Lehtonen-Pahkinen (2004)
Section 8.4

Table 8.7 Wald tests $X^2(\mathbf{b})$ for the significance of the interaction term SEX*PHYS in Model 2 under the design-based and unweighted SRS analysis options.

		Design-based		Unweighted SRS	
Term	df	X^2_{des}	*p*-value	X^2_{bin}	*p*-value
SEX*PHYS	1	2.39	0.1218	3.97	0.0463

Let us turn to the corresponding design-based analysis with a linear model for the proportions of Table 8.2. In this situation, logit and linear formulations of an ANOVA model lead to similar results because proportions do not deviate much from the value 0.5. The main effects model (Model 1) is chosen, and results on model fit, residuals, and on significance of the model terms, are close to those for the logit model. But the estimates of the model coefficients differ and are subject to different interpretations. For the logit model with the partial parametrization, an estimated coefficient indicates differential effect on a logit scale of the corresponding class from the estimated intercept being the fitted logit for the reference domain. And for the linear model, an estimated coefficient indicates differential effect on a linear scale of the corresponding class from the estimated intercept, which is now the fitted proportion for the reference domain.

The linear model formulation thus involves a more straightforward interpretation of the estimates of the model coefficients. Under Model 1, these estimates are as follows:

$$
\begin{aligned}
\hat{b}_1 &= 0.5705 && \text{(Intercept)}\\
\hat{b}_2 &= -0.1172 && \text{(Differential effect of SEX = Males)}\\
\hat{b}_3 &= -0.0355 && \text{(Differential effect of AGE} = -44)\\
\hat{b}_4 &= 0.0650 && \text{(Differential effect of PHYS} = 1).
\end{aligned}
$$

The fitted proportion for falling into the upper psychic strain group is thus 0.57 for females in the older age group whose working conditions are less hazardous, and for males in the same age group, $0.57 - 0.12 = 0.45$. The highest fitted proportion, $0.57 + 0.07 = 0.64$, is for the older age group females doing more hazardous work. Also, the fitted proportions are close to those obtained with the corresponding logit ANOVA model.

8.4 LOGISTIC AND LINEAR REGRESSION

The PML method of pseudolikelihood is often used on complex survey data for logit analysis in analysis situations similar to the GWLS method. But the applicability of the PML method is wider, covering not only models on domain proportions of a binary or polytomous response but also the usual regression-type settings with continuous measurements as the predictors. We consider in this section first a PML analysis on domain proportions and then a more general situation of logit modelling of a binary response with a mixture of continuous measurements and categorical variables as predictors. Finally, an example is given of linear modelling for a continuous response variable in an ANCOVA setting.

In PML estimation of model coefficients and their asymptotic covariance matrix, we use a modification of the maximum likelihood (ML) method. In the ML estimation for simple random samples, we work with unweighted observations and appropriate likelihood equations can be constructed, based on standard distributional assumptions, to obtain the ML estimates of the model coefficients and the corresponding covariance-matrix estimate. Using these estimates, standard likelihood ratio (LR) and binomial-based Wald test statistics can be used for testing the model adequacy and linear hypotheses on the model coefficients.

Under more complex designs involving element weighting and clustering, an ML estimator of the model coefficients and the corresponding covariance-matrix estimator are not consistent and, moreover, the standard test statistics are not asymptotically chi-squared with appropriate degrees of freedom. For consistent estimation of model coefficients, the standard likelihood equations are modified to cover the case of weighted observations. In addition to this, a consistent covariance-matrix estimator of the PML estimators is constructed such that the clustering effects are properly accounted for. Using these consistent estimators, appropriate asymptotically chi-squared test statistics are derived.

