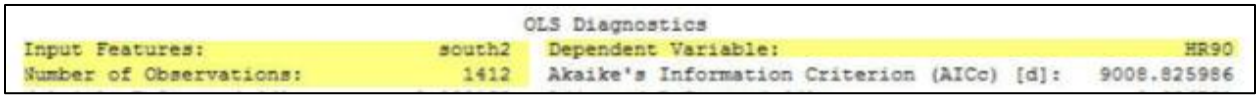
Interpreting Regression Output in Geoda and ArcMap

Summary Statistics:

Geoda:

```
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:



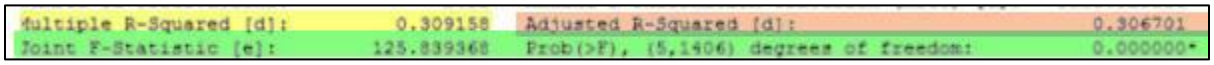
Traditional Measures of Regression Fit:

- F-statistics/Joint F-Statistic: typically rejects the null hypothesis and is not very useful.
- Adjusted R-squared: Tells you how much of the variation in the dependent variables is accounted for by the independent variables. Ranges from 0 (the independent variables are not related to the dependent variable) to 1 (the independent variables explain all variation in the dependent variable).

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:



Analysis of Individual Variables:

- Check the p-values of each variable to see if their coefficients are statistically significant.
- Look at the sign of the coefficient to determine whether the relationship is positive or negative.
- You can create the linear regression equation using these coefficients:
    - HR90 = 8.962 + 4.587(RD90) + 1.955(PS90) - .049(MA90) + .461(DV90) - .524(UE90)
- In ArcMap, if the Koenker statistic is significant, look at the robust p-value column.
- Try removing variables with high p-values from your model and observe the effect on R-squared.

Geoda:

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	8.962537	1.781336	5.031357	0.0000005
RD90	4.587789	0.2145701	21.38131	0.0000000
PS90	1.955899	0.2054009	9.522349	0.0000000
MA90	-0.04948188	0.04890147	-1.011869	0.3117676
DV90	0.46159	0.1151726	4.007811	0.0000645
UE90	-0.5244025	0.07002783	-7.488488	0.0000000

ArcMap:

Summary of OLS Results

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

VIF (Variance Inflation Factor):

- A VIF larger than 7.5 means there could be redundancy among variables. Remove variables with large VIFs one by one and observe the impact on R-squared.

Geoda: not included in output

ArcMap:

Summary of OLS Results

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

Measures of Comparability:

- When comparing models, look at the AICc value. A lower AICc value means the model is a better fit for the data.
- In Geoda, the higher the log likelihood, the better the fit. The lower the Schwartz criterion, the better the fit of the model.

Geoda:

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

ArcMap:

Akaike's Information Criterion (AICc) [d]: 9008.825986

Multicollinearity Condition Number:

- Values over 30 could indicate a problem with multicollinearity, meaning two predictor variables are highly correlated.

Geoda:

MULTICOLLINEARITY CONDITION NUMBER	30.863233

ArcMap: not provided in output

Jarque-Bera Test:

- Tests the normality of the errors.
- This should not be statistically significant. If it is, your model is biased and you may be missing an explanatory variable.

Geoda:

TEST ON NORMALITY OF ERRORS			
TEST	DF	VALUE	PROB
Jarque-Bera	2	2833.424	0.0000000

ArcMap:

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

Tests for Heteroskedasticity:

- The Breusch-Pagan tests the squares of the explanatory variables to determine if there is non-constant variance in the errors.
- The Koenker-Bassett is similar to the Breusch-Pagan, except the residuals are studentized, meaning they are made robust to outliers/non-normality.
- The White test tests the squares and cross-products of the explanatory variables. It tests for a more general form of heteroskedasticity.
- In ArcMap, if the Koenker statistic is significant, consult the Joint Wald statistic to determine the overall model significance. If it is not significant, consult the joint F statistic.
- When these tests are statistically significant, it indicates relationships between some or all of your explanatory variables and your dependent variables are non-stationary. One of your variables may be a strong predictor in some areas, but a weak one in others. In ArcMap, you may improve your model by running Geographically Weighted Regression (GWR).

