Logistic regression in experimental research

Regression-related applications

- Testing for differences in proportions (paired and unpaired)
- Comparing proportions adjusted for numerical covariates
- Analysing various types of contrasts over longitudinal models with binary response + analysis of the simple, main, interaction and marginal effects over the EM-means (estimated marginal means)
- Inverse Probability Weighting
- Propensity Score Matching (e.g. https://tinyurl.com/2ntx65da)

Adrian Olszewski, 2022

Logistic regression – a few facts. 1/2

- A regression model for categorical endpoints / responses (here: binary) used to analyze the impact (magnitude, direction, inference) of predictor variables on the probability of occuring an event.

- As every regression, it predicts a numerical outcome: **E(Y|X=x)**. For the Bernoulli's distribution (or binomial with k=1) it's a probability of success (or log-odds, dependeing on formulation). More precisely, it's a direct probability estimator.

- Logistic regression itself does NOT produce labels. More precisely, it's a direct probability estimator. To give classes, it needs additional post-fit activity: applying a decision rule to the predicted probability. This makes the **logistic classifier,** <u>which is a 3-step procedure:</u>
  1) Train the logistic regression model on existing data to to get the model coefficients
  2) Predict the probability using a new data and the obtained model coefficients
  3) Apply the decision rule to the probability to obtain the class

  I hope eveyone can see with own eyes that logistic classifier ≠ a logistic regression (it **<u>uses</u>** the LR)

- The ML community uses it entirely for classification. But, instead of using a proper name: „**logistic classifier**", they took over the name and completely replaced meanings. Today they claim that „*logistic regression is not a regression*", denying what thousands of people do at work every day. In my work, the LR is not used for classification.

Logistic regression – a few facts. 2/2

- LR is fit by the Maximum Likelihood (it's NOT just OLS with logit transformation) https://stats.stackexchange.com/questions/48485/what-is-the-difference-between-logit-transformed-linear...

- LR is part of the Generalized Linear Model (like linear, Poisson, gamma, beta, etc.). Technically there's no difference for the fitting algorithm which one you choose. Just specify the binomial conditional distribution and logit link (or probit – for a less convenient interpretation) and get the table of model coefficients.

- LR was invented in 1958 by sir **Cox** (traces of the logistic function reach the 19th century: https://papers.tinbergen.nl/02119.pdf *; Verhulst, Quetelet, Pearl*) and multinomial LR in 1966. It was then generalized by **Nelder** and **Wedderburn** into a general GLM framework. **Hastie** and **Tibshirani** invented then the GAM (Generalized Additive Model) allowing for non-linear predictor.

  Hastie authored the glm() command in the S-PLUS statistical suite (the father of R, origin of the R syntax), estimating the coefficients of the regression models and performing the inference. Additional extensions for handling repeated observations were made by **Liang** and **Zieger** in 1986 (GEE – Generalized Estimating Equations) and Breslow, Clayton and others around 1993 (GLMM – generalized linear mixed models).
- https://www.jmp.com/content/dam/jmp/documents/en/statistically-speaking-resources/statspkg-stroup-glmms.pdf

Logistic regression – a few facts. 2/2

Generalized additive models for medical research

Trevor Hastie and **Robert Tibshirani*** Department of Statistics and Division of Biostatistics, Stanford University, Stanford, California, USA

This article reviews flexible statistical methods that are useful for characterizing the effect of potential prognostic factors on disease endpoints. Applications to survival models and binary outcome models are illustrated.

1 Introduction

In the statistical analysis of clinical trials and observational studies, the identification and adjustment for prognostic factors is an important component. Valid comparisons of different treatments requires the appropriate adjustment for relevant prognostic factors. The failure to consider important prognostic variables, particularly in observational studies, can lead to errors in estimating treatment differences. In addition, incorrect modelling of prognostic factors can result in the failure to identify nonlinear trends or threshold effects on survival. This article describes flexible statistical methods that may be used to identify and characterize the effect of potential prognostic factors on disease endpoints. These methods are called 'generalized additive models'.

