```
R version 4.3.1 (2023-06-16) -- "Beagle Scouts"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: aarch64-apple-darwin20 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.79 (8238) aarch64-apple-darwin20]

[History restored from /Users/u1572955/.Rapp.history]

> setwd('/Users/u1572955/Dropbox/mexican-citizen-forecasting-2024/
replication')
> # Citizen forecasts of Mexican presidential elections, 2000-2024
> # Andreas Murr
> # Creates Figure 1 and Tables 1, 2
>
> # clear working memory
>
> rm(list = ls())
>
> # load data
>
> load("study-1-win-forecasts.rdata")
> load("study-1-vote-intentions.rdata")
>
> # ================================
> # = statistics mentioned in text =
> # ================================
>
> # compute number of surveys
>
> nrow(data.win)
[1] 55
>
> # compute range of survey respondents
>
```

```
> round(range(data.win$n))
[1]  909 4161
>
> # compute election coverage
>
> table(data.win$election)

2000 2006 2012 2018 2024
  12   16   14    5    8
>
> # compute lead time
>
> round(range(data.win$dist)) / 365
Time differences in days
[1] 0.00000 2.89589
>
> # ============
> # = figure 1 =
> # ============
>
> # solid line indicades 1st place
> # dashed line indicates 2nd place
> # dotted line indicates 3rd place
> pdf("figure-1.pdf", width = 10, height = 5)
> par(las = 1, mfcol = c(2, 5), mar = c(0,1,0,0), oma = c(2, 4, 2, 1))
> # 2000
>   ## citizen forecasts
>   sel = data.win[data.win$election == 2000,]
>   r = range(data.win[,1:5])
>   plot(sel$date, sel$PAN, ylim = r, type = "l", lty = 1, xlab = "", ylab
= "", main = "", xaxt = "n")
>   axis(2)
>   points(sel$date, sel$PRD, type = "l", lty = 3)
>   points(sel$date, sel$PRI, type = "l", lty = 2)
>   text(min(sel$date), c(.22, .38, .075), c("PAN", "PRI", "PRD"), pos =
4)
>   mtext("Winner forecasts", 2, 2.5, las = 3)
>   mtext("2000: PAN won", 3, 0.5)
>   ## vote intentions
>   sel = data.vot[data.vot$election == 2000,]
>   r = range(data.vot[,1:5], na.rm = TRUE)
>   plot(sel$date, sel$PAN, ylim = r, type = "l", lty = 1, xlab = "", ylab
= "", main = "", xaxt = "n")
>   axis.Date(1, format="%b %y")
>   axis(2)
>   points(sel$date, sel$PRD, type = "l", lty = 3)
>   points(sel$date, sel$PRI, type = "l", lty = 2)
>   text(min(sel$date), c(.3, .42, .18), c("PAN", "PRI", "PRD"), pos = 4)
>   mtext("Vote intentions", 2, 2.5, las = 3)
```

```
> # 2006
>   ## citizen forecasts
>   sel = data.win[data.win$election == 2006,]
>   r = range(data.win[,1:5])
>   plot(sel$date, sel$PAN, ylim = r, type = "l", lty = 1, xlab = "", ylab
= "", xaxt = "n", yaxt = "n")
>   points(sel$date, sel$PRD, type = "l", lty = 2)
>   points(sel$date, sel$PRI, type = "l", lty = 3)
>   text(min(sel$date), c(.20, .30, .26), c("PAN", "PRI", "PRD"), pos = 4)
>   mtext("2006: PAN won", 3, 0.5)
>   ## vote intentions
>   sel = data.vot[data.vot$election == 2006,]
>   r = range(data.vot[,1:5], na.rm = TRUE)
>   plot(sel$date, sel$PAN, ylim = r, type = "l", lty = 1, xlab = "", ylab
= "", main = "", xaxt = "n", yaxt = "n")
>   axis.Date(1, format="%b %y")
>   points(sel$date, sel$PRD, type = "l", lty = 2)
>   points(sel$date, sel$PRI, type = "l", lty = 3)
