Priyasri Sankaran
Project 2

1. Part I

The effects of region (A) and percent below poverty level (B) on the crime rate (Variable10/Variable5) are to be studied.

Dependent variable: crime_rate

Number of observations: 440

For purposes of t his ANOVA study, percent below poverty level is to be classified into threecategories: under 6%, 6-10%, 10% or more.

(i) State the ANOVA model for this case. Also state the equivalent regression model.

$Y_{ijk}= \mu_{ijk} + \epsilon_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$

Equivalent Regression Model:

We will use dummy variables for region and poverty in a multiple regression framework.

The Region has 4 levels and the Percent below poverty has 3 levels.

The equivalent regression model is written as follows:

$Y_{ijk}= \mu_{..} +\alpha_1X_{ijk1}+\alpha_2X_{ijk2}+\alpha_3X_{ijk3}+\beta_1X_{ijk4}$
$+\beta_2X_{ijk5}+(\alpha\beta)_{11}X_{ijk1}X_{ijk4}+(\alpha\beta)_{21}X_{ijk2}Xijk4+(\alpha\beta)_{31}X_{ijk3}X_{ijk4}+(\alpha\beta)_{11}X_{ijk1}X_{ijk4}+(\alpha\beta)_{21}X_{ijk2}X_{ijk4}+(\alpha\beta)_{31}$
$X_{ijk3}X_{ijk4}+(\alpha\beta)_{12}X_{ijk1}X_{ijk5}+(\alpha\beta)_{22}X_{ijk2}X_{ijk5}+(\alpha\beta)_{32}X_{ijk3}X_{ijk5}+\epsilon_{ijk}+(\alpha\beta)_{12}X_{ijk1}X_{ijk5}+(\alpha\beta)_{22}X_{ijk2}X_{ijk5}$
$+(\alpha\beta)_{32}X_{ijk3}X_{ijk5}+\epsilon_{ijk}$

$$X_1 = \begin{cases} 1, & if\ case\ from\ level\ 1\ for\ factor\ A \\ -1, & if\ case\ from\ level\ 4\ for\ factor\ A \\ 0, & otherwise \end{cases}$$

$$X_2 = \begin{cases} 1, & if\ case\ from\ level\ 2\ for\ factor\ A \\ -1, & if\ case\ fom\ level\ 4\ for\ factor\ A \\ 0, & otherwise \end{cases}$$

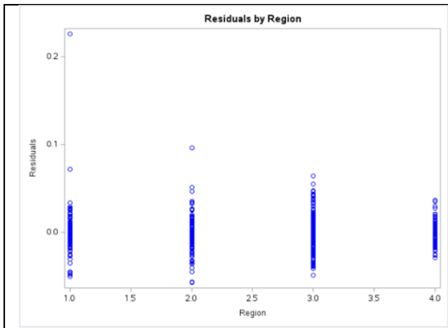
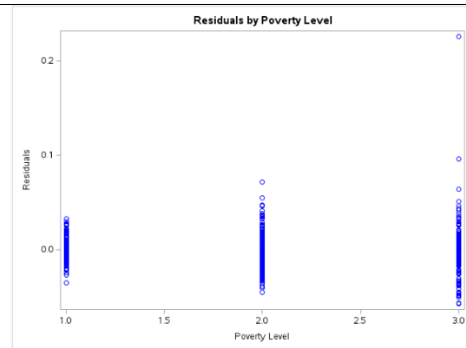
$$X_3 = \begin{cases} 1, & if\ case\ from\ level\ 3\ for\ factor\ A \\ -1, & if\ case\ from\ level\ 4\ for\ factor\ A \\ 0, & otherwise \end{cases}$$

$$X_4 = \begin{cases} 1, & if\ case\ from\ level\ 1\ for\ factor\ B(below\ Poverty) \\ -1, & if\ case\ from\ level\ 3\ for\ factor\ B(below\ Poverty) \end{cases}$$

$$X_5 = \begin{cases} 1, & if\ case\ from\ level\ 2\ for\ factor\ B(below\ Poverty) \\ -1, & if\ case\ from\ level\ 3\ for\ factor\ B(below\ Poverty) \end{cases}$$

(ii) Fit the regression model, and prepare aligned residual dot plots for the treatments. State your findings.

