

# SPATIAL STATISTICS: SPATIAL REGRESSION

**NAME:** OLUOMA CHRISTIAN CHIBUIKE

**REG NO:** 2018/245829

**COURSE:** GIS DESIGN AND DEVELOPMENT II (GSV 471) *ASSIGNMENT 2*

**DEPARTMENT / LEVEL:** GEOINFORMATICS & SURVEYING / 400 LEVEL

**LECTURER:** DR. I. N. MADUAKOR

**DATE:** 13TH OF MARCH, 2023

## Interpreting Regression Report Results

### Summary Statistics

Geoda:

```
>>03/14/23 03:43:01
REGRESSION
-----
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set      : south
Dependent Variable : HR90  Number of Observations: 1412
Mean dependent var : 9.5493  Number of Variables : 6
S.D. dependent var : 7.03637  Degrees of Freedom : 1406
```

ArcMap:

OLS Diagnostics			
Input Features:	south	Dependent Variable:	HR90
Number of Observations:	1412	Akaike's Information Criterion (AICc) [d]:	9008.825986

### Analysis Questions:

- 1) What summary statistics are provided in the output?

Ans: Number of observation, number of variables, dependent variables, mean and standard deviation of dependent variables.

- 2) How many observations were used in the regression?

Ans: 1412

- 3) How many variables were used and how many degrees of freedom are there?

Ans: Six (6) variables and 1406 degrees of freedom

## Traditional Measures of Regression Fit

Geoda:

R-squared	:	0.309158	F-statistic	:	125.839
Adjusted R-squared	:	0.306701	Prob(F-statistic)	:	0
Sum squared residual	:	48295.9	Log likelihood	:	-4497.37
Sigma-square	:	34.3499	Akaike info criterion	:	9006.75
S.E. of regression	:	5.86088	Schwarz criterion	:	9038.26
Sigma-square ML	:	34.2039			
S.E of regression ML	:	5.84841			

ArcMap:

Multiple R-Squared [d]:	0.309158	Adjusted R-Squared [d]:	0.306701
Joint F-Statistic [e]:	125.839368	Prob(>F), (5,1406) degrees of freedom:	0.000000*

4) What is R-squared?

Ans: The corrected/adjusted R-squared is 0.306701

5) How much variation in homicide rate (dependent variable) is accounted for by the variables in our model?

Ans: Approximately 31%

## Analysis of Individual Variables

Geoda:

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	8.96254	1.78134	5.03136	0.00000
RD90	4.58779	0.21457	21.3813	0.00000
PS90	1.9559	0.205401	9.52235	0.00000
UE90	-0.524402	0.0700278	-7.48849	0.00000
DV90	0.46159	0.115173	4.00781	0.00006
MA90	-0.0494819	0.0489015	-1.01187	0.31177

ArcMap:

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	8.962537	1.781336	5.031357	0.000001*	1.709932	5.241457	0.000000*	-----
RD90	4.587789	0.214570	21.381309	0.000000*	0.291864	15.718950	0.000000*	2.090407
PS90	1.955899	0.205401	9.522349	0.000000*	0.330308	5.921435	0.000000*	1.226016
UE90	-0.524402	0.070028	-7.488488	0.000000*	0.082467	-6.358941	0.000000*	1.892543
DV90	0.461590	0.115173	4.007811	0.000072*	0.115710	3.989207	0.000078*	1.101253
MA90	-0.049482	0.048901	-1.011869	0.311763	0.049898	-0.991655	0.321524	1.263381

6) Which variable(s) are not significant predictors of homicide rate?

Ans: MA90

7) Which variables are positive predictors? Which are negative predictors?

Ans: Positive: RD90, PS90, And DV90. Negative: MA90 and UE90

8) What is the regression equation?

$$HR90 = 4.59(RD90) + 1.96(PS90) - 0.05(MA90) + 0.46(DV90) - 0.52(UE90)$$

## **Measures of Comparability**

Geoda:

Log likelihood	:	-4497.37
Akaike info criterion	:	9006.75
Schwarz criterion	:	9038.26

ArcMap:

Akaike's Information Criterion (AICc) [d]:

9008.825986

## **Multicollinearity**

Geoda:

MULTICOLLINEARITY CONDITION NUMBER	30.863233
------------------------------------	-----------

ArcMap:

VIF [c]
-----
2.090407
1.226016
1.892543
1.101253
1.263381

9) Is there potential redundancy among variables based on the VIF?

Ans: No, all the VIF values are below 7.5

10) Is there potential correlation among variables?

