Journal of Integrated Knowledge
Vol. 2 No. 7 (2026)

Polynomial Complexity at Withheld Boundaries: A Paired Error Audit of a Public Reaction Benchmark

Jordan Allen[1] · Adrian Porter[2] · Ryan King[3]

[1] Greece · [2] Estonia · [3] Latvia

Published: July 08, 2026 | Article 859

Abstract

The same polynomial extension can improve interior predictions while worsening boundary predictions. We withhold each of 114 reported observations within 21 groups of a public reaction benchmark and compare fixed linear, quadratic and cubic fits for two percentage responses. For conversion, adding a cubic term changes mean absolute error by −0.567 percentage points at interior targets and +2.092 at boundaries. Their paired difference is 2.659, with a conditional group-composition interval of −0.414 to 6.543. The corresponding selectivity difference is 0.217, with interval −2.788 to 3.047. Both intervals include zero. Unrestricted predictions also exceed percentage bounds. These results concern observed-coordinate validation, without identifying laboratory repeatability, causal condition effects or optimal operating settings.

Polynomial Complexity at Withheld Boundaries: A Paired Error Audit of a Public Reaction Benchmark

Abstract

The same polynomial extension can improve interior predictions while worsening boundary predictions. We withhold each of 114 reported observations within 21 groups of a public reaction benchmark and compare fixed linear, quadratic and cubic fits for two percentage responses. For conversion, adding a cubic term changes mean absolute error by -0.567 percentage points at interior targets and +2.092 at boundaries. Their paired difference is 2.659, with a conditional group-composition interval of -0.414 to 6.543. The corresponding selectivity difference is 0.217, with interval -2.788 to 3.047. Both intervals include zero. Unrestricted predictions also exceed percentage bounds. These results concern observed-coordinate validation, without identifying laboratory repeatability, causal condition effects or optimal operating settings.

1. A withheld endpoint asks a different prediction question

A fitted curve is often judged by how closely it follows the observations used to estimate it. A more demanding assessment removes an observation before predicting it. Even then, the location of the removed point matters. An interior target remains between training coordinates; a boundary target lies outside the remaining training span. Adding a polynomial term may therefore have different consequences at these two locations, although both targets were originally present in the same released table.

Response-surface analysis has a long connection with experimental design and local approximation (Box and Wilson, 1951; Myers et al., 2016). Its practical use does not make every additional term predictively useful. Allen (1974) frames prediction as a criterion requiring attention beyond simply adding variables, while statistical learning treatments connect flexibility with the quality of prediction (Hastie et al., 2009). Here we examine a specific numerical consequence of that distinction, using existing observations rather than proposing experiments or optimizing conditions.

Zhao et al. (2023) combine ANOVA, polynomial regression, design-uniformity assessment and staged regression in a reaction-condition modeling workflow. Their polynomial and support arguments provide the direct motivation for this audit. The term mixed in that work describes a hybrid modeling architecture; it does not establish a random-effects model for the present data. We evaluate fixed polynomial degrees against withheld observations, without reproducing their significance-screening sequence, optimization calculation or experimental conclusions. The connection is the adequacy of a curve over its estimation support, rather than a generic shared use of regression.

Our research question is whether the cubic-minus-quadratic error increment differs between withheld boundary and interior coordinates within recorded groups. We retain both reported conversion and selectivity percentages as co-primary responses. A secondary comparison asks the same question when moving from linear to quadratic regression. Polynomial approximation is tied to its interval and approximation setting (Trefethen, 2019). An observation inside the original source range can still require extrapolation from a reduced training subset, so the source range alone does not answer our validation question.

2. The official benchmark supplies groups and coordinates, not independent repetitions

The evidence comes from the organizer's original 2021 Contemporary Undergraduate Mathematical Contest in Modeling B release (CUMCM Organizing Committee, 2021). We downloaded the national organizer's archive and extracted its B design statement and two attachments. The first attachment contains one data sheet with 114 observations. Its rows identify twenty-one recorded condition groups, a numerical temperature coordinate and several percentage responses. The analysis selects the group identifier, coordinate, reported conversion percentage and reported selectivity percentage. Detailed condition descriptions and other response channels are not predictors or published combinations.

