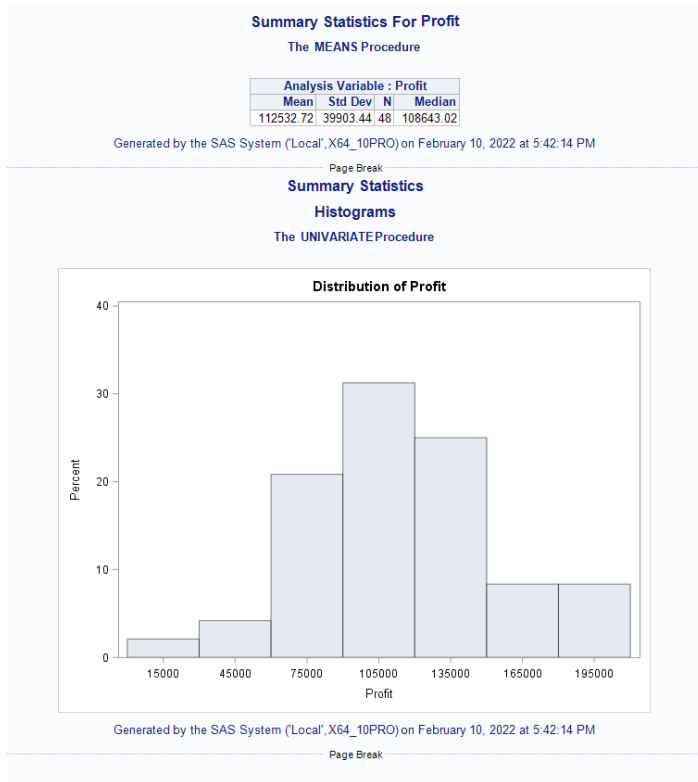


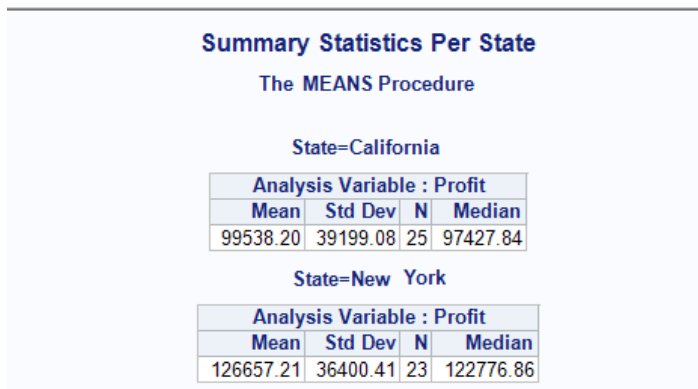
Alain Euksuzian (ID: 40070126)

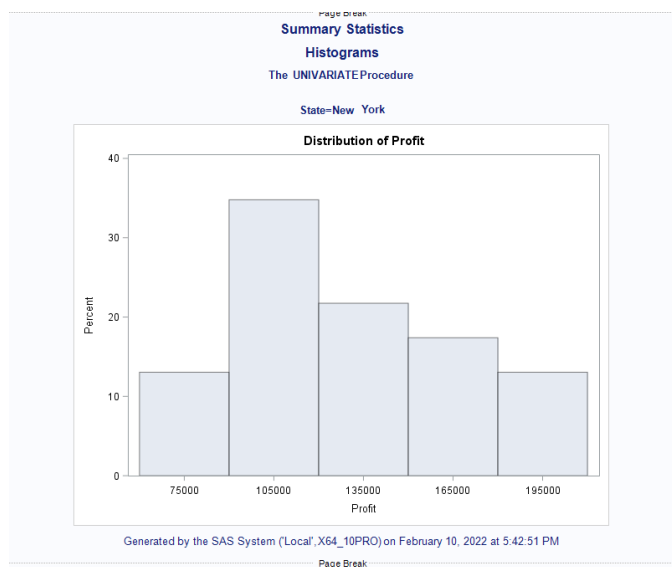
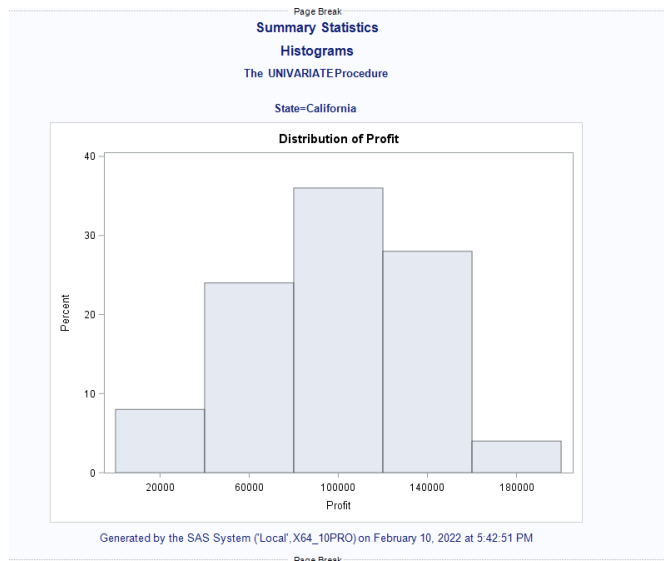
Question 1)