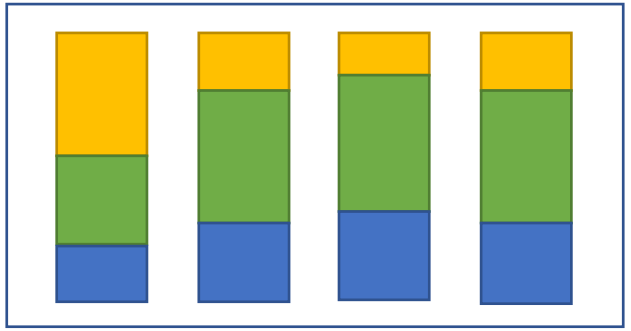
Two of the most commonly used statistical models in medical research are the proportional hazards regression model for survival data and the logistic regression model for binary data. Both of these techniques (and many others) model the effects of prognostic factors x_j in terms of a linear predictor of the form $\sum x_j\beta_j$, where the β_j are parameters. The generalized additive model replaces $\sum x_j\beta_j$ with $\sum f_j(x_j)$ where f_j is an unspecified ('nonparametric') function. This function is estimated in a flexible manner using a scatterplot smoother. The estimated function $\hat{f}_j(x_j)$ can reveal possible nonlinearities in the effect of the x_j.

LR can replace the classic statistical tests for comparison of % of success*:

(but not % being any ratio → then use beta regression, simplex regression, fractional logistic regression, GEE estimated linear model)

- **2/n-sample between independent groups** (ordinary)
- **2/n-sample between dependent groups** (via GEE estimation or a mixed model)
- **Over time** – (linear, quadratic, cubic) trend in % (via GEE estimation or a mixed model)

- to compare **percentage structures** over 2+ categories the **multinomal logistic regression** is used



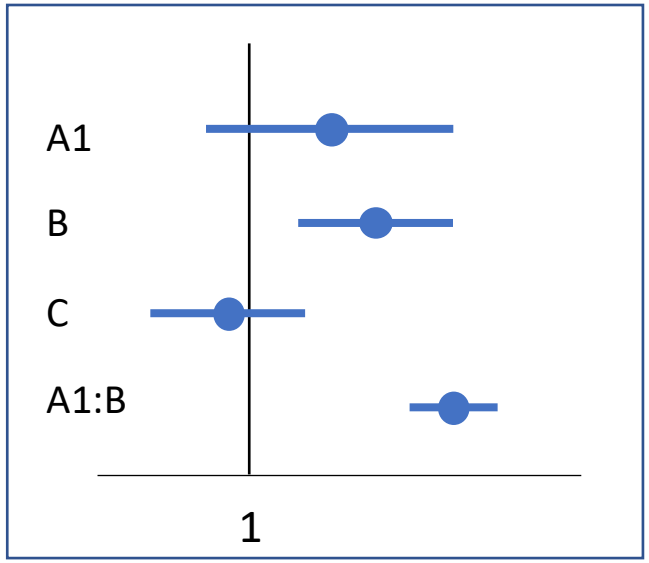
- If the compared classes are ordered, the **proportional-odds model** can be used (aka **ordinal LR**)
  - If the proportionality of odds doesn't hold, the Generalized Ordinal LR can be employed

- If the compared classes are dependend (nested), the **hierarchical multinomial LR** is used

- BTW: the **Ordinal Logistic Regression is a generalization of the Wilcoxon** (Mann-Whitney) test. https://www.fharrell.com/post/powilcoxon/ | https://www.fharrell.com/post/rpo/

LR allows one to analyze longitudinal studies with binary endpoint (observations repeated over time)