$Y_{ijk}= \mu_{..} +\alpha_1X_{ijk1}+\alpha_2X_{ijk2}+\alpha_3X_{ijk3}+\beta_1X_{ijk4}+\beta_2X_{ijk5}+(\alpha\beta)_{11}X_{ijk1}X_{ijk4}+(\alpha\beta)_{21}X_{ijk2}Xijk4+$
$(\alpha\beta)_{31}X_{ijk3}X_{ijk4}+(\alpha\beta)_{11}X_{ijk1}X_{ijk4}+(\alpha\beta)_{21}X_{ijk2}X_{ijk4}+(\alpha\beta)_{31}X_{ijk3}X_{ijk4}$
$+(\alpha\beta)_{12}X_{ijk1}X_{ijk5}+(\alpha\beta)_{22}X_{ijk2}X_{ijk5}+(\alpha\beta)_{32}X_{ijk3}X_{ijk5}+\epsilon_{ijk}+(\alpha\beta)_{12}X_{ijk1}X_{ijk5}+(\alpha\beta)_{22}X_{ijk2}X_{ijk5}+(\alpha\beta)_{32}X_{ijk3}$
$X_{ijk5}+\epsilon_{ijk}$

	
By Region: the residual plot shows differences across regions. Confirms the significant main effect of the region. Each region 1 and 3 seems to have residuals concentrated closer to the center. On the other hand, the 2 and 4 are a little more variability. The difference in variability could reflect the crime rate model fits differently. We also notice the outliers. Few outliers are observed.	By Poverty: The residuals are different across poverty levels. Under 6% category appears less spread and close model fit. 6=10% shows more variability. We noticed some outliers.
Main effect: Crime rates vary significantly across the region.	Main effect: Crime rates are significantly affected by the poverty categories.

(iii) State the reduced regression model for testing for the interaction effects. Fit the reduced regression model and test whether or not interaction effects are present.

For each test, use $\alpha = 0.05$. State the alternatives and P-values.

Residual Analysis for Crime Rate

The GLM Procedure

Dependent Variable: crime_rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	0.09674619	0.00879511	16.29	<.0001
Error	428	0.23110074	0.00053995		
Corrected Total	439	0.32784692			

R-Square	Coeff Var	Root MSE	crime_rate Mean
0.295096	40.56271	0.023237	0.057286

Source	DF	Type I SS	Mean Square	F Value	Pr > F
pov	2	0.06401565	0.03200783	59.28	<.0001
region	3	0.02794920	0.00931640	17.25	<.0001
pov*region	6	0.00478133	0.00079689	1.48	0.1848

Source	DF	Type III SS	Mean Square	F Value	Pr > F
pov	2	0.03098141	0.01549071	28.69	<.0001
region	3	0.02007679	0.00669226	12.39	<.0001
pov*region	6	0.00478133	0.00079689	1.48	0.1848

H_0: No interactions effects $(\alpha\beta)_{ij} = 0$

H_a: Interaction effects are present $(\alpha\beta)_{ij} \neq 0$

$F_{critical} = F_{(0.95,\ 6,428)} = 2.11976$

$F_{obs} = 1.48$

$F_{obs} < F_{critical}$.

Decision rule: If $F_{obs} < F_{critical}$, accept the null hyplothesis.

If $F_{obs} > F_{critical}$, reject the null hypothesis.

Conclusion: $F_{obs} < F_{critical}$, we accept the null hypothesis, and conclude that no strong evidence to reject the null hypothesis and no interaction present at 5% significant level.

P-value: P-value= 0.1848 >0.05, we accept the null hypothesis.

2)Part II

The effects of region (A), percent below poverty level (B) and percent of population 65 or older (C) on the crime rate are to be studied. For purposes of this ANOVA study, percent below poverty level is to be classified into two categories (< 8%, ≥ 8%) and percent of population 65 or older is to be classified into two groups (< 12%, ≥12%)

(i) Conduct a three-way ANOVA then do the diagnostic procedure.

Conducted a three-way ANOVA and included all possible interactions in the model. The model can be written as follows

$$Y_{ijk} = \mu_{ijk} + \epsilon_{ijk} = \mu_{\ldots} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijk}$$

Where i=1,2,3,4; j=1,2; and k= 1,2;

$$X_1 = \begin{cases} 1, & if\ case\ from\ level\ 1\ for\ factor\ A \\ -1, & if\ case\ from\ level\ 4\ for\ factor\ A \\ 0, & otherwise \end{cases}$$

$$X_2 = \begin{cases} 1, & if\ case\ from\ level\ 2\ for\ factor\ A \\ -1, & if\ case\ fom\ level\ 4\ for\ factor\ A \\ 0, & otherwise \end{cases}$$

$$X_3 = \begin{cases} 1, & if\ case\ from\ level\ 3\ for\ factor\ A \\ -1, & if\ case\ from\ level\ 4\ for\ factor\ A \\ 0, & otherwise \end{cases}$$

$$X_4 = \begin{cases} 1, & if\ case\ from\ level\ 1\ for\ factor\ B(below\ Poverty) \\ -1, & if\ case\ from\ level\ 2\ for\ factor\ B(below\ Poverty) \end{cases}$$

$$X_5 = \begin{cases} 1, & if\ case\ from\ level\ 1\ for\ factor\ C(65\ or\ older) \\ -1, & if\ case\ from\ level\ 2\ for\ factor\ C(65\ or\ older) \end{cases}$$

