Comparison of Cross-Validation Approaches

	Validation Set Approach	**LOOCV** (Leave One Out Cross Validation)	**PRESS** (Predicted Residual Sum of Squares)	**K-Fold** Cross Validation
Bias	**High.** Since it only uses a subset of data for training, it tends to overestimate the test error	**Low.** It uses `n - 1` observations for training, which is almost the entire data set, leading to a nearly unbiased estimate	**Low.** Functionally equivalent to LOOCV for linear models; it uses almost all data for each calculation.	Intermediate. Usually, `k = 5` or `k = 10` provides a balance between the high bias of the validation set and the low bias of LOOCV
Variance	**High.** The result can change significantly depending on exactly which observations end up in the training vs testing (validation) set	**High.** Because the `n` training sets are highly similar (correlated), the outputs of each fold are also correlated, leading to higher variance in the mean	**High.** Inherits the variance characteristics of LOOCV because it is a specific implementation of it	**Intermediate.** Offers a “sweet spot” with lower variance than LOOCV because the training sets are less overlapping.
Computational Cost	**Lowest.** The model is only trained `once`	**Highest.** Requires fitting the model `n` times (where `n` is the number of observations)	**Low (for linear models).** Can be calculated using a single model fit using a specific mathematical formula (`h_i` leverage values)	**Medium.** Requires fitting the model `k` times (usually 5 or 10)
Consistency	**Inconsistent.** Results vary with different random splits	**Consistent.** Will always produce the same result for the same dataset because there is no random sampling (***deterministic***)	**Consistent.** Like LOOCV, it is ***deterministic*** for a given data set	**Variable.** Results can change slightly depending on how the “folds” are randomly assigned

Comparison of Linear vs. Nonlinear Models

	Linear (Simple) Model	Nonlinear (Flexible) Model
Bias	**High:** these models make stronger assumptions and are more likely to have higher bias as they may be too simple to capture the true relationship	**Low:** These models are more flexible, allowing them to capture complex patterns and reduce the error introduced by simple approximations
Variance	**Low:** Simple models are stable; the estimated functions are less likely to change significantly with a different training dataset.	**High:** Flexible models are highly sensitive to the specific training data used, which can lead to significant changes in the model with different data sets
Training MSE	**Higher:** because they are less flexible, they cannot fit the training data points as closely as a complex model	**Lower:** highly flexible models (like degree 10 polynomials) will always have a training MSE less than or equal to a linear model
Test MSE	**Lower if the relationship is simple:** it generalizes better when the true relationship is close to linear	**Lower if the relationship is complex:** it performs better if it captures real nonlinear patterns, but becomes higher if the model overfits
Irreducible Error	**Fixed:** represents the noise (ϵ) in the data, which can NOT be reduced regardless of the model chosen.	

- ***Reducible Error:*** This consists of the squared Bias and the Variance. As you move from a linear model to a nonlinear model (increasing flexibility), bias generally decreases while variance generally increases.
- ***Overfitting:*** This occurs when a nonlinear model has extremely **low training MSE** but **high test MSE** because it has captured the irreducible noise in the training data rather than just the underlying pattern.
- ***Underfitting:*** This occurs when a linear model is used for a complex relationship, resulting in **high training MSE** and **high test MSE** due to high bias.

	Training Set Correlation	Effect of Variance
LOOCV	**Very High:** each of the n training sets is nearly identical, sharing $n - 2$ observations	**Higher Variance:** because the outputs of the n models are highly correlated, the mean of those outputs has higher variance than if they were independent
k-fold	**Lower:** the training sets have less overlap compared to LOOCV	**Lower Variance:** since the models are less correlated with each other, the average of their performance is more stable

```
> summary(model)

Call:
lm(formula = satisf ~ age + severe + anxiety, data = patient)

Residuals:
     Min       1Q   Median       3Q      Max 
-18.3524  -6.4230   0.5196   8.3715  17.1601 
```

Coefficients: $\hat{\beta}$ $SE(\hat{\beta})$

```
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 158.4913    18.1259   8.744 5.26e-11 ***
age          -1.1416     0.2148  -5.315 3.81e-06 ***
severe       -0.4420     0.4920  -0.898   0.3741    
anxiety     -13.4702     7.0997  -1.897   0.0647 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```