The official statement presents these numbers as laboratory results supplied for modeling. It does not provide a named laboratory, instrument records, run order, randomization schedule or independently repeated measurements at each group-coordinate cell. The second attachment describes repeated times within one experiment and is kept separate. It is not stacked with the first attachment to create additional independent observations. Experimental-design theory requires the distinction between an observational cell and independent experimental replication (Cox and Reid, 2000). Our validation calculation relies on the released table, without claiming to reconstruct its measurement process.

Thirteen groups have five recorded coordinate points, seven have six and one has seven. Every selected group-coordinate cell is unique. Both analyzed responses and every coordinate are finite; neither response has a missing value, a recorded zero or a value outside zero to one hundred. Original merged identifier cells are filled only inside their explicit merged ranges. We perform no generic forward filling through unknown rows, delete no unfavorable outcome and replace no source value with a fitted one. The resulting observation set is unchanged across polynomial degrees and responses.

Group labels are replaced by neutral codes, while a recorded two-category mode remains available in the provenance. The mode is not analyzed as a randomized treatment. The underlying condition components are not a complete factorial cross, and their separate effects are not estimated. Dataset documentation should describe such missing design information and the context of reuse (Gebru et al., 2021). A group code identifies which observations share a reported condition; it does not supply a causal interpretation for differences between groups.

The official release is public and retains an All Rights Reserved notice. Original archive and workbook hashes, returned download metadata and the source design statement are preserved locally. No unrestricted data license is inferred, and raw redistribution is not promised. This article reports original aggregate statistical calculations from that release. It supplies no preparation directions, condition recommendations or additional-run design. The source's description of reported measurements is retained together with its limited measurement provenance.

3. Each observed point becomes an explicit validation target

Within each group, we remove one row at a time and use every remaining row from that group for estimation. This is repeated for degrees one, two and three, yielding 342 polynomial fits and 684 response predictions. Each group contributes two boundary targets, its original minimum and maximum coordinate, and all other targets are interior. There are forty-two boundary and seventy-two interior target rows. The response at a target is unavailable to its fit, and predictions are evaluated only afterward.

Cross-validation constructs an assessment through a specified withholding operation (Stone, 1974). A survey of cross-validation methods distinguishes how that operation supports comparison and selection (Arlot and Celisse, 2010). Our degrees are fixed before errors are calculated; none is selected for deployment by this analysis. The withholding unit also limits the claim: structured validation depends on

which relationships are shared between training and evaluation (Roberts et al., 2017). Here a target shares its group with its training observations. This is neither a test on a new condition group nor an independent experimental replication.

For each group, the full known predictor range is centered and scaled to minus one through one. Actual unequal coordinate spacing is retained. Including the known target coordinate in that coding uses the released design, without using its response. We subsequently verify that training-only affine scaling gives the same predictions for the identical polynomial spaces. Clear information boundaries are essential in predictive analysis (Kapoor and Narayanan, 2023). No response, error or selected degree enters the scale transformation or the classification of a point as boundary or interior.

The polynomial regressions are linear in their fitted coefficients, despite including nonlinear powers of the coordinate. Seber and Lee (2003) provide the relevant distinction between regression structure, explanatory-variable transformations and the design matrix. Each fit includes an intercept and powers through its declared degree. The smallest group leaves four distinct training coordinates after withholding, which is sufficient for a full-rank cubic design. In such a four-point fit, however, exact interpolation of training responses offers no additional residual degrees of freedom for estimating experimental noise.

We solve the scaled Vandermonde systems with a singular-value-based least-squares routine. Numerical linear algebra distinguishes direct solution methods and the stability of least-squares computations (Golub and Van Loan, 2013; Björck, 1996). All 342 declared designs have full column rank, and the largest recorded design condition number is 45.265. Numerical accuracy is a different question from scientific predictive uncertainty (Higham, 2002). Agreement between solvers cannot establish that the fitted curve describes the unknown physical response between or beyond observations.

No regularization, degree search, stepwise testing or clipping is applied. Every prediction remains on its unconstrained numerical scale. Model selection performed on the same held-out errors would introduce an additional assessment problem, as the discussion of cross-validation selection bias by Varma and Simon (2006) illustrates. We retain the fixed comparisons, including errors that worsen, predictions outside the percentage range and results that fail to establish a stable direction.

