

The erythrocyte sedimentation rate (ESR) is the rate at which red blood cells (erythrocytes) settle out of suspension in blood plasma, when measured under standard conditions. If the ESR increases when the level of certain proteins in the blood plasma rise in association with conditions such as rheumatic diseases, chronic infections and malignant diseases, its determination might be useful in screening blood samples taken from people suspected of suffering from one of the conditions mentioned. A value of ESR less than 20mm/hr indicates a healthy individual.

The question of interest is whether there is any association between the probability of an ESR reading greater than 20mm/hr and the levels of the two plasma proteins. If not, then the determination of ESR would not be useful for diagnostic purposes. To assess whether the ESR is a useful diagnostic tool, use the dataframe `plasma` from library `HSAUR`.

- a) Fit a simple logistic regression model `m1` to predict `ESR > 20`, using `fibrinogen` as predictor.
- b) Interpret the regression coefficients
- c) Fit a multiple logistic regression model `m2` to predict `ESR > 20`, using `fibrinogen` and `globulin` as predictors.
- d) Use function `anova()` to test if the second model is a better fit.
- e) Use `m2` to predict `ESR > 20` for the full data set
- f) Create a bubbleplot to show the probabilities of `ESR > 20` for each individual in the dataset.

```

# plasma.r

str(plasma)
# 'data.frame': 32 obs. of 3 variables:
# $ fibrinogen: num 2.52 2.56 2.19 2.18 3.41 2.46 3.22 2.21 3.15 2.6 ...
# $ globulin : int 38 31 33 31 37 36 38 37 39 41 ...
# $ ESR : Factor w/ 2 levels "ESR < 20","ESR > 20": 1 1 1 1 1 1 1 1 1 1 ...

head(plasma)
# fibrinogen globulin ESR
#1 2.52 38 ESR < 20
#2 2.56 31 ESR < 20
#3 2.19 33 ESR < 20
#4 2.18 31 ESR < 20
#5 3.41 37 ESR < 20
#6 2.46 36 ESR < 20

table(plasma$ESR)
# ESR < 20 ESR > 20
# 26 6

# one predictor
# =====

m1 = glm(ESR~fibrinogen, plasma, family = binomial())
summary(m1)
# Coefficients:
# Estimate Std. Error z value Pr(>|z|)
#(Intercept) -6.8451 2.7703 -2.471 0.0135 *
#fibrinogen 1.8271 0.9009 2.028 0.0425 *

#(Dispersion parameter for binomial family taken to be 1)
# Null deviance: 30.885 on 31 degrees of freedom
#Residual deviance: 24.840 on 30 degrees of freedom
#AIC: 28.84

# interpret b1

# An increase of one unit in fibrinogen increases the log-odds ratio, in favour of ESR > 20
# by an estimated 1.827

# 95% confidence interval
confint(m1, parm = "fibrinogen")
# 2.5 % 97.5 %
# 0.3387619 3.9984921

# small dataset, low accuracy

```

```

# interpret exp(b1)

exp(coef(m1)["fibrinogen"])
# fibrinogen
# 6.215715

# confidence interval
exp(confint(m1, parm = "fibrinogen"))
# 2.5 % 97.5 %
# 1.403209 54.515884

# very large change (6.2157)
# in the odds ratio when fibrinogen increased by one unit

# two predictors
# =====

m2 = glm(ESR ~ fibrinogen + globulin, plasma, family = binomial())
summary(m2)

## Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
#(Intercept) -12.7921     5.7963  -2.207   0.0273 *
#fibrinogen    1.9104     0.9710   1.967   0.0491 *
#globulin      0.1558     0.1195   1.303   0.1925

#(Dispersion parameter for binomial family taken to be 1)
#   Null deviance: 30.885  on 31  degrees of freedom
#Residual deviance: 22.971  on 29  degrees of freedom
#AIC: 28.971

anova(m1, m2, test="Chisq")
#Analysis of Deviance Table
#Model 1: ESR ~ fibrinogen
#Model 2: ESR ~ fibrinogen + globulin
#  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
#1         30      24.840
#2         29      22.971  1    1.8692   0.1716

# p-value is not small, do not reject Ho: model 1 is the true model

# predict prob of ESR > 20
prob = predict(m2, type = "response")
d2 = data.frame(plasma, prob)

```

```
# plot
plot(globulin~fibrinogen,plasma,xlim = c(2,6),ylim = c(25, 55), pch = ".")
symbols(plasma$fibrinogen,plasma$globulin,circles=prob,add = T)
grid()
```