Three-Way ANOVA for Crime Rate

The GLM Procedure

Dependent Variable: Crime_Rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	0.09004016	0.00600268	10.70	<.0001
Error	424	0.23780677	0.00056087		
Corrected Total	439	0.32784692			

R-Square	Coeff Var	Root MSE	Crime_Rate Mean
0.274641	41.34065	0.023683	0.057286

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Region	3	0.05918092	0.01972697	35.17	<.0001
Poverty_Level	1	0.02613920	0.02613920	46.61	<.0001
Region*Poverty_Level	3	0.00069826	0.00023275	0.41	0.7423
Aged_Group	1	0.00093026	0.00093026	1.66	0.1985
Region*Aged_Group	3	0.00232416	0.00077472	1.38	0.2479
Poverty_L*Aged_Group	1	0.00006893	0.00006893	0.12	0.7261
Region*Povert*Aged_G	3	0.00069842	0.00023281	0.42	0.7423

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Region	3	0.03231691	0.01077230	19.21	<.0001
Poverty_Level	1	0.01268295	0.01268295	22.61	<.0001
Region*Poverty_Level	3	0.00118461	0.00039487	0.70	0.5500
Aged_Group	1	0.00103372	0.00103372	1.84	0.1753
Region*Aged_Group	3	0.00290920	0.00096973	1.73	0.1603
Poverty_L*Aged_Group	1	0.00022525	0.00022525	0.40	0.5266
Region*Povert*Aged_G	3	0.00069842	0.00023281	0.42	0.7423

Diagnostic Procedure: The diagnostic procedure includes Normality of Residuals, Homoscedasticity, and Independence. Please refer the below images.

Normality of residuals:

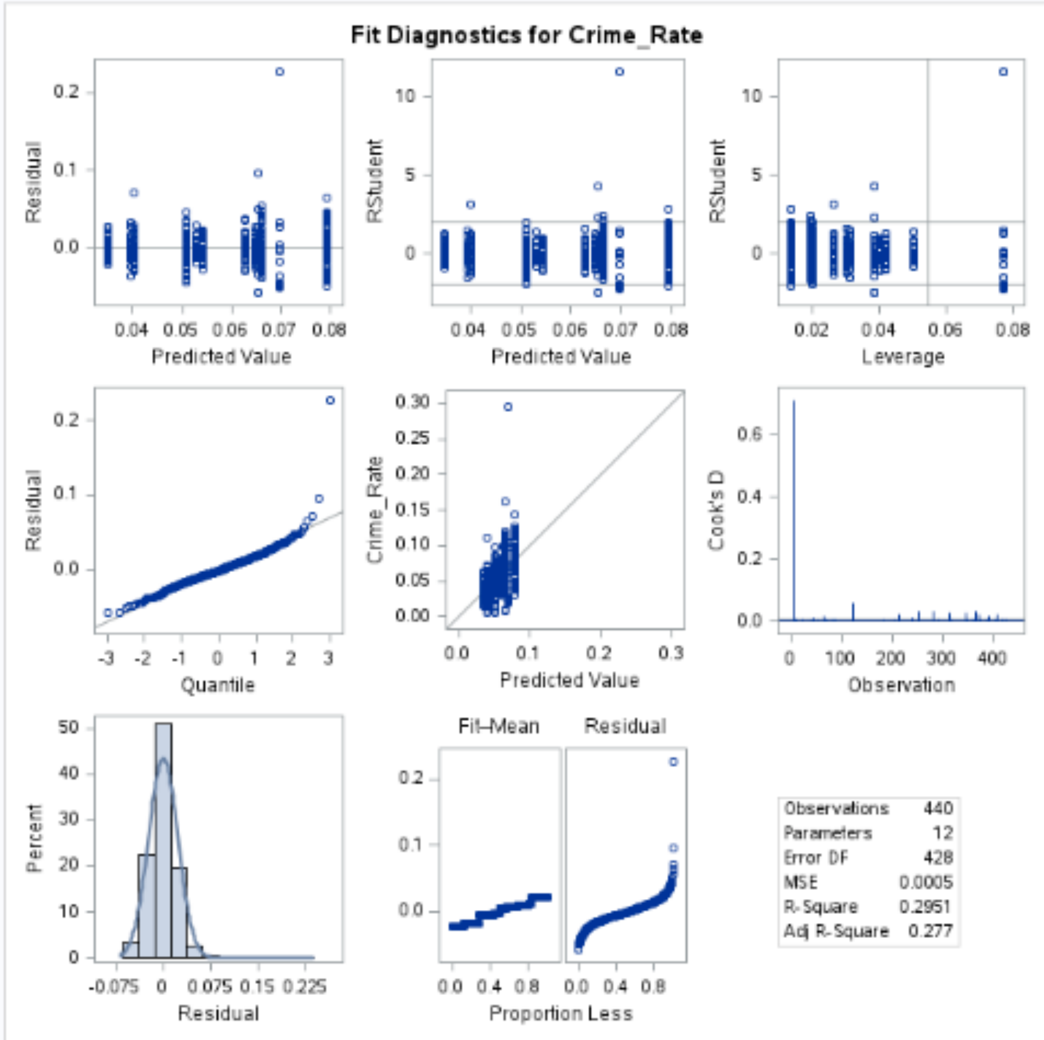
H_0: the residulas follow normal distribution.