The PML method can be conveniently introduced in a setting similar to the GWLS method, assuming again a binary response variable and a set of categorical predictors. The data set is arranged in a multidimensional table, such as Table 8.1, with u domains, and our aim is to model the variation of the domain proportion estimates $\hat{p}_j$ across the domains. The variation is modelled by a logit model of the type given in (8.1) and (8.2). A PML logit analysis for domain proportions, covering logit ANOVA, ANCOVA and regression models with categorical predictors can be carried out under any of the analysis options previously introduced by using the corresponding domain proportion estimator vector and its covariance-matrix estimate, and the steps in model-building are equivalent to those in the GWLS method. The design-based analysis option provides a generally valid PML logit analysis for complex surveys. In practice, a PML logit analysis under the design-based option requires access to specialized software for survey analysis.

Design-based and Binomial PML Methods

Under both design-based and weighted SRS options, a consistent *PML estimator* $\hat{\mathbf{b}}_{pml}$ for the vector $\mathbf{b}$ of the s model coefficients b_k in a logit model $F(\mathbf{p}) = \mathbf{Xb}$ is obtained by iteratively solving the PML estimating equations

$$\mathbf{X}'\mathbf{Wf}(\hat{\mathbf{b}}_{pml}) = \mathbf{X}'\mathbf{W}\hat{\mathbf{p}}, \tag{8.24}$$

where $\mathbf{W}$ is a $u \times u$ diagonal weight matrix with weights $w_j = \hat{n}_j$ on the main diagonal, and $\mathbf{f} = \exp(\mathbf{Xb})/(1 + \exp(\mathbf{Xb}))$ is the inverse function of the logit function. It is essential in (8.24) that the weighted domain sample sizes $\hat{n}_j$ and the weighted proportion estimates $\hat{p}_j$ be used, not their unweighted counterparts n_j and $\hat{p}_j^U$ as in the ML method, i.e. under the unweighted SRS option. This is for consistency of the PML estimators. The corresponding vector (8.5) of the GWLS estimates can be used as an initial value for the PML iterations. Note that under the linear formulation of the ANOVA model, the function vector $\mathbf{f}(\hat{\mathbf{b}}_{pml})$ would be linear in $\hat{b}_k$ and, thus, no iterations are needed. Henceforth, in this section we denote the vector of PML estimates of logit model coefficients by $\hat{\mathbf{b}}$ for short.

Because the vector $\hat{\mathbf{b}}$ of PML estimates is equal under the design-based and weighted SRS options, so also are the vectors $\hat{\mathbf{F}} = \mathbf{X}\hat{\mathbf{b}}$ and $\hat{\mathbf{f}} = F^{-1}(\mathbf{X}\hat{\mathbf{b}})$ of fitted logits and fitted proportions. The equality also holds for estimated odds ratios, which can be obtained as $\exp(\hat{b}_k)$ under the partial parametrization of the model. Fitted proportions $\hat{f}_j = f_j(\hat{\mathbf{b}})$ are estimated under both options by the formula

$$\hat{\mathbf{f}} = \mathbf{f}(\hat{\mathbf{b}}) = \exp(\mathbf{X}\hat{\mathbf{b}})/(1 + \exp(\mathbf{X}\hat{\mathbf{b}})). \tag{8.25}$$

Let us derive under the weighted SRS and design-based options the $s \times s$ covariance-matrix estimators of the PML estimator vector $\hat{\mathbf{b}}$ calculated by (8.24). Assuming simple random sampling, the covariance-matrix estimator is given by

$$\hat{\mathbf{V}}_{bin}(\hat{\mathbf{b}}) = (\mathbf{X}'\mathbf{W}\hat{\Delta}\mathbf{W}\mathbf{X})^{-1}, \tag{8.26}$$

where the diagonal elements of the diagonal $u \times u$ matrix $\hat{\Delta}$ are binomial-type variances $\hat{f}_j(1 - \hat{f}_j)/\hat{n}_j$. The binomial covariance-matrix estimator (8.26) is not consistent for complex sampling designs involving clustering. For these designs, we derive a more complicated consistent covariance-matrix estimator that is valid under the design-based option:

$$\hat{\mathbf{V}}_{des}(\hat{\mathbf{b}}) = \hat{\mathbf{V}}_{bin}(\hat{\mathbf{b}})\mathbf{X}'\mathbf{W}\hat{\mathbf{V}}_{des}\mathbf{W}\mathbf{X}\hat{\mathbf{V}}_{bin}(\hat{\mathbf{b}}). \tag{8.27}$$

