ps4

Xiaolin Cao

Question 1

```
df1 = read_excel("ps4_q1.xls") %>%
  rename(curvatures = ...1)
```

```
## New names:
## * `` -> ...1
```

```
knitr::kable(df1, caption = "Effects of Different Conditions on Curvatures")
```

Table 1: Effects of Different Conditions on Curvatures

curvatures	est_log	lwr	upr
tot_dist	-0.161	-0.234	-0.089
max_abs_dev	-0.510	-0.731	-0.289
avg_abs_dev	-0.651	-0.906	-0.396
auc	-0.350	-0.549	-0.150

Question 2

```
## use Stata output to form a table
df2 = read.csv("ps4_q2.csv")
knitr::kable(df2, caption = "Internet Disparity Between Urban and Rural Areas In 10 Different Divisions")
```

Table 2: Internet Disparity Between Urban and Rural Areas In 10 Different Divisions

division	diff	lwr	upr
1	0.0178010	-0.0248420	0.0604441
2	-0.0194982	-0.0908940	0.0518976
3	0.0004151	-0.0530456	0.0538758
4	0.0767983	-0.0245220	0.1781186
5	0.0325632	-0.0354069	0.1005334
6	0.0933476	-0.0139532	0.2006484
7	0.0510231	-0.0229654	0.1250116
8	0.0549587	-0.0619455	0.1718629
9	0.1851698	0.0719216	0.2984181
10	0.0343232	-0.0450827	0.1137291

By looking at the result table, division 9 has the largest disparity between urban and rural areas in terms of the proportions of homes with internet access. Division 9 has the largest upper bound.

Question 3

c

```
df3c = matrix(c(1.130136,   1.027747,   1.242726,   0.024121,   0.0053218,  0.0429203,
                1.109808,   1.009919,   1.219577,   0.0204689, 0.0019269,  0.0390109,
                1.158033,   1.128201,   1.188654,   0.0316543,  0.0256843,  0.0376244,
                0.9737323,  0.9648531,  0.9826932, NA, NA, NA,
                1.33266,    1.213307,   1.463753,   0.0564991,  0.0380799,  0.0749184,
                1.093849,   1.060388,   1.128366,   0.0176006,  0.0115375,  0.0236637),
                nrow = 6, byrow=TRUE)
rownames(df3c) = c("1.weekday", "2.winter", "cage", "cage^2", "2.gender", "cpir")
colnames(df3c) = c("Odds Ratio", "lwr", "upr", "marginal effect", "lwr_marginal_effect", "upr_marginal_
df3c = round(df3c, digits = 4)
df3c
```

```
##           Odds Ratio    lwr    upr marginal effect lwr_marginal_effect
## 1.weekday     1.1301 1.0277 1.2427          0.0241              0.0053
## 2.winter      1.1098 1.0099 1.2196          0.0205              0.0019
## cage          1.1580 1.1282 1.1887          0.0317              0.0257
## cage^2        0.9737 0.9649 0.9827              NA                  NA
## 2.gender      1.3327 1.2133 1.4638          0.0565              0.0381
## cpir          1.0938 1.0604 1.1284          0.0176              0.0115
##           upr_marginal_effect
## 1.weekday              0.0429
## 2.winter               0.0390
## cage                   0.0376
## cage^2                     NA
## 2.gender               0.0749
## cpir                   0.0237
```

d

```
df3d = matrix(c(1.239979,   1.10902,    1.386404,   0.0232385,  0.0111058,  0.0353713,
                1.077818,   0.9369008,  1.23993,    0.0080304,  -0.0069865, 0.0230472,
                1.362175,   1.308046,   1.418544,   0.0359009,  0.0309386,  0.0408632,
                0.9550781,  0.9420468,  0.9682896, NA, NA, NA,
                1.757915,   1.527092,   2.023627,   0.0605507,  0.0455947,  0.0755066,
                1.179018,   1.125907,   1.234633,   0.0176414,  0.0127545, 0.0225282),
                nrow = 6, byrow=TRUE)
rownames(df3d) = c("1.weekday", "2.winter", "cage", "cage^2", "2.gender", "cpir")
colnames(df3d) = c("Odds Ratio", "lwr", "upr", "marginal effect", "lwr_marginal_effect", "upr_marginal_
df3d = round(df3d, digits = 4)
df3d
```

```
##           Odds Ratio    lwr    upr marginal effect lwr_marginal_effect
## 1.weekday     1.2400 1.1090 1.3864          0.0232              0.0111
## 2.winter      1.0778 0.9369 1.2399          0.0080             -0.0070
## cage          1.3622 1.3080 1.4185          0.0359              0.0309
## cage^2        0.9551 0.9420 0.9683              NA                  NA
## 2.gender      1.7579 1.5271 2.0236          0.0606              0.0456
## cpir          1.1790 1.1259 1.2346          0.0176              0.0128
##           upr_marginal_effect
## 1.weekday              0.0354
```

```
## 2.winter              0.0230
## cage                  0.0409
## cage^2                    NA
## 2.gender              0.0755
## cpir                  0.0225
```

Summary: Those two logistic regressions are similar to each other in terms of odds ratio and marginal effects. And the corresponding confidence intervals also overlap. To answer the question: Yes, people in the US are more likely to drink water on a weekday than on a weekend day. The odds ratios for that are 1.130 and 1.240 respectively in the above models.