are thought to be correlated with each other), and selects a model for the within-subject (or within-group) correlation structure. Two useful choices for correlation structure are "exchangeable" and "AR(1)" models. In the exchangeable model, the observations for a given subject are taken to be correlated with each other and the correlation of any two of them (after accounting for the effects of the explanatory variables) is the same. In the AR(1) (autoregressive of lag 1) model, the correlation of observations within a subject is thought to be induced by an AR(1) time series (see Section 15.3). In this model, observations taken closer together in time are thought to be more highly correlated with each other than observations taken farther apart in time. In both of these models there is a single additional parameter—a correlation coefficient or a first serial correlation coefficient—which is assumed to be the same for all subjects or clusters.

A good starting point for modeling correlation is to use the AR(1) model if the correlated observations are repeated measures over time and to use the exchangeable model otherwise. In many data problems, it doesn't matter too much which model is used because even an approximate model for correlation structure solves the problem well enough to make useful conclusions about the logistic regression model.

21.8 SUMMARY

Binomial logistic regression is a special type of generalized linear model in which the response variable is binomial and the logit of the associated probability is a linear function of regression coefficients. The case in which all the binomial denominators are 1 was referred to as binary logistic regression and was discussed in Chapter 20.

In analyzing proportions, one should begin, if possible, with scatterplots of the sample logits versus explanatory variables. The model-building stage consists of checking for extra-binomial variation and then finding an appropriate set of explanatory variables. The main inferential statements are based on Wald's z-test or the associated confidence interval for a single coefficient, or on the drop-in-deviance chi-squared test for a hypothesis about one or several coefficients. If extra-binomial variation is suspected, these tests are replaced by t-tests and F-tests, respectively.

Krunnit Islands Data

The analysis comes into focus with a plot of empirical logits versus the log of island area. It is evident from Display 21.2 that the log odds of extinction are linear in the log of island area. This can be confirmed by fitting the model and observing that the deviance statistic is not large (the p-value from the deviance goodness-of-fit test is 0.74), and also by testing for the significance of a squared term in log island area. The remainder of the analysis only requires that the inferential statement about the association between island area and log odds of extinction be worded on the original scale.

Moth Coloration Data

A plot of the empirical logits versus distance from Liverpool, with different codes for the light and dark moths, revealed initial evidence of an interactive effect—particularly that the odds of light morph removal decrease with increasing distance from Liverpool while the odds of dark morph removal increase with increasing distance from Liverpool. Based on the scatterplot (Display 21.4), the linearity of the logistic regression on distance is questionable, particularly since the logits drop rather dramatically at Loggerheads. There seems to be good reason to believe that distance from Liverpool is not entirely adequate as a surrogate for darkness of trees. Two avenues for dealing with this problem were suggested. One involved using quasi-likelihood analysis, which incorporated the model inadequacy into a dispersion parameter. The other, more appealing model, however, included location as a factor with seven levels (rather than relying on their distance from Liverpool as a single numerical explanatory variable) but retained the product of the morph indicator variable and distance so that a single parameter represented the interactive effect of interest. This model fit well and maximum likelihood analysis was used to draw inferences about the effect.

21.9 EXERCISES

Conceptual Exercises

1. **Krunnit Islands Extinctions.** Logistic regression analysis assumes that, after the effects of explanatory variables have been accounted for, the responses are independent of each other. Why might this assumption be violated in these data?

2. **Moth Coloration.** (a) Why might one suspect extra-binomial variation in this problem? (b) Why does the question of interest pertain to an interaction effect? (c) How (if at all) can the response variable be redefined so that the question of interest is about a "main effect" rather than an interaction?

3. Which of the following are not *counted proportions*? (a) The proportion of cells with chromosome aberrations, out of 100 examined. (b) The proportion of body fat lost after a special diet. (c) The proportion of 10 grams of offered food eaten by a quail. (d) The proportion of a mouse's four limbs that have malformations.