This estimator is of a ‘sandwich’ form such that the design-based covariance-matrix estimator $\hat{\mathbf{V}}_{des}$ of the proportion vector $\hat{\mathbf{p}}$ acts as the ‘filling’.

Approximate confidence intervals for odds ratio estimates $\exp(b_k)$ under the design-based and weighted SRS options can be calculated by (8.7) using the corresponding variance estimates $\hat{v}_{des}(\hat{b}_k)$ and $\hat{v}_{bin}(\hat{b}_k)$ of the PML estimates $\hat{b}_k$, as in the GWLS method. Also, the design-effect estimates $\hat{d}(\hat{b}_k)$ of the model coefficients $\hat{b}_k$ can be obtained by (8.23), again analogously to the GWLS method.

Expressions for the consistent covariance-matrix estimators $\hat{\mathbf{V}}_{des}(\hat{\mathbf{F}})$ and $\hat{\mathbf{V}}_{des}(\hat{\mathbf{f}})$ of the vector $\hat{\mathbf{F}}$ of fitted logits and the vector $\hat{\mathbf{f}}$ of fitted proportions are similar under the design-based option to those of the GWLS method, as given in equations (8.8) and (8.9). The PML analogue $\hat{\mathbf{V}}_{des}(\hat{\mathbf{b}})$ from (8.27) and the corresponding matrix $\hat{\mathbf{H}}$ must of course be used in the equations. And under the weighted SRS option, the covariance-matrix estimators $\hat{\mathbf{V}}_{bin}(\hat{\mathbf{F}})$ and $\hat{\mathbf{V}}_{bin}(\hat{\mathbf{f}})$ are derived similarly by using the binomial estimator (8.26) in the equations in place of its design-based counterpart.

A residual covariance-matrix estimator is needed for conducting a proper residual analysis under the design-based option. This $u \times u$ estimator is given by

$$\hat{\mathbf{V}}_{res} = \mathbf{A}\hat{\mathbf{V}}_{des}\mathbf{A}', \tag{8.28}$$

where the matrix **A** is obtained by the formula

$$\mathbf{A} = \mathbf{I} - \hat{\Delta}\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\hat{\Delta}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}$$

with **I** being a $u \times u$ identity matrix. Using this estimate, design-based standardized residuals of the form (8.22) can then be calculated.

There are thus many similarities between the PML formulae and those derived for the GWLS method. The main differences lie in the way the estimates of model coefficients and their covariance-matrix estimate are calculated. More similarities are evident in the testing procedures. All the test statistics derived for the GWLS method are also applicable to the PML method.

Under the design-based option, goodness of fit of the model can be tested with the design-based Wald statistic X^2_{des} given by (8.11). When examining the model fit more closely, PML analogues to the Wald statistics $X^2_{des}(overall)$ and $X^2_{des}(gof)$ can be used. The Wald statistics (8.13) and (8.14) for linear hypotheses on model parameters are applicable as well. Finally, in unstable situations, the F-corrected Wald and Rao–Scott statistics (8.16)–(8.20) can be used. It should be noted that the PML estimates from (8.24) and the corresponding covariance-matrix estimate (8.27) must be used in the calculation of these test statistics under the design-based option. These test statistics are available in commonly used software products for logit analysis for complex survey data.