- **assessment of specific contrasts** (simple effects): Tukey (all-pairwise), Dunnett (all-vs. control), selected, trends.
- **n-way comparisons** across many categorical variables & their interactions.
- the comparisons can be **adjusted for numerical covariates.**
- followed by the LRT or Wald's testing → **AN(C)OVA** („**analysis of deviance**") for the **main & intereaction effects.**
- **marginal effects** express the predictions in „percentage-points" rather than „odds ratios" or „probabilities".
- allows for using **time-varying covariates** and **piecewise analysis.**
- **GEE estimation** allows for population-average comparisons.
  Mixed-effect model allows for comparisons conditional on subject.
  The two answer different questions and cannot be used interchangeably.
- In presence of **missing data**, the **Inverse Probability Weighting** can be employed to prevent biased estimates. The **IPW also** uses the LR internally ☺
- Both **using smoothing terms** (splines) **or employing the the GAM** approach (invented by *Hastie & Tibshirani*) can handle non-linearity in the predictor.
- **LASSO or Ridge** regularization can be used in case of **multi-collinearity.**



Analyses of contrasts over a longitudinal model with a binary endpoint (SUCCESS/FAILURE)

Study arm 1	Baseline numerical covariates to adjust for	2%	15%	30%	60%	78%
Study arm 2		0%	12%	18%	20%	45%
Time	**Baseline (T0)**	**T1**	**T2**	**T3**	**T4**	**….**
	Pre-treatment			Post-treatment		

- **All-pairwise comparisons** (rather exploratory, not much useful if not supported by some clinical justification):
  - Arm1-T1 vs. Arm1-T2
  - Arm1-T1 vs. Arm1-…
  - Arm1-T1 vs. Arm2-T1
  - Arm1-T1 vs. Arm2-T2, …
- **Between-treatment comparison** (typical analysis in clinical trials; particular focus on selected timepoint(s) → primary objective)
  - T1-Arm1 vs. T1-Arm2
  - T2-Arm1 vs. T2-Arm2
  - **T3-Arm1 vs. T3-Arm2**, …
- **Within-treatment comparison** (sometimes practiced, but much criticized as not a valid measure of clinical effect)
  - Arm1: T1 vs. T2, T1 vs. T3, T1 vs. T…, T2 vs. T3, T2 vs. T…
  - Arm2: T1 vs. T2, T1 vs. T3, T1 vs. T…, T2 vs. T3, T2 vs. T…
- **Comparison of difference in trends** (sometimes practiced, must be supported by valid clinical reasoning)
  - Arm1 – Linear (Quadratic, …) vs. Arm2 – Linear (Quadratic, …)

Analyses of deviance [Type-2 or Type-3 AN(C)OVA]

Success ~ Treatment * Time * Site * Numerical_covariate1 + Baseline_covariate_1 + …..

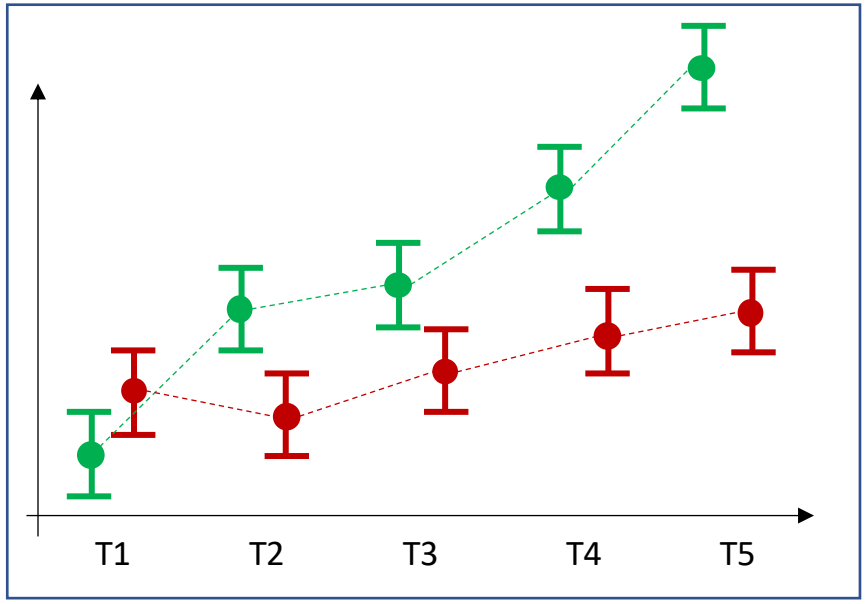
Term	Estimate	SE	p-value
Treatment	xxx	xxx	p=0.0032
Time	xxx	xxx	p=0.0001
Site	xxx	xxx	p=0.98
Numerical_covar_1	xxx	xxx	p=0.101
Treatment*Time	xxx	xxx	p=0.004
Treatment*Site	xxx	xxx	p=0.87
….			

