This is a classroom example related to the 2006 elections and should be viewed as a simple comparison of the relative poll standings of the top two candidates, versus how they fared in the vote, in 19 different gubanatorial races and 17 different national senatorial races.

For each race, data was available from a poll close to the election and the sample size of the poll was known. In all but one of these races the leader in the poll won the election. One gubanatorial poll was reported as a tie.

A number of the gubanatorial polls had sample size 500. Many of those bore the name of a small private polling organization and were conducted less close to the election.

11 of the 19 gubanatorial polls underestimated the final vote of the eventual winner.
8 of the 17 senatorial polls underestimated the final vote of the eventual winner.

The data format is {a, b, c, d, e} with

a = % of poll vote for the election winning candidate
b = % of poll vote for the candidate placing second
c = % of election vote for the election winning candidate
d = % of election vote for the candidate placing second
e = sample size n of the poll

I've converted these figures into what would be standard scores if the poll could be regarded as a with-replacement and equal-probability sampling of the votes recorded for voters whose choice placed first or second in the election and who would have polled for one of those two candidates.

$$\frac{\hat{p}-p}{\sqrt{\frac{p\,q}{n}}}$$

where

$$\hat{p} = \frac{a}{a+b},\ p = \frac{c}{c+d}.$$

Means and standard deviations of above standard scores are surprisingly alike when comparing the gubanatorial and senatorial elections. The appear to differ from 0 and 1

gubanatorial races: mean = -0.42, sample standard deviation = 1.42
senatorial races: mean = -0.43, sample standard deviation = 1.68.

One might ask whether the above differences from 0 and 1 are significant (using a test of normality), and further enquire as to whether the similarities of emans and sd above might easily arise by chance, but that is outside the scope of what we will do here.

These standard scores ignore many issues relating to the undecided or independent vote. When plotted against the standard normal they do show marked departure from normality, particularly on the low side where rather too many instances of a poll severely underestimating the fraction of the election winner are seen.

```
gov2 = {{42, 40, 48.52, 40.73, 747}, {51, 42, 55.77, 40.76, 549}, {53, 38, 55.77, 39.14, 500}, {51, 39, 56.33, 40.87, 500}, {50, 42, 52.18, 45.11, 500}, {52, 43, 53.99, 45.11, 756}, {50, 47, 53.21, 45.63, 663}, {53, 29, 55.56, 35.22, 500}, {52, 42, 56.35, 42.32, 1200}, {51, 43, 47.92, 43.92, 500}, {61, 26, 68.69, 31.31, 500}, {45, 45, 46.69, 45.73, 710}, {61, 33, 62.75, 35.24, 625}, {51, 38, 57.8, 38.35, 800}, {69, 24, 69.01, 29.15, 500}, {55, 38, 60.36, 29.15, 500}, {55, 38, 60.36, 36.84, 646}, {58, 35, 60.26, 39.74, 646}, {48, 45, 52.77, 45.33, 800}};

sen2 = {{48, 51, 47, 48, 677}, {45, 53, 41, 50, 520}, {48, 49, 44, 49, 602}, {47, 50, 46, 49, 594}, {38, 58, 34, 54, 747}, {41, 59, 44, 48, 800}, {44, 54, 46, 49, 663}, {44, 56, 42, 54, 436}, {49, 50, 44, 52, 741}, {40, 50, 38, 50, 676}, {39, 58, 43, 52, 800}, {38, 60, 35, 58, 1200}, {41, 57, 44, 50, 1200}, {35, 60, 31, 60, 553}, {47, 53, 45, 48, 880}, {44, 53, 41, 49, 625}, {31, 67, 31, 64, 1502}};

Length[gov2]

19

Length[sen2]

17

compare[x_] := Module[{phat = x[[1]] / (x[[1]] + x[[2]]), p = x[[3]] / (x[[3]] + x[[4]])},
  (phat - p) / Sqrt[p (1 - p) / x[[5]]]]

phi[r_, sig_] := Exp[-r^2 / (2 sig^2)] / (sig Sqrt[2 Pi])
```

```
gov2z = Table[compare[gov2[[i]]], {i, 1, 19}] 1.0
```

```
{-1.72552, -1.39283, -0.23583, -0.582536, 0.320318,
 0.141632, -1.18179, 1.57453, -1.25316, 0.930022, 0.687059, -0.276795,
 0.446179, -1.62373, 1.90366, -3.95756, -1.55021, 1.09361, -1.23627}
```

11 of 19 gubanatorial polls underpredicted the vote for the candidate leading in the poll (negative gov2z scores).

```
{Mean[gov2z], Sqrt[(19 / 18) (Mean[gov2z^2 - Mean[gov2z]^2])]}
```

```
{-0.416802, 1.41899}
```

```
sen2z = Table[compare[sen2[[i]]], {i, 1, 17}] 1.0
```

```
{-0.514603, 0.395721, 1.06772, 0.0158769, 0.531548, -3.86507, -1.81521, 0.105228, 2.00044,
 0.662757, -2.87358, 0.815924, -3.45159, 1.37751, -0.823387, -0.0977524, -0.825698}
```

