

# SPPH567 TUTORIAL V

# TODAY'S TASKS

- Make scatter plot
- Get Pearson's Correlation
- Do simple linear regression with a continuous independent variable
- Examine diagnostics plots
- Do multiple linear regression

# MAKE SCATTER PLOT -- I

1. Open Plot Builder and go to tab Geometric Elements

The screenshot shows the Plot Builder application window. The title bar reads "Plot Builder". Below the title bar is a "File Tools Window" with several tabs: "Templates", "Geometric Elements", "Statistics", "Scales", "Facets", "Coordinates", and "Other". The "Geometric Elements" tab is active, displaying a grid of plot types. The "geom point" option is highlighted with a blue background. An orange callout bubble points to this option with the text "2. Double click geom point". Below the grid is a section for "template" plots, including "histogram", "bar", "mean", "scatter", "grouped dotplot", "grouped line", "histogram 2d", and "bubble". At the bottom of the window are "Run", "Reset", and "Cancel" buttons.

2. Double click geom point

# MAKE SCATTER PLOT -- 2

The screenshot shows a software window with a 'Position' dropdown set to 'identity' and a 'point' dropdown. The 'dat' dropdown is set to 'dat'. A list of variables is shown on the left, with 'Radon\_subLODroot2.tr' selected. The 'Mappings' tab is active, showing 'HomeAge1990' assigned to the X-axis and 'Radon\_subLODroot2.tr' assigned to the Y-axis. The 'Options' tab shows settings for Shape (● : 19), Colour (Set colour), Size (2.0), Fill (Set colour), Alpha (1.0), and Group. An orange callout points to the X-axis variable, and another points to the OK button.

1. Put in X and Y variable

2. Click OK

# MAKE SCATTER PLOT -- 3

Plot Builder

File Tools Window

Templates Geometric Elements Statistics Scales Facets Coordinates Other

geom density 2d geom dotplot geom errorbar geom errorbarh geom freqpoly geom hex geom histogram geom hline geom jitter

Radon\_subLODroot2.tr

HomeAge1990

Components

geom point

2. Click Run

Run Reset Cancel

I. Will see the preview scatter plot. Change x or y label if needed as in previous tutorials

# GET PEARSON'S CORRELATION-- I

The image shows the R Studio interface with the 'Correlation' dialog box open. The dialog box is titled 'Correlation' and has a close button (X) in the top right corner. The main window behind it shows the 'Console' and a menu bar with 'File', 'Edit', 'Workspace', 'Data', 'Analysis', 'Plots', 'Extras', 'Packages & Data', 'Window', and 'Help'. The 'Correlation' dialog box contains the following elements:

- Variables:** A list box containing 'Radon\_subLODroot2.tr' and 'HomeAge1990'. A callout points to this list with the text '2. Put the two continuous variables here'.
- With (Optional):** An empty list box.
- Correlation type:** Three radio button options: 'Pearson's' (selected), 'Kendall's', and 'Spearman's'. A callout points to the 'Pearson's' option with the text '3. Select Pearson's'.
- Subset:** A text box containing 'HomeAge1990<=20'. A callout points to this box with the text '4. You can type in the subset criteria here if you want to get the correlation for a subset of the data'.
- Buttons:** 'Run', 'Reset', and 'Cancel' buttons at the bottom. A callout points to the 'Run' button with the text '5. Click Run'.
- Other elements:** A 'Plots' button and an 'Options' button on the right side of the dialog box.

Callout 1: '1. Go to Analysis > Correlation' points to the 'Analysis' menu item in the main window.

Callout 2: '2. Put the two continuous variables here' points to the 'Variables' list box.

Callout 3: '3. Select Pearson's' points to the 'Pearson's' radio button option.

Callout 4: '4. You can type in the subset criteria here if you want to get the correlation for a subset of the data' points to the 'Subset' text box.

Callout 5: '5. Click Run' points to the 'Run' button.

# GET PEARSON'S CORRELATION-- 2

1. You would see this 2 by 2 table in the Console

```
Radon_subLODroot2.tr      Radon_subLODroot2.tr | HomeAge1990
cor | 1
N | 631
CI* |
stat** |
p-value | 0.1352
```

3. Why is this correlation = 1?

2. Correlation between Home Age and Log Radon

4. P-value

```
-----
HomeAge1990      cor | 0.05954
N | 631
CI* | (-0.01859, 0.137)
stat** | 1.496 (629)
p-value | 0.1352
```

```
** t (df)
* 95% percent interval
```

```
HA: two.sided
```

5. Why are these two cells the same?

# SIMPLE LINEAR REGRESSION - I

1. Go to Analysis  
> Linear Model

Linear Regression Model

dat

Filter:

UID  
Community  
MainRadon  
Location  
Foundation  
Soil  
LowerWindow  
Separation  
Heating  
AC  
RockType  
Potential  
MainRadon.tr  
Radon\_subLOD  
Radon\_subLODhalf  
Radon\_subLODhalf.tr  
Radon\_subLODroot2

Outcome  
Radon\_subLODroot2.tr

As Numeric  
HomeAge1990

As Factor

Weights

Subset

Continue Reset Cancel

2. Put the two  
continuous  
variables here

3. Click Continue  
here and in the  
next window



# SIMPLE LINEAR REGRESSION - 2

The screenshot shows the 'Linear Regression Model Explorer' window. The 'Diagnostics' tab is selected, displaying the following model output:

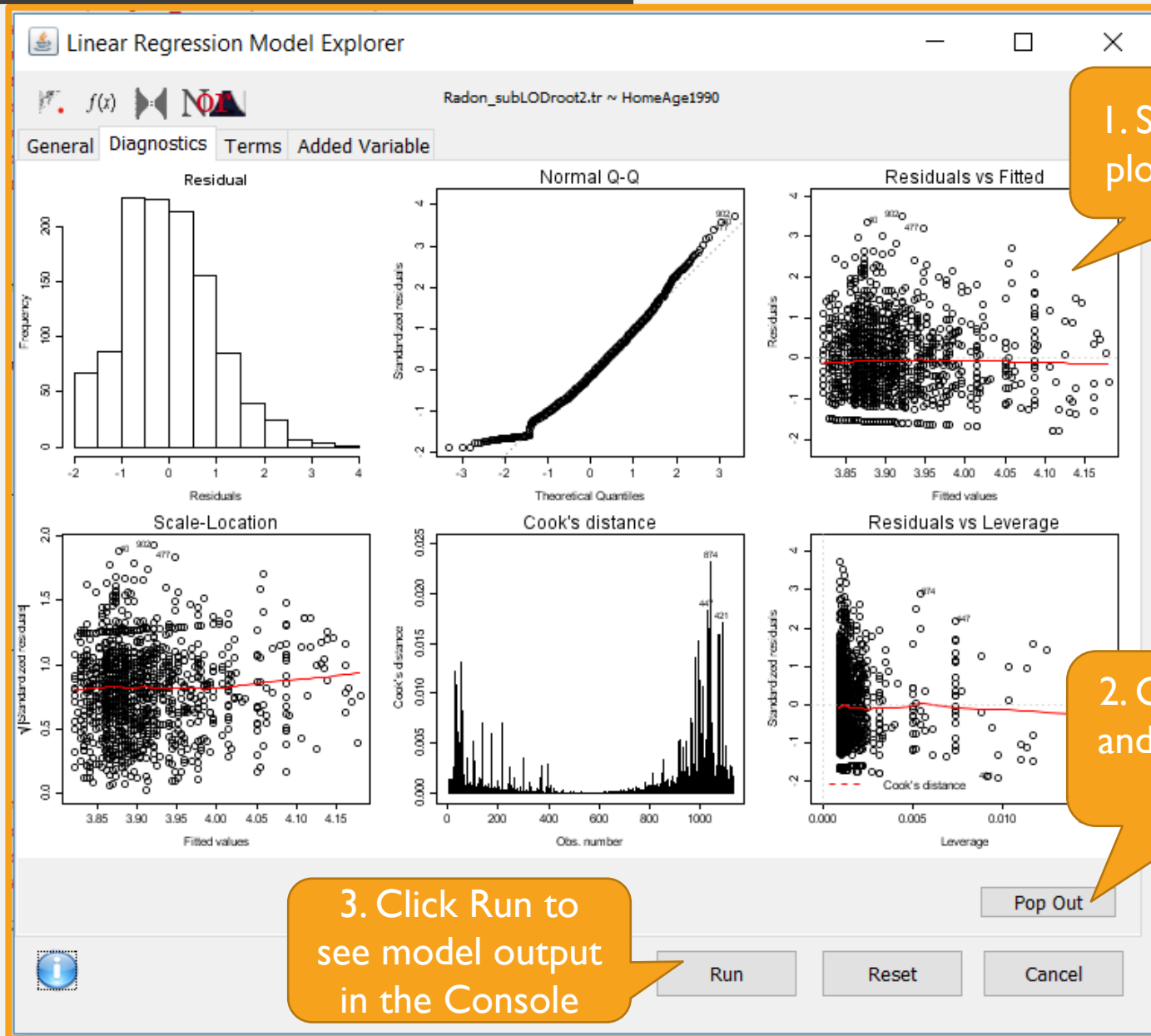
```
Radon_subLODroot2.tr ~ HomeAge1990  
General Diagnostics Terms Added Variable  
Previous  
Formula = Radon_subLODroot2.tr ~ HomeAge1990, data = .gui.wor  
dat,  
(.action = na.omit)  
Residuals:  
Min 1Q Median 3Q Max  
-1.7549 -0.6911 -0.0942 0.6016 3.5222  
Coefficients:  
Estimate Std. Error t value Pr(>|t|)  
(Intercept) 3.820963 0.046813 81.622 <2e-16 ***  
HomeAge1990 0.003685 0.001573 2.342 0.0193 *  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
Residual standard error: 0.9418 on 1132 degrees of freedom  
Multiple R-squared: 0.004824, Adjusted R-squared: 0.003945  
F-statistic: 5.487 on 1 and 1132 DF, p-value: 0.01933
```

Buttons on the right side of the window include: Options, Post Hoc, Tests, Plots, Means, Export, and Update Model. At the bottom are Run, Reset, and Cancel buttons.

2. Click the tab Diagnostics

1. Should see the model output in the preview

# DIAGNOSTICS PLOTS



1. Six diagnostics plots are shown

2. Click Pop Out and you can save the plots

3. Click Run to see model output in the Console

# MULTIPLE LINEAR REGRESSION

Linear Regression Model

dat

Filter:

UID  
Community  
MainRadon  
Location  
Soil  
LowerWindow  
Heating  
AC  
RockType  
Potential  
MainRadon.tr  
Radon\_subLOD  
Radon\_subLODhalf  
Radon\_subLODhalf.tr  
Radon\_subLODroot2

Outcome  
Radon\_subLODroot2.tr

As Numeric  
HomeAge1990

As Factor  
Separation  
Foundation

Weights

Subset

Continue Reset Cancel

I. Same as simple linear regression, just add more variables in Numeric or Factor