Delft University of Technology
Faculty EWI

Mekelweg 4
2628 CD Delft

Test exam wi3624 - wi4624 **June 2013**

It is allowed to use a pocket calculator and one self-made A4-sheet that may contain notes on both sides. It is *not* allowed to use a book or other notes during the exam.

1. In the appendix of this question a dataset is listed of measurements of the percentage of inhabitants of an American city or state that have difficulty speaking or writing English.

   (a) Compute the median and the lower- and upper hinges for the dataset.

   (b) Instead of a logit transformation (or another transformation meant for proportions) one decides to use one of the power transformations, for which the median and the lower- and upper hinges are given in the table in the appendix of this assignment. Decide on the basis of this information which of these transformations is most suitable to correct the skewness of the original dataset.

   (c) The dataset can be split into two groups: percentages for major cities and for states, or state-remainders. The boxplot in the appendix shows a difference in spread for both groups. The median and the lower- and upper hinges are given in the table below:

Region	H_L	Median	H_U
city	1.35	2.65	5.90
state	0.40	0.70	1.25

   Decide on the basis of this information, which power transformation is most suitable to correct the non-constant spread.

2. Suppose

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix} \right).$$

   (a) Determine $a \in \mathbb{R}$ such that $X_1 - aX_2$ is independent of X_2.

   (b) Use the fact that $X_1 = aX_2 + (X_1 - aX_2)$, to calculate the conditional expectation $\mathbb{E}(X_1 \mid X_2 = 3)$ and the conditional variance $\mathrm{Var}(X_1 \mid X_2 = 3)$.

3. In the appendix to this question the output is listed from regressing the logarithm of infant-mortality rate per 1000 births on the variable income and the factor region, which is coded by three dummy-variables as follows:

Region	D_1	D_2	D_3
Africa	1	0	0
Americas	0	1	0
Asia	0	0	1
Europe	0	0	0

Answer the following questions on the basis of this output.

(a) Specify the regression equation for each region separately. Furthermore, give the expected mortality rate (not the log-mortality rate!) for a country in Asia with income $= 10$.

(b) Investigate whether the interaction between income and region is of significant influence on log-infant mortality rate by testing the appropriate null hypothesis at significance level 5%.

(c) Investigate whether region is of significant influence on log-infant mortality rate in the common-slope model by testing the appropriate null hypothesis at significance level 5%.

4. We consider an outcome variable Y with outcomes in the set $\{1, 2, 3\}$. We have two explanatory variables: a regressor x and a factor `group` with two levels, "Yes" or "No". We used the following `R`-code to fit two different models to the data:

```
fit=glm(Y~x+group,family="multinomial")
summary(fit)

Y1=1*(Y==1)

fit1=glm(Y1~x+group,family="binomial")
summary(fit1)

Y2=1*(Y[Y!=1]==2)
x2=x[Y!=1]
group2=group[Y!=1]

fit2=glm(Y2~x2+group2,family="binomial")
summary(fit2)
```

The (relevant) output of this code can be found in the appendix to this question.

(a) Describe the two models that are being used here to describe the data as precisely as possible. Furthermore, use the output the determine which model you would prefer.

(b) In the appendix to this question you can also find the relevant estimated covariance matrices for `fit1` and `fit2`. Call the corresponding model Model 2. Suppose $x = 2$ and `group`="No". Give an estimate of the probability that $Y = 2$, using Model 2. Also give an estimate of this probability using Model 1 (the other model).

(c) It is also of interest to investigate the variable x. The values of x can be modeled as a sample from a Gamma distribution with two parameters $k, \theta > 0$:

$$f(x; k, \theta) = \frac{1}{\Gamma(k)}\, \theta^{-k} x^{k-1} e^{-x/\theta} \quad (x > 0)$$

Here,

$$\Gamma(k) = \int_0^\infty e^{-t} t^{k-1}\, dt.$$

Determine whether the maximum likelihood estimators of k and θ are asymptotically independent.