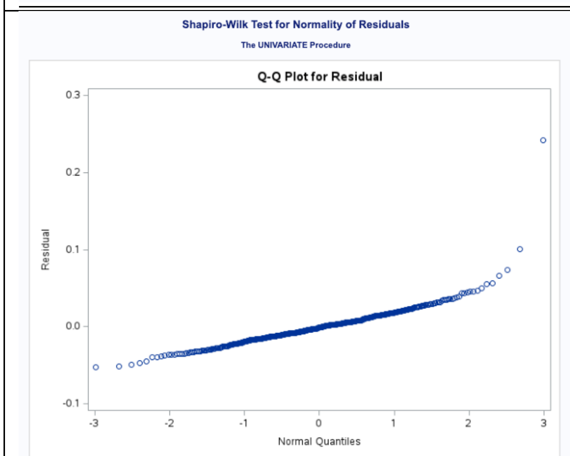
H_a: the residuals ar not normal distribution

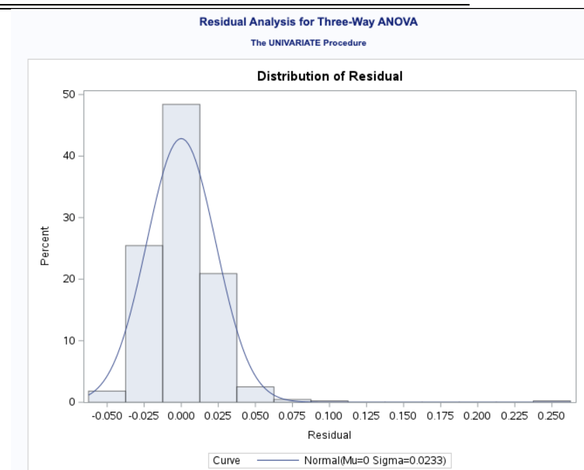
We performed Shapiro-Wilk test. Statistics W=0.852747 and the P-value = 0.0001 < 0.05. We reject the null hypothesis.

Q-Q plot: We can observe that on the Q-Q plot that the normality is violated. Especially in the tails. We also notice outliers on the data.

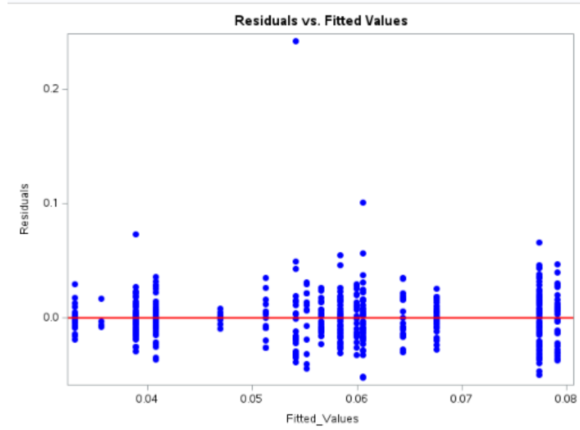
Histogram: However, the histogram appears approximately normal, symmetic and bell-shaped on visual observation.


Fit Diagnostics for Crime_Rate
Residual
Predicted Value
RStudent
Leverage
Quantile
Crime_Rate
Cook's D
Observation
Percent
Fit–Mean
Proportion Less
Observations 440
Parameters 12
Error DF 428
MSE 0.0005
R-Square 0.2951
Adj R-Square 0.277


Shapiro-Wilk Test for Normality of Residuals
The UNIVARIATE Procedure
Q-Q Plot for Residual
Residual
Normal Quantiles


Residual Analysis for Three-Way ANOVA
The UNIVARIATE Procedure
Distribution of Residual
Percent
Residual
Curve
Normal(Mu=0 Sigma=0.0233)

