R^2, A measure of Good Fit for the Multiple Regression Model

$H_0: \beta_1 = 0$

$H_a: \beta_1 > 0$

$$R^2 = \frac{\text{Explained Variable}}{\text{Unexplained variability}} = \frac{SS_{yy} - SSE}{SS_{yy}} = 1 - \frac{SSE}{SS_{yy}}$$

$= .925$

$$0 \le R^2 \le 1$$

So, $R^2 = 0$ means complete lack of fit

$R^2 = 1$ " a perfect fit

$R^2 = .925$ means that 92.5% of the variation in y around its mean is explained by the variables x_1 and x_2 in the model

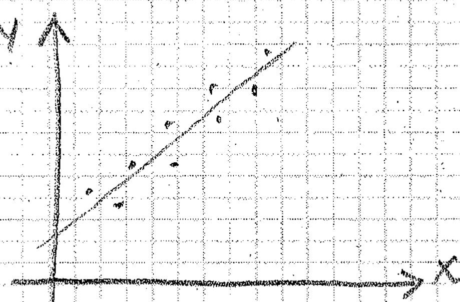
The Adjusted R^2 : R_a^2

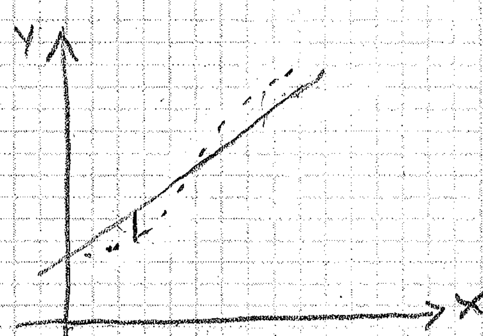
$$R_a^2 = 1 - \left[\frac{n-1}{n-(k+1)}\right]\left(\frac{SSE}{SS_{yy}}\right)$$

"adjusts" for the sample size and the number of parameters and cannot be forced to be 1 by adding more and more variables

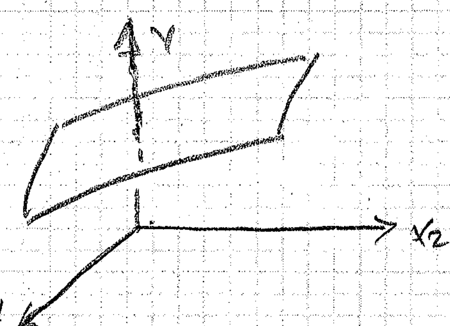
We can say that our model fits our data well if the differences between the predicted and observed values are <u>small</u> and <u>unbiased</u>

In the case of Simple Linear Regression we can easily get a visual confirmation by plotting Y vs x

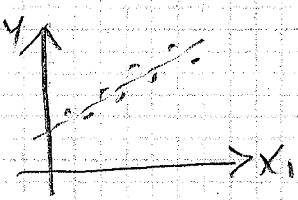


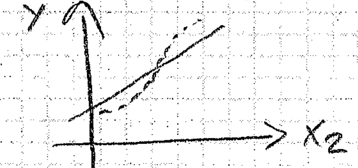


For multiple Regression, the situation is more complicated because we don't have a regression <u>line</u> but a regression <u>surface</u>

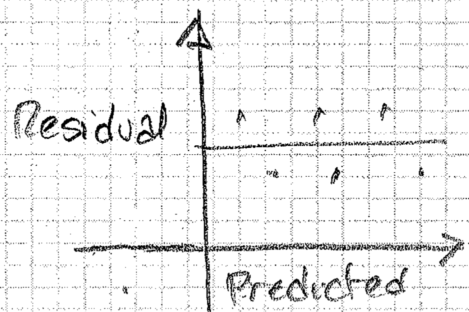


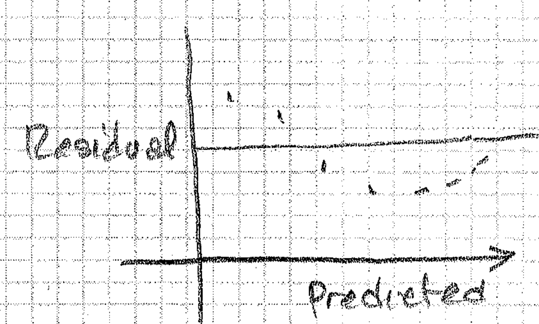
We could still do plots of Y vs X_1 and Y vs X_2





And we can do a plot of Residuals vs Predicted values which will give us an idea of the combined effects of x_1 and x_2





Bad Fit
Bias