Ans: Yes, the multicollinearity condition number is greater than 30.

## **Normality of Errors**

Geoda:

TEST ON NORMALITY OF ERRORS			
TEST	DF	VALUE	PROB
Jarque-Bera	2	2833.4241	0.00000

ArcMap:

Jarque-Bera Statistic [g]:	2833.424057	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
----------------------------	-------------	---	-----------

11) Is the jarque-Bera test significant? What does that mean?

Ans: Yes. The model could be missing variables

12) Does this agree with the R-squared value?

Ans: Yes. The R-squared indicated that our model only explained about 31% of the variance.

## **Tests for Heteroskedasticity**

Geoda:

DIAGNOSTICS FOR HETEROSKEDASTICITY			
RANDOM COEFFICIENTS			
TEST	DF	VALUE	PROB
Breusch-Pagan test	5	515.0765	0.00000
Koenker-Bassett test	5	124.2738	0.00000
SPECIFICATION ROBUST TEST			
TEST	DF	VALUE	PROB
White	20	242.8060	0.00000

ArcMap:

Joint F-Statistic [e]:	125.839368	Prob(>F), (5,1406) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	330.817288	Prob(>chi-squared), (5) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	64.758576	Prob(>chi-squared), (5) degrees of freedom:	0.000000*

13) Are any of the tests for heteroskedasticity significant? What does this mean?

Ans: Yes, all of them. Some variables might be a strong predictor in an area but weak in other areas.

## **Tests for Spatial Autocorrelation**

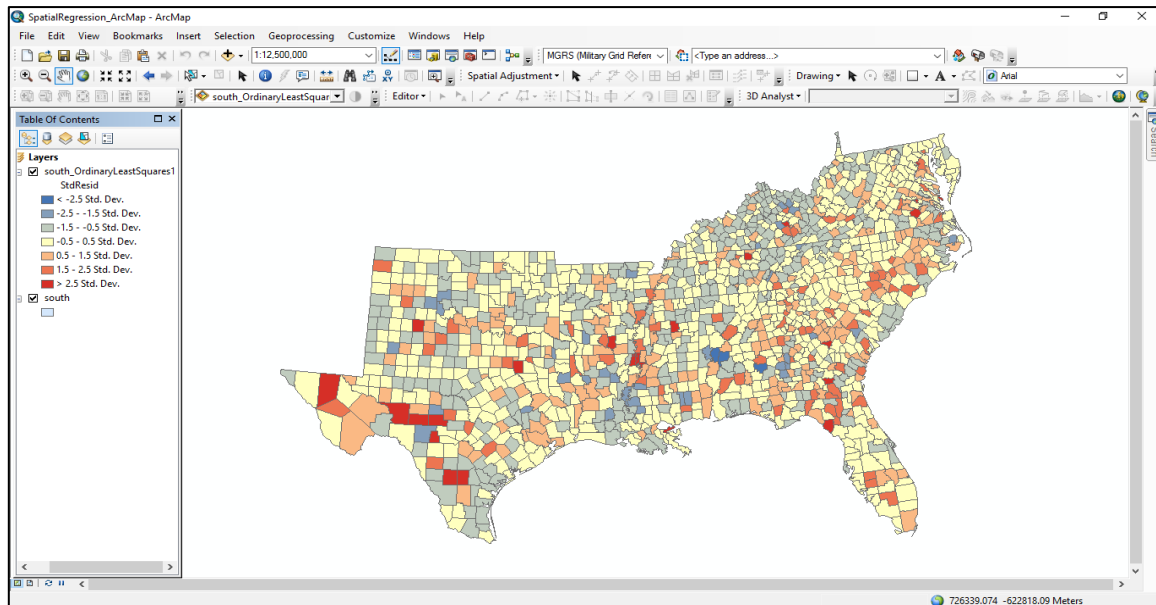
Geoda:

DIAGNOSTICS FOR SPATIAL DEPENDENCE			
FOR WEIGHT MATRIX : southrk			
(row-standardized weights)			
TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.0899	9.8643	0.00000
Lagrange Multiplier (lag)	1	71.6961	0.00000
Robust LM (lag)	1	4.7738	0.02890
Lagrange Multiplier (error)	1	89.3048	0.00000
Robust LM (error)	1	22.3825	0.00000

14) What test statistic should you use?

Ans: Since the standard version are significant, we should use the robust version.