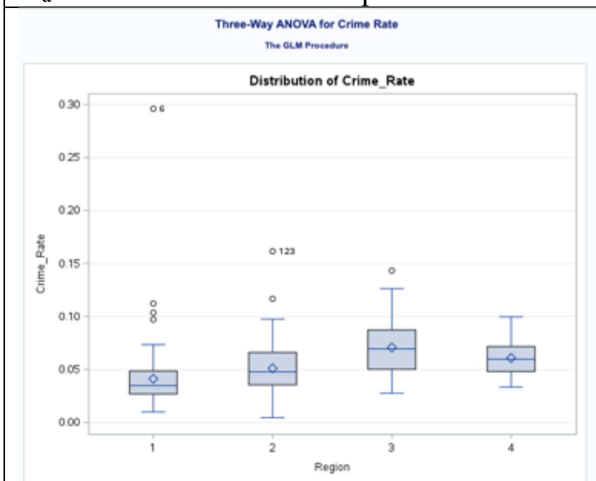
Homoscedasticity: the residual vs fitted vlaues does not show a randomness. We see evidence of Homoscedasticity. The residual values are scatter around the zero line. As the fitted value increses, we see variability. The



Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.852747	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.073255	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.641873	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	4.346596	Pr > A-Sq	<0.0050

Shapiro-Wilk's test P-value<0.0001. We reject the null hypothesis. The assumuption of the test does not hold. We conclude that the data does not follow a normal dictribution.

H_0: the variance are equal across the levels of Region
H_a: The variance are not equal across the levels of Region



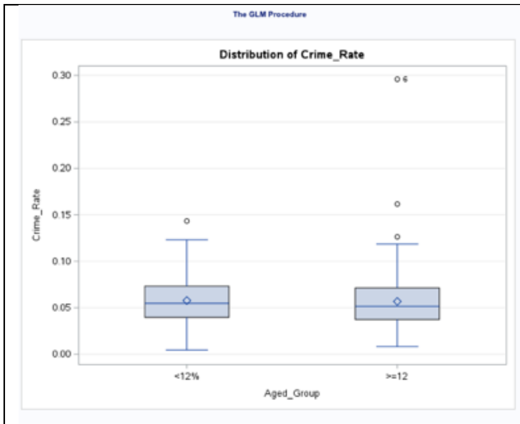
Levene's Test for Equality of Variances (Region)

The GLM Procedure

Levene's Test for Homogeneity of Crime_Rate Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Region	3	0.00254	0.000846	2.82	0.0387
Error	436	0.1308	0.000300		

The P value=0,0387<0.05, we reject the null hypothesis. We also observe that on the box plot, this indicates that the variances are not equal(homogeneity of variance is violated) across the regions.

H_0: The variances of crime_rate are equal across the levels of Age_Group (<12% and ≥ 12%)
H_a: The variances of crime_rate are not equal across the levels of Age_Group (<12% and ≥ 12%)



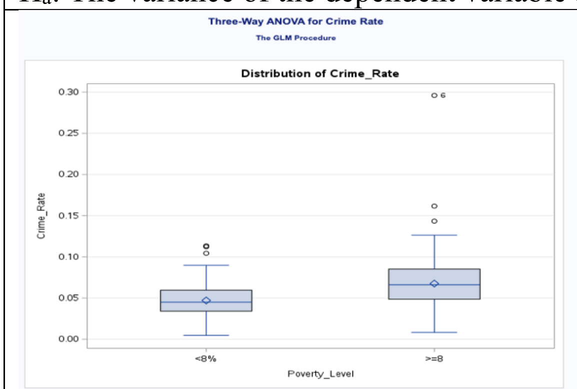
Levene's Test for Equality of Variances (Aged Group)

The GLM Procedure

Levene's Test for Homogeneity of Crime_Rate Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Aged_Group	1	0.000132	0.000132	0.41	0.5242
Error	438	0.1421	0.000324		

The P value=0.0001<0.5242, we reject the null hypothesis. We also observe that on the box plot, this indicates that the variances are equal(homogeneity of variance is satisfied) across the two (<12% and ≥12%)

H_0: The variance of the dependent variable are equal across the levels of Poverty_Level
H_a: The variance of the dependent variable are not equal across the levels of Poverty_Level



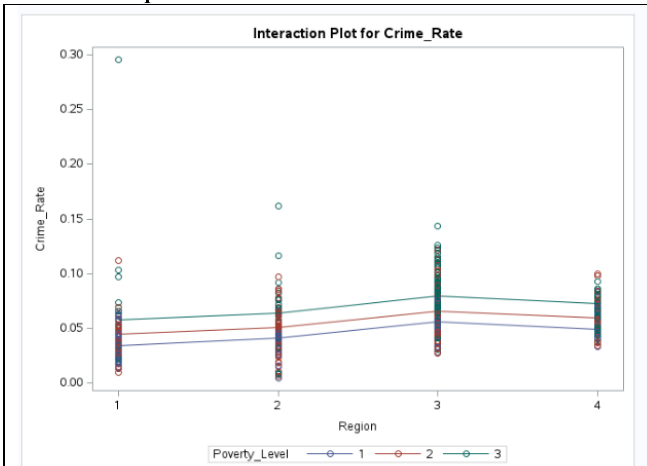
Levene's Test for Equality of Variances (Poverty Level)