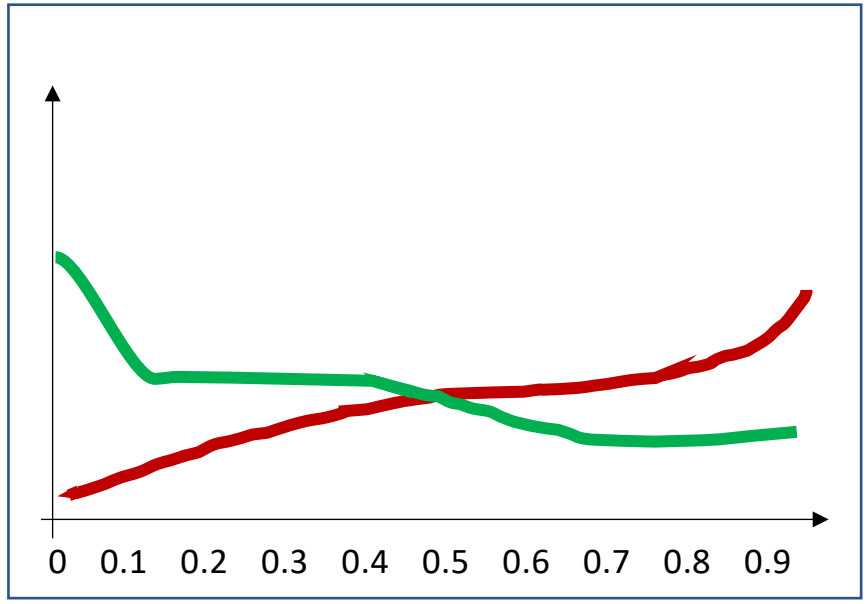
Comparing (nested) models – one per model term – via sequence of Likelihood Ratio Tests (LRT)

Using Wald's joint testing over appropriate model coefficients (Wald is less precise but faster and always available)

Analyses of interactions over the EM-means*

(*estimated marginal means = LS-means)





There are so many members of the Logistic Regression family!

✅ **Binary Logistic Regression** = binomial regression with logit link, case of the Generalized Linear Model, modelling the % of successes.

✅ **Multinomial Logistic Regression** (MLR) = if we deal with a response consisting of multiple non-ordered classes (e.g. colours).

✅ **Nested MLR** - when the classes in MLR are related

✅ **Ordinal LR** (aka Proportional Odds Model) = if we deal with multiple ordered classes, like responses in questionnaires, called Likert items, e.g. {bad, average, good}. The OLR is a generalization of the Mann-Whitney (-Wilcoxon) test, if you need a flexible non-parametric test, that: a) handles multiple categorical variables, b) adjusts for numerical covariates (like ANCOVA)

✅ **Generalized OLR** = Partial Proportional Odds M. when the proportionality of odds doesn't hold.

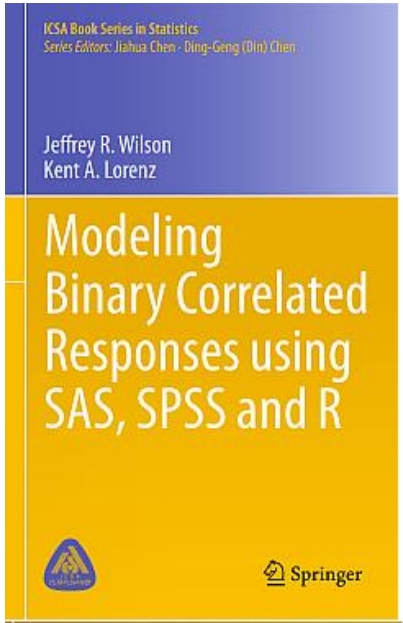
✅ **Alternating Logistic Regression** = if we deal with correlated observations, e.g. when we analyse repeated or clustered data. We have 3 alternatives: mixed-effect LR, LR fit via GEE (generalized estimating equations), or alternatively, the ALR. ALR models the dependency between pairs of observations by using log odds ratios instead of correlations (like GEE). It handles ordinal responses.

