Should a Predictor be Added or Transformed

- When choosing possible predictors for a regression model, one question that comes up is whether one predictor would be useful (i.e. explain the variability of the response) given the presence of other predictors.
- A simple plot of the response versus the predictor in question does not **take into account the effect of the other predictors**.
- The usual residual plot addresses the fit with all the predictors simultaneously.

Partial Regression Plots

- An **partial regression plot** illustrates the **marginal effect** of adding a predictor when the other predictors are already in the model.
- If **there is a pattern** in this plot, it indicates that the predictor will be a useful addition to the model.

Partial Regression Plots

To create an partial regression plot for predictor x_j:

1. We regress y against the predictors that are already in the model and obtain the residuals, $e(y|-x_j)$.
2. We regress the predictor in question against the other predictors in the model and obtain the residuals, $e(x_j|-x_j)$.
3. Then, we plot the residuals against each other, $e(y|-x_j)$ against $e(x_j|-x_j)$.

Partial Regression Plots

Example: for a data set with response variable y and two predictor variables x_1 and x_2. Given that x_1 is in the model, what is the marginal effect of x_2?

- Fit the regression function $y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$, and get $\hat{y}_i(x_1) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1}$ and $e_i\left(y|x_1\right) = y_i - \hat{y}_i(x_1)$,
- Fit the regression function $x_{i2} = \beta_0^* + \beta_1^* x_{i1} + \epsilon_i$, and get $\hat{x}_{i2}(x_1) = \hat{\beta}_0^* + \hat{\beta}_1^* x_{i1}$ and $e_i\left(x_2|x_1\right) = x_{i2} - \hat{x}_{i2}(x_1)$,
- Fit the regression function $e_i\left(y|x_1\right) = \beta_2 e_i\left(x_2|x_1\right) + \epsilon_i$, then $\hat{\beta}_2$ will be the same as the estimated β_2 when fitting the regression function $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$.

Partial Regression Plots

- $e_i\,(y|x_1)$ contains the variation in y that cannot be explained by x_1.
- $e_i\,(x_2|x_1)$ contains the variation in x_2 that cannot be explained by x_1.
- Since x_1 is used in both models, the relationship between $e_i\,(y|x_1)$ and $e_i\,(x_2|x_1)$ informs us about the relationship between y and x_2, beyond what can be explained by x_1.

Properties of Partial Regression Plots

- The estimated slope is the **coefficient** of x_2 in the full model.
- The estimated intercept is 0.

Patterns in Partial Regression Plots

- The prototype of partial regression plots includes: (1), Random horizontal band (no pattern); (2), Linear pattern; (3), Quadratic pattern.
- (1) Indicates the predictor is not needed in the model.
- (2) Indicates a linear term for the predictor is appropriate.
- (3) Indicate a quadratic term for the predictor should be used.

t tests in MLR

An insignificant t test for x_j in MLR can mean two things:

- Predictor x_j can be removed from the model.
- Predictor x_j may have a marginal non-linear relationship with y (after accounting for other predictors).

Partial regression plots can give us insight into whether the 2nd issue is present.

Worked Example: NYC Italian Restaurants

Data from Zagat in 2001, on 168 NYC Italian Restaurants:

- y: Price
- x_1: Food Rating
- x_2: Decor Rating
- x_3: Service Rating
- x_4: East of 5th Avenue (or not)

Worked Example: NYC Italian Restaurants

Worked Example: NYC Italian Restaurants

	Price	Food	Decor	Service
Price	1.0000000	0.6270435	0.7243525	0.6411402
Food	0.6270435	1.0000000	0.5039161	0.7945248
Decor	0.7243525	0.5039161	1.0000000	0.6453306
Service	0.6411402	0.7945248	0.6453306	1.0000000

Worked Example: NYC Italian Restaurants

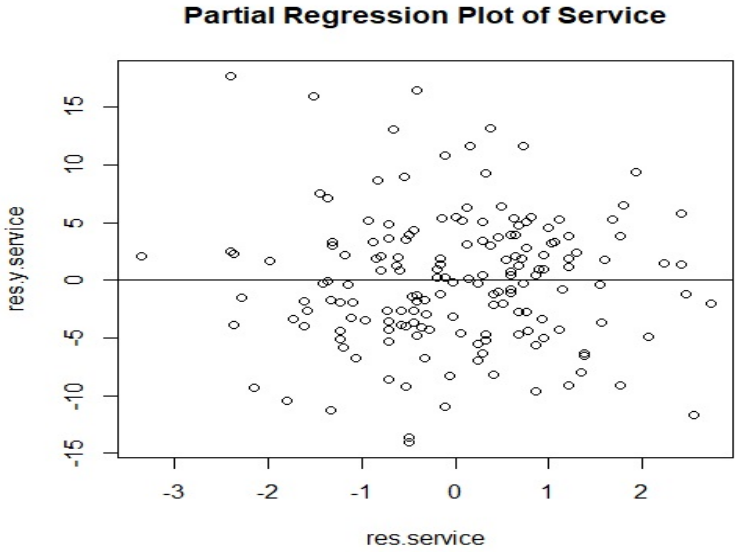
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-24.023800	4.708359	-5.102	9.24e-07	***
Food	1.538120	0.368951	4.169	4.96e-05	***
Decor	1.910087	0.217005	8.802	1.87e-15	***
Service	-0.002727	0.396232	-0.007	0.9945	
East	2.068050	0.946739	2.184	0.0304	*

Worked Example: NYC Italian Restaurants

The predictor Service is insignificant. So,

- Service can be dropped from the model.
- Perhaps Service should be a non-linear term instead?

Worked Example: NYC Italian Restaurants

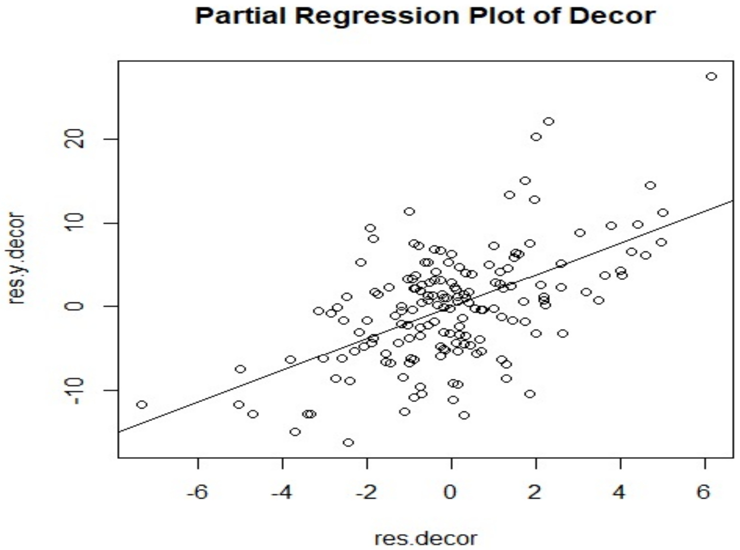


Worked Example: NYC Italian Restaurants

From partial regression plot of Service,

- Plots are in a horizontal band around 0.
- No curvature to indicate that a non-linear term for Service should be considered.
- Remove Service from model, no need to use a non-linear term for Service.

Worked Example: NYC Italian Restaurants



Worked Example: NYC Italian Restaurants

From partial regression plot of Decor,

- Plots are increasing linearly.
- A linear term for Decor should be used.
- Absence of curvature means no need to consider a quadratic term for Decor.
- The slope of this partial regression plot is equal to the coefficient for Decor in MLR.
- The intercept of this partial regression plot is 0.