>   text(min(sel$date), c(.28, .37, .22), c("PAN", "PRI", "PRD"), pos = 4)
> # 2012
>   ## citizen forecasts
>   sel = data.win[data.win$election == 2012,]
>   r = range(data.win[,1:5])
>   plot(sel$date, sel$PAN, ylim = r, type = "l", lty = 3, xlab = "", ylab
= "", main = "", xaxt = "n", yaxt = "n")
>   points(sel$date, sel$PRD, type = "l", lty = 2)
>   points(sel$date, sel$PRI, type = "l", lty = 1)
>   text(min(sel$date), c(.23, .44, .14), c("PAN", "PRI", "PRD"), pos = 4)
>   mtext("2012: PRI won", 3, 0.5)
>   ## vote intentions
>   sel = data.vot[data.vot$election == 2012,]
>   r = range(data.vot[,1:5], na.rm = TRUE)
>   plot(sel$date, sel$PAN, ylim = r, type = "l", lty = 3, xlab = "", ylab
= "", main = "", xaxt = "n", yaxt = "n")
>   axis.Date(1, format="%b %y")
>   points(sel$date, sel$PRD, type = "l", lty = 2)
>   points(sel$date, sel$PRI, type = "l", lty = 1)
>   text(min(sel$date), c(.30, .47, .12), c("PAN", "PRI", "PRD"), pos = 4)
> # 2018
>   ## citizen forecasts
>   sel = data.win[data.win$election == 2018,]
>   r = range(data.win[,1:5])
>   plot(sel$date, sel$PAN, ylim = r, type = "l", lty = 2, xlab = "", ylab
= "", main = "", xaxt = "n", yaxt = "n")
>   points(sel$date, sel$MRN, type = "l", lty = 1)
>   points(sel$date, sel$PRI, type = "l", lty = 3)
>   text(min(sel$date), c(.17, .23, .3), c("PAN", "PRI", "MRN"), pos = 4)
>   mtext("2018: MRN won", 3, 0.5)
>   ## vote intentions
```

```
>   sel = data.vot[data.vot$election == 2018,]
>   r = range(data.vot[,1:5], na.rm = TRUE)
>   plot(sel$date, sel$PAN, ylim = r, type = "l", lty = 2, xlab = "", ylab
= "", main = "", xaxt = "n", yaxt = "n")
>   axis.Date(1, format="%b %y")
>   points(sel$date, sel$MRN, type = "l", lty = 1)
>   points(sel$date, sel$PRI, type = "l", lty = 3)
>   text(min(sel$date), c(.33, .25, .39), c("PAN", "PRI", "MRN"), pos = 4)
> # 2024
>   ## citizen forecasts
>   sel = data.win[data.win$election == 2024,]
>   r = range(data.win[,1:5])
>   plot(sel$date, sel$PAN, ylim = r, type = "l", lty = 2, xlab = "", ylab
= "", main = "", xaxt = "n", yaxt = "n")
>   points(sel$date, sel$MRN, type = "l", lty = 1)
>   points(sel$date, sel$MC, type = "l", lty = 3)
>   text(min(sel$date), c(.28, .12, .6), c("PAN", "MC", "MRN"), pos = 4)
>   mtext("2024: MRN won", 3, 0.5)
>   ## vote intentions
>   sel = data.vot[data.vot$election == 2024,]
>   r = range(data.vot[,1:5], na.rm = TRUE)
>   plot(sel$date, sel$PAN, ylim = r, type = "l", lty = 2, xlab = "", ylab
= "", main = "", xaxt = "n", yaxt = "n")
>   axis.Date(1, format="%b %y")
>   points(sel$date, sel$MRN, type = "l", lty = 1)
>   points(sel$date, sel$MC, type = "l", lty = 3)
>   text(min(sel$date), c(.22, .05, .60), c("PAN", "MC", "MRN"), pos = 4)
> # close device
> dev.off()
null device
          1
>
> # ===========
> # = table 1 =
> # ===========
>
> # count successes
>
> y.win = rep(NA, 5)
> y.win[1] = with(data.win[data.win$election==2000,], sum(PAN * n))
> y.win[2] = with(data.win[data.win$election==2006,], sum(PAN * n))
> y.win[3] = with(data.win[data.win$election==2012,], sum(PRI * n))
> y.win[4] = with(data.win[data.win$election==2018,], sum(MRN * n))