5. (a) In the appendix to this question several diagnostics (in alphabetical order) are displayed to identify unusual and influential observations in a linear regression of a response Y on explanatory variables X_1 and X_2. In each plot horizontal lines are added which refer either to cut-off values or centered values for the different diagnostics, and some of the points are labeled by their observation number. On the basis of these plots, determine

- which observations are leverage point and regression outlier,
- which observations are influential on the value and/or the standard error of coefficients, and indicate in what way they influence the value of the coefficient.

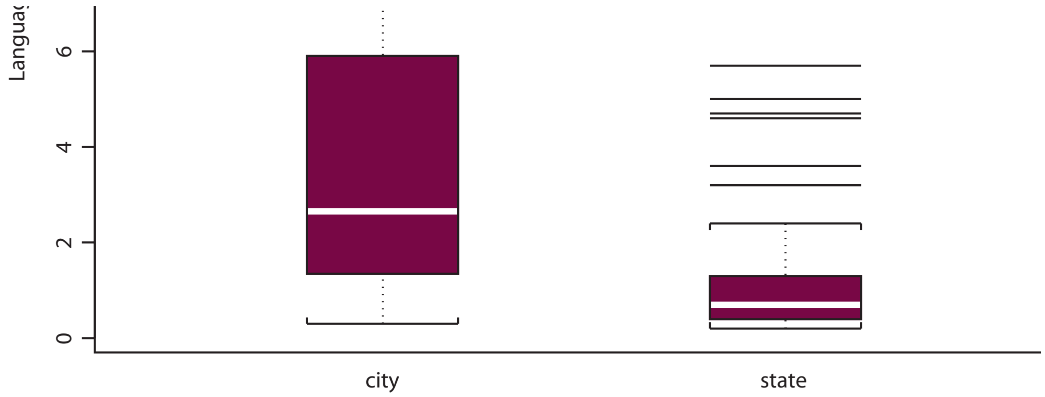
Appendix to question 1.[1] Percentages of inhabitants having difficulty speaking or writing English:

```
0.2  0.2  0.2  0.2  0.2   0.2  0.2  0.2  0.3  0.3
0.3  0.3  0.4  0.4  0.4   0.5  0.5  0.5  0.5  0.5
0.5  0.5  0.5  0.5  0.5   0.6  0.7  0.7  0.7  0.7
0.8  0.8  0.8  0.9  1.0   1.0  1.0  1.0  1.1  1.1
1.2  1.3  1.6  1.6  1.6   1.6  1.7  2.1  2.2  2.4
3.1  3.2  3.2  3.6  3.6   4.2  4.4  4.6  4.7  5.0
5.1  5.7  6.7  8.9  9.2  12.7
```

Hinges and median for several power transformations of the above dataset:

Transformation	H_U	Median	H_L
$\sqrt{X}$	1.55	0.92	0.71
$\log X$	0.38	-0.07	-0.30
$-1/\sqrt{X}$	-0.65	-1.09	-1.41
$-1/X$	-0.42	-1.18	-2.00

Boxplot of percentages for major cities and states:



[1] Data taken from Ericksen, E. P., Kadane, J. B. and Tukey, J. W. (1989) Adjusting the 1980 Census of Population and Housing. *Journal of the American Statistical Association* **84**, 927–944.

Appendix to question 3.[2]

```
Coefficients:
             Value Std. Error t value Pr(>|t|)
(Intercept)  7.027  1.729      4.064   0.000
     income -0.532  0.219     -2.426   0.017
         d1 -2.088  1.842     -1.134   0.260
         d2 -0.522  1.997     -0.261   0.794
         d3 -0.825  1.821     -0.453   0.652
  income:d1  0.520  0.252      2.069   0.041
  income:d2  0.123  0.268      0.457   0.648
  income:d3  0.141  0.240      0.586   0.560

Analysis of Variance Table

Terms added sequentially (first to last)
          Df Sum of Sq Mean Sq F Value   Pr(F)
   income  1    47.083  47.083  132.08 0.00000
       d1  1     7.833   7.833   21.97 0.00001
       d2  1     1.332   1.332    3.74 0.05631
       d3  1     1.541   1.541    4.32 0.04038
income:d1  1     2.705   2.705    7.59 0.00706
income:d2  1     0.000   0.000    0.00 0.97515
income:d3  1     0.122   0.122    0.34 0.55951
Residuals 93    33.152   0.356
```

