**Causal Factor**: x, **Criterion**: y

Pre-Experimental (observational) study: subjects assign themselves to different groups on their own

True-experimental designs (randomized controlled) study: scientist assigns subjects to treatment and control groups at random

Quasi-experimental designs: scientist unable to achieve complete control over scheduling of treatments or cannot randomly assign respondents to experimental treatment conditions, but has stringer controls than with basic observational studies

---

**Notation to Represents Experimental Designs**

**X**: exposure of individual, group, or other entity to the experimental treatment

**O**: observation or measurement of the test unit

**R**: randomized assignment

Movement through time represented by horizontal arrangement of X's and O's from left to right. Simultaneous exposure or measurement by a vertical arrangement

Factors other than the experimental variable that affect the dependent variable are called nuisance, confounding, or extraneous factors/vars. When nuisance factors are not properly controlled, they are said to cofound the effects of the experimental variable

**Internal validity:** The ability of the experiment to unambiguously show a cause-and-effect relationship, i.e., to what extent can we attribute the effect that was observed to the experimental variable and not the other (cofounding) factors?

**External validity:** The extent to which the results of the experiment can be generalized to other people, settings (e.g., geography), and time (seasonality)

---

**After-only design** (one-shot case study)

| | |
|---|---|
| *X* | *O* |

**Threats to IV**: history, maturation, selection

**One-group Pretest-Posttest Design** (Before-after)

| | | |
|---|---|---|
| $O_1$ | *X* | $O_2$ |

**Result of interest:** D_hat = $O_2 - O_1$

**Analysis:** Paired-sample t test

**Threats to IV:** history, maturation, pre-measurement, placebo effect

**Threats to Internal Validity (One-group PP Design)**

**Interaction** (or interactive testing) **effect:** when a pre-measure changes the respondent's sensitivity or responsiveness to the independent variable(s). This is only a threat to external validity

**Placebo effect:** respondents acts differently because they know that they are being exposed to the treatment

**Static group comparison:**

| | | |
|---|---|---|
| *TG:* | *X* | $O_1$ |
| *CG:* | | $O_2$ |

**Result of interest:** D_hat = $O_2 - O_1$

**Threats to IV:** selection, maturation / mortality (if treatment is unpleasant)

**Threats to Internal Validity (Static group compare)**

**Experimental mortality:** Differential loss of respondents from different groups

**Selection bias:** When the groups formed for the purposes of the experiment are initially unequal with respect to the dependent variable or in the propensity to respond to the independent variable

Remedies:

**- Randomization:** assign subjects to treatment and control group using a random procedure.

- **Matching:** match treatment and control groups with respect to variables which you suspect influence response (used with small sample sizes). * Form blocks of units that are similar * Randomly assign units within a block to treatment and control groups * Blocking is closely related to stratification. Blocking is used in experiments and stratification in surveys

- ***Control*** for other causal factors (forks) with regression

- ***Propensity score*** models for observational studies. Find matched "twin(s)" for each treated case that is as similar as possible prior to self-selection into treatment

**Threats to Internal Validity (General)**

**Statistical regression:** When individuals are assigned to groups because of their scores on some measurement, such as initial attitude towards a brand. (Also called the *regression effect*)

**Threats to External Validity (General)**

All previous threats to internal validity are also threats to external validity. In addition, there are the following:

**Surrogate situation:** All situational specifics (e.g. treatment conditions, time, location, lighting, noise, treatment administration, investigator, timing, scope and extent of measurement, etc.) of a study potentially limit generalizability.

**Measurement timing:** When pre-measurements are made at an inappropriate time to indicate the effect of the experimental treatment, e.g., effect of temporary price cut on forward buying

**True Design experiments**

**After-only vs Before-after with control group**

**After-only:**

| | | | |
|---|---|---|---|
| *TG (R):* | | *X* | $O_1$ |
| *CG (R):* | | | $O_2$ |

**Before-after:**

| | | | |
|---|---|---|---|
| *TG (R):* | $O_1$ | *X* | $O_2$ |
| *CG (R):* | $O_3$ | | $O_4$ |

**Result of interest:** D_hat = $(O_2 - O_1) - (O_4 - O_3)$

**Analysis:** Compute differences between post and pre measures and compare with independent sample *t* test

**Threats to IV:** Interactive testing effect

**Quasi-Experimental Designs**

**Before-after with Control Quasi Design**

| | | | |
|---|---|---|---|
| *TG:* | $O_1$ | *X* | $O_2$ |
| *CG:* | $O_3$ | | $O_4$ |

**Result of interest:** D_hat = $(O_2 - O_1) - (O_4 - O_3)$

**Analysis:** Independent-sample *t* test on differences

- Sometimes, treatment and control *matched*: units assigned to treatment and control based on key factors