4. The paired contrast gives every recorded group the same influence

Absolute error is measured in percentage points. It has a direct interpretation as the magnitude of a discrepancy in the response's units (Willmott and Matsuura, 2005). Accuracy measures still depend on the scientific comparison being made (Hyndman and Koehler, 2006); we do not interpret absolute error as an operational cost or as evidence of optimal conditions. Table 1 reports separate errors for each response, degree and target location, rather than merging the two responses into a selected composite.

For each degree and response, we first average the two boundary errors within a group and separately average its interior errors. We then average those twenty-one group means equally. Consequently, a seven-point group has the same contribution as a five-point group. Observation-weighted values are retained as a diagnostic with a different target. Weighting choices define the quantity being summarized, rather than simply repairing a calculation (Solon et al., 2015). Boundary weighting happens to coincide because every group has exactly two such targets; interior weighting need not coincide.

For group g, let Delta_B be its cubic MAE minus its quadratic MAE for boundary targets, and Delta_I the corresponding difference for interior targets. The co-primary location contrast averages Delta_B minus Delta_I across groups, separately for conversion and selectivity. A positive value means the cubic addition is relatively more adverse at the boundary. It does not necessarily mean that cubic predictions are

worse at both locations, or worse for both responses. The quadratic-minus-linear comparison uses exactly the same definition and is reported as secondary.

P1 concerns these two cubic-addition contrasts. P2 records the secondary comparison and all degree-specific errors. Figure 1 connects the original observation groups, withheld prediction tasks and paired calculations. The pathways are estimation steps. There is no arrow from a polynomial coefficient to an identified causal condition effect, and no step selects a reported condition for experimental use.

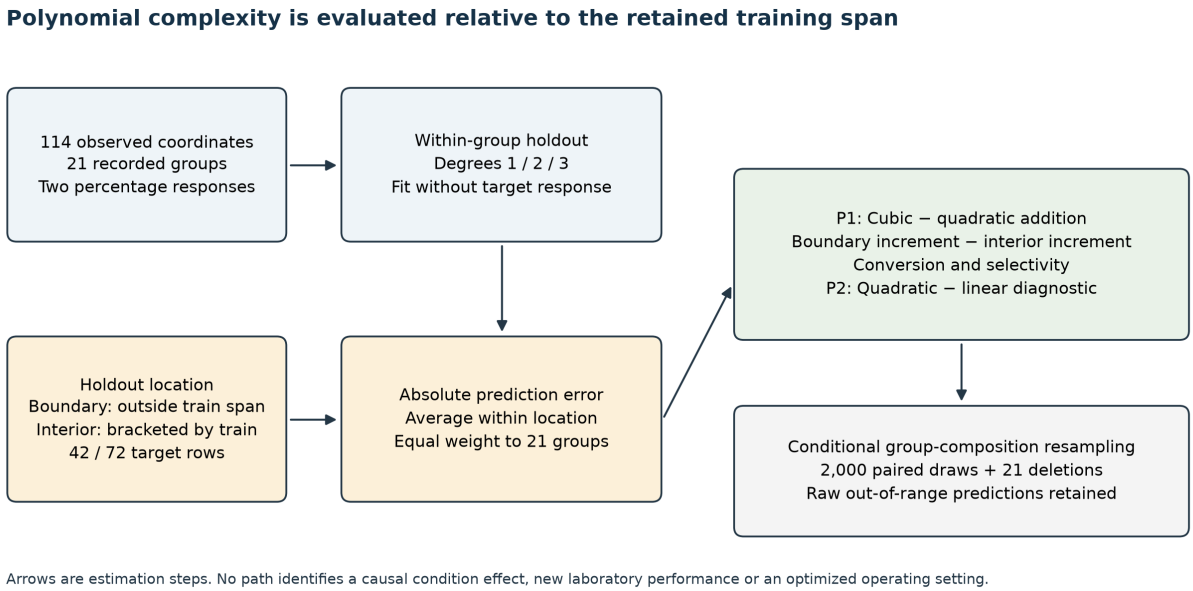


Figure 1. Study model. *Boundary and interior refer to the coordinate relative to the retained within-group training span. Both percentage responses and all three degrees are fixed. Equal-group losses feed paired location contrasts, conditional resampling and group deletion. The model contains no experimental optimization path.*