4. Suppose that Y has a binomial(m, π) distribution. (a) Is the variance of the binomial count Y an increasing or decreasing function of m? (b) Is the variance of the binomial proportion Y/m an increasing or decreasing function of m?

5. To confirm the appropriateness of the logistic regression model logit$(\pi) = \beta_0 + \beta_1 x$, it is sometimes useful to fit logit$(\pi) = \beta_0 + \beta_1 x + \beta_2 x^2$ and test whether β_2 is zero. (a) Does the reliability of this test depend on whether the denominators are large? (b) Is the test more relevant when the denominators are small?

6. What is the quasi-likelihood approach for extending the logistic regression model to account for a case in which more variation appears than was predicted by the binomial distribution?

7. **Malformed Forelimbs.** In a randomized experiment, mouse fetuses were injected with 500, 750, or 1,000 mg/kg of acetazolamide on day 9 of gestation. (Data from L. B. Holmes, H. Kawanishi, and A. Munoz, "Acetazolamide: Maternal Toxicity, Pattern of Malformations, and Litter Effect," *Teratology* 37 (1988): 335–42.) On day 18 of gestation, the fetus was removed and examined for malformations in the front legs. Display 21.17 shows the results. (a) Suppose that the response

variable for each mouse fetus is the number of forelimbs with malformations. Explain how logistic regression can be used to study the relationship between malformation probability and dose. (b) What is the sample size? (c) Is it necessary to worry about extra-binomial variation? (d) Is it appropriate to use the deviance goodness-of-fit test for model checking? (e) Is it appropriate to obtain a confidence interval for the effect of dose? (f) How would the resulting confidence interval be used in a sentence summarizing the statistical analysis?

DISPLAY 21.17 Dose of acetazolamide and number of malformed forelimbs in mouse fetuses

Dose (mg/kg)	Number of malformed forelimbs			Sample size
	0	1	2	
500	137	40	15	192
750	81	68	42	191
1,000	87	49	69	205

8. If your computer software always produces binary response deviances, how can you determine the lack-of-fit chi-square for the moth data? (See Section 21.7.1.)

Computational Exercises

9. Moth Coloration. For the moth coloration data in Section 21.1.1, consider the response count to be the number of light moths removed and the binomial denominator to be the total number of moths removed (light and dark) at each location. (a) Plot the logit of the proportion of light moths removed versus distance from Liverpool. (b) Fit the logistic regression model with distance as the explanatory variable; then report the estimates and standard errors. (c) Compute the deviance goodness-of-fit test statistic, and obtain a p-value for testing the adequacy of the model.

10. Death Penalty and Race of Victim. Reconsider the data of Display 19.2 involving death penalty and race of victim. Reanalyze these data using logistic regression. The response variable is the number of convicted murderers in each category who receive the death sentence, out of the m convicted murderers in that category. (a) Plot the logits of the observed proportions versus the level of aggravation. The logit, however, is undefined for the rows where the proportion is 0 or 1, so compute the empirical logit $= \log[(y + 0.5)/(m - y + 0.5)]$ and plot this versus aggravation level, using different plotting symbols to distinguish proportions based on white and black victims. (b) Fit the logistic regression of death sentence proportions on aggravation level and an indicator variable for race of victim. (c) Report the p-value from the deviance goodness-of-fit test for this fit. (d) Test whether the coefficient of the indicator variable for race is equal to 0, using the Wald's test. (e) Construct a confidence interval for the same coefficient, and interpret it in a sentence about the odds of death sentence for white-victim murderers relative to black-victim murderers, accounting for aggravation level of the crime.

11. Death Penalty and Race of Victim. Fit a logistic regression model as in Exercise 10, but treat aggravation level as a factor. Include the products of the race variable with the aggravation level indicators to model interaction, and determine the drop in deviance for including them. This constitutes a test of the equal odds ratio assumption, which could be used to check the assumption for the Mantel–Haenszel test. What can you conclude about the assumption?