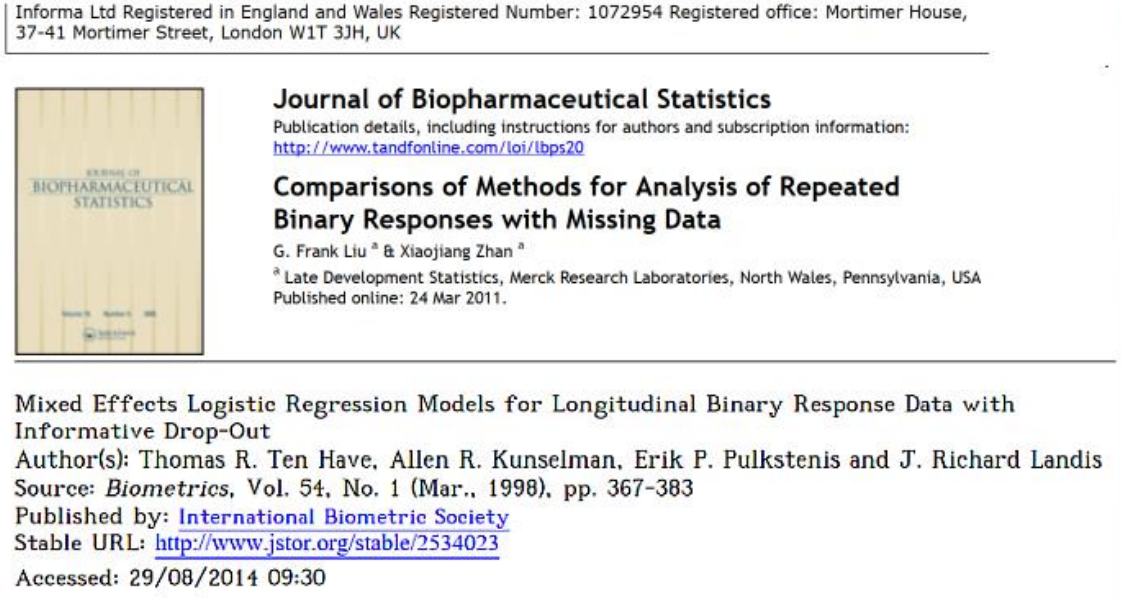
✅ **Fractional LR** = if we deal with a bounded range. Typically used with [0-100] percentages rather than just [TRUE] and [FALSE]. More flexible than beta reg., but not as powerful as the simplex reg. or 0-1-inflated beta r.

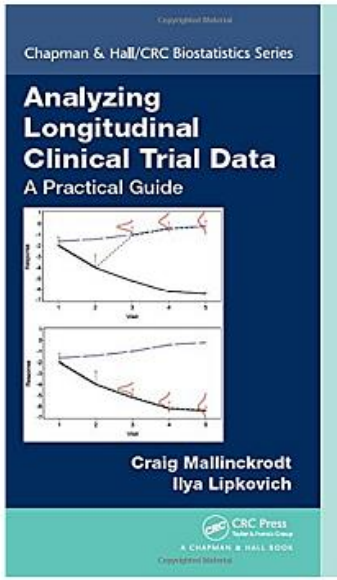
✅ **Logistic Quantile Regression** - application as above.

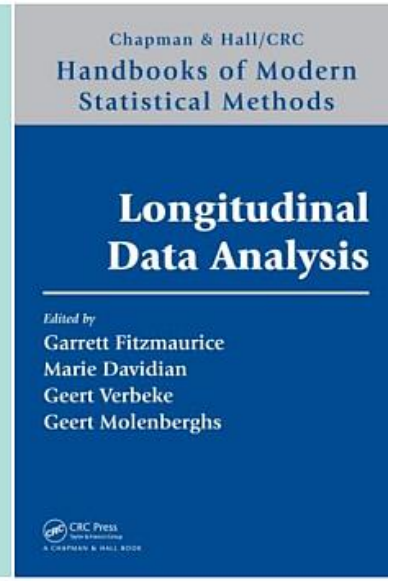
✅ **Conditional LR** = if we deal with stratification and matching groups of data, e.g. in observational studies without randomization, to match subjects by some characteristics and create homogenous "baseline".

