

選舉民調準嗎？

To do: 1.Change regressions into charts 2.Do a bump chart for the rank changes in different specification

Sample Selection

我選取了 2010, 2012, 2014, 2016, 2018 五次大選中，總統選舉及直轄市長的民意調查。資料中總共有 2144 個候選人民調數據，有 59 個不同的民調資料來源。

```
dt[, .N] # Number of Election-Candidate Count
```

```
## [1] 1924
```

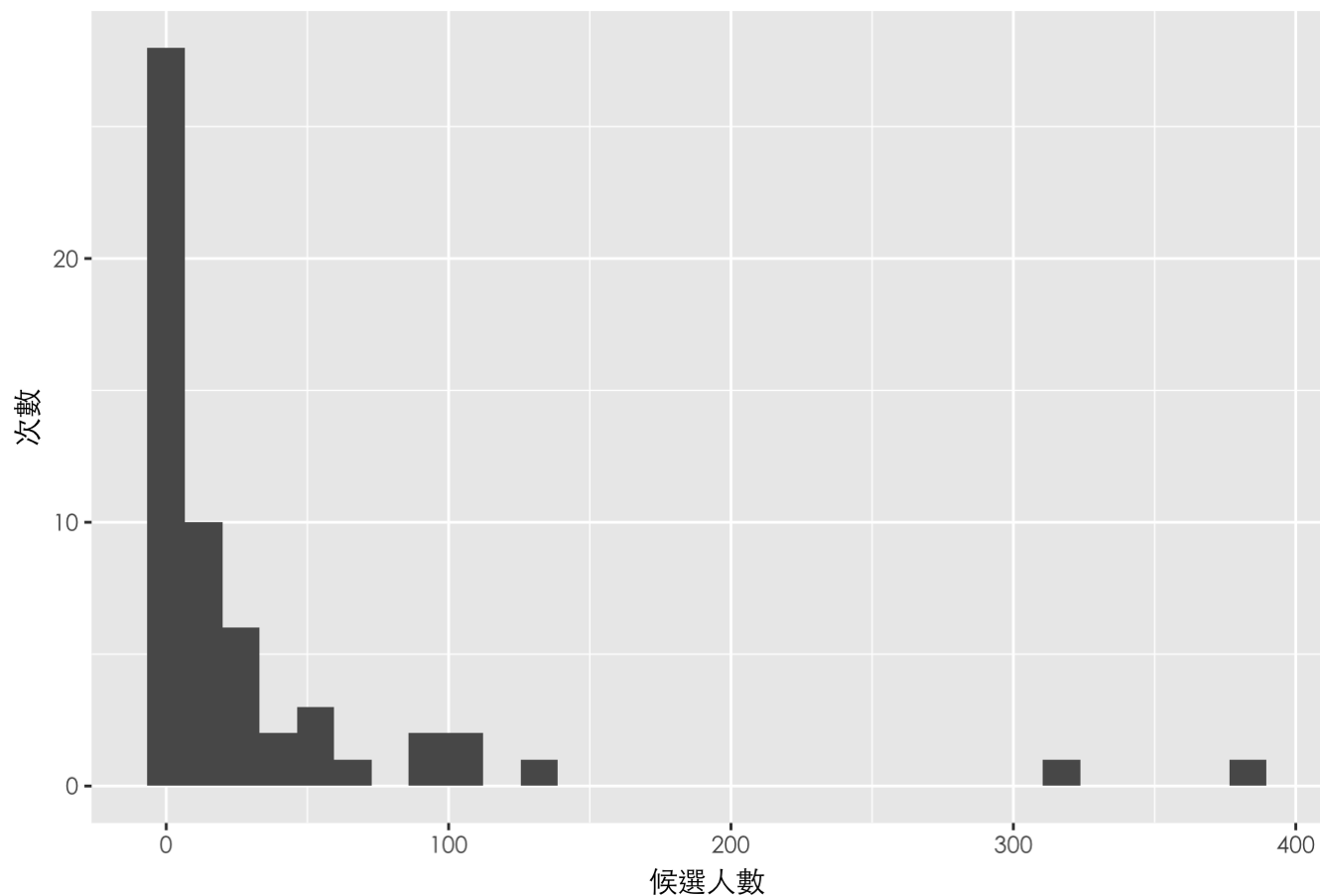
```
length(unique(dt$Institute)) # Number of unique institute
```

```
## [1] 57
```

```
dt[, Institute := as.factor(Institute)]
dtInstituteCount <- dt[, list(N = .N, noElection = length(unique(Election))), by = Institute]
ggplot(dtInstituteCount, aes(x=N)) +
  theme(text = element_text(family = "Heiti TC Light")) +
  geom_histogram() + xlab("候選人數") + ylab("次數") +
  ggtitle("原始民調機構民調次數分佈")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

原始民調機構民調次數分佈



前 15 大民調機構如下：