12. Vitamin C and Colds. Reconsider the data in Display 18.2 from the randomized experiment on 818 volunteers. Let Y_i represent the number of individuals with colds in group i, out of m_i. Let X_i be an indicator variable that takes on the value 1 for the placebo group and 0 for the vitamin C group. Then the observed values for the two groups are:

Group (i)	Individuals with Colds (Y)	Group size (m)	Placebo indicator (X)
1	335	411	1
2	302	407	0

Consider the logistic regression model in which Y_i is binomial(π_i, m_i) and logit$(\pi_i) = \beta_0 + \beta_1 X_i$. (a) Using a computer package, obtain an estimate of and a confidence interval for β_1. (b) Interpret the results in terms of the odds of cold for the placebo group relative to the odds of a cold for the vitamin C group. (c) How does the answer to part (b) compare to the answer in Display 18.9?

13. Vitamin C. Between December 1972 and February 1973, a large number of volunteers participated in a randomized experiment to assess the effect of large doses of vitamin C on the incidence of colds. (Data from T. W. Anderson, G. Suranyi, and G. H. Beaton, "The Effect on Winter Illness of Large Doses of Vitamin C," *Canadian Medical Association Journal* 111 (1974): 31–36.) The subjects were given tablets to take daily, but neither the subjects nor the doctors who evaluated them were aware of the dose of vitamin C contained in the tablets. Shown in Display 21.18 are the proportion of subjects in each of the four dose categories who did not report any illnesses during the study period.

DISPLAY 21.18 Vitamin C and colds

Daily dose of vitamin C (g)	Number of subjects	Number with no illnesses	Proportion with no illnesses
0	1,158	267	0.231
0.25	331	74	0.224
1	552	130	0.236
2	308	65	0.211

(a) For each of the four dose groups, calculate the logit of the estimated proportion. Plot the logit versus the dose of vitamin C. (b) Fit the logistic regression model, logit$(\pi) = \beta_0 + \beta_1$ *dose*. Report the estimated coefficients and their standard errors. Report the p-value from the deviance goodness-of-fit test. Report the p-value for a Wald's test that β_1 is 0. Report the p-value for a drop-in-deviance test for the hypothesis that β_1 is 0. (c) What can be concluded about the adequacy of the binomial logistic regression model? What evidence is there that the odds of a cold are associated with the dose of vitamin C?

14. Spock Conspiracy Trial. Reconsider the proportions of women on venires in the Boston U.S. district courts (case study 5.2). Analyze the data by treating the number of women out of 30 people on a venire as a binomial response. (a) Do the odds of a female on a venire differ for the different judges? Answer this with a drop-in-deviance chi-square test, comparing the full model with judge as a factor to the reduced model with only an intercept. (b) Do judges A–F differ in their probabilities of selecting females on the venire? Answer this with a drop-in-deviance chi-square test by comparing the full model with judge as a factor to the reduced model that has an intercept and an indicator variable for Spock's judge. (c) How different are the odds of a woman on Spock's judge's venires from the odds on the other judges? Answer this by interpreting the coefficients in the binomial logistic regression model with an intercept and an indicator variable for Spock's judge.

Data Problems

15. Belief Accessibility. Increasingly, politicians look to public opinion surveys to shape their public stances. Does this represent the ultimate in democracy? Or are seemingly scientific polls being rigged by the manner of questioning? Psychologists believe that opinions—expressed as answers to questions—are usually generated at the time the question is asked. Answers are based on a quick sampling of relevant beliefs held by the subject, rather than a systematic canvas of all such beliefs. Furthermore, this sampling of beliefs tends to overrepresent whatever beliefs happen to be most accessible at the time the question is asked. This aspect of delivering opinions can be abused by the pollster. Here, for example, is one sequence of questions: (1) "Do you believe the Bill of Rights protects personal freedom?" (2) "Are you in favor of a ban on handguns?" Here is another: (1) "Do you think something should be done to reduce violent crime?" (2) "Are you in favor of a ban on handguns?" The proportion of yes answers to question 2 may be quite different depending on which question 1 is asked first.