**Threats to Validity:** selection, interactive testing effects

---

**Regression Terms and Symbols:**

| Term | ACT | JWHT | Other |
|---|---|---|---|
| Sum of squared errors | **SSE** | **RSS** | |
| Total sum of squares | **SST** | **TSS** | |
| Mean squared error | **MSE** | | $S_e^2$ |
| Residual Standard Error | | **RSE** | |

**SSE** = $\sum_{i=1}^{n}[y_i - f(X_i)]^2$

**MSE** = $\frac{SSE}{n}$

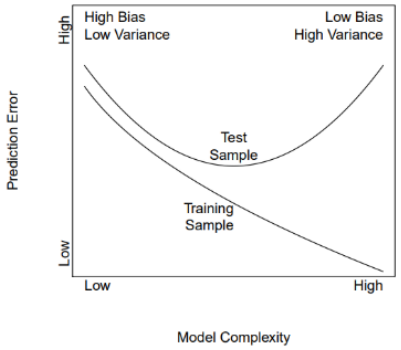
$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{RSS}{(n-1)var(y)}$$

**Penalized Estimates**

$R^2 = 1 - \frac{\frac{SSE}{n-p-1}}{\frac{SST}{n-1}}$ or AIC = deviance + 2p

---

**Model Complexity**




**Training data:** Used to estimate model parameters

**Validation data:** Used to select model hyperparameters (or use K-fold cross validation if you are data poor)

**Test data:** Used for final, inter-model comparisons

**Model fit (SSE and deviance) adding predictors will:**

- Always improve SSE on *estimation* sample

- Not necessarily improve SSE on *validation* data

**Iterative Model Selection**

**Forward selection:** Begin with no vars. Add var the yields greatest significant improvement in SSE. Repeat until no significant improvement in SSE.

**Backward elimination:** Begin with all candidate vars. Drop variable that causes smallest non-significant increase in SSE.

**Stepwise selection:** (Usually) begin with no vars. Drop var that causes smallest *non-significant* increase in SSE. Add variable that yields greatest significant improvement in SSE. Repeat both previous steps until no improvement in SSE.

**Shrinkage Estimation**

$$\boldsymbol{MSE}(\hat{\boldsymbol{\beta}}) = variance + bias^2$$

The Gauss-Markov Theorem tells us that the OLS estimates are BLUE, and thus have the smallest MSE among unbiased estimates

Shrinkage estimation introduces bias that reduces the variance to give an estimate with lower overall mean squared error

**Ridge Regression**

Ridge penalizes $\Sigma_j \beta_j^2$ the **squared Euclidean length** of the slope vector

$$\hat{\beta}_\lambda = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left[ \sum_{i=1}^{k} (y_i - \beta_0 - \mathbf{x}_i^\mathsf{T}\boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right]$$

$$\hat{\beta}_\lambda = (\mathbf{X}^\mathsf{T}\mathbf{X} + \lambda \mathbf{I})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y}$$

$\lambda \geq 0$ is a constant, which determines how much to penalize large regression coefficients. When $\lambda = 0$ we get OLS and then $\lambda$ is big the penalty is great and the coeffects will be close to 0

**Ridge existence theorem:** There exist values of $\lambda$ so that $\hat{\beta}_\lambda$ has smaller mean squared error than OLS

Simulations have shown that ridge regression produces $\hat{y}$ values that are closer to the true values than PCR and stepwise regression

Scaling of *X* variables: standardize when units are incommensurate (glmnet does this by default)

**Criticisms of ridge:**

- The optimal value of $\lambda$ depends on $\beta$ and $\sigma^2$ (error variance), which are being estimated by the regression

- Lack of theoretical justification for particular penalty term. Why unweighted sum of squares?