```
setorder(dtInstituteCount, -N)
pander(dtInstituteCount[1:15,])
```

Institute	N	noElection
中國時報	384	5
TVBS	314	5
聯合報	137	5
自由時報	107	5
趨勢	102	4
蘋果日報	99	5
世新大學	92	5
民主進步黨	61	5
東森	55	2
台灣指標	52	3
山水	51	2

Institute	N	noElection
年代	42	4
展欣科技	41	3
美麗島	33	1
決策	32	3

此 15 大機構佔的民調比例達所有民調的 79 %。

```
sum(dtInstituteCount[1:15,N])/nrow(dt)
```

```
## [1] 0.8326403
```

選取前 15 大民調機構，並且去除部分未包含未表態或者候選人最後退選的民調，

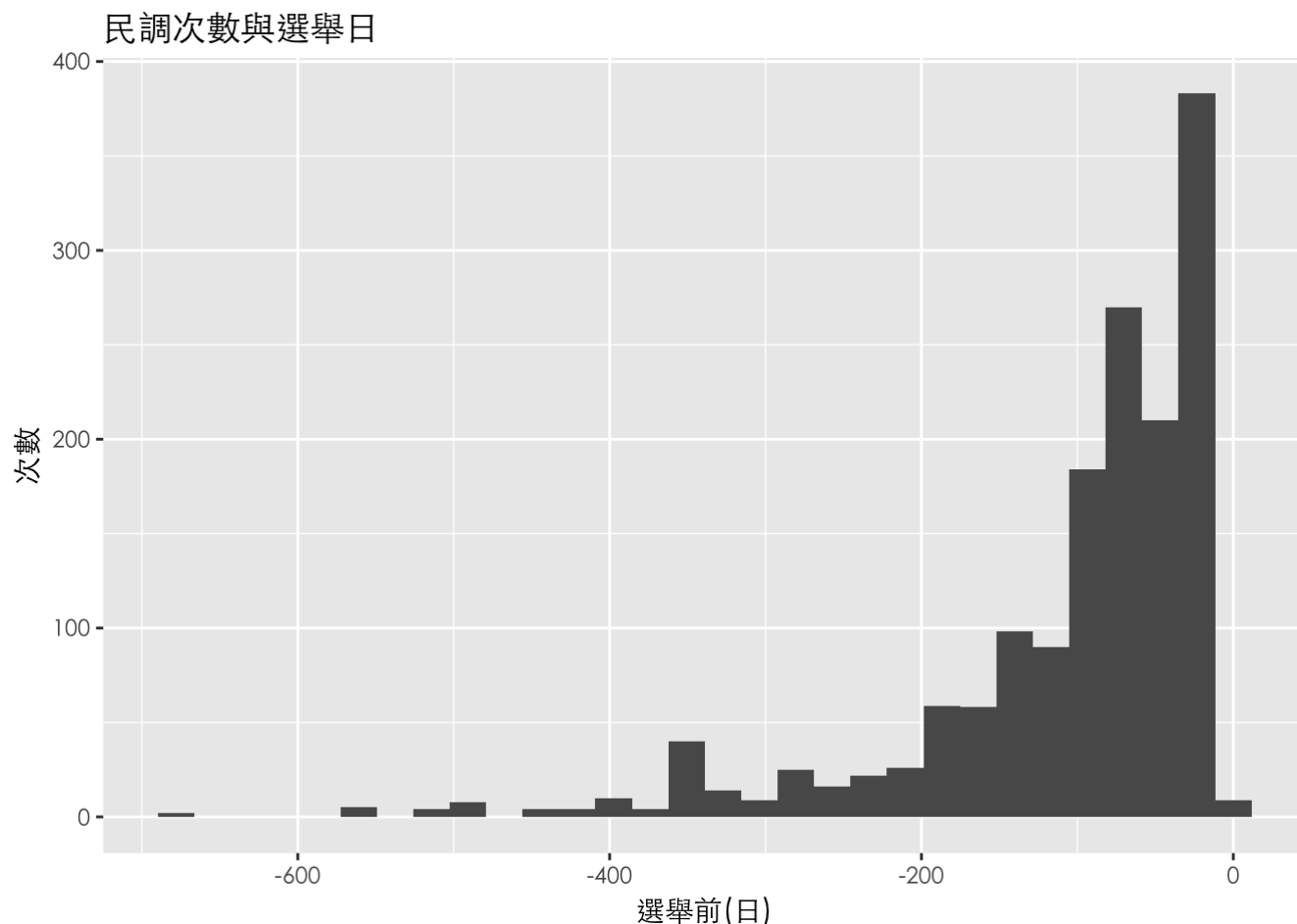
```
dtSample <- dt[Institute %in% dtInstituteCount[1:15,Institute]]
dtSample <- dtSample[!is.na(NoResponse)]
dtSample <- dtSample[!is.na(voteShare)]
```

民調的時間

經過上述的樣本選擇後，民調所涵蓋的時間依然極為廣泛。最早一組在選舉 2 年前即開始進行民調，最晚的民調於封關之後仍然持續。我們排除民調封關 (選舉 10 天前) 之後的樣本，也排除選舉 1 年以前的樣本。經過這樣的樣本選擇之後，我們留下 76 % 的樣本。