Geoda:

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	5	515.0765	0.0000000
Koenker-Bassett test	5	124.2738	0.0000000

SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	20	242.806	0.0000000

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*

Tests for Spatial Autocorrelation:

Geoda:

- Only look at the robust versions when the standard versions are significant.
- Use flow chart below to determine what additional model to run, if any.

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : southrk12.gal
(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.089930	9.8642967	0.0000000
Lagrange Multiplier (lag)	1	71.6961448	0.0000000
Robust LM (lag)	1	4.7738150	0.0288957
Lagrange Multiplier (error)	1	89.3048170	0.0000000
Robust LM (error)	1	22.3824873	0.0000022
Lagrange Multiplier (SARMA)	2	94.0786320	0.0000000

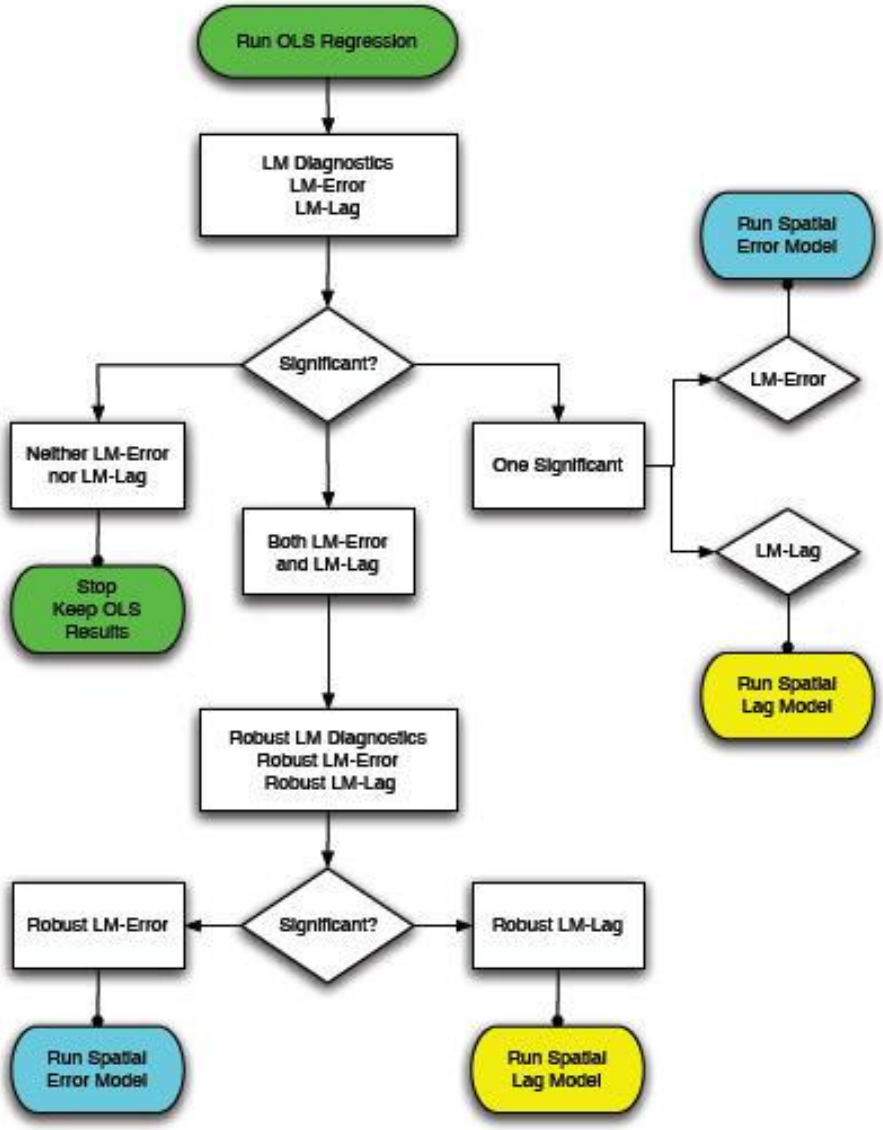


Figure 23.24: Spatial regression decision process.

From *Exploring Spatial Data with Geoda: A Workbook*, https://geodacenter.asu.edu/system/files/geodaworkbook.pdf, p. 199