5. Error patterns differ across the two responses

Table 1. Held-out absolute errors for fixed within-group least-squares polynomials. *Values are percentage points and average the twenty-one group means equally. Every row uses forty-two boundary targets and seventy-two interior targets from 114 original observations. No response clipping or degree selection is applied.*

Response	Polynomial degree	Boundary MAE (42 targets)	Interior MAE (72 targets)	Groups
Conversion	1	13.283	4.652	21
Conversion	2	8.564	2.445	21
Conversion	3	10.656	1.878	21
Selectivity	1	6.087	2.480	21
Selectivity	2	6.013	1.707	21
Selectivity	3	5.544	1.021	21

For conversion, the cubic extension reduces interior MAE from 2.445 to 1.878 percentage points, but increases boundary MAE from 8.564 to 10.656. The resulting increments are -0.567 and +2.092. Their difference is +2.659 percentage points. This is the anticipated form of a boundary-specific penalty in the finite comparison: the same addition has opposite error directions at the two locations. The sign describes the observed paired losses, rather than a universal property of cubic regression.

Selectivity gives a less pronounced contrast. Its cubic extension reduces interior MAE from 1.707 to 1.021 and boundary MAE from 6.013 to 5.544. Both increments are favorable. The boundary reduction is smaller, giving a location contrast of +0.217 percentage points. Reporting only the shared positive contrast sign would conceal this substantive difference from conversion. There is no single finding that cubic flexibility uniformly damages boundary prediction across both outcomes.

Table 2. Paired degree increments and location differences. *All effects and sensitivity endpoints are percentage points. The model is fixed group-specific polynomial least squares. Each comparison uses 114 source rows, twenty-one paired groups, forty-two boundary targets and seventy-two interior targets. Percentile intervals and SE describe 2,000 conditional group-composition resamples with fixed predictions; they are not laboratory confidence intervals. The deletion range omits each group once.*

Response	Degree addition	ΔBoundary	ΔInterior	ΔBoundary - ΔInterior	Conditional SE	Conditional percentile interval	Delete-group range
Conversion	3 minus 2 (co-primary)	2.092	-0.567	2.659	1.814	[-0.414, 6.543]	[1.455, 3.113]
Conversion	2 minus 1 (secondary)	-4.719	-2.207	-2.512	1.404	[-5.153, 0.380]	[-3.259, -1.861]
Selectivity	3 minus 2 (co-primary)	-0.469	-0.686	0.217	1.468	[-2.788, 3.047]	[-0.458, 0.930]
Selectivity	2 minus 1 (secondary)	-0.074	-0.773	0.699	0.874	[-0.923, 2.466]	[0.190, 0.949]

The secondary linear-to-quadratic comparison also resists a universal interpretation. Conversion MAE decreases more at boundaries than at interior targets, producing a location difference of -2.512 percentage points. Selectivity's difference is +0.699. These additional fixed comparisons show why model complexity cannot be summarized by counting terms alone. The observed response, target location and chosen complexity increment all enter the empirical answer. No secondary result is promoted to a newly selected primary outcome.

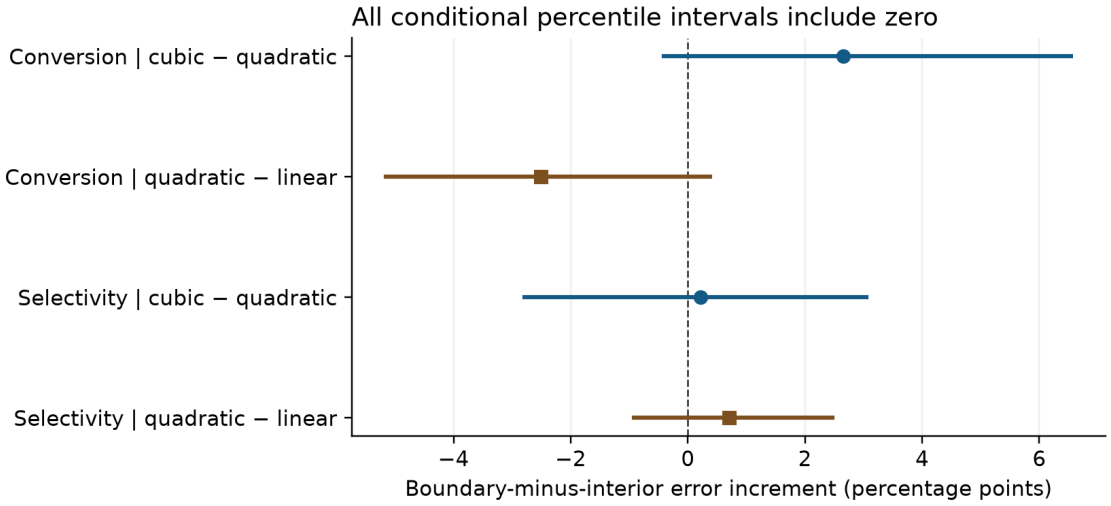
6. Broad composition intervals limit a general boundary claim