```
ggplot(data = dtSample, aes(x = preDate)) +
  geom_histogram() + xlab("選舉前(日)") + ylab("次數") + ggtitle("民調次數與選舉日")+
  theme(text = element_text(family = "Heiti TC Light"))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
dtSample <- dtSample[preDate <= -10 & preDate >= -366]
```

最後剩下的樣本百分比：

```
dtSample[,.N]/dt[,.N]
```

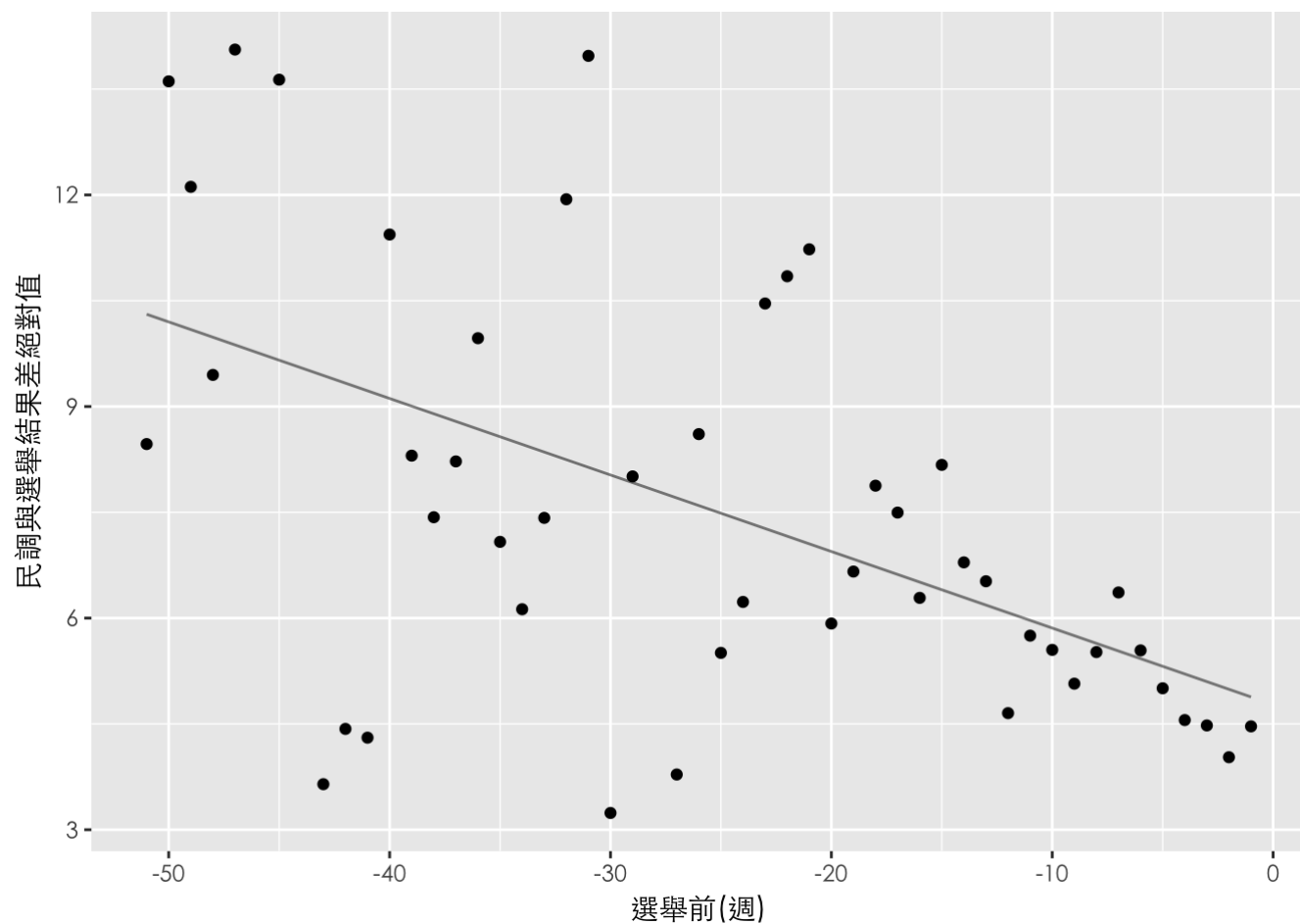
```
## [1] 0.783264
```

假設真實的支持度在民調的過程中沒有變化

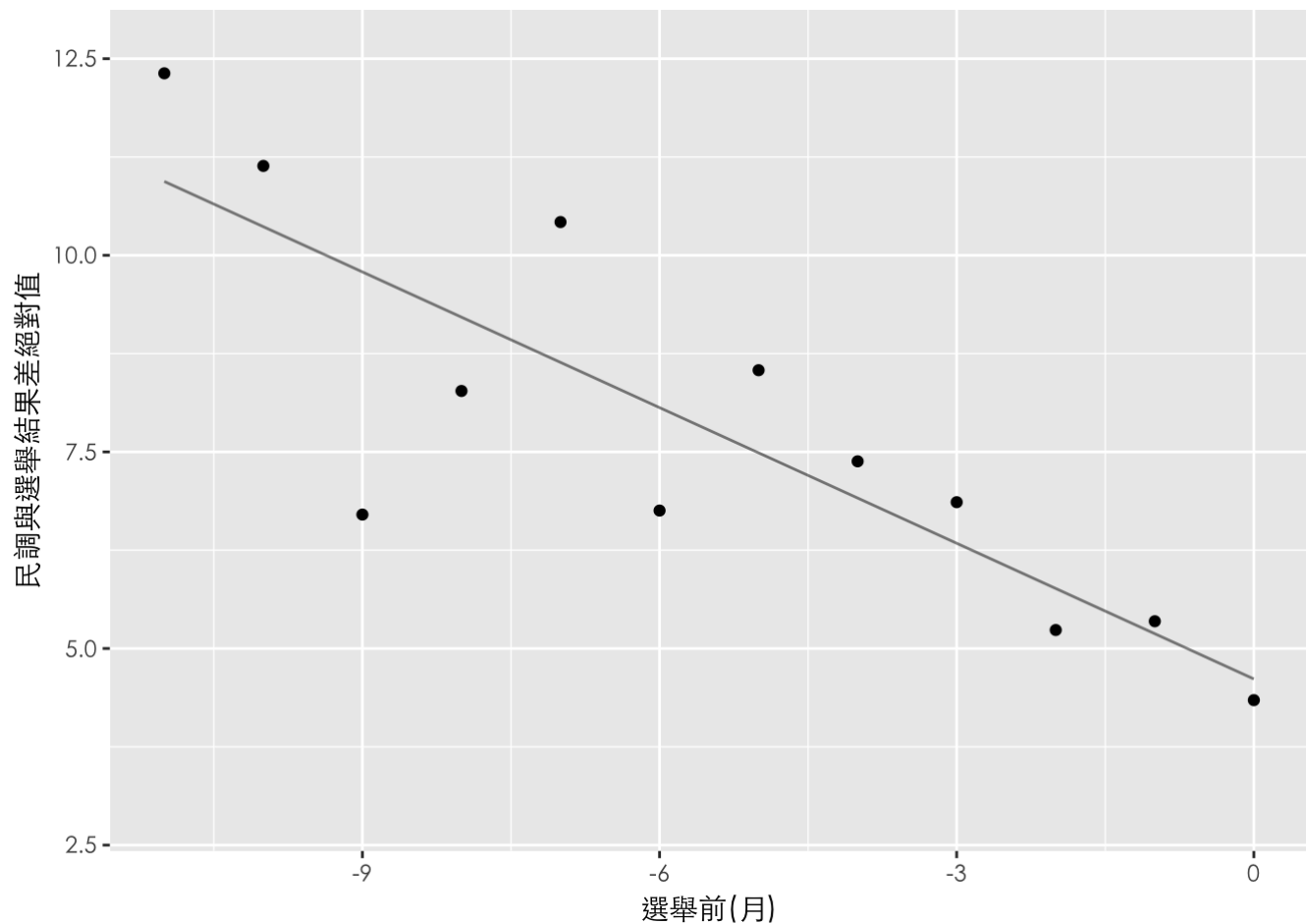
民調越靠近選舉越準嗎？

我先將選舉前每週與每月加總：