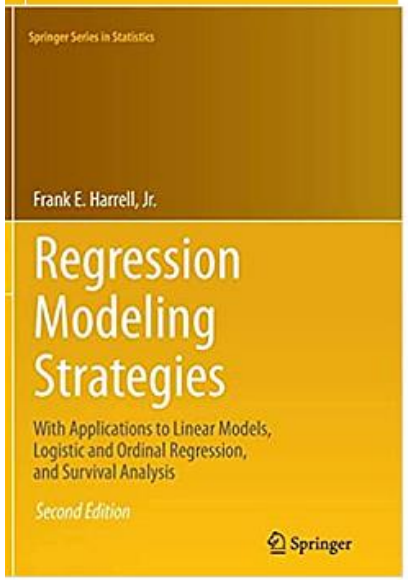
A few books about the logistic regression in experimental research 1/2

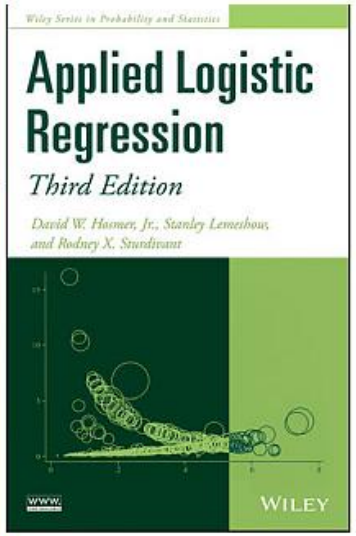


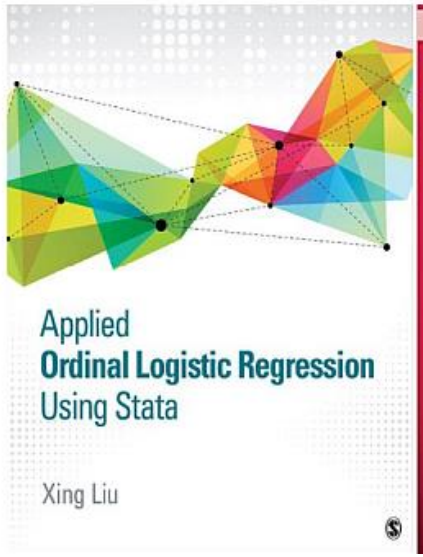


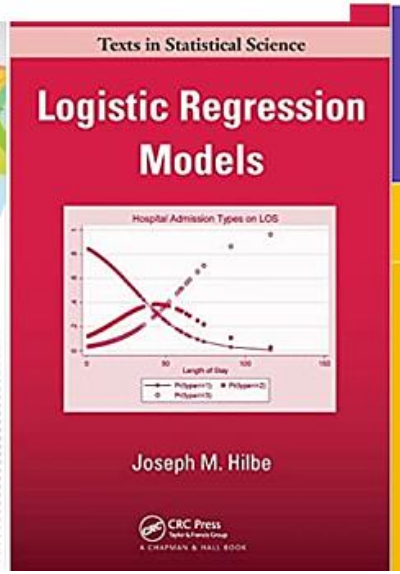


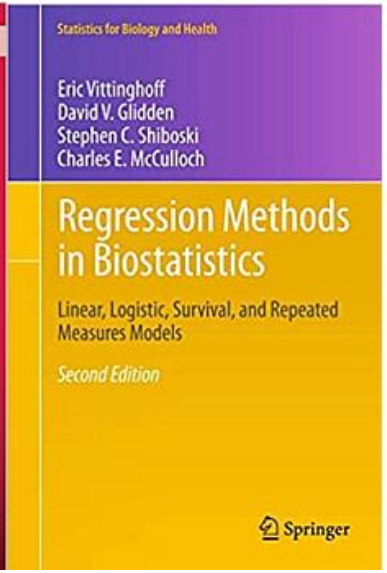


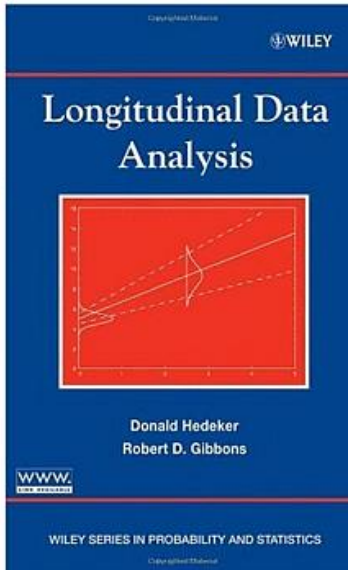




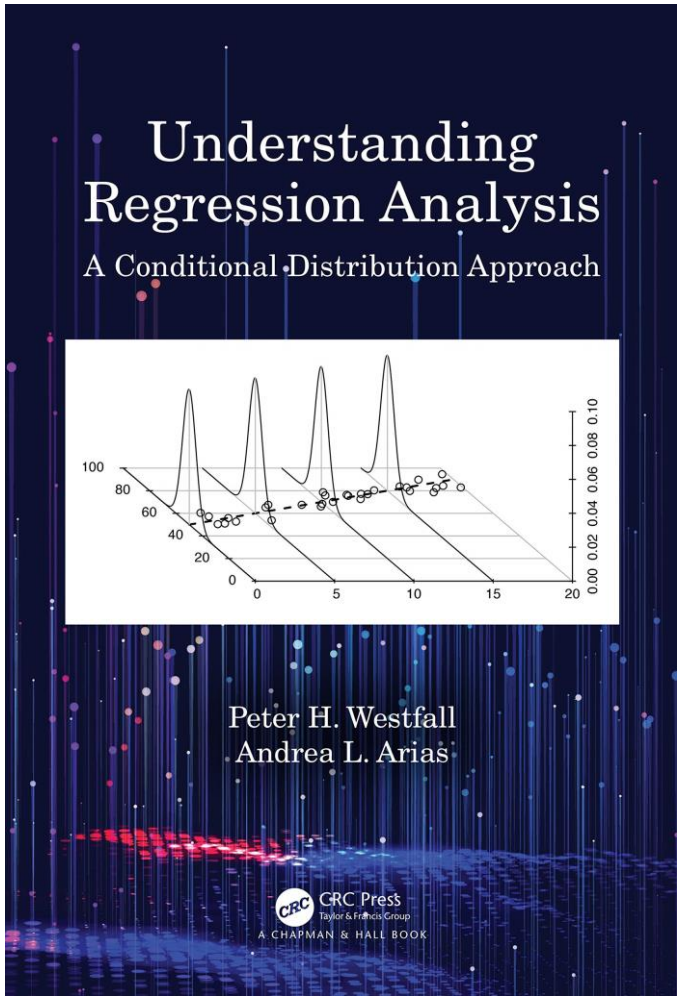


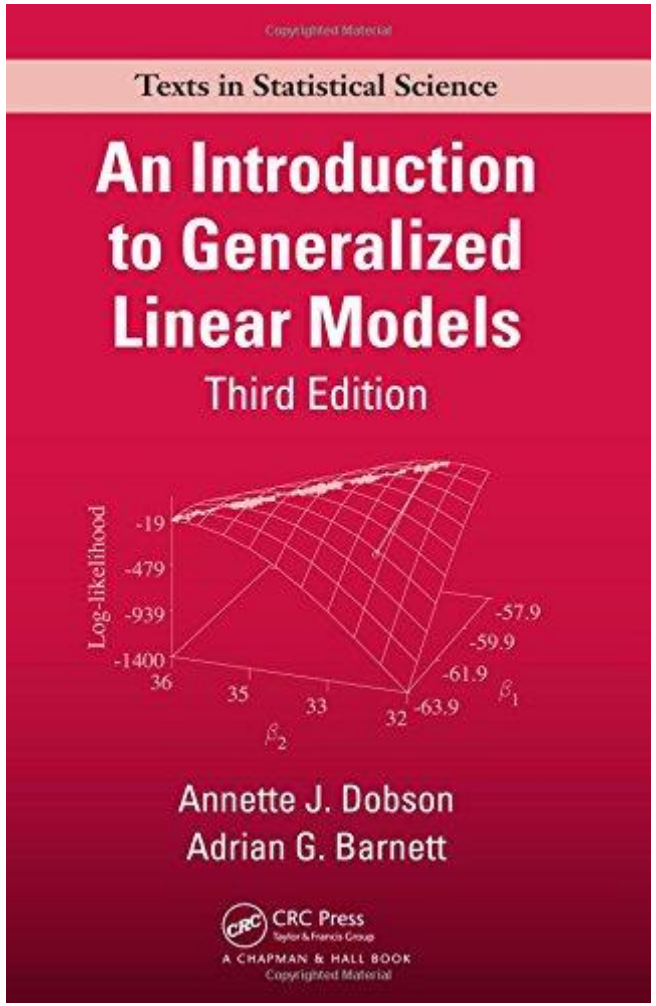