The GLM Procedure

Levene's Test for Homogeneity of Crime_Rate Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Poverty_Level	1	0.00574	0.00574	19.72	<.0001
Error	438	0.1274	0.000291		

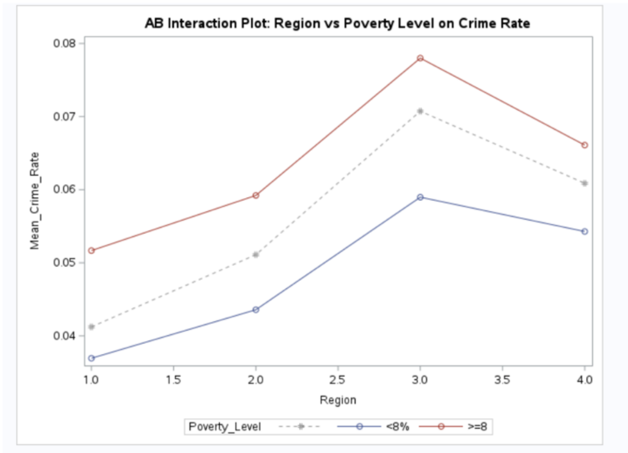
The P value=0.0001<0.05, we reject the null hypothesis. We also observe that on the box plot, this indicates that the variances are not equal(homogeneity of variance is violated) across the two Poverty_Levels(<8% and≥8%)

(ii) Prepare AB interaction plots of the estimated treatment means. Does it appear that any factor effects are present?

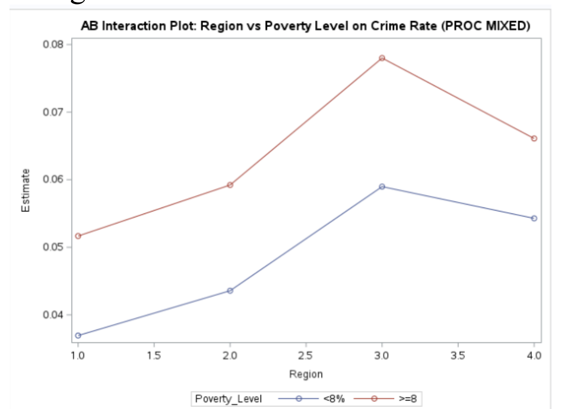


Interaction: Interaction plots display the relationship between Region vs Poverty Level on Crime Rate. The interaction plot lines are not parallel. This indicates an interaction effect between Region and Poverty level on crime rate. We also observe that as the poverty level goes up the crime rate is increasing and goes down after 3.0 Region. The difference in crime rates between the two poverty levels varies depending on the region, further supporting the presence of an interaction effect.

Main effect: Region and poverty level likely to have individual effects on the crime rate.



Using Proc mixed:



(iii) Test for three-factor interactions and for AB, AC and BC interactions. For each test, use =
:025
and state the alternatives, reduced regression model and P-values.

Full Model: $Y_{ijk} = \mu_{ijk} + \epsilon_{ijk} = \mu_{\ldots} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijk}$

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Region	3	0.03231691	0.01077230	19.21	<.0001
Poverty_Level	1	0.01268295	0.01268295	22.61	<.0001
Aged_Group	1	0.00103372	0.00103372	1.84	0.1753
Region*Poverty_Level	3	0.00118461	0.00039487	0.70	0.5500
Region*Aged_Group	3	0.00290920	0.00096973	1.73	0.1603
Poverty_L*Aged_Group	1	0.00022525	0.00022525	0.40	0.5266
Region*Povert*Aged_G	3	0.00069842	0.00023281	0.42	0.7423

ANOVA for Main Effects and Interaction Terms

The GLM Procedure

Dependent Variable: Crime_Rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	0.09004016	0.00600268	10.70	<.0001
Error	424	0.23780677	0.00056087		
Corrected Total	439	0.32784692			

R-Square	Coeff Var	Root MSE	Crime_Rate Mean
0.274641	41.34065	0.023683	0.057286

Three-factor interactions:Region*Poverty_Level*Aged_Group	**P-Value**
H_0: There is no three-factor interaction effect on Crime_Rate H_a: There is a three-factor interaction effect on Crime_Rate	$F_{obs} = 0.42$ $F_{critical(0.25,3,427)} = 3.14699 > 0.42$, accept the null hypothesis. The P-value=0.7423>0.025. Accept the null hypothesis. The three-factor interaction effect is not significant.
Two-Factor Interactions	**P-Value**
Region * Poverty_Level H_0:The interaction between Region and Poverty_Level has no significant effect on Crime_Rate Ha: The interaction between Region and Poverty_Level has a significant effect on Crime_Rate	$F_{obs} = 0.70$ $F_{critical().025,3,424)} = 3.1467 > F_{obs} = 0.42$, accept the null hypothesis. The P-value=0.55>0.025. Accept the null hypothesis. The three-factor interaction effect is not significant.
Region * Aged_Group H_0:The interaction between Region and Aged_Group has no significant effect on Crime_Rate	$F_{obs} = 1.73$, $F_{critical().025,1,424)} = 3.1467 > F_{obs} = 1.73$, accept the null hypothesis.

