

High-Profile Shootings Since 1997
Tragedy after tragedy with guns...
 We want solutions.

Click the map above for details
<http://www.stategunlaws.org/>

Chapter 8 – Linear Regression

Today's Data: Detroit Homicides

- The data are on the homicide rate in Detroit for the years 1961-1973.
 - <http://www.canvons.edu/faculty/morrowa/140/datasets/>

FTP - Full-time police per 100,000 population
 UEMP - % unemployed in the population
 MAN - number of manufacturing workers in thousands
 LIC - Number of handgun licenses per 100,000 population
 GR - Number of handgun registrations per 100,000 population
 CLEAR - % homicides cleared by arrests
 WM - Number of white males in the population
 NMAN - Number of non-manufacturing workers in thousands
 GOV - Number of government workers in thousands
 HE - Average hourly earnings
 WE - Average weekly earnings
 HOM - Number of homicides per 100,000 of population
 ACC - Death rate in accidents per 100,000 population
 ASR - Number of assaults per 100,000 population

- Is there an association between the number of homicides and any other variable?

	FTP	UEMP	MAN	LIC	GR	CLEAR	WM	NMAN	GOV	HE	WE	HOM	ACC	ASR
1	260.35	11.0	455.5	178.15	215.98	93.4	558724	538.1	133.9	2.98	117.18	8.60	39.17	306.18
2	269.80	7.0	480.2	156.41	180.48	88.5	538584	547.6	137.6	3.09	134.02	8.90	40.27	315.16
3	272.04	5.2	506.1	198.02	209.57	94.4	519171	562.8	143.6	3.23	141.68	8.52	45.31	277.53
4	272.96	4.3	535.8	222.10	231.67	92.0	500457	591.0	150.3	3.33	147.98	8.89	49.51	234.07
5	272.51	3.5	576.0	301.92	297.65	91.0	482418	626.1	164.3	3.46	159.85	13.07	55.05	230.84
6	261.34	3.2	601.7	391.22	367.62	87.4	465029	659.8	179.5	3.60	157.19	14.57	53.90	217.99

Examine HOM vs CLEAR, FTP, and WE

Correlations: HOM, CLEAR, FTP, WE

	HOM	CLEAR	FTP	WE
CLEAR	-0.968			
FTP	0.964	-0.974		
WE	0.888	-0.936	0.922	

Goal:
 If we suspect there's a linear relationship, we find it...

The Linear Model

- If we suspect there is a linear relationship, we try to fit the **linear model**.

$$\hat{y} = b_0 + b_1x$$
- Minitab > Stat > Regression > Regression
 - Enter the Y-variable in "Response"
 - Enter the X-variable in "Predictors"
 - Results: Select the second option

Regression analysis: HOM versus FTP

The regression equation is
 HOM = - 77.6 + 0.337 FTP

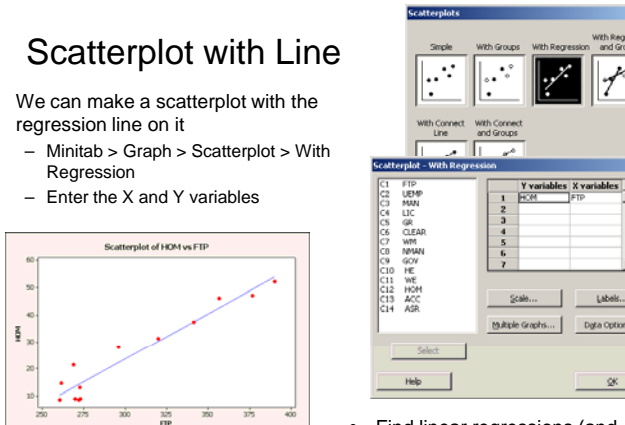
Predictor	Coef	SE Coef	T
Constant	-77.630	8.631	-8.99
FTP	0.33745	0.02804	12.03

S = 4.54687 R-Sq = 92.9% R-Sq(adj) = 92.7%

Source	DF	SS	MS
Regression	1	2994.4	2994.4
Residual Error	11	227.4	20.7

Scatterplot with Line

- We can make a scatterplot with the regression line on it
 - Minitab > Graph > Scatterplot > With Regression
 - Enter the X and Y variables