> y.win[5] = with(data.win[data.win$election==2024,], sum(MRN * n))
>
> # count trials
>
> n.win = rep(NA, 5)
> n.win[1] = sum(data.win$n[data.win$election==2000])
```

```
> n.win[2] = sum(data.win$n[data.win$election==2006])
> n.win[3] = sum(data.win$n[data.win$election==2012])
> n.win[4] = sum(data.win$n[data.win$election==2018])
> n.win[5] = sum(data.win$n[data.win$election==2024])
>
> # simulate from posterior
>
> n.sim = 100000
> alpha = 1
> beta = 1
> r.win = matrix(NA, nrow = n.sim, ncol = length(y.win))
> set.seed(12394)
> for (i in 1:length(y.win)){
+   r.win[,i] = rbeta(n.sim, y.win[i] + alpha, n.win[i] - y.win[i] + beta)
+ }
>
> # create table
>
> theta.p = formatC(apply(r.win, 2, mean) * 100, format = "f", 0)
> theta.q = apply(t(apply(r.win, 2, function(x){quantile(x, c(0.025,
0.975))*100})), 1, function(x){
+   paste("[", paste(formatC(x, format = "f", 0), collapse = "; "), "]",
sep = "")
+ })
>
> diff.p = formatC(apply(r.win - 1/3, 2, mean) * 100, format = "f", 0)
> diff.q = apply(t(apply(r.win - 1/3, 2, function(x){quantile(x, c(0.025,
0.975))*100})), 1, function(x){
+   paste("[", paste(formatC(x, format = "f", 0), collapse = "; "), "]",
sep = "")
+ })
>
> o = matrix(NA, nrow = 5 * 2, ncol = 3)
> o[seq(1, 10, 2), 1] = theta.p
> o[seq(2, 10, 2), 1] = theta.q
> o[seq(1, 10, 2), 2] = 33
> o[seq(2, 10, 2), 2] = ""
> o[seq(1, 10, 2), 3] = diff.p
> o[seq(2, 10, 2), 3] = diff.q
> colnames(o) = c("Correct forecast", "Random guess", "Difference")
> rownames(o) = rep("", 10)
> rownames(o)[seq(1, 10, 2)] = seq(2000, 2024, 6)
> noquote(o)
     Correct forecast Random guess Difference
2000 22               33           -11
     [21; 23]                      [-12; -11]
2006 30               33           -4
     [29; 30]                      [-4; -3]
2012 46               33           12
```

```
     [45; 46]                    [11; 13]
2018 46              33          12
     [44; 47]                    [11; 13]
2024 62              33          28
     [61; 63]                    [27; 29]
>
> # ===========
> # = table 2 =
> # ===========
>
> # compare success rate of surveys
>
> # forecast of expectation survey
> e1 = apply(data.win[data.win$election==2000,1:7], 1, which.max)
> e2 = apply(data.win[data.win$election==2006,1:7], 1, which.max)
> e3 = apply(data.win[data.win$election==2012,1:7], 1, which.max)
> e4 = apply(data.win[data.win$election==2018,1:7], 1, which.max)
> e5 = apply(data.win[data.win$election==2024,1:7], 1, which.max)
> # compute number of correct expectation surveys
> x = rep(NA, 5)
> x[1] = sum(ifelse(e1==1, 1, 0))
> x[2] = sum(ifelse(e2==1, 1, 0))
> x[3] = sum(ifelse(e3==3, 1, 0))
> x[4] = sum(ifelse(e4==4, 1, 0))
> x[5] = sum(ifelse(e5==4, 1, 0))
> # compute number of expectation surveys
> n.x = rep(NA, 5)
> n.x[1] = length(e1)
> n.x[2] = length(e2)
> n.x[3] = length(e3)
> n.x[4] = length(e4)
> n.x[5] = length(e5)
> # forecast of intention survey
> i1 = apply(data.vot[data.vot$election==2000,1:5], 1, which.max)
> i2 = apply(data.vot[data.vot$election==2006,1:5], 1, which.max)
> i3 = apply(data.vot[data.vot$election==2012,1:5], 1, which.max)
