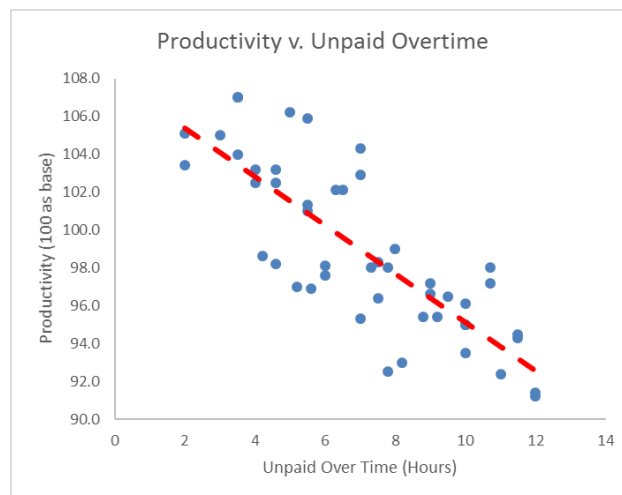


## Topic 4 Tutorial – Simple Linear Regression

### Brief Answers

#### Step 2. Scatter diagrams and Correlation analysis

- (a) Construct a **scatter diagram** between Productivity and Unpaid Overtime along with a **correlation analysis**.



<b>Correlation</b>	<b>-0.808</b>
--------------------	---------------

- (b) Using your diagram in (a) and the fact that  $r = -0.808$ , how would you describe the relationship between the variables.

Very strong negative linear relationship. It seems management's suspicions are correct.

#### Step 3. Regression Analysis

- (a) Build a **regression model** between Productivity and Unpaid Overtime.

## SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.808
R Square	0.652
Adjusted R Square	0.645
Standard Error	2.622
Observations	48

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	593.228	593.228	86.305	0.000
Residual	46	316.187	6.874		
Total	47	909.415			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	107.961	1.050	102.803	0.000	105.847	110.075	105.847	110.075
UOvTime	-1.283	0.138	-9.290	0.000	-1.560	-1.005	-1.560	-1.005

(b) Explain, in practical terms, the values of  $b_0$  and  $b_1$  in the above equation.

$b_0$ : On average, an employee who does no unpaid overtime will have a productivity of approximately 108%

$b_1$ : On average, for each extra hour of unpaid o/time an employee does, their productivity goes down by 1.3%

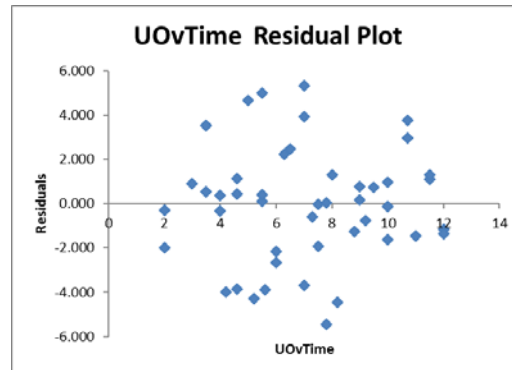
(c) From your regression output, write down the practical interpretations of for  $R^2$  and  $s_{yx}$ .

$R^2$ : Approximately 65% of the variation in productivity can be explained by the variation in the amount of overtime an employee does. The remaining 35% of variation would be explained by other factors not in the model.

$S_{yx}$ : On average, the error we will make when using the model is 2.6%.

#### Step 4. Residual Analysis

- (a) Check the model you created in question 3 does not violate any of the underlying assumptions of regression by doing a **residual plot(s)**.



- (b) Interpret the residual plot.

All points are randomly scattered – no obvious problems with assumptions.

- (b) Are there any outliers (influential values)?

Two potential outliers (observations 4 and 10).

#### Step 5. Use the Regression model

- (a) Use a calculator to predict the productivity on an employee who works 10 hours of unpaid overtime. (That is, determine a point estimate).

95.1%

- (b) Interpret the following interval estimates for an employee who works 10 hours of unpaid overtime.

	95% Lower	95% Upper
Confidence Interval	94.026	96.245
Prediction Interval	89.743	100.528

We are 95% confident that the average productivity of all BLITZ employees who work 10 hours of unpaid overtime will be somewhere between 94.0% and 96.2%.

We are 95% confident that the productivity of a BLITZ employee who work 10 hours of unpaid overtime will be somewhere between 89.7% and 100.5%.

- (c) Comment on the reliability of the result if we were to use the model to predict the productivity of an employee working 20 hours of unpaid overtime.

Extrapolation. Not reliable! Data set ranges from 2 to 12 hours. Thus we are trying to extrapolate and we can't be sure the same model holds outside the original domain.