In testing procedures for the weighted and unweighted SRS options, the corresponding binomial covariance-matrix estimates are used in the test statistics in place of those from the design-based option. As an alternative to the Wald statistics, LR test statistics can be used, which for the design-based option should be adjusted using the Rao–Scott methodology. A second-order adjustment to LR test statistics similar to (8.14) for the binomial-based Wald statistic provides asymptotically chi-squared test statistics. The residual covariance-matrix estimate (8.28) can be used in deriving an appropriate generalized design-effects matrix estimate for the adjustments.

The main application area of the PML method for complex surveys is under the design-based option, and the weighted and unweighted SRS options are used as the reference when examining the effects of weighting and intra-cluster correlation on standard-error estimates of model coefficients and on p-values of Wald test statistics.

Logistic Regression

The PML method can also be used in strictly regression-type logit analyses on a binary response variable from a complex survey, where the predictors are continuous measurements. In logistic regression, we work with an element-level data set without aggregating these data into a multidimensional table. So, the measured values of the continuous predictor variables constitute the columns in an $n \times s$ model matrix $\mathbf{X}$ for a logistic regression model. But all the other elements of the PML estimation remain unchanged, and consistent PML estimates with their consistent covariance-matrix estimate are obtained in a way similar to that described for the design-based analysis option. Moreover, a logistic ANCOVA can be performed by incorporating categorical predictors into the logistic regression model. Then, interaction terms of the continuous and categorical predictors can also be included.

A logistic regression model is usually built by entering predictors into the model using subject-matter criteria or significance measures of potential predictors. In this, t-tests $\mathrm{t}_{des}(b_k)$, or the corresponding Wald tests $X^2_{des}(b_k)$, on model coefficients can be used as previously and, under the design-based option, asymptotic properties of these test statistics remain unchanged.

Instability of an estimate $\hat{\mathbf{V}}_{des}(\hat{\mathbf{b}})$ from (8.27) can destroy the distributional properties of the test statistics on model coefficients in such small-sample situations where the number of sample clusters is small. Usual degrees-of-freedom, F-corrections to the Wald and t-test statistics can then be used.

The GEE methodology of generalized estimating equations can also be used for logistic modelling on complex survey data. In this method, the model coefficients are estimated using the multivariate quasilikelihood technique, and intra-cluster correlations are taken as nuisances. Using an estimated intra-cluster correlation structure, a 'robust' estimator of the covariance matrix of the model coefficients can be obtained, basically similar to the 'sandwich' form in the PML method. Thus, the GEE method can be used to account for the clustering effects. We describe only briefly the method and give an example for logistic ANCOVA in the OHC Survey.

The GEE method was originally developed for accounting for the possible correlation of observations in fitting generalized linear models in the context of longitudinal surveys (Liang and Zeger 1986). The methodology has been further described and illustrated in Liang *et al.* (1992) and Diggle *et al.* (2002).

Two alternatives of the GEE method have been presented. A preliminary GEE method with an independent correlation assumption relates to the standard PML method where observations are assumed independent within clusters for the estimation of the regression coefficients, but are allowed to be correlated for the estimation of the covariance matrix of the estimated regression coefficients. In covariance-matrix estimation, a 'sandwich' form of estimator is used. In a more advanced GEE method, assuming an exchangeable correlation structure, observations are allowed to be correlated within clusters in the estimation of both regression coefficients and the covariance matrix of estimated regression coefficients. There, a 'working' intra-cluster correlation is estimated and incorporated in the estimation procedure of regression coefficients and the covariance matrix of estimated coefficients.

A generalized linear model can be compactly written as

$$\mathrm{E_M}(\mathrm{g}(\mathbf{y})) = \mathbf{Xb}, \tag{8.29}$$

where $\mathrm{E_M}$ refers to the expectation under the model and the function g refers to the so-called link function postulating a relationship between the expectation of the response variable vector **y** and the linear part **Xb** of the model. Special cases of link functions are identity, logistic and logarithmic functions used in linear models for continuous responses, logistic models for binary responses and log-linear models for count data, respectively.

