 54.39 | 30.36
				
				WA | 13 | 17 | 6 | 3 | 39
				9.8363 | 9.6637 | 9.8363 | 9.6637
				1.0176 | 5.5694 | 1.4962 | 4.595
				5.75 | 7.52 | 2.65 | 1.33 | 17.26
				33.33 | 43.59 | 15.38 | 7.69
				22.81 | 30.36 | 10.53 | 5.36
				
				Total | 57 | 56 | 57 | 56 | 226
				25.22 | 24.78 | 25.22 | 24.78 | 100.00
				
				Statistics for Table of state by phy
				
				Statistic | DF | Value | Prob
				Chi-Square | 9 | 151.4035 | <.0001
					The p-value is <0.001 which means that we reject the null hypothesis (based on significance level 0.05) and conclude that the distribution of physical inactivity differs among the states of CA, FL, NY, and WA.
					
					• CA had the highest percentage (68.97%) of counties in the first quartile of physical inactivity.
					• FL had the highest percentage (53.73%) of states in the fourth quartile of physical inactivity.
					• CA had no counties in the fourth quartile of physical inactivity.

STA5176: Spring 2023 Group 4 Alzheimer's Project

13.Give a summary statement

Name	Var	Conclusion
Brad Lipson	Cancer	The cancer data for the four states of California, Florida, New York, and Washington has a distribution that is not normal. Also, they have at least one distinct mean as well as being statistically significant. Furthermore, it has a positive correlation with obesity, and at least one state has a distinct variance from the others. This means that some states have higher cancer rates than others.
	Obesity	The obesity data for the four states of California, Florida, New York, and Washington is not normally distributed. This suggests that the data is not evenly distributed and may be skewed or have outliers. Additionally, the data has at least one mean that is different from the others, meaning that the average obesity rates for the four states are not the same. Furthermore, the data has a significant and positive correlation with cancer, meaning that there is a positive relationship between obesity and cancer rates. Finally, at least one state has a different variance than the others, meaning that the spread of obesity rates is not the same for all four states
Mary Morrow	Heart Disease	The heart disease data for the four states chosen (CA, FL, NY, and WA) is not normally distributed, has at least one mean that is different, has a significant and positive correlation with diabetes, and that at least one state has a different variance than the others. When performing a linear regression of Alzheimer's rate on heart disease rate, we find that there is a significant association between the two and that 5.88% of the variability in Alzheimer's rate is explained by the model. When testing for homogeneity for heart disease, we found that the distribution is not the same for each state, Washington had the highest percentage for the lowest heart disease rate, and New York had the highest percentage for the highest heart disease rate.
	Diabetes	The diabetes data for the four states chosen (CA, FL, NY, and WA) is not normally distributed, has at least one mean that is different, has a significant and positive correlation with heart disease, and that at least one state has a different variance than the others. When performing a linear regression of Alzheimer's rate on diabetes rate, we find that there is not a significant association between the two and that 0.58% of the variability in Alzheimer's rate is explained by the model. When testing for homogeneity for diabetes, we found that the distribution is not the same for each state, California had the highest percentage for the lowest diabetes rate, and Florida had the highest percentage for the highest diabetes rate.
Matthew McGehee	Mercury	The mercury data measured in tons per year for the four states chosen (CA, FL, NY, and WA) is not normally distributed. If we consider an alpha of 0.05 for each test, we would conclude that at least one of the state's mean was significantly different, that mercury levels have a significant and positive correlation with lead, and that at least one state has a different variance than the others. Furthermore, the test for homogeneity shows us that the levels were not evenly distributed across the four states.
	Lead	The lead data measured in tons per year for the four states chosen (CA, FL, NY, and WA) is not normally distributed. If we consider an alpha of 0.05 for each test, we would conclude that at least one of the state's mean was significantly different, that lead levels have a significant and positive correlation with mercury, and that at least one state has a different variance than the others. Furthermore, the test for homogeneity shows us that the levels were not evenly distributed across the four states.

STA5176: Spring 2023 Group 4 Alzheimer's Project

Name	Var	Conclusion
Pamela Mishaw	Mental Distress	The data on mental distress levels are not normally distributed according to the results of the Shapiro Wilk test. At significance level 0.05, it was found that at least two of the state means of mental distress levels differ significantly. Mental distress and the percentage of the population that hold bachelor's degrees are strongly negatively correlated. At least two of the states have differing variances in mental distress level. Notably, Florida was found to have a significantly different mean from the other states. A simple linear regression model found that there is not a statistically significant influence of mental distress on Alzheimer's death rates. The Chi Square homogeneity test results suggest that Florida has the highest concentration of higher mental distress scores than the other states analyzed while New York appeared to have the lowest.
	% Bachelor Degree	The data on the percentages of those with bachelor's degrees are not normally distributed according to the results of the Shapiro Wilk test. At significance level 0.05, it was found that at least two of the state means of percentages of those with bachelor's degrees differ significantly. Mental distress and the percentage of the population that hold bachelor's degrees are strongly negatively correlated. All of the states have statistically the same variances in percentages of those with bachelor's degrees. Notably, Florida was found to have a significantly different mean from the other states. A simple linear regression model found that there is not a statistically significant influence of the percentage of those with a bachelor's degree on Alzheimer's death rates. The Chi Square homogeneity test results suggest that Florida has the highest percentage of counties in the lowest quartile of percentage of those with a bachelor's degree while California had the highest percentage of counties in the highest quartile in comparison to the other states included in the analysis.
Prince Nhliziyo	Smoking	The data on the four states was not normally distributed based on the results of the Shapiro-Wilks test. Based on the ANOVA tests, we concluded that at least one mean is significantly different for alpha levels 0.1 and 0.01. There is a strong positive correlation between smoking rate and physical inactivity. Based on the Brown-Forsythe test, we concluded that there was not enough evidence to suggest that at least two state variances were not equal. Based on the results of the linear regression model, there is enough evidence to suggest that alzheimer's is influenced by smoking rate . The test for homogeneity showed that the distribution of smoking rate differed among the states of CA, FL, NY, and WA.
	Inactivity	The data on the four states was not normally distributed based on the results of the Shapiro-Wilks test. Based on the ANOVA tests, we concluded that at least one mean is significantly different for alpha levels 0.1 and 0.01.There is a strong positive correlation between smoking rate and physical inactivity.Based on the Brown-Forsythe test, we concluded that at least two state variances were not equal.Based on the results of the linear regression model, there is enough evidence to suggest that alzheimer's is influenced by physical inactivity.The test for homogeneity showed that the distribution of physical inactivity differed among the states of CA, FL, NY, and WA.

STA5176: Spring 2023 Group 4 Alzheimer's Project

Our Shared Dataset:

alz4full.xlsx (shared under sheets - I just removed the first column of county names)
https://docs.google.com/spreadsheets/d/1pGyqof0WVoqUON6dcLPkaFjqHtxVZNFdJ9LgpzjjK2s/edit#gid=98690290

Initial Code (to set up data table in SAS on-demand):

```
FILENAME REFFILE '/home/u63032368/sasuser.v94/alz4full.xlsx';

PROC IMPORT DATAFILE=REFFILE DBMS=XLSX OUT=WORK.IMPORT5;
GETNAMES=YES;
RUN;

PROC CONTENTS DATA=WORK.IMPORT5;
RUN;

%web_open_table(WORK.IMPORT5);
```