```
dtSample[,preWeek := ceiling(preDate/7)]
ggplot(data = dtSample[,mean(abs(Error)), by = preWeek],
  aes(x = preWeek, y = V1 ))+ geom_line(stat="smooth",method = "lm",
    alpha = 0.5)+ geom_point() +
  xlab("選舉前(週)") + ylab("民調與選舉結果差絕對值") +
  theme(text = element_text(family = "Heiti TC Light"))
```



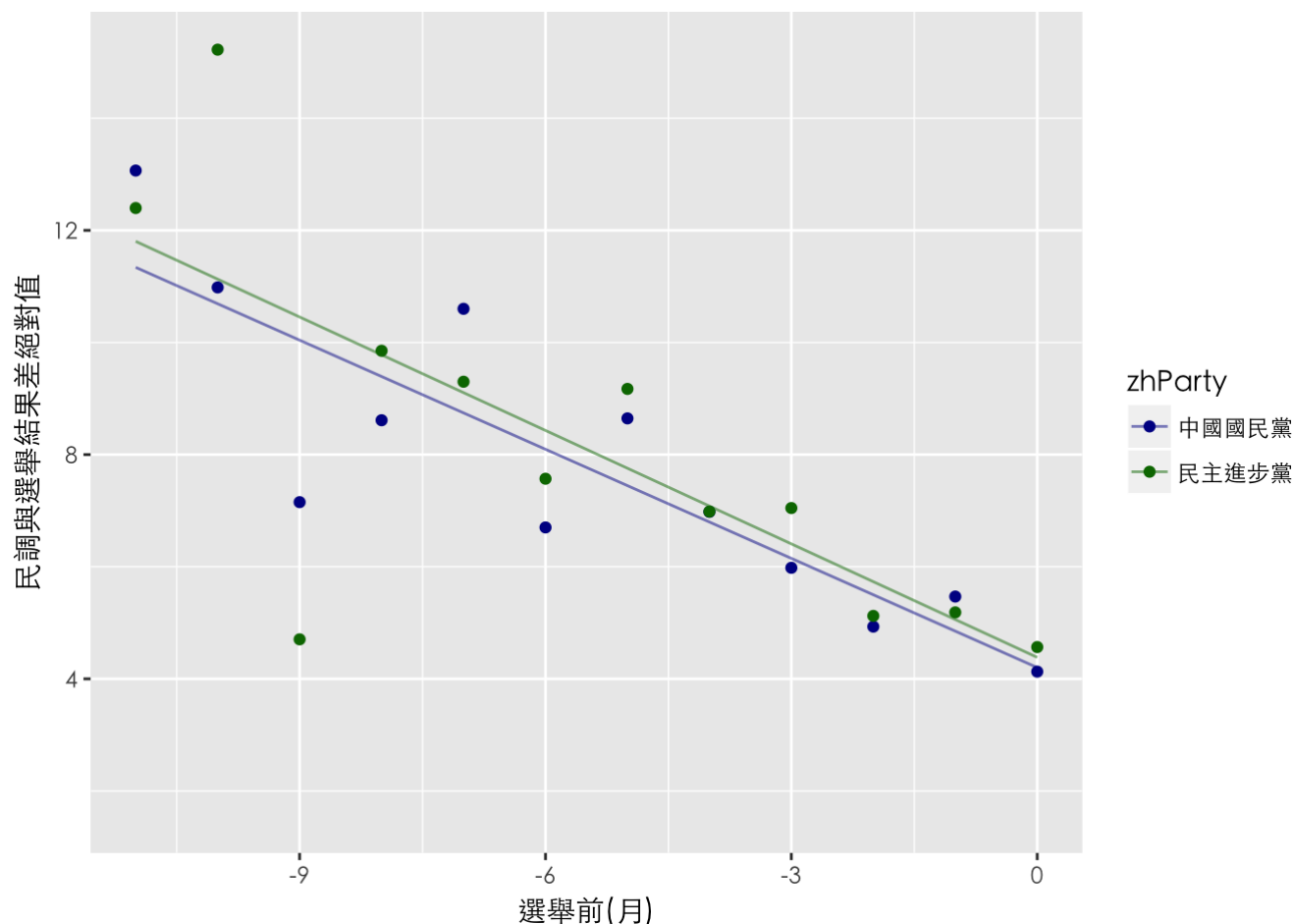
```
dtSample[,preMonth := ceiling(preDate/30)]
ggplot(data = dtSample[,mean(abs(Error)), by = preMonth],
  aes(x = preMonth, y = V1 )) + geom_line(stat="smooth",method = "lm",
    alpha = 0.5)+ geom_point()+
  xlab("選舉前(月)") + ylab("民調與選舉結果差絕對值") +
  theme(text = element_text(family = "Heiti TC Light"))
```



兩大黨哪個的民調比較難猜？