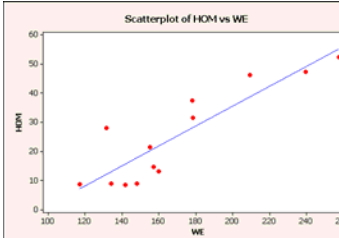
The screenshot shows the Minitab 'Scatterplot' dialog box with 'With Regression' selected. Below it is a 'Scatterplot - With Regression' window showing a list of variables (C1-FTP, C2-LEMP, C3-MAN, C4-LIC, C5-CLEAR, C6-WH, C7-IPMAN, C8-GOV, C9-HE, C10-WE, C11-HOM, C12-ACC, C13-GR, C14-AGR) and a table for Y and X variables. A scatterplot of HOM vs FTP is shown with a blue regression line.

- Note what happens when you hover over the regression line...
- Find linear regressions (and graphs) for HOM with each of CLEAR and WE.

Just Checking...

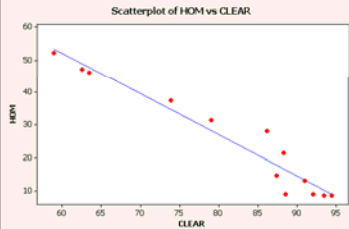
Regression Analysis: HOM versus WE

The regression equation is
 $HOM = -33.1 + 0.342 WE$



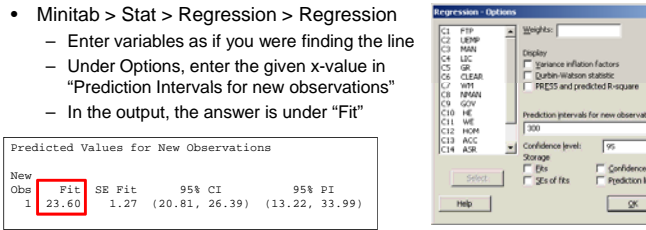
Regression Analysis: HOM versus CLEAR

The regression equation is
 $HOM = 127 - 1.25 CLEAR$



Making Predictions

- We can use the equation of the line to **make predictions** by substituting values in for the x-variable and calculating the corresponding y-variable.
- Use the equation to find the expected number of homicides per 100,000 for a year with 300 full-time police per 100,000.
- Minitab > Stat > Regression > Regression
 - Enter variables as if you were finding the line
 - Under Options, enter the given x-value in "Prediction Intervals for new observations"
 - In the output, the answer is under "Fit"

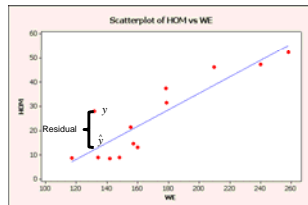


The screenshot shows the 'Regression - Options' dialog box with 'Prediction intervals for new observations' set to 300. Below it is a table of predicted values for new observations.

New Obs	Fit	SE Fit	95% CI	95% PI
1	23.60	1.27	(20.81, 26.39)	(13.22, 33.99)

So... What Exactly Makes The Line?

- The estimate made from the model is called the **predicted value**.
 - These are the y-values given by the line
- A **residual**, or error, is the difference between an observed value and its predicted value.



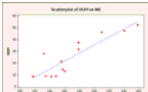
- The **regression line** gives the best fit in that it minimizes the sum of the squared lengths of the residuals
 - AKA **least squares line**
 - AKA **line of best fit**

$\hat{y} = b_0 + b_1x$ **The Slope and The Intercept**

- The slope: b_1 $b_1 = r \frac{s_y}{s_x}$
 - Tells how rapidly the predicted value for y changes with respect to x
 - The units of slope are "units of y per unit of x"
 - Interpretation: An increase in the x-variable by 1 corresponds to an increase in the y-variable by b_1 .
- The intercept: b_0 $b_0 = \bar{y} - b_1\bar{x}$
 - Tells where the line hits the y-axis
 - Interpretation: If 0 is a plausible value for x, then the intercept tells us the corresponding predicted y-value.
- Interpret the slope/intercept for the regression of HOM on WE

