APPLIED STATSTICS(50015)

ASSIGNMENT-6

NARASIMHA NAIDU MOPIDEVI

KSU ID: 811229705

E-MAIL: nmopidev@kent.edu

5.1 For a factor X with d categories, the one-factor mean function is E | (Y U2, , … U U d) = + β β 0 2 2 + + - βdUd (5.17) where Uj is a dummy variable equal to 1 for the jth level of the factor and 0 otherwise.

5.1.1 Show that μ1 = β0 is the mean for the first level of X and that μj = β0 + βj is the mean for all the remaining levels, j = 2, . . . , d.

Sol: Now, The given model is state as below as:

$$E(Y|U2,\ldots,Ud) = \beta 0 + \beta 2U2 + \ldots + \beta dud$$

Also, given Uj is a dummy variable which is equal to 0 and one.

Let us assume that Uj be 0.

$$Uj= 0 \text{ for all } J=2,3,\ldots,d$$

$$\text{Therefore, } E(Y|U2,U3,\ldots,Ud)=\beta 0+\beta 2U2+\beta 3U3+\ldots +\beta dud$$

$$E(Y|U2=0,U3=0,\ldots,Ud=0)=\beta 0+\beta 2*(0)+\beta 3*(0)+\ldots + \beta d*(0)$$

$$E (Y| U2=0,U3=0,\ldots,Ud=0)=\beta 0$$

$$E(Y|U1)=\beta 0$$

So, μ1=β0

Let us assume that Uj be 1 for all j=2,3,4,….,d

Therefore,

$$E(Y|U2,U3,\ldots,Ud)=\beta 0+\beta 2U2+\beta 3U3+\ldots +\beta dud$$

$$E(Y|U2=1,U3=1,\ldots,Ud=1)=\beta 0+\beta 2*(1)+\beta 3*(1)+\ldots + \beta d*(1)$$

$$E(Y|Uj=1,Uk=0,k!=j)=\beta 0+\beta d(1)$$

$$E(Y|Ud)=\beta 0+\beta d$$

Thus, $\mu_j = \beta_0 + \beta_d$

5.3 (Data file: UN11)

5.3.1 In the fit of lifeExpF ∼ group, verify the results of Table 5.2.

Sol:

```
> library("lsmeans")
Loading required package: emmeans
The 'lsmeans' package is now basically a front end for 'emmeans'.
Users are encouraged to switch the rest of the way.
See help('transition') for more information, including how to
convert old 'lsmeans' objects and scripts to work with 'emmeans'.
Warning messages:
1: package ‘lsmeans’ was built under R version 4.2.2
2: package ‘emmeans’ was built under R version 4.2.2
> naidu<-lm(lifeExpF~group,UN11)
> lsmeans(naidu,pairwise~group)
$lsmeans
 group  lsmean    SE  df lower.CL upper.CL
 Oecd     82.4 1.128 196     80.2     84.7
 Other    75.3 0.586 196     74.2     76.5
 Africa   59.8 0.863 196     58.1     61.5

Confidence level used: 0.95

$contrasts
 contrast       estimate   SE  df t.ratio p.value
```

oecd - other 7.12 1.27 196 5.602 <.0001

oecd - africa 22.67 1.42 196 15.968 <.0001

other - africa 15.55 1.04 196 14.918 <.0001

P value adjustment: tukey method for comparing a family of 3 estimates

```
> library("lsmeans")
Loading required package: emmeans
The 'lsmeans' package is now basically a front end for 'emmeans'.
Users are encouraged to switch the rest of the way.
See help('transition') for more information, including how to
convert old 'lsmeans' objects and scripts to work with 'emmeans'.
Warning messages:
1: package ‘lsmeans’ was built under R version 4.2.2
2: package ‘emmeans’ was built under R version 4.2.2
> naidu<-lm(lifeExpF~group,UN11)
> lsmeans(naidu,pairwise~group)
$lsmeans
 group  lsmean    SE  df lower.CL upper.CL
 oecd     82.4 1.128 196     80.2     84.7
 other    75.3 0.586 196     74.2     76.5
 africa   59.8 0.863 196     58.1     61.5

Confidence level used: 0.95

$contrasts
 contrast       estimate   SE  df t.ratio p.value
 oecd - other       7.12 1.27 196   5.602  <.0001
 oecd - africa     22.67 1.42 196  15.968  <.0001
 other - africa    15.55 1.04 196  14.918  <.0001

P value adjustment: tukey method for comparing a family of 3 estimates

> |
```

Thus, the values are verified.

5.3.2 Compare all adjusted mean differences in the levels of group in the model lifeExpF ∼ group + log(ppgpd) with the results in Table 5.2.

Sol:

> library(lsmeans)

Loading required package: emmeans

The 'lsmeans' package is now basically a front end for 'emmeans'.

Users are encouraged to switch the rest of the way.

See help('transition') for more information, including how to

convert old 'lsmeans' objects and scripts to work with 'emmeans'.