The covariance structure of observations within clusters is modelled by

$$\mathbf{V}_i = \phi \mathbf{A}_i^{1/2} \mathbf{R}(\alpha) \mathbf{A}_i^{1/2}, \qquad i = 1, \ldots, m, \tag{8.30}$$

where $\mathbf{A}_i$ is a diagonal matrix of variances $V(y_k)$ in cluster i and $\mathbf{R}(\alpha)$ is the 'working' correlation matrix specified by the (possibly vector-valued) correlation parameter α of observations in cluster i. The parameter ϕ denotes the dispersion parameter of the corresponding member of the exponential family of distributions. Under an independent correlation assumption, all off-diagonal elements α of the 'working' correlation matrix are set to zero. Under an exchangeable correlation of pairs of observations within a cluster, the parameter α is a scalar and requires estimation. In an estimation procedure to obtain an estimate $\hat{\mathbf{b}}$, Newton–Raphson-type algorithms are usually used. The covariance-matrix estimate $\hat{\mathbf{V}}_{des}(\hat{\mathbf{b}})$ is obtained using a 'sandwich' type estimator (see equation (8.27)). Element weights can be incorporated in a GEE estimation procedure. GEE and the weighted analogue can be applied using suitable software for the analysis of complex surveys.

The GEE method has been shown to produce consistent estimates of model parameters and their covariance matrices, independently of a correct specification of the 'working' correlation structure. In the next two examples, we apply logistic ANCOVA first with the PML method and then with the GEE method assuming an exchangeable intra-cluster correlation structure. For further training on the PML and GEE methods in logistic modelling on the OHC Survey data, the reader is advised to visit the web extension of the book.

Example 8.2

Logistic ANCOVA with the PML method. Let us consider in a slightly more general setting the analysis situation of Example 8.1, where a logit ANOVA model was fitted by the GWLS method to proportions in a multidimensional table. We now fit a logistic ANCOVA model using the PML method, by entering some of the predictors as continuous measurements in the model. The design-based analysis option is applied, providing valid PML analysis.

The binary response variable PSYCH measures high psychic strain, and we take the variables AGE, PHYS (physical working conditions) and CHRON (chronic morbidity) as continuous predictors such that AGE is measured in years and PHYS and CHRON are binary. Thus there are four predictors, of which SEX is taken as a qualitative predictor. So, the interaction of SEX with AGE, PHYS and CHRON can also be examined.

A model with SEX, AGE, PHYS and CHRON as the main effects and an interaction term of SEX and AGE was taken as the final model, because the other interactions appeared nonsignificant at the 5% level. Results of the model coefficients are displayed in Table 8.8.

The fitted logit ANCOVA model can be written using the estimated coefficients $\hat{b}_k$ and the corresponding model matrix $\mathbf{X}$ similar to the ANOVA modelling in Example 8.1:

$$
\begin{aligned}
F(\hat{f}_l) = \hat{b}_1 &+ \hat{b}_2(\text{SEX})_l + \hat{b}_3(\text{AGE})_l + \hat{b}_4(\text{PHYS})_l \\
&+ \hat{b}_5(\text{CHRON})_l + \hat{b}_6(\text{SEX} * \text{AGE})_l,
\end{aligned}
$$

where $l = 1, \ldots, 7841$, and $F(\hat{f}_l) = \log(\hat{f}_l/(1 - \hat{f}_l))$. The values for the model terms are obtained from the corresponding columns of the 7841×6 model matrix $\mathbf{X}$. There, SEX, PHYS and CHRON are binary, and AGE has its original values (age

Table 8.8 Design-based logistic ANCOVA on overall psychic strain with the PML method.