```
group.colors <- c("民主進步黨" = "#007500", "中國國民黨" = "#000093")

ggplot(data = dtSample[zhParty %in% c("民主進步黨", "中國國民黨"), mean(abs(Error)), by = c(
  "preMonth", "zhParty")],
  aes(x = preMonth, y = V1, color = zhParty)) + geom_line(stat="smooth", method =
  "lm",
  alpha = 0.5) + geom_point()+
  xlab("選舉前(月)") + ylab("民調與選舉結果差絕對值") +
  theme(text = element_text(family = "Heiti TC Light")) +
  scale_color_manual(values=group.colors)
```



```
pander(lm(data = dtSample[zhParty %in% c("民主進步黨", "中國國民黨")],
  formula = I(abs(Error)) ~ zhParty), caption = "誤差民進黨與國民黨無顯著差異",
  add.significance.stars = TRUE)
```

誤差民進黨與國民黨無顯著差異

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	6.043	0.202	29.91	2.033e-145	***
zhParty民主進步黨	0.09565	0.2937	0.3257	0.7447	

```
dtSample[,underPredicted := (Error > 0)]
pander(lm(data = dtSample[zhParty %in% c("民主進步黨", "中國國民黨")],
  formula = underPredicted ~ zhParty), caption = '民進黨民調被低估的其實比較少',
  add.significance.stars = TRUE)
```

民進黨民調被低估的其實比較少

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.7028	0.01923	36.54	6.789e-194	***
zhParty民主進步黨	-0.08016	0.02796	-2.867	0.004222	**

哪家民調比較準？

```
dtSample[,underPredicted := (Error > 0)]
pander(lm(data = dtSample[,zhParty %in% c("民主進步黨","中國國民黨")],
  formula = underPredicted ~ zhParty),caption = '民進黨民調被低估的其實比較少',
  add.significance.stars = TRUE)
```

民進黨民調被低估的其實比較少

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.7028	0.01923	36.54	6.789e-194	***
zhParty民主進步黨	-0.08016	0.02796	-2.867	0.004222	**

```
fit <- lm(data = dtSample,
  formula = I(abs(Error)) ~ 0+ Institute)

dtAbsError <- data.table(Institute = names(fit$coefficients), AbsError = fit$coefficients,
  SE = summary(fit)$coefficient[,2])
dtAbsError[,Institute := gsub(pattern = "Institute",replacement = "",x = Institute)]
setorder(dtAbsError,AbsError)