[2] Original data from Leinhardt, S. and Wasserman, S. S. (1979) Exploratory data analysis: An introduction to selected methods. In Schuessler, K. (Ed.) *Sociological Methodology 1979* Jossey-Bass.

Appendix to question 4.

```
>summary(fit)

Coefficients:
                Value Std. Error t value
(Intercept):1  2.7814    0.35403  7.8566
(Intercept):2  2.1414    0.40001  5.3534
x:1           -2.5838    0.33517 -7.7088
x:2           -2.1707    0.41823 -5.1901
groupYes:1    -1.0245    0.29088 -3.5219
groupYes:2    -2.1380    0.34859 -6.1334

Number of linear predictors:  2

Residual Deviance: 693.6994 on 794 degrees of freedom



>summary(fit1)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.03933    0.23623   4.400 1.08e-05 ***
x           -1.82908    0.28295  -6.464 1.02e-10 ***
groupYes    -0.02609    0.22116  -0.118    0.906

Null deviance: 552.27  on 399  degrees of freedom
Residual deviance: 498.09  on 397  degrees of freedom



>summary(fit2)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.9292     0.4067   4.743 2.10e-06 ***
x2           -2.0210     0.4431  -4.561 5.10e-06 ***
group2Yes    -1.9566     0.3507  -5.579 2.42e-08 ***

Null deviance: 274.16  on 214  degrees of freedom
Residual deviance: 202.86  on 212  degrees of freedom
```