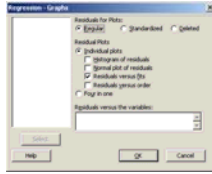
Regression Analysis: HOM versus WE

The regression equation is
 $HOM = -33.1 + 0.342 WE$

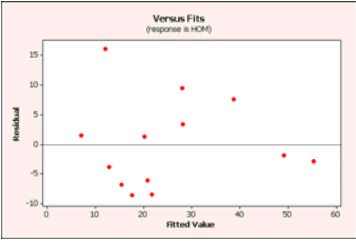


Diagnostics: Model Checking

- Once we fit the linear model, we examine the **residuals** to see how well it fit.
 - Residual = Data – Model
 - $e = y - \hat{y}$
- Diagnostic Tool: Residuals vs. Fitted Scatterplot
 - Minitab > Stat > Regression > Regression
 - Enter variables as if you were finding the line
 - Under Graphs, select "Individual plots" and "Residuals versus fits"
 - Note: Slightly different plot from book...
- Try it for HOM and WE




Just Checking...

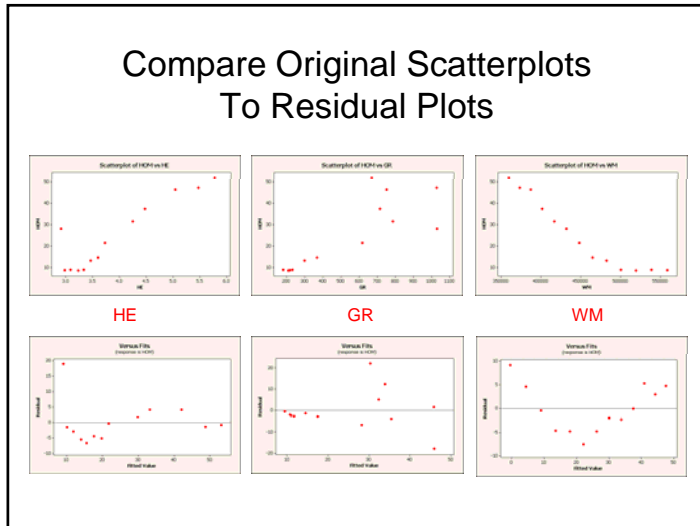


- This plots the residual against the fitted value.
- There should be no relationship between the two.
- Look For
 - Scattered points with no apparent pattern or interesting features
- For contrast, take a look at residual plots for
 - HOM with HE
 - HOM with GR
 - HOM with WM

Just Checking...



Residual plot examples where the linear model we've studied is not good.



R²

- **R²** = "R-squared" = r²: the squared correlation
 - R² gives the fraction of variation in the y-data accounted for by the model
 - 1 – R² gives the fraction of original variation left in the residuals
 - In Minitab, it's part of the regression output
- What do you mean "accounted for versus leftover"?

From the regression of HOM on FTP

Regression Analysis: HOM ver

The regression equation is
 $HOM = -77.6 + 0.337 FTP$

Predictor	Coef	SE Coef
Constant	-77.630	8.631
FTP	0.33745	0.02804

S = 4.54687 **R-Sq = 92.9%**

- ### Recap: Is the Regression Reasonable?
- Before Performing a Regression
 - Check the scatterplot (linear, no outliers)
 - After Regression
 - Check residual plot (no obvious patterns, no bends, no changes in spread, no obvious outliers)
 - Report R²
 - Notes on Predictions
 - Make sure the values make sense
 - Don't extrapolate beyond the data
 - Don't forget! Correlation is not causation!!

- ## Class Work
- To get credit, it is your responsibility to get checked off.
 - Chapter 8 Handout
 - Rules for checking answers: No Pens in the Front!!!
- ## Homework
- Textbook/Routine Homework
 - Due Next Week (25% chance of collection)
 - Read Chapter 8
 - Pg 216-225: #1, 3, 5, 7, 13, 21, 39, 45, 47, 57
 - Don't forget! If the question has a yellow T, the data can accessed through the link on our dataset webpage!
 - Project/Exploration Homework
 - Project #1