MATH 523 - ASSIGNMENT 1: SOLUTIONS

1 (a) For the model

$$\mathbb{E}[Y|X=x] = \beta_0 + \beta_1 x$$

the usual least-squares formulae apply for the estimates, that is

$$\widehat{\beta} = \begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\boldsymbol{y}$$

with (under the assumption of correct model specification) the usual properties of *unbiasedness* $\mathbb{E}[\widehat{\beta}] = \beta$, and so that

$$\text{Var}[\widehat{\beta}] = \sigma^2(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}$$

Now, using the centered covariate $z_i = x_i - \overline{x}$, recalling that the x values are fixed constants in the analysis, the model may be re-written

$$\mathbb{E}[Y|Z=z] = \beta_0^* + \beta_1^* z = \beta_0^* + \beta_1^*(x-\overline{x}) = (\beta_0^* - \beta_1^*\overline{x}) + \beta_1^* x$$

Thus we deduce that for the estimates resulting from the centered model,

$$\widehat{\beta}_0^* = \widehat{\beta}_0 + \widehat{\beta}_1\overline{x} \qquad\qquad \widehat{\beta}_1^* = \widehat{\beta}_1$$

that is, the estimated slope is the same in the two analysis. However, note that if $\mathbf{z} = (z_1, \ldots, z_n)^\mathsf{T}$

$$\mathbf{1}^\mathsf{T}\mathbf{z} = \mathbf{1}^\mathsf{T}(\mathbf{x} - \overline{x}\mathbf{1}) = \sum_{i=1}^n (x_i - \overline{x}) = 0$$

that is, the columns of the $n \times 2$ matrix $\mathbf{Z} = [\mathbf{1}\ \mathbf{z}]$ are orthogonal. Hence we have that

$$\text{Var}[\widehat{\beta}^\star] = \sigma^2(\mathbf{Z}^\mathsf{T}\mathbf{Z})^{-1} = \sigma^2\mathbf{I}_n$$

and the estimators are **uncorrelated**.

4 MARKS

(b) For the model

$$\mathbb{E}[Y|X=x] = \beta_0 + \beta_1 x + \beta_2 x^2$$

the same general estimation results apply. However, for the centered covariate, we have

$$\begin{aligned} \mathbb{E}[Y|Z=z] &= \beta_0^\dagger + \beta_1^\dagger z + \beta_2^\dagger z^2 \\ &= \beta_0^\dagger + \beta_1^\dagger(x-\overline{x}) + \beta_2^\dagger(x-\overline{x})^2 \\ &= (\beta_0^\dagger - \beta_1^\dagger\overline{x} + \beta_2^\dagger(\overline{x})^2) + (\beta_1^\dagger - 2\beta_2^\dagger\overline{x})x + \beta_2^\dagger x^2 \end{aligned}$$

so again we have an identical model fitting procedure available, and there is a 1-1 mapping $(\beta_0, \beta_1, \beta_2) \longleftrightarrow (\beta_0^\dagger, \beta_1^\dagger, \beta_2^\dagger)$. Again, we have for the new design matrix $\mathbf{Z} = [\mathbf{1}\ \mathbf{z}_1\ \mathbf{z}_2]$, where

$$\mathbf{z}_1 = (z_1, \ldots, z_n)^\mathsf{T} \qquad\qquad \mathbf{z}_2 = (z_1^2, \ldots, z_n^2)^\mathsf{T}$$

that $\mathbf{1}^\mathsf{T}\mathbf{z}_1 = 0$, but that

$$\mathbf{1}^\mathsf{T}\mathbf{z}_2 = \sum_{i=1}^n z_i^2 = \sum_{i=1}^n (x_i - \overline{x})^2 = S_2 > 0 \qquad\qquad \mathbf{z}_1^\mathsf{T}\mathbf{z}_2 = \sum_{i=1}^n z_i^3 = \sum_{i=1}^n (x_i - \overline{x})^3 = S_3 > 0$$

in general. Hence the form of the variance for the estimators is

$$\mathbf{Z}^{\mathsf{T}}\mathbf{Z} = \begin{bmatrix} n & 0 & S_2 \\ 0 & S_2 & S_3 \\ S_2 & S_3 & S_4 \end{bmatrix}$$

and therefore there is no general orthogonality; the estimators of $(\beta_1^\dagger, \beta_2^\dagger)$ are in general **correlated**.

