... inference

```
 > head(shuttle2)
  m r temperature pressure
1 6 0          66       50
2 6 1          70       50
3 6 0          69       50
4 6 0          68       50
5 6 0          67       50
6 6 0          72       50
> logitmodcorrect2 <- glm(cbind(r,m-r) ~ temperature + pressure, family = binomial, data = shu
> summary(logitmodcorrect2)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.520195   3.486784   0.723   0.4698
temperature -0.098297   0.044890  -2.190   0.0285 *
pressure     0.008484   0.007677   1.105   0.2691
---

    Null deviance: 24.230  on 22  degrees of freedom
Residual deviance: 16.546  on 20  degrees of freedom
AIC: 36.106

Number of Fisher Scoring iterations: 5
> anova(logitmodcorrect,logitmodcorrect2)
Analysis of Deviance Table

Model 1: cbind(r, m - r) ~ temperature
Model 2: cbind(r, m - r) ~ temperature + pressure
  Resid. Df Resid. Dev Df Deviance
1        21     18.086
2        20     16.546  1   1.5407
```

family = bin, data = sh 2

$\hat{\beta}_2 / s.e(\hat{\beta})$

$2\{l(\hat{\beta}_a) - l(\hat{\beta}_R)\}$

... inference

- Model A: $\text{logit}(p_i) = \beta_0 + \beta_1 \texttt{temp}_i + \beta_2 \texttt{pressure}_i$
- Model B: $\text{logit}(p_i) = \beta_0 + \beta_1 \texttt{temp}_i$
- nested: Model B is obtained by setting $\beta_2 = 0$
- Under Model B, the change in deviance is (approximately) an observation from a χ^2_1
- $\Pr(\chi^2_1 \geq 1.5407) = 0.22$
  this is a p-value for testing $H_0 : \beta_2 = 0$
- so is $1 - \Phi\{\frac{\hat{\beta}_2}{\widehat{\text{s.e.}}(\hat{\beta}_2)}\} = 1 - \Phi(1.105) = 0.27$

ELM p.30

... inference

- confidence intervals for β_1

? logitmodcorrect(2) ?

- based on normal approximation: $\hat{\beta}_1 \pm \widehat{\text{s.e.}}(\hat{\beta}_1) * 1.96$
- `(-0.208, -0.023)`

- based on profile log-likelihood $\ell_p(\beta_1)$, details to follow
- `confint(logitmodcorrect):`
  `( -0.2122262, -0.0244701 )`

ELM p. 31

... inference

- confidence intervals for β_1
- based on normal approximation: $\hat{\beta}_1 \pm \widehat{\text{s.e.}}(\hat{\beta}_1) * 1.96$
- `(-0.208, -0.023)`
- based on profile log-likelihood — $\ell_p(\beta_1)$, details to follow
- `confint(logitmodcorrect)`:
  `( -0.2122262, -0.0244701 )`

ELM p. 31

knitr (Sweave) || markdown

Rstudio

Special to the binomial

and Poisson

- likelihood ratio test for logistic model
  $p_i = p_i(\beta) = \text{expit}(x_i^T\beta), \quad \hat{p}_i = p_i(\hat{\beta})$
- this model is nested in the saturated model $\tilde{p}_i = y_i/n_i$
- residual deviance compares fitted model to saturated model
- under the fitted model, approximately distributed as χ^2_{n-q} if each n_i "large"

nested

$[\hat{p}_i = y_i/n_i]$

fitted $p_i(\hat{\beta})$ to $\tilde{p} = y_i/n_i$

$= \hat{p}_i$

$2\{l(\tilde{p}) - l(\hat{p})\} \sim \chi^2_{n-q}$ n_i's $\to \infty$

```
> summary(logitmodcorrect)

...
    Null deviance: 24.230  on 22  degrees of freedom
Residual deviance: 18.086  on 21  degrees of freedom
AIC: 35.647
Number of Fisher Scoring iterations: 5
```

$p(\chi^2_{21} \geq 18.086) = .75$?

Actually, null model ($\beta_1 = 0$) also fits: `pchisq(24.23,22, lower.tail = F) = 0.33`, but improvement is statistically significant

Logistic regression

- read §2.4 for one motivation of logistic regression model
- read §2.5 (and AS I) for interpretation of parameters in terms of log odds
- see Example `mdl` in §2.5 for logistic regression with qualitative covariates
- what is the algebraic form of the model? how are the dummy covariates coded?
- in other words, what is x_i^{T}?

$$\text{logit}(p_i) = \beta_0 + \beta_1 \_\_ + \beta_2 \_\_ + \ldots$$

“What are the principles of applied statistics?” CD Ch. 1

- “formulation and clarification of focused research questions of subject-matter importance
- design of individual investigations and sequences of investigations that produce secure answers and open up new possibilities
- production of effective and reliable measurement procedures
- development of simple, and where appropriate, not-so-simple methods of analysis, with suitable software, that address the primary research questions, often through a skilful choice of statistical model, and give some assessment of uncertainty
- effective presentation of conclusions
- structuring of analyses to facilitate their interpretation in subject matter terms and their relationship to the knowledge base of the field.”

“What are the principles of applied statistics?” CD Ch. 1

- “formulation and clarification of focused research questions of subject-matter importance
- design of individual investigations and sequences of investigations that produce secure answers and open up new possibilities
- production of effective and reliable measurement procedures
- development of simple, and where appropriate, not-so-simple methods of analysis, with suitable software, that address the primary research questions, often through a skilful choice of statistical model, and give some assessment of uncertainty
- effective presentation of conclusions
- structuring of analyses to facilitate their interpretation in subject matter terms and their relationship to the knowledge base of the field.”

“What are the principles of applied statistics?” CD Ch. 1

- “formulation and clarification of focused research questions of subject-matter importance
- design of individual investigations and sequences of investigations that produce secure answers and open up new possibilities
- production of effective and reliable measurement procedures
- development of simple, and where appropriate, not-so-simple methods of analysis, with suitable software, that address the primary research questions, often through a skilful choice of statistical model, and give some assessment of uncertainty
- effective presentation of conclusions
- structuring of analyses to facilitate their interpretation in subject matter terms and their relationship to the knowledge base of the field.”

“What are the principles of applied statistics?” CD Ch. 1

- “formulation and clarification of focused research questions of subject-matter importance
- design of individual investigations and sequences of investigations that produce secure answers and open up new possibilities
- production of effective and reliable measurement procedures
- development of simple, and where appropriate, not-so-simple methods of analysis, with suitable software, that address the primary research questions, often through a skilful choice of statistical model, and give some assessment of uncertainty
- effective presentation of conclusions
- structuring of analyses to facilitate their interpretation in subject matter terms and their relationship to the knowledge base of the field.”

“What are the principles of applied statistics?” CD Ch. 1

- “formulation and clarification of focused research questions of subject-matter importance
- design of individual investigations and sequences of investigations that produce secure answers and open up new possibilities
- production of effective and reliable measurement procedures
- development of simple, and where appropriate, not-so-simple methods of analysis, with suitable software, that address the primary research questions, often through a skilful choice of statistical model, and give some assessment of uncertainty
- effective presentation of conclusions
- structuring of analyses to facilitate their interpretation in subject matter terms and their relationship to the knowledge base of the field.”