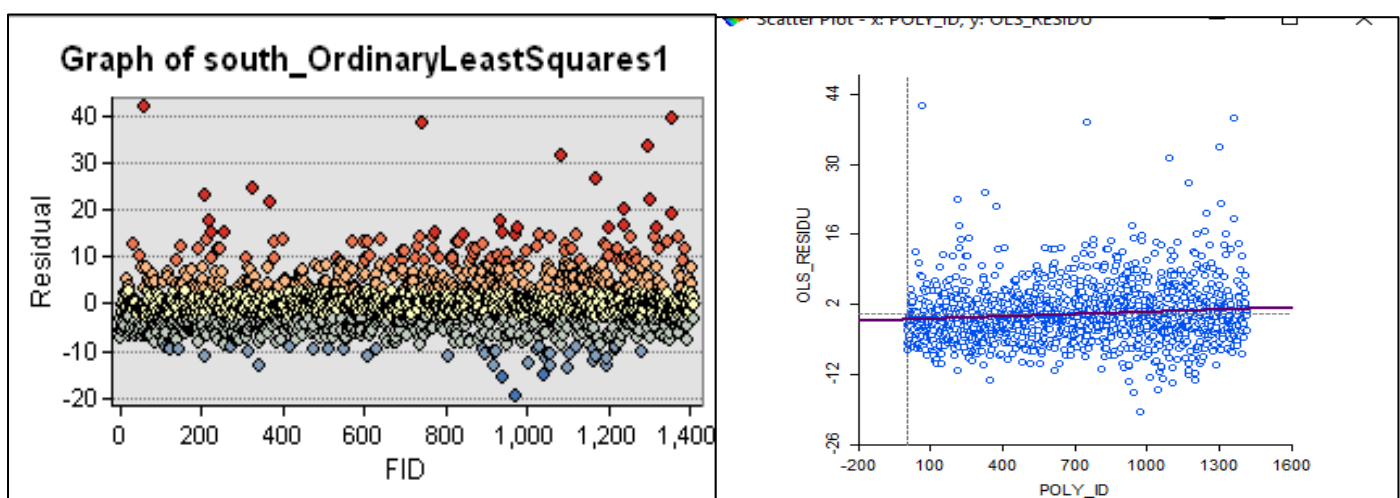
## Mapping Residuals



15) Do you see any patterns?

Ans: There are few clusters of high and low values.