Ha: The interaction between Region and Aged_Group has a significant effect on Crime_Rate	The P-value=0.1603>0.025. Accept the null hypothesis. The three-factor interaction effect is not significant.
Poverty_Level * Aged_Group H_0:The interaction between Poverty_Level and Aged_Group has no significant effect on Crime_Rate Ha: The interaction between Poverty_Level and Aged_Group has a significant effect on Crime_Rate	F_{obs}= 0.40 $F_{critical().025,3,424)}$= 5.0597> F_{obs}=0.40, accept the null hypothesis. The P-value=0.5266>0.025. Accept the null hypothesis. The three-factor interaction effect is not significant.
Reduced Regression Models $Y_{ijk} = \mu_{ijk} + \epsilon_{ijk} = \mu_{..} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijk}$	
Region*Poverty_Level H_0: All $(\alpha\beta)_{ij} = 0$ H_0: All $(\alpha\beta)_{ij} \neq 0$	F_{obs} = 0.60 P-Value= 0.6137>0.025. We accept the null hypothesis and conclude that the no interaction effect, not significant.
Region*Aged_Group H_0: All $(\alpha\gamma)_{ik} = 0$ H_0: All $(\alpha\gamma)_{ik} \neq 0$	F_{obs} = 1.41 P-Value= 0.2379>0.025. We accept the null hypothesis and conclude that the no interaction effect, not significant.
Poverty_Level*Aged_Group H_0: All $(\beta\gamma)_{jk} = 0$ H_0: All $(\beta\gamma)_{jk} \neq 0$	F_{obs} = 0.12 P-Value= 0.7255>0.025. We accept the null hypothesis and conclude that the no interaction effect, not significant.

(iv) Test for A, B, C main effects. For each test, use $\alpha = 0.025$ and state the alternatives, reduced regression model and P-values.

$Y_{ijkm} = \mu_{..} + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijkm}$

Reduced Regression Model:

$Y_{ijk} = \mu_{..} + \epsilon_{ijk}$

Testing Main Effects for Region, Poverty Level, and Aged Group

The GLM Procedure

Dependent Variable: Crime_Rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	0.08605161	0.01721032	30.89	<.0001
Error	434	0.24179531	0.00055713		
Corrected Total	439	0.32784692			

R-Square	Coeff Var	Root MSE	Crime_Rate Mean
0.262475	41.20284	0.023604	0.057286

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Region	3	0.05918092	0.01972697	35.41	<.0001
Poverty_Level	1	0.02613920	0.02613920	46.92	<.0001
Aged_Group	1	0.00073149	0.00073149	1.31	0.2525

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Region	3	0.03841764	0.01280588	22.99	<.0001
Poverty_Level	1	0.02325998	0.02325998	41.75	<.0001
Aged_Group	1	0.00073149	0.00073149	1.31	0.2525

Region (A) H_0: All α_i= 0; H_0: not all α_i =0	F_{obs} = 22.99, $F_{Critical(0.025,1,424)}$= 5.0597$F_{obs}$ =22.99> $F_{Critical}$ = 5.0597 P-value =0.0001<0.025, we reject the null hypothesis and conclude that region has significant effect on crime_rate.
Poverty_Level (B) H_0: All β_i = 0 H_0: not all β_i =0	F_{obs} = 41.75, $F_{Critical(0.025,1,424)}$= 5.0597$F_{obs}$ =41.75> $F_{Critical}$)= 5.0597 P-value =0.0001<0.025, we reject the null hypothesis and conclude that region has significant effect on crime_rate.
Aged_Group (C) H_0: All γ_k= 0 H_0: Not all γ_k =0	F_{obs} = 1.31, $F_{Critical(0.025,1,424)}$= 5.0597 F_{obs} =1.31< $F_{Critical}$)= 5.0597. P-value =0.2525>0.025, we accept the null hypothesis and conclude that region is does not significantly affect crime_rate.

3. Part III

This time assuming a mixed effects model with mixed factors A and C (< 12%,   12%) and a random factor B (1: under 4%, 2: 4-8%, 3: 8-12%). Derive a suitable model and conduct the suitable tests.

Fixed and Random Factors:Factor A(Region) and Factor C(Aged_group_2) are fixed factors and the Factor B(Proverty_Group) as the random factor.