We generate 2,000 paired resamples of the twenty-one group-loss vectors with a fixed seed. The same draw is used across responses, degrees and target locations. The resampling principle follows Efron (1979), and the computational summaries follow the bootstrap framework discussed by Davison and Hinkley (1997). Models and their cross-validation predictions remain fixed. These percentile ranges describe sensitivity to the composition of recorded groups; they are not laboratory confidence intervals, estimates of independent measurement repeatability or guarantees of coverage for a scientific population.

Conversion's co-primary standard error under this conditional resampling is 1.814 percentage points, with percentile endpoints -0.414 and 6.543. Selectivity's is 1.468, with endpoints -2.788 and 3.047. Both ranges cross zero, as do both secondary ranges in Table 2. The numerical evidence consequently does not support a confidently general boundary disadvantage across alternative recorded-group compositions. It also does not establish zero difference or practical equivalence. Statistical interpretation should retain effect size, design and uncertainty rather than turn a threshold into the conclusion (Wasserstein and Lazar, 2016).

A separate diagnostic omits each group once and recalculates the mean with the remaining twenty. The conversion cubic-addition contrast stays positive, from 1.455 to 3.113 percentage points. Selectivity ranges from -0.458 to 0.930, changing sign. Deletion analysis examines influence, a concern developed in regression diagnostics by Cook (1977), although these reported values are neither Cook's distance nor a sampling interval. Removing one group and repeatedly changing group composition answer different

sensitivity questions; the positive conversion deletion range does not invalidate its zero-crossing resampling range.



Lines: 2.5th–97.5th percentiles of 2,000 paired group-composition resamples.
These intervals hold observed cross-validation predictions fixed; no laboratory confidence guarantee.

Figure 2. Conditional sensitivity of location contrasts. *Points are the four predeclared equal-group contrasts. Lines mark the 2.5th and 97.5th percentiles of paired recorded-group resampling with fixed predictions. All four lines cross zero. This display supplies no laboratory confidence guarantee.*

7. Unconstrained predictions expose another limit

Table 3. Raw prediction failures against the percentage bounds. *Every original prediction remains in the losses. Counts are outside zero through one hundred, and ranges retain all raw predictions rather than only the failures. Boundary and interior denominators are 42 and 72 per row respectively.*

Response	Degree	Boundary outside / 42	Boundary prediction range (%)	Interior outside / 72	Interior prediction range (%)
Conversion	1	19	[-12.276, 101.944]	1	[-0.310, 67.333]
Conversion	2	6	[-10.070, 122.453]	4	[-1.149, 67.380]
Conversion	3	9	[-24.482, 146.024]	0	[0.648, 67.984]
Selectivity	1	10	[-3.317, 70.045]	0	[0.948, 45.736]
Selectivity	2	3	[-9.242, 74.593]	0	[0.832, 47.386]
Selectivity	3	1	[-0.600, 55.930]	0	[0.874, 53.364]

The response bounds are not enforced by ordinary polynomial least squares. For conversion, the cubic boundary predictions range from -24.482 to 146.024 percent, with nine of forty-two outside zero through one hundred. All cubic interior conversion predictions remain inside that range. Linear conversion fits generate nineteen out-of-range boundary predictions and one interior prediction, while quadratic fits generate six and four. Thus the presence of impossible percentage predictions is not unique to the largest degree, and a count alone does not rank all aspects of performance.

Selectivity has ten out-of-range linear boundary predictions, three quadratic boundary predictions and one cubic boundary prediction. Its interior predictions remain within the percentage range for every degree.

These failures are retained in the error calculations and source-level exports. Clipping might reduce some numerical errors, but would define a different predictor and was not part of the frozen comparison. The audit therefore identifies both the limitation of the declared regression class and the limits of a simple higher-degree failure narrative.