4 MARKS

(c) For the model

$$\mathbb{E}[Y|X = x] = \beta_0 + \beta_2 x^2$$

we can again use least-squares. However, for the centered covariate

$$\begin{aligned} \mathbb{E}[Y|Z = z] &= \beta_0^\ddagger + \beta_2^\ddagger z^2 \\ &= \beta_0^\ddagger + \beta_2^\ddagger (x - \overline{x})^2 \\ &= (\beta_0^\ddagger + \beta_2^\ddagger (\overline{x})^2) + (-2\beta_2^\ddagger \overline{x})x + \beta_2^\ddagger x^2 \end{aligned}$$

which includes a linear term in x, unlike the original model. Hence this model **does not admit centering** of the covariate, that is, the model is specific to the original measurement scale of the covariate.

This is a powerful reason not to fit polynomial models that include higher-order terms but omit lower-order ones.

4 MARKS

2 The `R` code for fitting the three basic models of interest is included below:

```
1 data.set<-read.csv("Accidents.csv",header=T)
2
3 #Null model
4 mod0<-glm(Y~1,family=poisson,data=data.set)
5
6 #Covariate only
7 mod1<-glm(Y~X1,family=poisson,data=data.set)
8
9 #Factor only
10 mod2<-glm(Y~X2,family=poisson,data=data.set)
11
12 #Additive model
13 mod3<-glm(Y~X1+X2,family=poisson,data=data.set)
14
15 #Interaction model
16 mod4<-glm(Y~X1+X2+X1:X2,family=poisson,data=data.set)
```

This fits the five models

$$\begin{aligned} \texttt{mod0} &: 1 \\ \texttt{mod1} &: 1 + X_1 \\ \texttt{mod2} &: 1 + X_2 \\ \texttt{mod3} &: 1 + X_1 + X_2 \\ \texttt{mod4} &: 1 + X_1 + X_2 + X_1.X_2 \end{aligned}$$

All of the models are nested inside `mod4`. Results for the models are included below (output edited for space).

```
> summary(mod0)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.73092    0.06075   28.50   <2e-16 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 414.04  on 47  degrees of freedom
Residual deviance: 414.04  on 47  degrees of freedom
AIC: 536.99

Number of Fisher Scoring iterations: 6

##############################################################
> summary(mod1)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.17582    0.09534   12.33   <2e-16 ***
X1           0.68843    0.04279   16.09   <2e-16 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 414.041  on 47  degrees of freedom
Residual deviance:  74.198  on 46  degrees of freedom
AIC: 199.14

Number of Fisher Scoring iterations: 5

##############################################################
> summary(mod2)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.0380     0.1443   7.192 6.39e-13 ***
X2            0.9352     0.1591   5.878 4.16e-09 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 414.04  on 47  degrees of freedom
Residual deviance: 372.51  on 46  degrees of freedom
AIC: 497.46

Number of Fisher Scoring iterations: 5

##############################################################
> summary(mod3)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.67817    0.15464   4.385 1.16e-05 ***
X1           0.65816    0.04195  15.689  < 2e-16 ***
X2           0.71989    0.16038   4.489 7.17e-06 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
```

```
63
64 (Dispersion parameter for poisson family taken to be 1)
65
66     Null deviance: 414.041  on 47  degrees of freedom
67 Residual deviance:  51.025  on 45  degrees of freedom
68 AIC: 177.97
69
70 Number of Fisher Scoring iterations: 5
71
72 ##################################################################
73
74 > summary(mod4)
75
76 Coefficients:
77             Estimate Std. Error z value Pr(>|z|)
78 (Intercept)   0.5356     0.2267   2.363 0.018133 *
79 X1            0.7596     0.1177   6.455 1.08e-10 ***
80 X2            0.8908     0.2493   3.573 0.000353 ***
81 X1:X2        -0.1173     0.1259  -0.931 0.351714
82 ---
83 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
84
85 (Dispersion parameter for poisson family taken to be 1)
86
87     Null deviance: 414.041  on 47  degrees of freedom
88 Residual deviance:  50.127  on 44  degrees of freedom
89 AIC: 179.07
90
91 Number of Fisher Scoring iterations: 5
92
93 ##################################################################
```

From the output, we can see that the interaction between factor and covariate does not seem to be significant in `mod4` (line 81, p-value in the Wald test of 0.351714). As all other coefficients in this model appear to be significant, we can conclude that `mod3` seems to be the most suitable model.