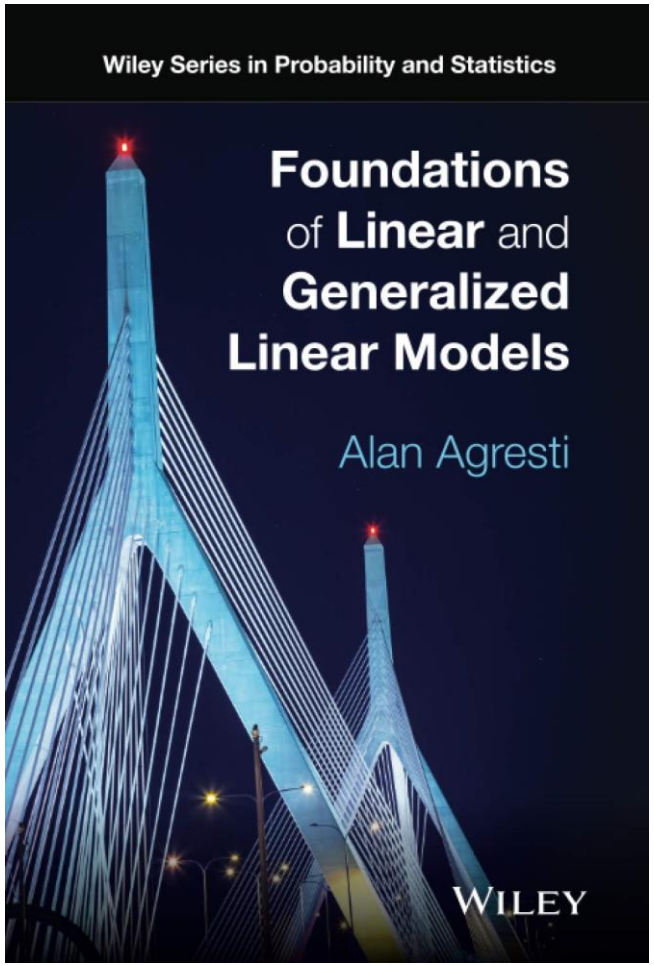




A few books about the logistic regression in experimental research 2/2







Sad reality…

A pharmaceutical company asks two guys to analyze the outcome of a clinical trial comparing 2 drugs using a binary endpoint.

TODO:

1) Compare the proportion (%) of the treatment successes between the two arms at month 24. Adjust for sex and age.
2) We want you to use the population-average logistic regression fit via GEE
3) Check, if there is a linear or quadratic trend in the % of successes in both arms over time.
4) Check, whether the categorical factors A & B and numerical D & E as well as their mutual interactions affect the % of successes. Adjust for sex and age.
5) Please account for potentially missing data. Expect the MAR pattern.

Questions?