The study does not attribute these numerical patterns to a chemical mechanism, run order or laboratory instability. There is one reported response pair per group-coordinate cell and no verified randomization or pure-error replication structure. Error differences may reflect curve form, response measurement variation and the geometry of the withheld coordinates together. Their components cannot be separated from this archive. In particular, a boundary holdout remains an observed original coordinate, so our calculation does not validate behavior beyond the entire released range.

8. Recalculation and conclusion

The archived analytical file preserves source-row locators, neutral group codes, both responses and the original coordinate spacing in coded form. NumPy supplies the actual polynomial matrices, least-squares fits and paired resampling (Harris et al., 2020). pandas supplies loss aggregation and exports (McKinney, 2010), and Matplotlib produces the two study figures (Hunter, 2007). An independent workbook reader reconciles all 114 selected original rows with the XML-based extraction. A fifty-digit decimal solver reproduces all 684 predictions, with a maximum absolute discrepancy below 0.000000000001 percentage points. Training-only scaling gives the same predictions to that tolerance.

Independent scalar aggregation also reproduces the twelve MAE rows, all four contrasts, their 2,000 resampling calculations, the eighty-four group deletions and every out-of-range summary. Reproducible computation requires preserving those transformations and intermediate numerical objects (Peng, 2011; Sandve et al., 2013). These checks support computational traceability. Accountable source-claim review, authorship and final scientific approval remain pending; no human approval or new experimental work is claimed.

The public benchmark shows a clear finite conversion pattern: a cubic addition improves interior predictions while worsening boundary predictions. The analogous selectivity change improves both locations and produces a much smaller contrast. Broad conditional composition ranges include zero for both responses, and unconstrained predictions can violate percentage bounds. Those findings justify reporting support location, response identity and uncertainty together when assessing polynomial complexity. They do not select an operating condition, establish a chemical mechanism or guarantee performance on new experiments.

References

1. Peng, Roger D.. (2011). Reproducible Research in Computational Science. Science. https://doi.org/10.1126/science.1213847
2. Roberts, David R.; Bahn, Volker; Ciuti, Simone; Boyce, Mark S.; Elith, Jane; Guillera-Arroita, Gurutzeta; Hauenstein, Severin; Lahoz-Monfort, José J.; Schröder, Boris; Thuiller, Wilfried; Warton, David I.; Wintle, Brendan A.; Hartig, Florian; Dormann, Carsten F.. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography. https://doi.org/10.1111/ecog.02881.
3. Cook, R. Dennis. (1977). Detection of Influential Observation in Linear Regression. Technometrics. https://doi.org/10.1080/00401706.1977.10489493.
4. Wasserstein, Ronald L.; Lazar, Nicole A.. (2016). The ASA Statement on p -Values: Context, Process, and Purpose. The American Statistician. https://doi.org/10.1080/00031305.2016.1154108.
5. Seber, George A. F.; Lee, Alan J.. (2003). Linear Regression Analysis. Wiley Series in Probability and Statistics. https://doi.org/10.1002/9780471722199.

6. Stone, M.. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. Journal of the Royal Statistical Society Series B: Statistical Methodology. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x

7. Hunter, John D.. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. https://doi.org/10.1109/mcse.2007.55

8. Gebru, Timnit; Morgenstern, Jamie; Vecchione, Briana; Vaughan, Jennifer Wortman; Wallach, Hanna; III, Hal Daumé; Crawford, Kate. (2021). Datasheets for datasets. Communications of the ACM. https://doi.org/10.1145/3458723.

9. Björck, Åke. (1996). Numerical Methods for Least Squares Problems. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611971484.

10. Higham, Nicholas J.. (2002). Accuracy and Stability of Numerical Algorithms: Second Edition. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898718027

11. Harris, Charles R.; Millman, K. Jarrod; van der Walt, Stéfan J.; Gommers, Ralf; Virtanen, Pauli; Cournapeau, David; Wieser, Eric; Taylor, Julian; Berg, Sebastian; Smith, Nathaniel J.; Kern, Robert; Picus, Matti; Hoyer, Stephan; van Kerkwijk, Marten H.; Brett, Matthew; Haldane, Allan; del Río, Jaime Fernández; Wiebe, Mark; Peterson, Pearu; Gérard-Marchant, Pierre; Sheppard, Kevin; Reddy, Tyler; Weckesser, Warren; Abbasi, Hameer; Gohlke, Christoph; Oliphant, Travis E.. (2020). Array programming with NumPy. Nature. https://doi.org/10.1038/s41586-020-2649-2.