**Lasso Regression**

Lasso penalizes $\Sigma_j |\beta_j|$ the **taxi-cab length** of the slope vector

$$\hat{\boldsymbol{\beta}}_\lambda = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left[ \sum_{i=1}^{k} (y_i - \beta_0 - \mathbf{x}_i^\mathsf{T}\boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right]$$

We can think of the methods as having different constraints:

- Subset selection: $\min_{\beta}(SSE)$ subject to $\Sigma_j \mathbf{1}(\beta_j \neq 0) \leq s$

- Ridge: $\min_{\beta}(SSE)$ subject to $\Sigma_j \beta_j^2 \leq s$

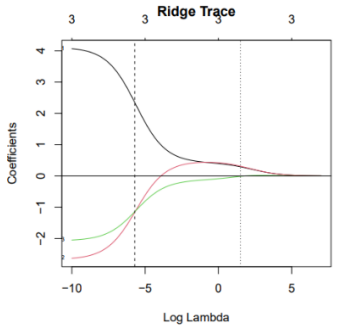
- Lasso: $\min_{\beta}(SSE)$ subject to $\Sigma_j |\beta_j| \leq s$

Lasso and ridge regression shrink coefficients towards zero, but the lasso tends to force some coefficients to equal zero, similar to variable subset selection

**Rule of thumb:** use lasso if you think some variables should be dropped

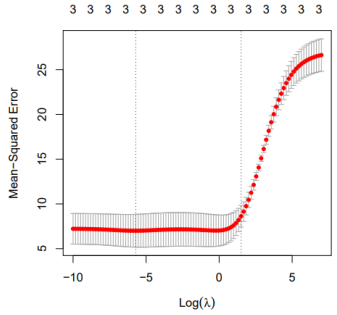
---

**Ridge Trace:**



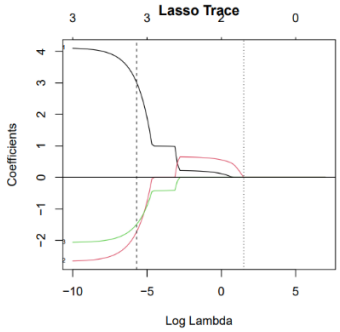

- The lines "gently" approach 0 as $\lambda \to \infty$

**Ridge: MSE vs Log($\lambda$)**

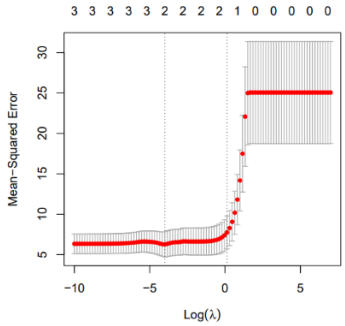



- Vertical lines show lambda.min and lambda.1se

**Lasso Trace:**




**Lasso: MSE vs Log($\lambda$)**




**Shrinkage model information**

Min = MSE + lambda * Penalty

Lambda -> infinity, big punishment for non 0 betas

$\lambda \to \infty$; big punishment for non 0 $\beta s$**,** high bias and low variance

$\lambda \to 0$; max strength model, low bias and high variance

---

**Multicollinearity**

**Pipe:** e.g., $x_1 \to x_2 \to y$. If you are studying $x_1 \to y$ then do not control for $x_2$

**Latent construct:** predictors manifestations of common, underlying **latent construct**, e.g., $w \to x_1$ *and* $w \to x_2$. Often, estimate *w* and use it instead of $x_1$ and $x_2$

**Back door confound** (**fork**): include control to block back-door path, e.g., if $w \to y$ and $w \to x \to y$ then control for *w* to study $x \to y$

**Collider:** usually do not control for colliders, e.g., if $x \to w$ and $y \to w$, then do not control for collider *w* when studying $x \to y$

---

**Extras**

$F = \frac{\frac{SST-SSE}{p}}{\frac{SSE}{n-p-1}} = \frac{\left(\frac{\Delta SSE}{\Delta df}\right)}{S_e^2}$, bottom from full model

**History:** other events that happened in between pre and post measures

**Maturation:** the possibility that mental or physical changes occur within the participants themselves that could account for the evaluation results

**Collider bias:** when an exposure and outcome each influence a common third variable (the collider) and that variable has been controlled for in the statistical analysis of the study data