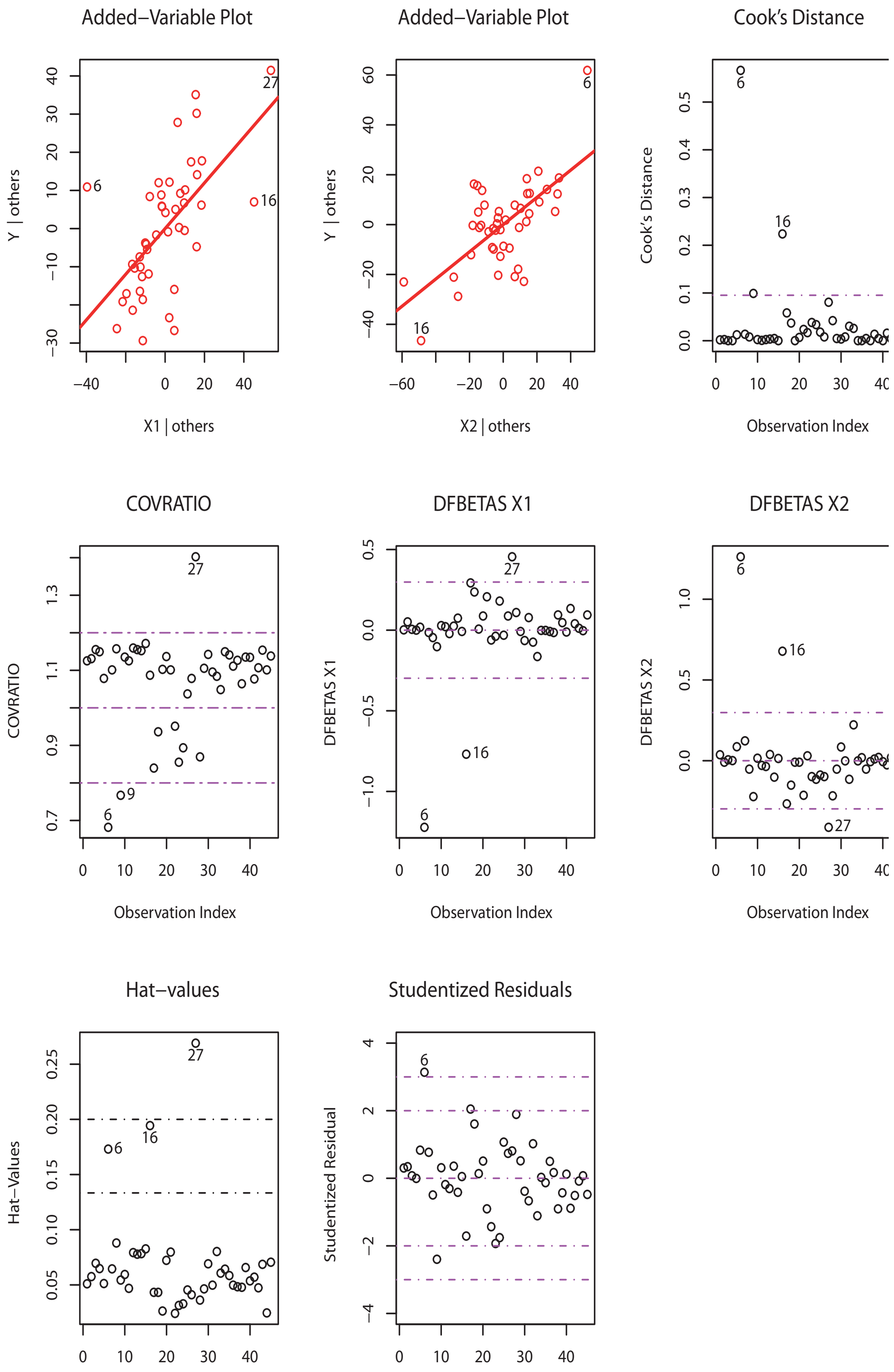
The two covariance matrices of `fit1` and `fit2`:

$$\texttt{vcov(fit1)} = \begin{pmatrix} 0.05580426 & -0.04537966 & -0.02766459 \\ -0.04537966 & 0.08006287 & -0.00426687 \\ -0.02766459 & -0.00426687 & 0.04891262 \end{pmatrix}$$

and

$$\texttt{vcov(fit2)} = \begin{pmatrix} 0.16542103 & -0.13960254 & -0.07656707 \\ -0.13960254 & 0.19636479 & 0.01462065 \\ -0.07656707 & 0.01462065 & 0.12301372 \end{pmatrix}.$$

Appendix to question 5.



Added−Variable Plot

Added−Variable Plot

Cook’s Distance

COVRATIO

DFBETAS X1

DFBETAS X2

Hat−values

Studentized Residuals

	$\longleftarrow \nu_1 \longrightarrow$								
ν_2	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97
110	3.93	3.08	2.69	2.45	2.30	2.18	2.09	2.02	1.97
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96

Tabel 1: Rechter kritieke waarden $F_\alpha(\nu_1, \nu_2)$ van de $F(\nu_1, \nu_2)$ verdeling bij $\alpha = 0.05$.

$m \backslash p$	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
∞	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Tabel 2: Right critical values $t_{m,p}$ of the t-distribution with m degrees of freedom corresponding to right tail probability p: $\mathrm{P}\left(T_m \geq t_{m,p}\right) = p$. The last row in the table are right critical values of the $N(0,1)$ distribution: $t_{\infty,p} = z_p$

Full solutions:

1a We have $n = 66$ so that depth$(M) = \frac{n+1}{2} = 33\frac{1}{2}$. This implies that

$$M = \frac{x_{(33)} + x_{(34)}}{2} = \frac{0.8 + 0.9}{2} = 0.85$$

Furthermore, depth$(H) = \frac{\lfloor \text{depth}(M) \rfloor + 1}{2} = \frac{33+1}{2} = 17$. Hence,

$$\begin{aligned} H_L &= x_{(17)} = 0.5 \\ H_U &= x_{(50)} = 2.4 \end{aligned}$$

1b Compute

$$\frac{H_U - M}{M - H_L}$$

for each transformation. The one which has a ratio close to 1 is most suitable. We find

$$\begin{aligned} \sqrt{X}: &\quad \frac{1.55 - 0.90}{0.90 - 0.71} = 3.00 \\ \log X: &\quad \frac{0.36 - (-0.07)}{-0.07 - (-0.30)} = 1.96 \\ -1/\sqrt{X}: &\quad \frac{-0.65 - (-1.09)}{-1.09 - (-1.41)} = 1.38 \\ -1/X: &\quad \frac{-0.44 - (-1.18)}{-1.18 - (-2.00)} = 0.93 \end{aligned}$$

It follows that $1/X$ is the most suitable transformation.

1c Plot log hinge spread against log median for both groups and determine the least squares regression line through the scatterplot. The slope b of this line suggests a suitable power transformation with $p = 1 - b$. First compute log hinge spread and log median for both groups. We find

	city	state
$\log M$	0.97	-0.36
$\log(H_U - H_L)$	1.54	-0.16

When we plot log hinge spread against log median for both groups, the least squares regression line must go through both points. Its slope is therefore

$$b = \frac{1.54 - (-0.16)}{0.97 - (-0.36)} = 1.28$$

This suggest a power transformation with power $p = 1 - b = -0.28$.

2a Since $(X_1 - aX_2, X_2)$ is a normally distributed vector, it is enough to choose a such that $\text{Cov}(X_1 - aX_2, X_2) = 0$. Since

$$\begin{aligned} \text{Cov}(X_1 - aX_2, X_2) &= \text{Cov}(X_1, X_2) - a\text{Cov}(X_2, X_2) \\ &= 1 - a, \end{aligned}$$

we find $a = 1$.

2b Define $Z = X_1 - X_2$. We have seen that Z is independent of X_2, so conditioning on $X_2 = 3$ has no effect on Z. Since $X_1 = X_2 + Z$, we get

$$\mathbb{E}(X_1 \,|\, X_2 = 3) = \mathbb{E}(X_2 + Z \,|\, X_2 = 3) = 3 + \mathbb{E}(Z) = 4,$$

and

$$\mathrm{Var}(X_1 \,|\, X_2 = 3) = \mathrm{Var}(X_2 + Z \,|\, X_2 = 3) = \mathrm{Var}(Z) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) - 2\mathrm{Cov}(X_1, X_2) = 3.$$

3a The dummy-variable regression model is

$$Y_i = \alpha + \beta X_i + \gamma_1 D_{i1} + \gamma_2 D_{i2} + \gamma_3 D_{i3} + \delta_1 X_i D_{i1} + \delta_2 X_i D_{i2} + \delta_3 X_i D_{i3} + \varepsilon_i$$

with Y is log infant mortality rate and X is income. This leads to the following regression equation for Africa

$$\begin{aligned} Y_i &= A + BX_i + C_1 + \Delta_1 X_i \\ &= (A + C_1) + (B + \Delta_1)X_i \\ &= (7.027 - 2.088) + (-0.532 + 0.520) \times X_i \\ &= 4.939 - 0.012 \times X_i \end{aligned}$$

Similarly

$$\begin{array}{ll} \text{Americas:} & Y_i = (7.027 - 0.522) + (-0.532 + 0.123)X_i = 6.505 - 0.409 \times X_i \\ \text{Asia:} & Y_i = (7.027 - 0.825) + (-0.532 + 0.141)X_i = 6.202 - 0.391 \times X_i \\ \text{Europe:} & Y_i = 7.027 - 0.532 \times X_i \end{array}$$

For the country in Asia with income = 10 we get $Y = 6.202 - 0.391 \times 10 = 2.292$, so the expected mortality rate equals $e^2.292 = 9.9$ per 1000 children.