To study the effect of *context questions* prior to a *target question*, researchers conducted a poll involving 1,054 subjects selected randomly from the Chicago phone directory. To include possibly unlisted phones, selected numbers were randomly altered in the last position. (Data from R. Tourangeau, K. A. Rasinski, N. Bradburn, and R. D'Andrade, "Belief Accessibility and Context Effects in Attitude Measurement," *Journal of Experimental Social Psychology* 25 (1989): 401–21.) The data in Display 21.19 show the responses to one of the questions asked concerning continuing U.S. aid to the Nicaraguan Contra rebels. Eight different versions of the interview were given, representing all possible combinations of three factors at each of two levels. The experimental factors were CONTEXT, MODE, and LEVEL. CONTEXT refers to the type of context questions preceding the question about Nicaraguan aid. Some subjects received a context question about Vietnam, designed to elicit reticence about having the United States become involved in another foreign war in a third-world country. The other context question was about Cuba, designed to elicit anti-communist sentiments. MODE refers to whether the target question immediately followed the context question or whether there were other questions scattered in between. LEVEL refers to two versions of the context question. In the HIGH LEVEL the question was worded to elicit a higher level of agreement than in the LOW LEVEL wording.

Analyze these data to answer the following questions of interest: (a) Does the proportion of favorable responses to the target question depend on the wording (LEVEL) of the question? If so, by how much? (b) Does the proportion depend on the context question? If so, by how much? (c) Does the proportion depend on the context question to different extents according to whether the target and context questions are scattered? (*Hint:* Use indicator explanatory variables for the factors.)

16. Aflatoxicol and Liver Tumors in Trout. An experiment at the Marine/Freshwater Biomedical Sciences Center at Oregon State University investigated the carcinogenic effects of aflatoxicol, a metabolite of Aflatoxin B1, which is a toxic by-product produced by a mold that infects cottonseed meal, peanuts, and grains. Twenty tanks of rainbow trout embryos were exposed to one of five doses

DISPLAY 21.19 Opinion polls and belief accessibility: CONTEXT refers to the context of the question preceding the target question about U.S. aid to the Nicaraguan Contra rebels; MODE is "scattered" if the target question was not asked directly after the context question; and LEVEL refers to the wording of the question ("high" designed to elicit a higher favorable response)

GROUP:	1	2	3	4	5	6	7	8
CONTEXT:	Vietnam	Cuba	Vietnam	Cuba	Vietnam	Cuba	Vietnam	Cuba
MODE:	scattered	scattered	scattered	scattered	not	not	not	not
LEVEL:	high	high	low	low	high	high	low	low
NUMBER (m):	132	132	132	131	132	131	132	132
PERCENTAGE IN FAVOR OF CONTRA AID:	20.5	31.1	28.0	38.9	34.1	48.9	23.5	45.5

of Aflatoxicol for one hour. The data in Display 21.20 (from George Bailey and Jerry Hendricks) represent the numbers of fish in each tank and the numbers of these that had liver tumors after one year. Describe the relationship between dose of Aflatoxicol and odds of liver tumor. It is also of interest to determine the dose at which 50% of the fish will get liver tumors. (*Note: Tank effects* are to be expected, meaning that tanks given the same dose may have slightly different π's. Thus, one should suspect extra-binomial variation.)

DISPLAY 21.20 Aflatoxicol and liver tumors in trout (four tanks at each dose)

Dose (ppm)	Number of trout with liver tumors (Y)/Number in tank (m)			
0.010	9/87	5/86	2/89	9/85
0.025	30/86	41/86	27/86	34/88
0.050	54/89	53/86	64/90	55/88
0.100	71/88	73/89	65/88	72/90
0.250	66/86	75/82	72/81	73/89

