# Impact of Moisture on Seed Germination

Parikshit Patil

2024-09-01

## Impact of Moisture on Seed Germination

```r
library(faraway)

data(seeds)
## creating a new predictor describing the box:
seeds$box <- factor(x=rep(1:8, c(6,6,6,6,6,6,6,6)),
levels=c("1","2","3","4","5","6","7","8"))
## removing one observation with missing data
(seeds[is.na(seeds$germ),])
```

```
##    germ moisture covered box
## 47   NA        9     yes   8
```

```r
seeds <- seeds[!is.na(seeds$germ),]
str(seeds)
```

```
## 'data.frame':    47 obs. of  4 variables:
##  $ germ    : num  22 41 66 82 79 0 25 46 72 73 ...
##  $ moisture: num  1 3 5 7 9 11 1 3 5 7 ...
##  $ covered : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ box     : Factor w/ 8 levels "1","2","3","4",..: 1 1 1 1 1 1 2 2 2 2 ...
```

a)

```r
#Fit binomial model with moisture and box as predictors
binomial_mod = glm(cbind(germ,100-germ) ~ moisture + box, family=binomial,seeds)
summary(binomial_mod)
```

```
##
## Call:
## glm(formula = cbind(germ, 100 - germ) ~ moisture + box, family = binomial,
##     data = seeds)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.593872   0.098319   6.040 1.54e-09 ***
## moisture    -0.110487   0.008813 -12.537  < 2e-16 ***
```

```
## box2          -0.041493   0.117609  -0.353    0.724
## box3          -0.041493   0.117609  -0.353    0.724
## box4           0.020724   0.117544   0.176    0.860
## box5          -0.090012   0.117697  -0.765    0.444
## box6          -0.062269   0.117643  -0.529    0.597
## box7          -0.069200   0.117655  -0.588    0.556
## box8           0.067297   0.123410   0.545    0.586
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1791.0  on 46  degrees of freedom
## Residual deviance: 1624.4  on 38  degrees of freedom
## AIC: 1832.1
##
## Number of Fisher Scoring iterations: 5
```

b)

```r
#Interpret using odds
1-exp(-0.1104873)
```

```
## [1] 0.1046023
```

```r
1-exp(0.0207244)
```

```
## [1] -0.02094064
```

Keeping all other predictors fixed, 1 unit increase in moisture reduces the odds of seed germination by 10%.

Keeping all other predictors fixed, 1 unit increase in box 4 relative to box 1 increases the odds of seed germination by 2%.

c)

The first hypothesis test we can do to assess the model's goodness of fit is the binomial deviance method. The other test is Pearson's chi-squared method.

Below is the binomial deviance method.

```r
#p-chi sqaured test
pchisq(deviance(binomial_mod),df.residual(binomial_mod),lower = FALSE)
```

```
## [1] 6.957246e-317
```

The null hypothesis states that the model is adequate, however since the p-value is extremely small («<0.5), we reject the null hypothesis indicating strong evidence for lack of fit.

d)

The other common causes for a deviance value to be larger than expected are:

2

- Wrong form of the model such as not including right predictors or transformations;
- Data being sparse;
- Presence of outliers.

e)

```
# sigma squared
sigma2 = sum(residuals(binomial_mod,type="pearson")^2)/(df.residual(binomial_mod))

sigma2
```

```
## [1] 35.71223
```

The dispersion parameter = 35.71223, indicating that there is a problem of over-dispersion.

f)

```
#F-test
drop1(binomial_mod,scale=sigma2,test="F")
```

```
## Warning in drop1.glm(binomial_mod, scale = sigma2, test = "F"): F test assumes
## 'quasibinomial' family
```

```
## Single term deletions
##
## Model:
## cbind(germ, 100 - germ) ~ moisture + box
##
## scale:  35.71223
##
##           Df Deviance    AIC F value  Pr(>F)
## <none>          1624.4 1832.2
## moisture   1    1786.5 1834.7  3.7918 0.05892 .
## box        7    1627.0 1818.2  0.0086 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(binomial_mod,dispersion=sigma2)
```

```
##
## Call:
## glm(formula = cbind(germ, 100 - germ) ~ moisture + box, family = binomial,
##     data = seeds)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.59387    0.58755   1.011   0.3121
## moisture    -0.11049    0.05266  -2.098   0.0359 *
## box2        -0.04149    0.70283  -0.059   0.9529
## box3        -0.04149    0.70283  -0.059   0.9529
## box4         0.02072    0.70244   0.030   0.9765
```

```
## box5          -0.09001     0.70335   -0.128    0.8982
## box6          -0.06227     0.70303   -0.089    0.9294
## box7          -0.06920     0.70310   -0.098    0.9216
## box8           0.06730     0.73750    0.091    0.9273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 35.71223)
##
##     Null deviance: 1791.0  on 46  degrees of freedom
## Residual deviance: 1624.4  on 38  degrees of freedom
## AIC: 1832.1
##
## Number of Fisher Scoring iterations: 5
```

g)

```
summary(binomial_mod)
```

```
##
## Call:
## glm(formula = cbind(germ, 100 - germ) ~ moisture + box, family = binomial,
##     data = seeds)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.593872   0.098319   6.040 1.54e-09 ***
## moisture    -0.110487   0.008813 -12.537  < 2e-16 ***
## box2        -0.041493   0.117609  -0.353    0.724
## box3        -0.041493   0.117609  -0.353    0.724
## box4         0.020724   0.117544   0.176    0.860
## box5        -0.090012   0.117697  -0.765    0.444
## box6        -0.062269   0.117643  -0.529    0.597
## box7        -0.069200   0.117655  -0.588    0.556
## box8         0.067297   0.123410   0.545    0.586
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1791.0  on 46  degrees of freedom
## Residual deviance: 1624.4  on 38  degrees of freedom
## AIC: 1832.1
##
## Number of Fisher Scoring iterations: 5
```