## Plotting Residuals

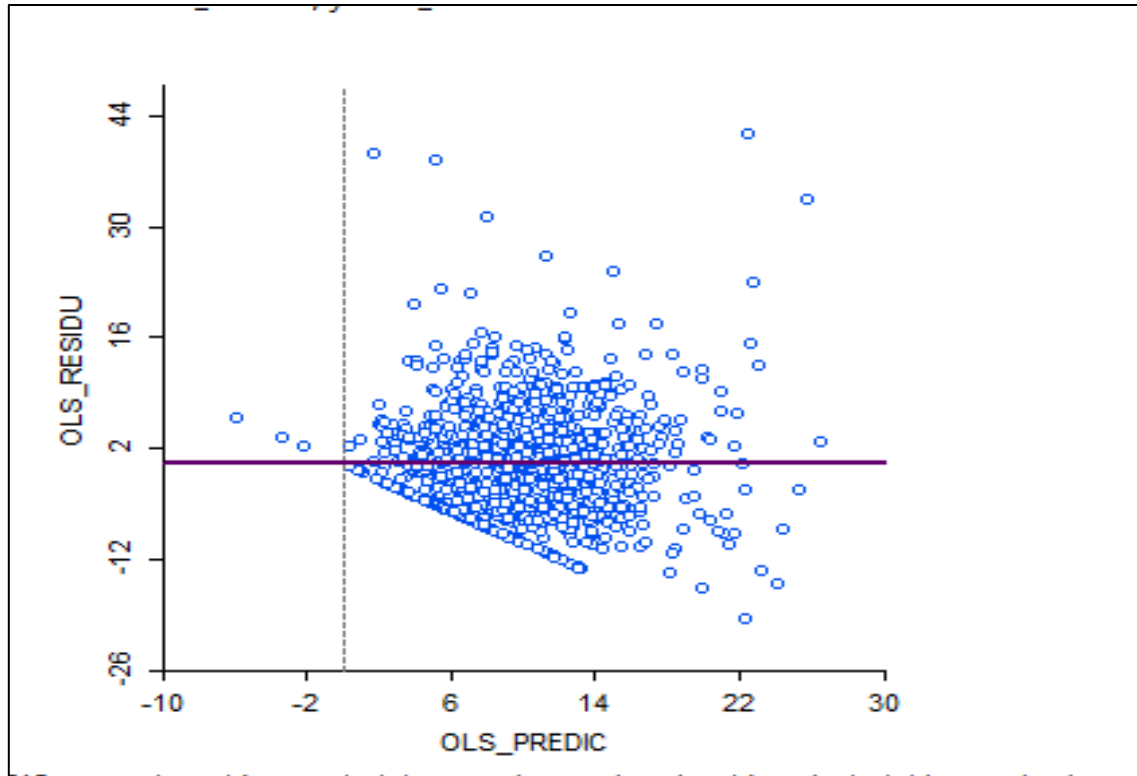


16) Are the large residual positive or negative? What does this mean?

Ans: They are positive residual. The actual crime rate is much higher than the model predicts.

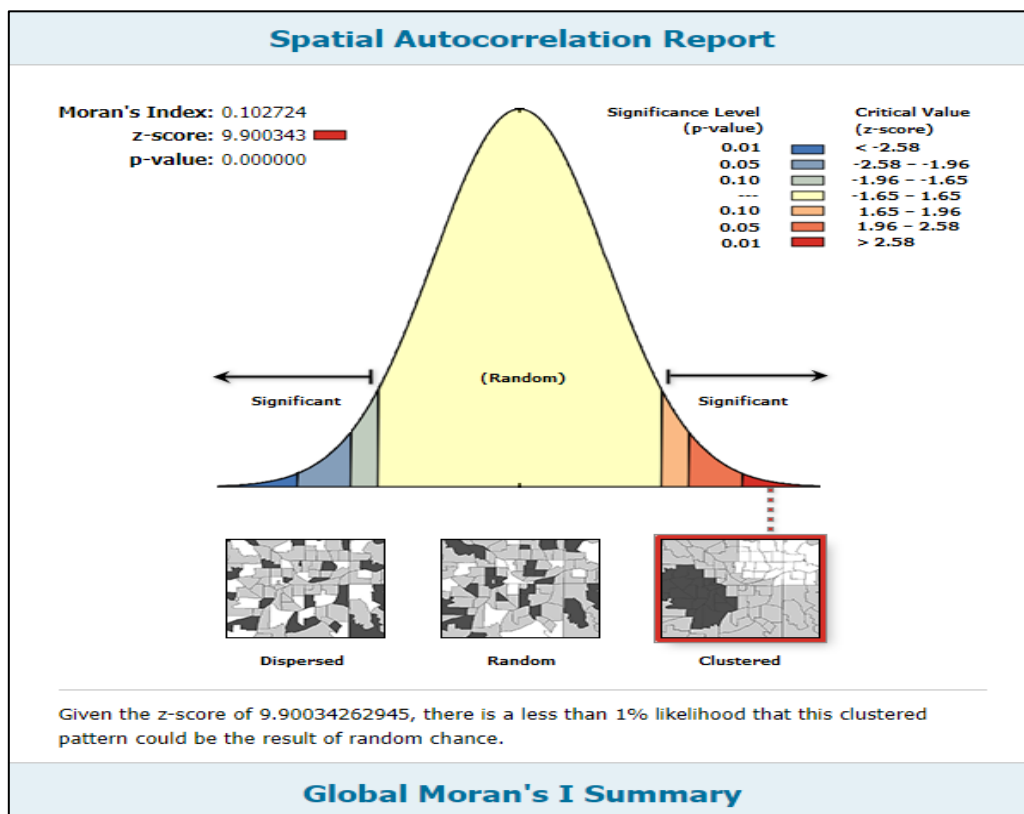
- 17) Where are the large residuals located?  
Ans: They are scattered across the map.

### **Plotting Residuals vs Predicted Values**



- 18) Do the residuals seem to be scattered with no visible pattern?  
Ans: Yes
- 19) Can you tell what that straight line of residual represents that runs diagonally across the bottom of the plot?  
Ans: They are variables with no homicide rate

## Testing Residuals for Spatial Autocorrelation



20) What does this tell you? Is there autocorrelation?

Ans: The low p-value indicates significant clustering or dispersion.



## Alternative Models

Geoda:

### Eliminating Variables

R-squared	:	0.308655	F-statistic	:	157.041
Adjusted R-squared	:	0.306689	Prob(F-statistic)	:	0
Sum squared residual	:	48331.1	Log likelihood	:	-4497.89
Sigma-square	:	34.3504	Akaike info criterion	:	9005.77
S.E. of regression	:	5.86093	Schwarz criterion	:	9032.04
Sigma-square ML	:	34.2288			
S.E of regression ML	:	5.85054			

- 21) How do the adjusted R-squared, log likelihood and AICc compare to the previous model?  
Ans: They are very similar.