17. Effect of Stress During Conception on Odds of a Male Birth. The probability of a male birth in humans is about 0.51. It has previously been noticed that lower proportions of male births are observed when offspring are conceived at times of exposure to smog, floods, or earthquakes. Danish researchers hypothesized that sources of stress associated with severe life events may also have some bearing on the sex ratio. To investigate this theory they obtained the sexes of all 3,072 children who were born in Denmark between January 1, 1980 and December 31, 1992 to women who experienced the following kinds of severe life events in the year of the birth or the year prior to the birth: death or admission to hospital for cancer or heart attack of their partner or of their other children. They also obtained sexes on a sample of 20,337 births for mothers who did not experience these life stress episodes. Shown in Display 21.21 are the percentages of boys among the births, grouped according to when the severe life event took place. Notice that for one group the exposure is listed as taking place during the first trimester of pregnancy. The rationale for this is that the stress associated with the cancer or heart attack of a family member may well have started before the recorded time of death or hospital admission. Analyze the data to investigate the researchers' hypothesis. Write a summary of statistical findings. (Data from D. Hansen et al., "Severe Periconceptional Life Events and the Sex Ratio in Offspring: Follow Up Study Based on Five National Registers," *British Medical Journal,* 319 (1999): 548–49.)

18. HIV and Circumcision. Researchers in Kenya identified a cohort of more than 1,000 prostitutes who were known to be a major reservoir of sexually transmitted diseases in 1985. It was determined that more than 85% of them were infected with human immunodeficiency virus (HIV)

DISPLAY 21.21 Percentages of boys born to mothers in a control group and to mothers exposed to severe life events, at various times in relation to the birth

Group	Time of stress event	Number of births	% boys
Control	(none)	20,337	51.2
Exposed	13–16 mo. prior	71	52.1
Exposed	7–12 mo. prior	789	49.6
Exposed	0–6 mo. prior	1,922	48.9
Exposed	1st trimester	290	46.0

in February 1986. The researchers then identified men who acquired a sexually transmitted disease from this group of women after the men sought treatment at a free clinic. Display 21.22 shows the subset of those men who did not test positive for the HIV on their first visit and who agreed to participate in the study. The men are categorized according to whether they later tested positive for HIV during the study period, whether they had one or multiple sexual contacts with the prostitutes, and whether they were circumcised. Describe how the odds of testing positive are associated with number of contacts and with whether the male was circumcised. (Data from D. W. Cameron et al., "Female to Male Transmission of Human Immunodeficiency Virus Type 1: Risk Factors for Seroconversion in Men," *The Lancet* (1989): 403–07.)

DISPLAY 21.22 Number of Kenyan men who tested positive for HIV, categorized according to two possible risk factors

	Single contact with prostitutes		Multiple contact with prostitutes	
	Circumcised	Uncircumcised	Circumcised	Uncircumcised
Tested positive for HIV	1	5	5	13
Number of men	46	27	168	52

19. Meta-Analysis of Breast Cancer and Lactation Studies. Meta-analysis refers to the analysis of analyses. When the main results of studies can be cast into 2×2 tables of counts, it is natural to combine individual odds ratios with a logistic regression model that includes a factor to account for different odds from the different studies. In addition, the odds ratio itself might differ slightly among studies because of different effects on different populations or different research techniques. One approach for dealing with this is to suppose an underlying common odds ratio and to model between-study variability as extra-binomial variation. Display 21.23 shows the results of 10 separate case–control studies on the association of breast cancer and whether a woman had breast fed children. How much greater are the odds of breast cancer for those who did not breast feed than for those who did breast feed? (Data gathered from various sources by Karolyn Kolassa as part of a Master's project, Oregon State University.)