Model term	Beta coefficient	Design effect	Standard error	t-test	p-value	Odds ratio	95% confidence interval for OR Lower	 Upper
Intercept	0.1964	1.56	0.1572	1.25	0.2127	1.22	0.89	1.66
Sex								
Males	−0.9926	1.43	0.2033	−4.88	0.0000	0.37	0.25	0.55
Females*	0	n.a.	0	n.a.	n.a.	1	1	1
Age	−0.0046	1.55	0.0041	−1.12	0.2624	1.00	0.99	1.00
Physical health hazards	0.2765	1.39	0.0596	4.64	0.0000	1.32	1.17	1.48
Chronic morbidity	0.5641	1.17	0.0575	9.82	0.0000	1.76	1.57	1.97
Sex, Age								
Males	0.0131	1.41	0.0051	2.56	0.0111	1.01	1.00	1.02
Females*	0	n.a.	0	n.a.	n.a.	1	1	1

* Reference class; parameter value set to zero.
n.a. not available.

in years). Note the difference in the ANCOVA model matrix when compared with that for the ANOVA model.

The *t*-tests on model coefficients indicate that the coefficients for the interesting predictors, physical working conditions and chronic morbidity are strongly associated with experiencing psychic strain. Persons in hazardous work, and chronically ill persons are more likely to suffer from psychic strain than healthy persons and persons whose working conditions are less hazardous. Note that the sex–age adjusted coefficient $\hat{b}_5$ for CHRON is larger than $\hat{b}_4$ for PHYS. Thus, in the model, chronic morbidity is more important as a predictor of psychic strain. This can also be seen in the odds ratio (OR) estimates provided in Table 8.8.

Odds ratios with their approximative 95% confidence intervals (in parenthesis) thus are

$$\text{PHYS: Odds ratio} = \exp(0.2765) = 1.32 \quad (1.17,\ 1.48),$$

$$\text{CHRON: Odds ratio} = \exp(0.5641) = 1.76 \quad (1.57,\ 1.97).$$

We may thus conclude that odds for experiencing a higher level of psychic strain, adjusted for sex, age and chronic morbidity, is about 1.3 times higher for those in more hazardous work than for those in less hazardous work. This conclusion was similar in Example 8.1, where a closely related odds ratio and confidence interval were obtained. Furthermore, the odds of experiencing much psychic strain, adjusted for sex, age and working conditions, are about 1.8 times higher for chronically ill persons than for healthier persons. Because neither of the 95% confidence intervals covers the value one, the corresponding odds ratios differ significantly (at the 5% level) from one. It should be noted that the binomial-based confidence intervals would be narrower especially for the predictor PHYS, for which the design-effect estimate is larger than for CHRON.

An analysis under the SRS options yield the same final model as the design-based analysis, but the observed values of the test statistics are somewhat larger and thus more liberal test results are attained.

Finally, let us examine more closely the fitted proportions $\hat{f}_l$ for the upper psychic strain group under the present model. The results are summarized in Figure 8.2 by plotting the proportions against the predictors included in the model. Fitted proportions increase with increasing age for males, and decrease for females. At a given age, the proportions are larger for the chronically ill and for those in more hazardous work than in the reference groups. Also, in females the fitted proportions tend to be larger than in males in all the corresponding domains, although the differences decline with increasing age.

Example 8.3

Logistic ANCOVA with the GEE method. Let us consider further the analysis situation of Example 8.2, where a logistic ANCOVA model was fitted by the PML method. We now fit a logistic ANCOVA model using the GEE method with

