

## **Top Five Statistical Modeling Problems: Nonissues for the Machine-learning GenIQ Model**

Bruce Ratner, Ph.D.

### **GenIQ: FAQs**

- **Q1. What kind of data preparation, EDA are required?**
  - **None!** Only insuring there are no impossible or improbable values (e.g., age of 120 years, or a boy named Sue).
- **Q2. How does GenIQ handle missing data?**
  - ▶ Statistical Complete-case Analysis - reduced sample size
  - ▶ **Genetic Imputation** - full sample size; missingness < 35%
- **Q3. How does GenIQ handle MultiCollinearity (MC)?**
  - **MC is a nonissue** for GenIQ, as it is a statistical "data" problem, affecting SEs of coefficients. GenIQ has no coefficients.
  - GenIQ can estimate its equation with a k-level categorical variable using all k-levels; cf. statistical models cannot do.
    - ▶ **Error Message:** One of the dummy variables is a perfect linear combination of the remaining k-1 dummy variables.

## GenIQ: FAQs

### ■ Q4. How does GenIQ handle outliers?

- GenIQ moderates/eliminates outliers, instead of discarding them, by **pulling the outliers into** a straight-line relationship.

### ■ Q5. How does GenIQ handle overfitting?

- GenIQ is **not more susceptible** to overfitting than any other modeling technique, which seeks a solution by optimization.
- GenIQ is potentially **less prone to overfitting** as it has a fitness-function smoothing component to moderate overfitting.
- GP itself **dampens overfitting**, as it has an optimization method Tournament versus fitness only, and fitness & beauty.
- Optimizing the **decile table innately is not prone** to overfitting.

**Do you want to know more about the GenIQ Model?  
Sign-up for a GenIQ Model Webcast: Click [here](#).**