12. Cox, D. R.; Reid, Nancy. (2000). The Theory of the Design of Experiments. Chapman & Hall/CRC. https://www.routledge.com/The-Theory-of-the-Design-of-Experiments/Cox-Reid/p/book/9781584881957

13. Kapoor, Sayash; Narayanan, Arvind. (2023). Leakage and the reproducibility crisis in machine-learning-based science. Patterns. https://doi.org/10.1016/j.patter.2023.100804

14. CUMCM Organizing Committee. (2021). 2021 Contemporary Undergraduate Mathematical Contest in Modeling: Problem B and original data attachments. Official organizer archive. Accessed 17 November 2026. https://www.mcm.edu.cn/html_cn/node/90d223833c1eb50f899aa096a66c6896.html

15. Arlot, Sylvain; Celisse, Alain. (2010). A survey of cross-validation procedures for model selection. Statistics Surveys. https://doi.org/10.1214/09-ss054

16. Hyndman, Rob J.; Koehler, Anne B.. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting. https://doi.org/10.1016/j.ijforecast.2006.03.001

17. Myers, Raymond H.; Montgomery, Douglas C.; Anderson-Cook, Christine M.. (2016). Response Surface Methodology: Process and Product Optimization Using Designed Experiments. Wiley,4th edition. https://www.wiley.com/en-ie/Response%2BSurface%2BMethodology%3A%2BProcess%2Band%2BProduct%2BOptimization%2BUsing%2BDesigned%2BExperiments%2C%2B4th%2BEdition-p-9781118916018

18. Golub, Gene H.; Van Loan, Charles F.. (2013). Matrix Computations. Johns Hopkins University Press, fourth edition. https://doi.org/10.56021/9781421407944

19. Willmott, CJ; Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research. https://doi.org/10.3354/cr030079

20. Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. (2009). The Elements of Statistical Learning. Springer Series in Statistics. https://doi.org/10.1007/978-0-387-84858-7

21. Solon, Gary; Haider, Steven J.; Wooldridge, Jeffrey M.. (2015). What Are We Weighting For?. Journal of Human Resources. https://doi.org/10.3368/jhr.50.2.301.

22. Trefethen, Lloyd N.. (2019). Approximation Theory and Approximation Practice, Extended Edition. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611975949.

23. Zhao, Runhao; Zhang, Jiayu; Wang, Xianrui; Su, Bolun; Tan, Junze; Ben, Niu. (2023). Research on optimal design of reaction conditions based on mixed statistical model. Second International Conference on Applied Statistics, Computational Mathematics, and Software Engineering (ASCMSE 2023). https://doi.org/10.1117/12.2692066.

24. Efron, B.. (1979). Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics. https://doi.org/10.1214/aos/1176344552

25. Sandve, Geir Kjetil; Nekrutenko, Anton; Taylor, James; Hovig, Eivind. (2013). Ten Simple Rules for Reproducible Computational Research. PLoS Computational Biology. https://doi.org/10.1371/journal.pcbi.1003285

26. McKinney, Wes. (2010). Data Structures for Statistical Computing in Python. Proceedings of the Python in Science Conference. https://doi.org/10.25080/majora-92bf1922-00a

27. Davison, A. C.; Hinkley, D. V.. (1997). Bootstrap Methods and their Application. Cambridge University Press. https://doi.org/10.1017/cbo9780511802843

28. Varma, Sudhir and Richard Simon. “Bias in error estimation when using cross-validation for model selection.” *BMC Bioinformatics*, vol. 7, no. 1, 2006. https://doi.org/10.1186/1471-2105-7-91.

29. Allen, David M.. (1974). The Relationship Between Variable Selection and Data Agumentation and a Method for Prediction. Technometrics. https://doi.org/10.1080/00401706.1974.10489157.

30. Box, G. E. P.; Wilson, K. B.. (1951). On the Experimental Attainment of Optimum Conditions. Journal of the Royal Statistical Society Series B: Statistical Methodology. https://doi.org/10.1111/j.2517-6161.1951.tb00067.x