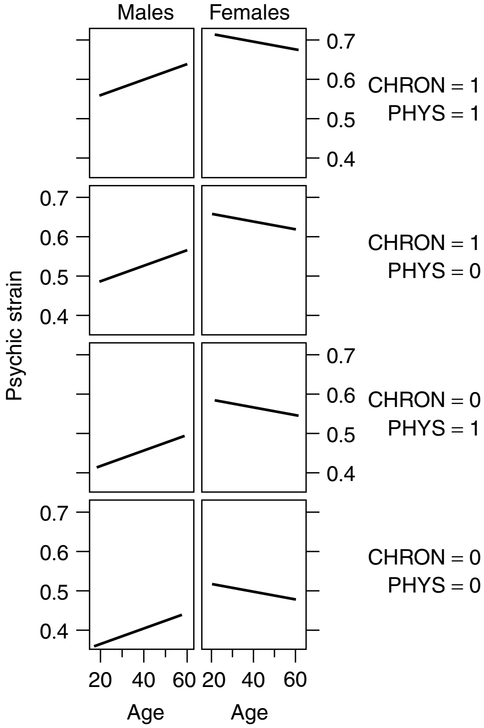

Figure 8.2 Fitted proportions of falling into the high psychic strain group for the final logistic ANCOVA model.

an assumed exchangeable correlation of pairs of observations within a cluster. Similarly as in Example 8.2, our response variable is the binary PSYCH measuring psychic strain. The variable SEX is included in the model as a categorical predictor and AGE, PHYS (physical working conditions) and CHRON (chronic morbidity) as continuous predictors such that AGE is measured in years and PHYS and CHRON are binary. We fit the same model as in Example 8.2.

Results are shown in Table 8.9. A comparison with logistic ANCOVA with the PML method in Example 8.2 indicates that the results are quite similar, and our inferential conclusions remain the same. There are, however, certain differences. First, the estimated beta coefficients have changed. Absolute values of estimates are larger than in the PML application, except for the CHRON effect. Standard-error estimates are somewhat smaller than the PML counterparts. Hence, the observed *t*-statistics tend to be larger involving slightly more liberal tests than in the PML case. These differences are due to the fact that in the GEE method with an exchangeable correlation structure, the correlation of observations also contributes to the estimation of the beta parameters. The 'working' intra-cluster

Table 8.9 Design-based logistic ANCOVA on overall psychic strain with the GEE method under exchangeable intra-cluster correlation structure.

Model Term	Beta coefficient	Design effect	Standard error	*t*-test	*p*-value
Intercept	0.2292	1.44	0.1524	1.50	0.1338
Sex					
Males	−1.0290	1.36	0.2000	−5.14	0.0000
Females*	0	n.a.	0	n.a.	n.a.
Age	−0.0057	1.43	0.0039	−1.45	0.1489
Physical health hazards	0.3011	1.31	0.0587	5.13	0.0000
Chronic morbidity	0.5569	1.14	0.0568	9.81	0.0000
Sex, Age					
Males	0.0144	1.33	0.0050	2.88	0.0044
Females*	0	n.a.	0	n.a.	n.a.

* Reference class; parameter value set to zero.
n.a. not available.

correlation is estimated as $\hat{\alpha} = 0.0189$. Using the expression $\text{deff} = 1 + (\overline{m} - 1)\hat{\alpha}$, where $\overline{m}$ is the average cluster size, this corresponds to an average design effect of 1.57.

Linear Modelling on Continuous Responses

We have extensively considered the modelling of binary response variables from complex surveys. The GWLS, PML and GEE methods were used, covering logit and linear modelling on categorical data and logit modelling with continuous predictors. These types of multivariate models are most frequently found in analytical surveys, for example, in social and health sciences. But in some instances it is appropriate to model a quantitative or continuous response variable, such as the number of physician visits or blood pressure. We discuss briefly the special features of multivariate analysis in such cases, and give an illustrative example of a special case of linear ANCOVA.

Linear modelling provides a convenient analysis methodology for analysis situations with a continuous response variable and a set of predictors. This situation was present in Examples 8.2 and 8.3, where the dichotomized PSYCH was analysed with a logistic ANCOVA model. There the original continuous variable on psychic strain could be taken as the response variable as well, leading to linear ANCOVA modelling. For a simple random sample, the analysis would be based on ordinary least squares (OLS) estimation with a standard program for linear modelling. For the OHC Survey data set, which is based on cluster sampling, the design-based approach with weighted least squares (WLS) estimation provides proper linear modelling.