Note that here `X2` is a binary factor predictor with numeric levels 0 and 1, so in fact can be fitted as a numeric continuous covariate - the resulting estimates are identical. If `X2` was not coded in this way, or had more levels, then `X2` would have to be fitted as a factor predictor, say

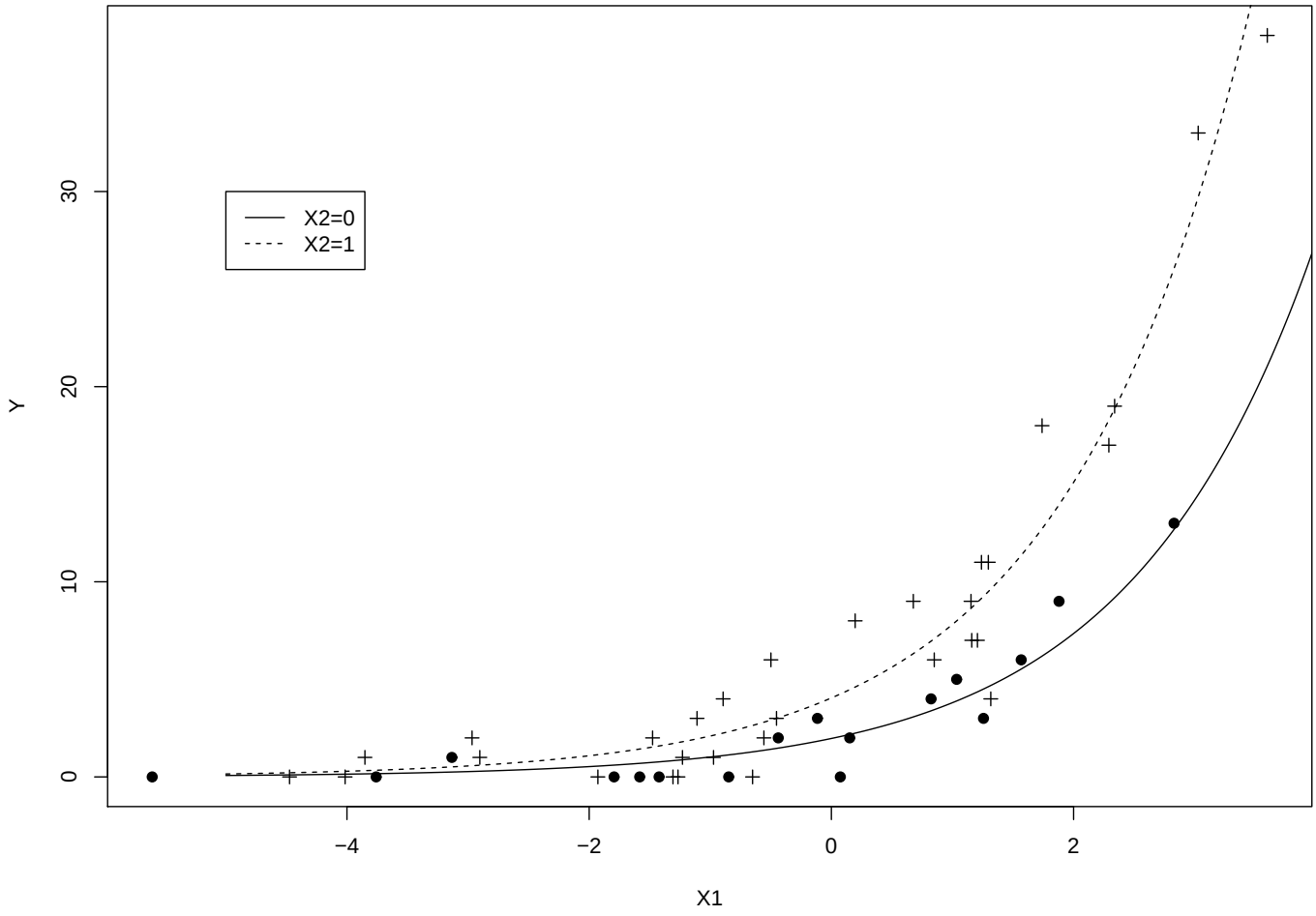
```
mod2<-glm(Y~factor(X2),family=poisson,data=data.set)
```

8 MARKS

The `predict` function can be used to obtain a fitted response function for these data:

```
1 plot(X1,Y,type="n")
2 points(X1[X2==0],Y[X2==0],pch=19)
3 points(X1[X2==1],Y[X2==1],pch=3)
4
5 Y1<-predict(mod3,newdata=new.data1,type="response")
6 Y2<-predict(mod3,newdata=new.data2,type="response")
7 lines(x1v,Y1)
8 lines(x1v,Y2,lty=2)
9 legend(-5,30,c("X2=0","X2=1"),lty=c(1,2))
```

Here the `predict` function acts on the `glm` object by predicting at the new data `new.data1` and `new.data2`. The `type="response"` argument makes the prediction on the response scale, whereas the default is to predict on the linear predictor scale. The resulting plot is below, and demonstrates a qualitatively adequate fit.


X2=0
X2=1
Y
0
10
20
30
−4
−2
0
2
X1