Index of Lab/Tutorial 4

Page	Title
1	Logistic regression menu (Minitab)
2	Logistic regression results (Minitab)

Notes for Exercises in Session 4

- add2:9; VER:15; add4:1; VER:16.1 (add2:10; add4:3)

Outline of lab session:

- brief demonstration of Minitab facilities for logistic regression (using `mice` and `nocardia` datasets),
- (upon request only) further demonstration/discussion of Minitab/Stata analysis for linear (regression) models,
- individual work/discussions on the exercises (using Minitab and/or Stata),
  - skip logistic regression parts not covered yet: model building/diagnostics,
- any questions on linear and logistic regression models (conceptual and/or practical).

Logistic Regression Menu (Minitab)

- Menu item `Stat-Regression-Binary`,
- Model setup:
  - choice between binary and grouped data formats,
  - model terms entered in box (factors[1] in *both* boxes),
- Graphs: diagnostics per covariate pattern plotted either against predicted probability or leverage,
- Results: recommended to request full display of results, incl. log-likelihood values,
- Options — useful features:
  - choice of reference level for factors,
  - input previous estimates (for validation),
  - choice of # groups for Hosmer-Lemeshow (H-L) test,
- Storage — useful variables:
  - diagnostics/residuals + fitted prob.,
  - log-likelihood value for model,
  - variance/covariance matrix of estimates,
    - diagonal: variances (squared SE's) of estimates,
    - off-diagonal: covariances[2] between estimates.

[1] Here, factors are parametrized with a reference category (as in Stata).
[2] Correlations can be computed as: covariance/product of SE's.

Logistic Regression Results (Minitab)

Notes on interpretation of output:

- Wald test for factors/coefficients, but likelihood-ratio (LR) test for all coefficients combined (except intercept),
- goodness-of-fit tests:
  - Pearson and Deviance (LR) tests against model with one parameter per covariate pattern,
  - forget about Brown's tests,
  - table of observed and expected counts for H-L test,
- measures of association or model fit[3]:
  - total number of pairs of obs. with different outcomes = # 0's × # 1's,
  - concordant pairs (c): $\hat{p}_1 > \hat{p}_0$, ($\hat{p}_1 \sim$ "event")
  - discordant pairs (d): $\hat{p}_1 < \hat{p}_0$, ($\hat{p}_0 \sim$ "non-event")
  - tied pairs: $\hat{p}_1 = \hat{p}_0$,
  - summary measures $\sim$ single value for concordance (higher is better), e.g. G-K gamma $= (c-d)/(c+d)$.

Features missing for logistic regression in Minitab:

- automated variable selection procedures,
- tools for assessing linearity,
- fit statistics (pseudo R^2, AIC, BIC), ROC curves.

[3] Not covered in VER; Stata `somersd` command gives Somer's D & Kendall's tau.