Warning messages:

1: package ‘lsmeans’ was built under R version 4.2.2

2: package ‘emmeans’ was built under R version 4.2.2

> siddhu<-lm(lifeExpF~group+log(ppgdp),UN11)

> lsmeans(siddhu,pairwise~group)

$lsmeans

group	lsmean	SE	df	lower.CL	upper.CL
Oecd	79.6	0.959	195	77.7	81.5
Other	78.1	0.550	195	77.0	79.2
Africa	67.5	1.038	195	65.4	69.5

Confidence level used: 0.95

$contrasts

contrast	estimate	SE	df	t.ratio	p.value
oecd - other	1.53	1.174	195	1.308	0.3927
oecd - africa	12.17	1.557	195	7.814	<.0001
other - africa	10.64	0.979	195	10.862	<.0001

P value adjustment: tukey method for comparing a family of 3 estimates

Hence the difference between the values are given:

Oecd and other =7.12

Oecd and Africa =22.67

Other and Africa =15.55

Hence, the mean values are much higher than the values in the above model.

```
> library(lsmeans)
Loading required package: emmeans
The 'lsmeans' package is now basically a front end for 'emmeans'.
Users are encouraged to switch the rest of the way.
See help('transition') for more information, including how to
convert old 'lsmeans' objects and scripts to work with 'emmeans'.
Warning messages:
1: package ‘lsmeans’ was built under R version 4.2.2
2: package ‘emmeans’ was built under R version 4.2.2
> siddhu<-lm(lifeExpF~group+log(ppgdp),UN11)
> lsmeans(siddhu,pairwise~group)
$lsmeans
 group  lsmean    SE  df lower.CL upper.CL
 oecd     79.6 0.959 195     77.7     81.5
 other    78.1 0.550 195     77.0     79.2
 africa   67.5 1.038 195     65.4     69.5

Confidence level used: 0.95

$contrasts
 contrast       estimate    SE  df t.ratio p.value
 oecd - other       1.53 1.174 195   1.308  0.3927
 oecd - africa     12.17 1.557 195   7.814  <.0001
 other - africa    10.64 0.979 195  10.862  <.0001

P value adjustment: tukey method for comparing a family of 3 estimates

> |
```

5.5 Interpreting parameters with factors and interactions Suppose we have a regression problem with a factor A with two levels (a_1, a_2) and a factor B with three levels (b_1, b_2, b_3), so there are six treatment combinations. Table 5.8 Minnesota Agricultural Land Sales Variable Definition acrePrice Sale price in dollars per acre, adjusted to a common date within year year Year of sale acres Size of property, acres tillable Percentage of farm rated arable improvements Percentage of property value due to buildings and other improvements financing Type of financing either title transfer or seller finance crp Enrolled of any part of the acreage is enrolled in the U.S. Conservation Reserve Program (CRP), and none otherwise crpPct Percentage of land in CRP productivity A numeric score between 1 and 100 with larger values indicating more productive land, calculated by the University of Minnesota problems 125 5.7 Suppose the response is Y, and further that $E(Y|A = a_i, B = b_j) = \mu_{ij}$. The estimated μ_{ij} are the quantities that are used in effects plots. The purpose of this problem is to relate the μ_{ij} to the parameters that are actually fit in models with factors and interactions.

5.5.2 The model in Problem 5.5.1 has six regression coefficients, including an intercept. Express the βs as functions of the μ_{ij}.

Sol: The given model is depicted as:

$$E(Y \mid A = a_i, B = b_j) = \beta_0 + \beta_1 A_2 + \beta_2 B_2 + \beta_3 B_3 + \beta_4 A_2 B_2 + \beta_5 A_2 B_3$$

So, The six coefficient regression expressions are given:

$$\mu_{11} = E(Y \mid A = a_1, B = b_1) = \beta_0 \qquad -1$$

$$\mu_{12} = E(Y \mid A = a_1, B = b_2) = \beta_0 + \beta_2 B_2 \qquad -2$$

$$\mu_{13} = E(Y \mid A = a_1, B = b_3) = \beta_0 + \beta_3 B_3 \qquad -3$$

$$\mu_{21} = E(Y \mid A = a_2, B = b_1) = \beta_0 + \beta_1 A_2 \qquad -4$$

$$\mu_{22} = E(Y \mid A = a_2, B = b_2) = \beta_0 + \beta_1 A_2 + \beta_2 B_2 + \beta_4 A_2 B_2 \qquad -5$$

$$\mu_{23} = E(Y \mid A = a_2, B = b_3) = \beta_0 + \beta_1 A_2 + \beta_3 B_3 + \beta_5 A_2 B_3 \qquad -6$$

In the above 6 expressions, we can derive the 6 unknowns:

$$\beta_0 = \mu_{11}$$

$$\beta_1 = \mu_{21} - \mu_{11}$$

$$\beta_2 = \mu_{12} - \mu_{11}$$

$$\beta_3 = \mu_{13} - \mu_{11}$$

$$\beta_4 = \mu_{11} - \mu_{21} - \mu_{12} + \mu_{22}$$

$$\beta_5 = \mu_{11} - \mu_{13} - \mu_{21} + \mu_{23}$$

Hence, The intercept it can be used to directly translate the set of factor levels into a mean. All primary effects are different from the mean and μ_{11} in such a way.

5.10 (Data file: MinnLand) Refer to Problem 5.4. Another variable in this data file is the region, a factor with six levels that are geographic identifiers.

5.10.1 Assuming both year and region are factors, consider the two mean functions given in Wilkinson–Rogers notation as: (a) log(acrePrice) ∼ year + region (b) log(acrePrice) ∼ year + region + year:region Explain the difference between these two models (no fitting is required for this problem).

Sol: The second model into the account of relationship between year and region. It indicates that combined influence of a year and area is greater than the product of those two factors' separate effects.