3b Within the full model

$$Y_i = \alpha + \beta X_i + \gamma_1 D_{i1} + \gamma_2 D_{i2} + \gamma_3 D_{i3} + \delta_1 X_i D_{i1} + \delta_2 X_i D_{i2} + \delta_3 X_i D_{i3} + \varepsilon_i$$

we test the null hypothesis $H_0 : \delta_1 = \delta_2 = \delta_3 = 0$. This means we compare the full model with the model under the null hypothesis

$$Y_i = \alpha + \beta X_i + \gamma_1 D_{i1} + \gamma_2 D_{i2} + \gamma_3 D_{i3} + \varepsilon_i$$

The regression sum of squares in the full model is

$$\mathrm{RegSS}_1 = 47.083 + \cdots + 0.122 = 60.616.$$

Because the interaction terms have been added after the main effects the regression sum of squares in the null model can also be read from the output:

$$\mathrm{RegSS}_0 = 47.083 + \cdots + 1.541 = 57.789.$$

Furthermore, the difference in number of parameters is $q = 3$, so that the F-test statistic becomes

$$F_0 = \frac{(\mathrm{RegSS}_1 - \mathrm{RegSS}_0)/q}{RSS_1/(n - rc)} = \frac{2.827/3}{0.356} = 2.647.$$

The critical value is $F_{0.05}(3, 93) \approx F_{0.05}(3, 90) = 2.71$. This means we do not reject the null hypotheses at significance level 5% and conclude that the interaction between income and region is not of significant influence on log-infant mortality rate.

3c Within the the common-slope model

$$Y_i = \alpha + \beta X_i + \gamma_1 D_{i1} + \gamma_2 D_{i2} + \gamma_3 D_{i3} + \varepsilon_i$$

we test the null hypothesis $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = 0$. This means that we compare the common-slope model with the model under the null hypothesis

$$Y_i = \alpha + \beta X_i + \varepsilon_i.$$

The regression sum of squares in the common-slope model is (see also part b)

$$\mathrm{RegSS}_1 = 47.083 + \cdots + 1.541 = 57.789.$$

Since income has been added to the model first, the regression sum of squares in the null model can also be read from the output

$$\mathrm{RegSS}_0 = 47.083.$$

The residual sum of squares in the common-slope model is equal to the squared error that remains in the common-slope model:

$$\mathrm{RSS}_1 = 33.152 + 2.705 + 0.000 + 0.122 = 35.979$$

and has $93 + 1 + 1 + 1 = 96$ degrees of freedom. Furthermore, the difference in number of parameters between the common-slope model and the null model is $q = 3$, so that the F-test statistic becomes

$$F_0 = \frac{(\mathrm{RegSS}_1 - \mathrm{RegSS}_0)/q}{RSS_1/96} = \frac{10.706/3}{35.979/96} = 9.522.$$

The critical value is $F_{0.05}(3, 96) \approx F_{0.05}(3, 100) = 2.70$. This means that we reject the null hypothesis and conclude that region is of significant influence on log-infant mortality rate in the common-slope model.

4a The first model is a polytomous model for Y: define $D_i = 1$ if `group` equals “Yes” for outcome i. Then the model states that

$$\begin{aligned}
\mathbb{P}(Y_i = 1) &= \frac{\exp(\mu_1 + \beta_1 x_i + \alpha_1 D_i)}{1 + \exp(\mu_1 + \beta_1 x_i + \alpha_1 D_i) + \exp(\mu_2 + \beta_2 x_i + \alpha_2 D_i)} \\
\mathbb{P}(Y_i = 2) &= \frac{\exp(\mu_2 + \beta_2 x_i + \alpha_2 D_i)}{1 + \exp(\mu_1 + \beta_1 x_i + \alpha_1 D_i) + \exp(\mu_2 + \beta_2 x_i + \alpha_2 D_i)} \\
\mathbb{P}(Y_i = 3) &= \frac{1}{1 + \exp(\mu_1 + \beta_1 x_i + \alpha_1 D_i) + \exp(\mu_2 + \beta_2 x_i + \alpha_2 D_i)}.
\end{aligned}$$