A)

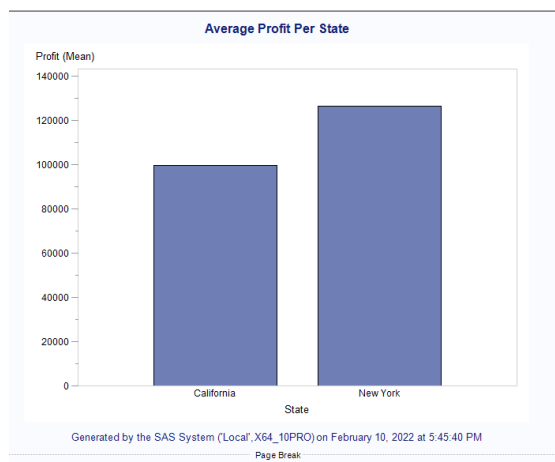


B)

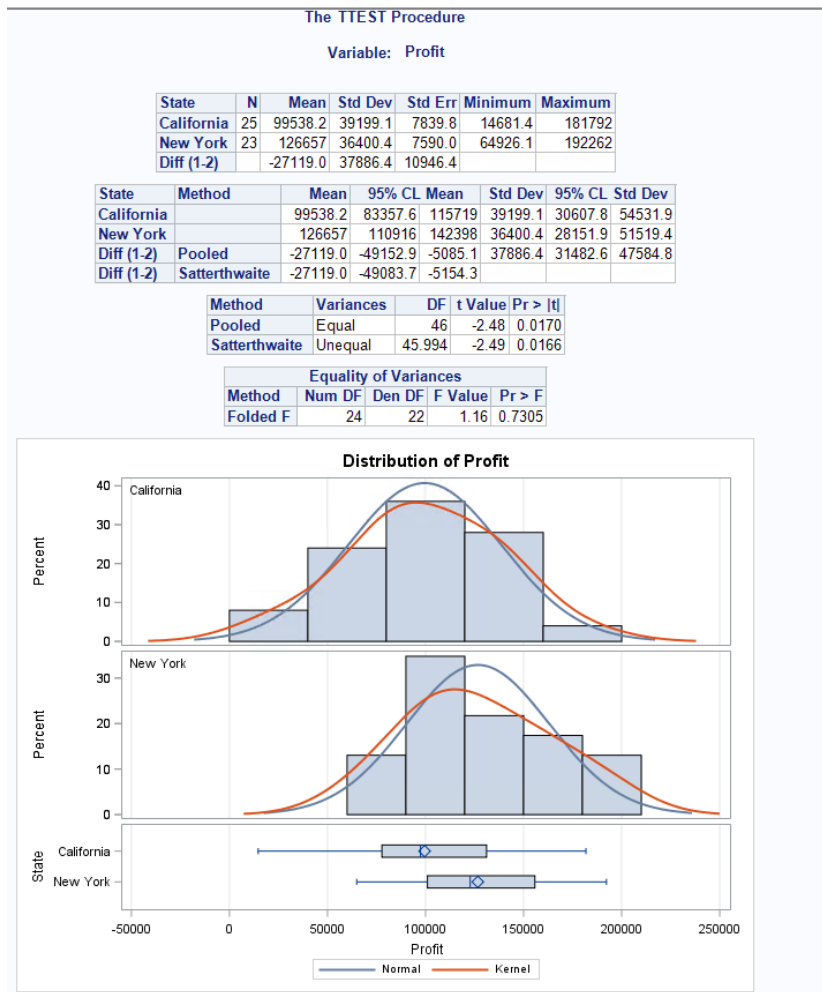




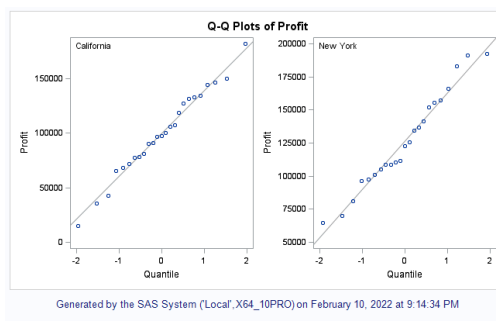
C)