20. Clever Hans Effect. Exercise 14.19 described a randomized block experiment conducted on psychology students to examine the effects of expectations—induced in the experiment by falsely telling some of the students that their rats were bred to be good maze learners—on the students' success in training the rats. The number of successful maze runs, out of 50, was recorded for each student on each of five consecutive days. Data file ex2120 repeats the data but in a revised format, with the responses on the five days listed on separate rows in the variable *Success*. There are 60 rows corresponding to five responses for each of the 12 students in the study. Also listed are the students reported expectation of success (*PriorExp*) prior to the randomization and commencement

DISPLAY 21.23 Counts from 10 case–control studies investigating the association between cancer and whether women had breast fed children (lactated); first 4 of 20 rows

Study		Breast cancer	No breast cancer
1	Did not lactate	107	226
	Lactated	352	865
2	Did not lactate	244	489
	Lactated	574	1,059

of the study, a student ID (*Student*), *Treatment* (*bright* if students were told their rats were bright and *dull* if not), and *Day* of the experiment. Analyze the data to judge the effect of the treatment on the students' success. (It isn't possible to simultaneously estimate a *Treatment* Effect and a *Student* Effect. Try including *PriorExp* instead of *Student* as a way of modeling the possible student effect. To more specifically deal with dependence of observations on the same student, use GEE, as discussed in Section 21.7.5. Be alert for possible overdispersion.)

21. Predicting Quality of Spring Desert Wildflower Display (Ordered Categorical Response). Review the description of the desert wildflower data problem in Exercise 12.21. The response for each desert–year combination is the subjectively rated quality of the spring wildflower display, with ordered response categories *poor, fair, good, great*, and *spectacular*. Use a computer routine for regression on ordered categorical responses (such as proportional odds logistic regression with polr in R or proc logistic in SAS) to fit a regression model with wildflower display quality as the ordered categorical response and total rainfall in the months September to March as an explanatory variable.

22. Environmental Voting. The data file ex0126 includes the number of pro-environment and anti-environment votes cast by each member of the U.S. House of Representatives in 2005, 2006, and 2007 (according to the League of Conservation Voters). Describe the disparity between Republican and Democratic representatives on the probability of casting a pro-environment vote after accounting for *Year*. Ignore representatives from parties other than *D* and *R*.

Answers to Conceptual Exercises

1. Neighboring islands are likely to be more similar in the types of species present and in the likelihood of extinction for each species, than islands farther apart. Another possibility is that extinctions of (migratory) shorebirds can occur in another part of the world and simultaneously affect all or several of the Krunnit Islands populations.

2. (a) Sites the same distance from Liverpool may have different probabilities of removal for dark and light morphs because the blackening of the trees may depend on more than simple distance from Liverpool. (b) The question is not whether the probability of removal depends on the covariate distance from Liverpool or on the factor morph, but whether the probability of removal depends on distance from Liverpool to a different extent for light morphs than for dark morphs. (c) Let m be the number of moths removed at a location. Let Y be the number of these that are dark morphs. How does the proportion of removed moths that are dark morphs depend on distance from Liverpool?

3. Choices (b) and (c).

4. (a) Increasing. (b) Decreasing.

5. (a) Partly. The Wald's test and the drop-in-deviance test both can be used if either the sample size is large or the binomial denominators are large. (b) The test is useful as an informal device for assessing the model in each case. It may be more relevant when the denominators are small, however, since few alternatives are available for model checking in this case.

6. The assumed variance is $\psi m_i \pi_i (1 - \pi_i)$. The dispersion parameter ψ is introduced to allow for variance greater than (or possibly less than) that expected from a binomial response.

7. (a) Suppose that Y_i is binomial$(2, \pi_i)$ with $\text{logit}(\pi_i) = \beta_0 + \beta_1 dose_i$ for $i = 1, \ldots, 588$. Explore the adequacy of this model. If it fits, make inferences about β_1. (b) 588. (c) Probably not, since the denominators are all so small. (d) No, the denominators are too small. (e) Yes. (f) "The odds of a malformation are estimated to increase by (lower limit) to (upper limit) for each 100 mg/kg of acetazolamide (95% confidence interval)." (*Note:* If L and U are the lower and upper endpoints of a confidence interval for β_1, then the interval in this sentence is $\exp(100L)$ and $\exp(100U)$.)

8. Compare the deviance from the model of interest to the deviance from the full model, with a separate mean for each morph at each location.