pander(dtAbsError)
```

Institute	AbsError	SE
世新大學	3.592	0.4845
決策	4.773	0.8216
民主進步黨	4.791	0.595
趨勢	5.252	0.4647
中國時報	5.699	0.2406
TVBS	5.823	0.2739
山水	6.358	0.6508
年代	6.367	0.7258
美麗島	6.622	0.8347
蘋果日報	6.81	0.4768
聯合報	6.872	0.3971
自由時報	7.26	0.4557
東森	8.99	0.6445
台灣指標	9.942	0.6572

控制距離選舉的時間


```
fit <- lm(data = dtSample,
  formula = I(abs(Error)) ~ 0+ Institute + preDate)
dtAbsError <- data.table(Institute = names(fit$coefficients), AbsError = fit$coefficients,
  SE = summary(fit)$coefficient[,2])
dtAbsError <- dtAbsError[grepl(pattern = "Institute",x = Institute)]
dtAbsError[,Institute := gsub(pattern = "Institute",replacement = "",x = Institute)]
setorder(dtAbsError,AbsError)
pander(dtAbsError)
```

Institute	AbsError	SE
世新大學	2.611	0.4671
民主進步黨	3.165	0.5801
決策	3.601	0.7865
年代	3.605	0.7238
趨勢	3.608	0.4606
美麗島	3.873	0.8228
TVBS	3.884	0.3022
蘋果日報	4.038	0.5036
中國時報	4.152	0.2595
山水	5.068	0.627
聯合報	5.184	0.4004
自由時報	5.804	0.4483
東森	5.941	0.6585
台灣指標	7.368	0.6572

每家最後一個預測

```
dtSample[,lastForElection := (Date == max(Date)) ,by = c("Institute","Candidate","Election","District")]
fit <- lm(data = dtSample[lastForElection == TRUE],
  formula = I(abs(Error)) ~ 0+ Institute + preDate)
dtAbsError <- data.table(Institute = names(fit$coefficients), AbsError = fit$coefficients,
  SE = summary(fit)$coefficient[,2])
dtAbsError <- dtAbsError[grepl(pattern = "Institute",x = Institute)]
dtAbsError[,Institute := gsub(pattern = "Institute",replacement = "",x = Institute)]
setorder(dtAbsError,AbsError)
pander(dtAbsError)
```

Institute	AbsError	SE
世新大學	2.838	0.7955

Institute	AbsError	SE
年代	3.313	1.191
山水	3.651	2.13
民主進步黨	3.679	0.9826
決策	4.203	1.771
TVBS	4.608	0.7305
中國時報	4.651	0.838
蘋果日報	4.728	0.9515
趨勢	5.498	1.082
聯合報	5.634	0.7884
自由時報	6.47	0.7957
東森	7.309	1.115
美麗島	7.425	1.681
台灣指標	10.06	1.157

有沒有機構效應？