The second model is a dichotomous tree model, where the first split is between the outcome $\{1\}$ and the outcomes $\{2, 3\}$. If we define

$$p_i = \frac{\exp(\mu_1 + \beta_1 x_i + \alpha_1 D_i)}{1 + \exp(\mu_1 + \beta_1 x_i + \alpha_1 D_i)}$$

and

$$q_i = \frac{\exp(\mu_2 + \beta_2 x_i + \alpha_2 D_i)}{1 + \exp(\mu_2 + \beta_2 x_i + \alpha_2 D_i)},$$

then

$$\begin{aligned}\mathbb{P}(Y_i = 1) &= p_i \\ \mathbb{P}(Y_i = 2) &= (1 - p_i)q_i \\ \mathbb{P}(Y_i = 3) &= (1 - p_i)(1 - q_i).\end{aligned}$$

We prefer the polytomous model, since the residual deviance is less than the sum of the two residual deviances of the dichotomous tree model (the number of parameters is equal).

4b Model 2: Define

$$p = \frac{1}{1 + \exp(-\mu_1 - 2\beta_1)} = 0.068$$

and

$$q = \frac{1}{1 + \exp(-\mu_2 - 2\beta_2)} = 0.11,$$

where μ_1 is the intercept of `fit1`, β_1 is the coefficient corresponding to `x`, μ_2 is the intercept of `fit2` and β_2 is the coefficient corresponding to `x2`. We have that

$$\mathbb{P}(Y = 2) = (1 - p)q =: \phi(\mu_1, \beta_1, \mu_2, \beta_2).$$

Our estimate is therefore $\mathbb{P}(Y = 2) = 0.10$.
Model 1: Here we calculate two weights:

$$W_1 = \exp(\mu_1 + 2\beta_1) = 0.092 \text{ and } W_2 = \exp(\mu_2 + 2\beta_2) = 0.11.$$

Then

$$\mathbb{P}(Y = 2) = \frac{W_2}{1 + W_1 + W_2} = 0.092.$$

4c The log-likelihood is given by

$$l(k, \theta) = -k \log(\theta) + (k - 1)x - x/\theta - \log(\Gamma(k)).$$

To determine the Fisher information, we need to calculate the second derivative of l. To see whether the two estimators are asymptotically independent, it is enough to see whether the off-diagonal element of the Fisher information matrix is zero or not. So we calculate:

$$\mathbb{E}\left(\frac{\partial^2 l}{\partial k \partial \theta}\right) = -1/\theta.$$

This shows that the two estimators are asymptotically correlated, for all values of $\theta > 0$.

5a From Chapter 11 we have that

- leverage points can be identified as observations with exceptional large hat-values. In the first plot observations 6, 16 and 27 are identified as possible leverage points. This confirmed by the added-variable plots.

- regression outliers can be identified as observations with an exceptional large studentized residual. The only observation (although not extreme) that is identified as such is observation 6.

- observations that are influential on the value of coefficients can be identified by having large values for Cook's distance. Observations 6 and 16 are identified as being influential. This is confirmed by the added-variable plots, where observations 6 and 16 cause the coefficient of X_1 to be smaller and (although less severe) cause the coefficient of X_2 to be larger.

- observations that are influential on the standard error of coefficients can be identified by having a large value of the COVRATIO. Observation 27 and (less severe) observations 6 and 9 are identified as such.

- observation that jointly influential appear to be outlying on the added-variable plots. Observations 6, 16, and 27 cause the coefficient of X_1 to be smaller and (although less severe) observations 6 and 16 cause the coefficient of X_2 to be larger.