### Spatial Lag Model

```
REGRESSION
-----
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set      : south
Spatial Weight : southrk
Dependent Variable : HR90      Number of Observations: 1412
Mean dependent var : 9.5493    Number of Variables : 7
S.D. dependent var : 7.03637   Degrees of Freedom : 1405
Lag coeff. (Rho) : 0.328975

R-squared      : 0.339353    Log likelihood      : -4471.24
Sq. Correlation : -          Akaike info criterion   : 8956.47
Sigma-square   : 32.7089    Schwarz criterion    : 8993.24
S.E of regression : 5.71917

-----
Variable      Coefficient      Std.Error      z-value      Probability
-----
W_HR90        0.328975      0.0447276     7.35508      0.00000
CONSTANT      3.51822      1.82602       1.92672      0.05401
RD90          3.83328      0.233595     16.41        0.00000
PS90          1.7169      0.20245       8.48063      0.00000
UE90         -0.403438     0.0687694    -5.86653     0.00000
DV90          0.474457     0.112469     4.21856     0.00002
MA90          0.00212062    0.0479761     0.0442016    0.96474
-----

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST      DF      VALUE      PROB
Breusch-Pagan test      5      631.0990     0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : southrk
TEST      DF      VALUE      PROB
Likelihood Ratio Test    1      52.2735     0.00000
===== END OF REPORT =====
```

- 22) How do the AICc value and the log likelihood compare to those from the OLS model?  
Ans: The AICc of the OLS model was 9006.75 compared to 8956.47 and the log likelihood was -4497.37 compared to 4471.24. The spatial lag model is a better fit.

## Spatial Error Model

REGRESSION				
-----				
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION				
Data set	:	south		
Spatial Weight	:	southrk		
Dependent Variable	:	HR90	Number of Observations:	1412
Mean dependent var	:	9.549301	Number of Variables	6
S.D. dependent var	:	7.036366	Degrees of Freedom	1406
Lag coeff. (Lambda)	:	0.485372		
R-squared	:	0.339860	R-squared (BUSE)	-
Sq. Correlation	:	-	Log likelihood	-4478.101679
Sigma-square	:	32.6838	Akaike info criterion	8968.2
S.E of regression	:	5.71698	Schwarz criterion	8999.72
-----				
Variable	Coefficient	Std.Error	z-value	Probability
-----				
CONSTANT	3.63152	2.07466	1.75041	0.08005
RD80	3.62419	0.22812	15.8872	0.00000
UE90	0.0282279	0.0745218	0.378788	0.70485
DV90	0.55578	0.130415	4.26162	0.00002
PS90	1.8655	0.239213	7.79847	0.00000
MA90	-0.0187204	0.0548882	-0.341064	0.73306
LAMBDA	0.485372	0.0511675	9.48595	0.00000
-----				
REGRESSION DIAGNOSTICS				
DIAGNOSTICS FOR HETEROSKEDASTICITY				
RANDOM COEFFICIENTS				
TEST		DF	VALUE	PROB
Breusch-Pagan test		5	634.5644	0.00000
DIAGNOSTICS FOR SPATIAL DEPENDENCE				
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : southrk				
TEST		DF	VALUE	PROB
Likelihood Ratio Test		1	73.3241	0.00000
===== END OF REPORT =====				

23) How do the AICc value and the log likelihood compare to those from the earlier models?

Ans: The AICc value and Log Likelihood value are smaller than the previous models. The spatial error model is a better fit.

## ArcMap:

### Geographically weighted Regression

Neighbors	: 145
ResidualSquares	: 37466.172664414189
EffectiveNumber	: 157.7136661693433
Sigma	: 5.4653920398246747
AICc	: 8896.719563447683
R2	: 0.46407026759790959
R2Adjusted	: 0.39710986875709275

24) How many neighbours were used in the calculations of the regression equations?

Ans: 145 neighbors were used.

25) How does this model compare to the OLS model?

Ans: The R-squared value for GWR (almost 0.4) is higher than that of the OLS model (0.31). The AICc for the GWR (8897) is lower than that of the OLS model (9008). The GWR is a better fit.

26) How does this model compare to the spatial lag model and the spatial error model?

Ans: The AICc of the GWR is lower than that of the spatial lag model and the spatial error model. The GWR is the overall best fit.