D & E)



- 1) They are both independent observations
- 2) Both distributions appear to be normal



= to analyze outliers (1 in California)

H_0 : New York population Profit = California population Profit

H_a : New York population profit > California population Profit

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	24	22	1.16	0.7305

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	46	-2.48	0.9915
Satterthwaite	Unequal	45.994	-2.49	0.9917

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	24	22	1.16	0.7305

F-Value Greater than 0.05, hence we cannot reject the Null Hypotheses. We can assume with a 95% confidence that the population profit in New York is equal to the population profit of California. Other factors to be considered in profit analysis would be internal costs such as marketing, R&D, Administration etc. Also considering external factors such as the state economy, customer average spending, taxation etc. would yield a more trustworthy conclusion.

Part 2

Linear Regression Results

The REG Procedure
Model: Linear_Regression_Model
Dependent Variable: Profit

Number of Observations Read	48
Number of Observations Used	48

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	52546792686	26273396343	53.04	<.0001
Error	45	22290593928	495346532		
Corrected Total	47	74837386615			

Root MSE	22256	R-Square	0.7021
Dependent Mean	112533	Adj R-Sq	0.6889
Coeff Var	19.77770		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	23767	15209	1.56	0.1251
Administration	1	0.25595	0.11885	2.15	0.0367
Marketing Spend	1	0.26509	0.02704	9.81	<.0001

Generated by the SAS System ('Local', X64_10PRO) on February 10, 2022 at 7:23:36 PM

Page Break

A)

Correlation Analysis

The CORR Procedure

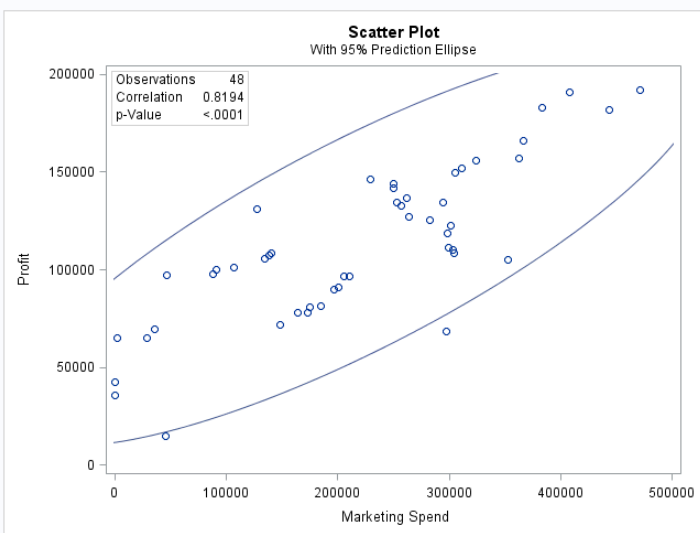
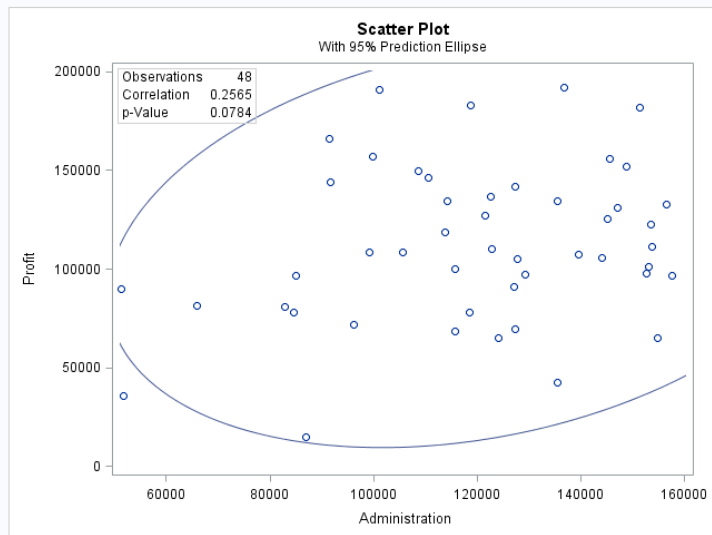
1 With Variables:	Profit
2 Variables:	Administration Marketing Spend