> i4 = apply(data.vot[data.vot$election==2018,1:5], 1, which.max)
> i5 = apply(data.vot[data.vot$election==2024,1:5], 1, which.max)
> # compute number of correct intention surveys
> y = rep(NA, 5)
> y[1] = sum(ifelse(i1==1, 1, 0))
> y[2] = sum(ifelse(i2==1, 1, 0))
> y[3] = sum(ifelse(i3==3, 1, 0))
> y[4] = sum(ifelse(i4==4, 1, 0))
> y[5] = sum(ifelse(i5==4, 1, 0))
> # compute number of intention surveys
> n.y = rep(NA, 5)
> n.y[1] = length(i1)
> n.y[2] = length(i2)
```

```
> n.y[3] = length(i3)
> n.y[4] = length(i4)
> n.y[5] = length(i5)
> # posterior analysis (beta binomial)
> ## each year
> theta.x = matrix(NA, n.sim, 5)
> theta.y = matrix(NA, n.sim, 5)
> set.seed(1249)
> for (i in 1:5){
+   theta.x[,i] = rbeta(n.sim, x[i] + alpha, n.x[i] - x[i] + beta)
+   theta.y[,i] = rbeta(n.sim, y[i] + alpha, n.y[i] - y[i] + beta)
+ }
> ## overall
> set.seed(1249)
> o.x = rbeta(n.sim, sum(x) + alpha, sum(n.x) - sum(x) + beta)
> o.y = rbeta(n.sim, sum(y) + alpha, sum(n.y) - sum(y) + beta)
> # store results
> ## each year
> q.x = t(round(apply(theta.x, 2, function(x){c(mean(x), quantile(x,
c(0.025, 0.975)))}) * 100))
> q.y = t(round(apply(theta.y, 2, function(x){c(mean(x), quantile(x,
c(0.025, 0.975)))}) * 100))
> q.d = t(round(apply(theta.x - theta.y, 2, function(x){c(mean(x),
quantile(x, c(0.025, 0.975)))}) * 100))
> ## overall
> o.d = o.x - o.y
> q.o.x = c(mean(o.x), quantile(o.x, c(0.025, 0.975)))
> q.o.y = c(mean(o.y), quantile(o.y, c(0.025, 0.975)))
> q.o.d = c(mean(o.d), quantile(o.d, c(0.025, 0.975)))
> # tabulate results
> tab = matrix(NA, nrow = 12, ncol = 3)
> ## point estimates
> ### each year
> tab[seq(1, 9, 2),1] =  q.x[,1]
> tab[seq(1, 9, 2),2] =  q.y[,1]
> tab[seq(1, 9, 2),3] =  q.d[,1]
> ### overall
> tab[11,] = round(c(q.o.x[1], q.o.y[1], q.o.d[1]) * 100)
> ## intervals
> ### each year
> tab[seq(2,10, 2),1] = apply(q.x[,2:3], 1, function(x){paste("[",
paste(x, collapse = ";"), "]", sep = "")})
> tab[seq(2,10, 2),2] = apply(q.y[,2:3], 1, function(x){paste("[",
paste(x, collapse = ";"), "]", sep = "")})
> tab[seq(2,10, 2),3] = apply(q.d[,2:3], 1, function(x){paste("[",
paste(x, collapse = ";"), "]", sep = "")})
> ### overall
> tab[12,1] = paste("[", paste(round(q.o.x[2:3]*100), collapse = ";"),
"]", sep = "")
```

```
> tab[12,2] = paste("[", paste(round(q.o.y[2:3]*100), collapse = ";"),
"]", sep = "")
> tab[12,3] = paste("[", paste(round(q.o.d[2:3]*100), collapse = ";"),
"]", sep = "")
> ### annotation
> colnames(tab) = c("CF", "VI", "DIF")
> rownames(tab) = rep("", nrow(tab))
> rownames(tab)[seq(1, nrow(tab), 2)] = c(seq(2000, 2024, 6), "Overall")
> ## return table
> noquote(tab)
        CF       VI       DIF
2000    7        36       -28
        [0;25]   [14;61]  [-57;-2]
2006    39       56       -17
        [18;62]  [33;77]  [-47;15]
2012    94       94       0
        [78;100] [78;100] [-18;18]
2018    71       86       -14
        [36;96]  [54;100] [-55;26]
2024    90       90       0
        [66;100] [66;100] [-27;27]
Overall 58       72       -14
        [45;70]  [60;83]  [-31;3]
>
> # ====================
> # = end source code =
> # ====================
>
```