```
fit <- lm(data = dtSample[zhParty %in% c("民主進步黨","中國國民黨"),],
  formula = Error ~ 0 + Institute + Institute*zhParty + zhParty + preDate + Election)

pander(fit,add.significance.stars = TRUE)
```

Fitting linear model: Error ~ 0 + Institute + Institute * zhParty + zhParty + preDate + Election (continued below)

	Estimate	Std. Error	t value
InstituteTVBS	-0.4024	0.8234	-0.4887
Institute世新大學	0.4364	1.238	0.3524
Institute中國時報	-0.4128	0.8195	-0.5037
Institute台灣指標	3.736	1.696	2.203
Institute山水	1.505	1.865	0.8071
Institute年代	3.429	1.845	1.858
Institute東森	4.496	1.653	2.72
Institute民主進步黨	1.531	1.408	1.087
Institute決策	4.894	2.261	2.164
Institute美麗島	8.02	2.135	3.756

	Estimate	Std. Error	t value
Institute聯合報	-2.045	1.06	-1.929
Institute自由時報	6.323	1.191	5.311
Institute蘋果日報	-1.654	1.278	-1.294
Institute趨勢	3.411	1.31	2.603
zhParty民主進步黨	1.365	0.9631	1.417
preDate	0.003032	0.002963	1.023
Election2012 Election	4.639	0.694	6.684
Election2014 Election	-0.3812	0.6916	-0.5513
Election2016 Election	0.04677	0.91	0.0514
Election2018 Election	-0.0334	0.7884	-0.04236
Institute世新大學:zhParty民主進步黨	-2.115	1.917	-1.103
Institute中國時報:zhParty民主進步黨	-0.6075	1.3	-0.4672
Institute台灣指標:zhParty民主進步黨	-5.348	2.454	-2.18
Institute山水:zhParty民主進步黨	-2.84	2.628	-1.081
Institute年代:zhParty民主進步黨	-5.821	2.698	-2.157
Institute東森:zhParty民主進步黨	-6.4	2.312	-2.768
Institute民主進步黨:zhParty民主進步黨	-3.901	2.183	-1.787
Institute決策:zhParty民主進步黨	-8.084	3.189	-2.535
Institute美麗島:zhParty民主進步黨	-6.042	2.957	-2.043
Institute聯合報:zhParty民主進步黨	4.733	1.678	2.821
Institute自由時報:zhParty民主進步黨	-13.07	1.874	-6.979
Institute蘋果日報:zhParty民主進步黨	3.116	2.029	1.536
Institute趨勢:zhParty民主進步黨	-5.546	1.968	-2.818

	Pr(> t)
InstituteTVBS	0.6252
Institute世新大學	0.7246
Institute中國時報	0.6146
Institute台灣指標	0.02782 *
Institute山水	0.4198
Institute年代	0.06342

	Pr(> t)	
Institute東森	0.00663	* *
Institute民主進步黨	0.2773	
Institute決策	0.03064	*
Institute美麗島	0.000182	* * *
Institute聯合報	0.05398	
Institute自由時報	1.32e-07	* * *
Institute蘋果日報	0.1959	
Institute趨勢	0.009371	* *
zhParty民主進步黨	0.1567	
preDate	0.3064	
Election2012 Election	3.678e-11	* * *
Election2014 Election	0.5816	
Election2016 Election	0.959	
Election2018 Election	0.9662	
Institute世新大學:zhParty民主進步黨	0.2701	
Institute中國時報:zhParty民主進步黨	0.6404	
Institute台灣指標:zhParty民主進步黨	0.0295	*
Institute山水:zhParty民主進步黨	0.2801	
Institute年代:zhParty民主進步黨	0.0312	*
Institute東森:zhParty民主進步黨	0.005734	* *
Institute民主進步黨:zhParty民主進步黨	0.07418	
Institute決策:zhParty民主進步黨	0.01137	*
Institute美麗島:zhParty民主進步黨	0.0413	*
Institute聯合報:zhParty民主進步黨	0.004868	* *
Institute自由時報:zhParty民主進步黨	5.128e-12	* * *
Institute蘋果日報:zhParty民主進步黨	0.1248	
Institute趨勢:zhParty民主進步黨	0.004919	* *