$Y_{ijkm} = \mu_{..} + \alpha_i + \beta_j + (ab)_{ij} + \epsilon_{ijkm}$ where $\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$

Full Model:

Full model-mixed effect

Fit Statistics	
-2 Res Log Likelihood	-1875.84
AIC (smaller is better)	-1871.84
AICC (smaller is better)	-1871.81
BIC (smaller is better)	-1873.07
CAIC (smaller is better)	-1871.07
HQIC (smaller is better)	-1874.54
Generalized Chi-Square	0.21
Gener. Chi-Square / DF	0.00

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Aged_Group	1	410	0.17	0.6829
Poverty_Group	3	410	13.81	<.0001
Aged_Grou*Poverty_Gr	3	410	0.08	0.9723
Region	3	410	6.82	0.0002
Aged_Group*Region	3	410	0.64	0.5926
Poverty_Group*Region	9	410	2.23	0.0194
Aged_G*Povert*Region	7	410	0.37	0.9198

Reduced Model:

Fit Statistics	
-2 Res Log Likelihood	-1937.78
AIC (smaller is better)	-1933.78
AICC (smaller is better)	-1933.75
BIC (smaller is better)	-1935.00
CAIC (smaller is better)	-1933.00
HQIC (smaller is better)	-1936.47
Generalized Chi-Square	0.21
Gener. Chi-Square / DF	0.00

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Aged_Group	1	420	0.47	0.4954
Poverty_Group	3	420	11.74	<.0001
Region	3	420	7.81	<.0001
Aged_Grou*Poverty_Gr	3	420	0.38	0.7670
Poverty_Group*Region	9	420	2.39	0.0120

Chi-Square

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Region	3	430.7	17.47	<.0001
Aged_Group_2	1	431.1	0.65	0.4201
Region*Aged_Group_2	3	429.6	1.05	0.3710

Tests of Covariance Parameters Based on the Restricted Likelihood

Label	DF	-2 Res Log Like	ChiSq	Pr > ChiSq	Note
No G-side effects	1	-1944.41	51.19	<.0001	MI

Fit Statistics	
-2 Res Log Likelihood	-1995.61
AIC (smaller is better)	-1991.61
AICC (smaller is better)	-1991.58
BIC (smaller is better)	-1992.83
CAIC (smaller is better)	-1990.83
HQIC (smaller is better)	-1994.30
Generalized Chi-Square	0.23
Gener. Chi-Square / DF	0.00

Final model Decision: After perfoming a thorough analysis, we observed that the Poverty_Group has a high F-value and a small P-value, we conclude that it plays a significant role. The next high F-value is the Region, we can't drop that either. We decide to drop the Age_Group which has a smallest F-value with high P-value. Please refer the full model result in the table.

Our Final model has the Region as fixed effect(C). Random effect as Poverty_Group(B).

Final model result:

Fit Statistics	
-2 Res Log Likelihood	-2027.0
AIC (Smaller is Better)	-2023.0
AICC (Smaller is Better)	-2023.0
BIC (Smaller is Better)	-2024.2

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	60.12	<.0001

Final Model Result:

Solution for Fixed Effects

Effect	Region	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		0.05856	0.007568	3.65	7.74	0.0022
Region	1	-0.01315	0.003547	435	-3.71	0.0002
Region	2	-0.00650	0.003443	434	-1.89	0.0597
Region	3	0.008181	0.003239	433	2.53	0.0119
Region	4	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Region	3	434	17.73	<.0001

Random effect:

Covariance Parameter estimates: AIC =-1995.61, BIC=-1992,83

Poverty_Group variance=0.000179, Residual Variance=0.000524

Likelihood Ratio Test for Random Effects:

Chi-square=60.12, P-value<0.0001<0.05: Significant

Tests of Covariance Parameters Based on the Restricted Likelihood

Label	DF	-2 Res Log Like	ChiSq	Pr > ChiSq	Note
No G-side effects	1	-1944.41	51.19	<.0001	MI

MI: P-value based on a mixture of chi-squares.

Covariance Test Results (Corrected Output)

Obs	Label	DF	-2 Res Log Like	ChiSq	Pr > ChiSq	Note
1	No G-side effects	1	-1944.41	51.19	<.0001	MI

Fit Statistics	
-2 Res Log Likelihood	-1995.61
AIC (smaller is better)	-1991.61
AICC (smaller is better)	-1991.58
BIC (smaller is better)	-1992.83
CAIC (smaller is better)	-1990.83
HQIC (smaller is better)	-1994.30
Generalized Chi-Square	0.23
Gener. Chi-Square / DF	0.00

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	Poverty_Group	0.000179	0.000154
Residual		0.000524	0.000036

Interpretation: Variance due to poverty group is significant, which indicates that the random effects captures meaning full variability in crime rate.

Fixed effect:Region (A) remains in the model. We will exclude Aged_group(C) is excluded due to lack of significance.

Random Effects: Poverty_Group(B) is retained as a significant source of variability.

The Reduced model has the lower AIC and BIC value. That will be the best model.