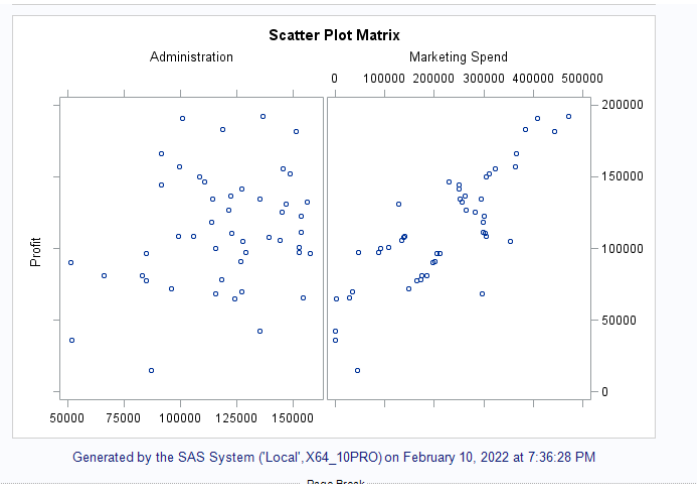
Pearson Correlation Coefficients, N = 48
Prob > |r| under H0: Rho=0

Profit	Marketing Spend	Administration
	0.81942	0.25650
	<.0001	0.0784

Correlation Analysis

The CORR Procedure





B)

Estimated equation = intercept Variable + (Slope Parameter * Predictor Variable)

$$\text{Profit} = 3767 + (0.26509 * \text{Marketing expense}) + (0.25595 * \text{Administration})$$

C)

H₀: B₁ = B₂ = 0 where B₁ & B₂ are the slopes for marketing and administration

H_a: At least one of the slopes for Marketing or Administration is not 0

At 5% alpha, p-value as per SAS = <0.05

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	52546792686	26273396343	53.04	<.0001
Error	45	22290593928	495346532		
Corrected Total	47	74837386615			

Root MSE	22256	R-Square	0.7021
Dependent Mean	112533	Adj R-Sq	0.6889
Coeff Var	19.77770		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	23767	15209	1.56	0.1251
Administration	1	0.25595	0.11885	2.15	0.0367
Marketing Spend	1	0.26509	0.02704	9.81	<.0001

Hence, we reject H_0 , meaning one of the slopes does not have a 0 value, meaning our model fits the data better than the baseline model. With the slope not having a value of 0, the regression graph would have a significant amount of variation with marketing and administration. In other words, the graph would not be horizontal at the y intercept. Both administration and Marketing have a P-value < 0.05

D)

For each unit of increase in profit, Administration accounts for an increase of 0.25595 and Marketing accounts for an increase of 0.26509, with both having a p value below 0.05, meaning they are both significant in the behavior of profit

3)

Linear Regression Results

The REG Procedure

Model: Linear_Regression_Model

Dependent Variable: Profit

Number of Observations Read	48
Number of Observations Used	48

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	70764583033	23588194344	254.83	<.0001
Error	44	4072803582	92563718		
Corrected Total	47	74837386615			

Root MSE	9621.00399	R-Square	0.9456
Dependent Mean	112533	Adj R-Sq	0.9419
Coeff Var	8.54952		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	46602	6773.07271	6.88	<.0001
R&D Spend	1	0.68975	0.04917	14.03	<.0001
Administration	1	-0.00020426	0.05453	-0.00	0.9970
Marketing Spend	1	0.06723	0.01832	3.67	0.0007

A)

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	46602	6773.07271	6.88	<.0001
R&D Spend	1	0.68975	0.04917	14.03	<.0001
Administration	1	-0.00020426	0.05453	-0.00	0.9970
Marketing Spend	1	0.06723	0.01832	3.67	0.0007

$H_0 = B_1 = B_2 = B_3 = 0$

$H_a =$ At least one of the slopes for Marketing or Administration or R&D is not 0

At 5% confidence, p-value for Administration 0.9970, hence we do not have enough evidence to reject the null hypothesis. Given Marketing and R&D are in the model, Administration is not significant in predicting profit trends, and not linearly related to profit.

B)

Estimated equation = intercept Variable + Slope Parameter * Predictor Variable

Profit = 46602 + (0.68975 * R&D expense) – (0.0002 * Administration Expense) + (0.06723 * Marketing Expense)

C)

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	46602	6773.07271	6.88	<.0001
R&D Spend	1	0.68975	0.04917	14.03	<.0001
Administration	1	-0.00020426	0.05453	-0.00	0.9970
Marketing Spend	1	0.06723	0.01832	3.67	0.0007

For each unit of increase in profit, an increase of 0.68975 increase in R&D, -0.0002 decrease in Administration and 0.06723 increase in marketing. In part 2-D, we concluded that administration was significant and did not have a slope of 0. But in this analysis, administration is not significant in predicting the trends of profit. This is due to collinearity with a total R value of 0.94 as per SAS