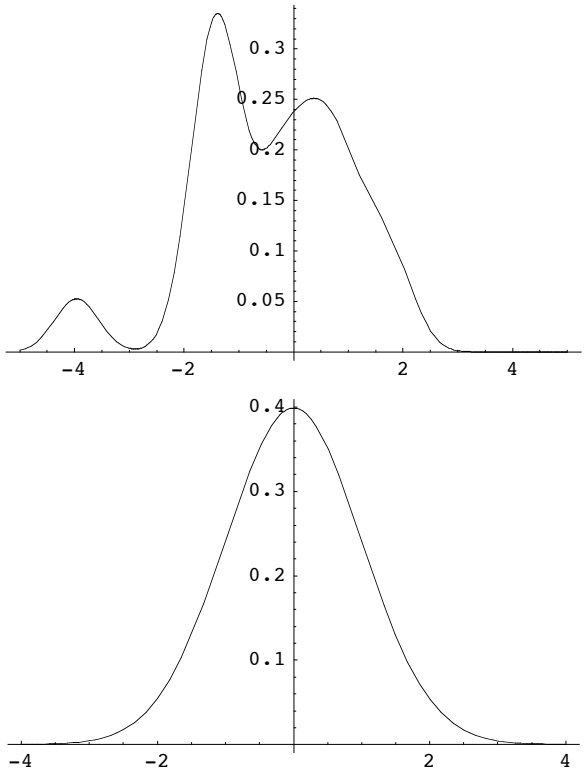
8 of 17 senatorial polls underpredicted the vote for the candidate leading in the poll (negative sen2z scores).

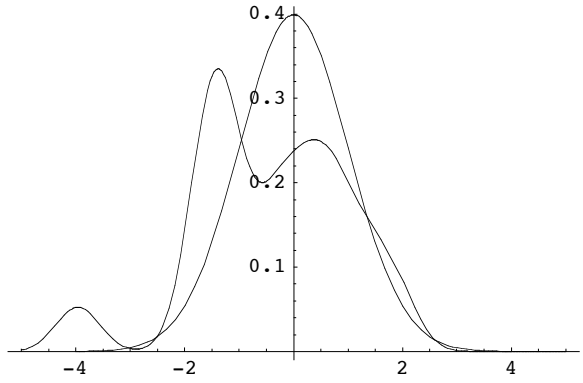
```
{Mean[sen2z], Sqrt[(19 / 18) (Mean[sen2z^2 - Mean[sen2z]^2])]}
```

```
{-0.429069, 1.68089}
```

Here is a plot of the gubanatorial standard scores shown together with the standard normal.

```
Show[Plot[Mean[phi[r - gov2z, 0.4]], {r, -5, 5}], Plot[phi[r, 1], {r, -4, 4}]]
```

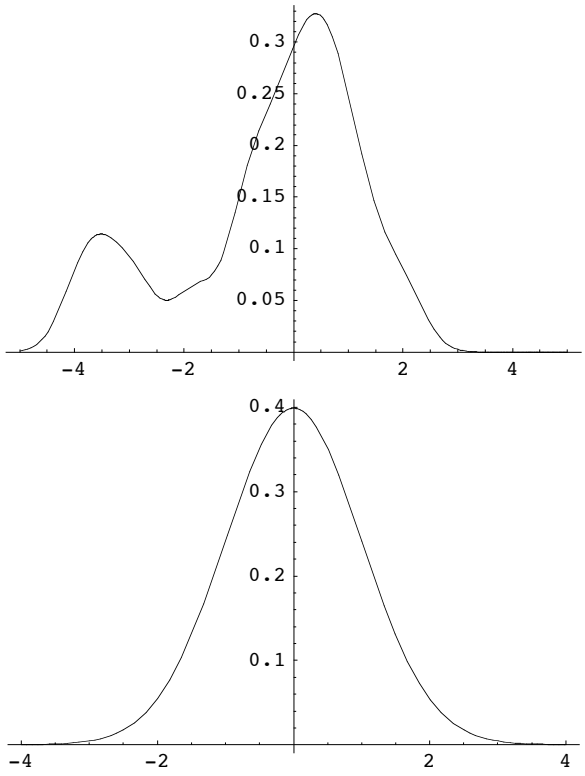


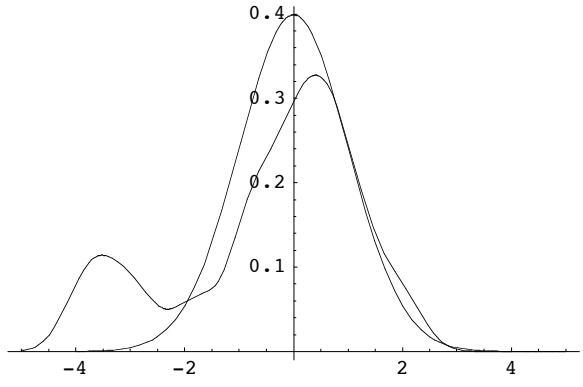


- Graphics -

Here is a plot of the senatorial standard scores shown together with the standard normal.

```
Show[Plot[Mean[phi[r - sen2z, 0.4]], {r, -5, 5}], Plot[phi[r, 1], {r, -4, 4}]]
```





- Graphics -