

# **ACCURACY OF POLLS IN PORTUGAL**

## **ABSTRACT**

The accuracy of pre-electoral polls has been the subject of studies in several countries, however in Portugal the issue is rarely studied, and even the few works on this matter are case studies without a general framework. This is mainly due to the fact that these polls are relatively recent in Portugal; it was only in 1991 that the law was modified allowing for the publication of pre-electoral polls.

This study has a vast analysis, studying almost all the pre-electoral polls published or issued in Portugal in the month previous to each of the elections, since 1991 until the last one that took place in February 2005. The accuracy measures I used were adapted from the study carried out by Frederick Mosteller in the report to the Committee on Analysis of Pre-election Polls, regarding the USA elections of 1948.

In the first part of this paper I present a contextualization of polls reality in Portugal, when have they been published for the first time, how they evolves, which laws regulate their utilization. I also present some examples of several studies of how we can measure the polls accuracy.

The second part is dedicated to the presentation of the results, beginning by a description of the sample involved in this study regarding some characteristics of the polls included in this sample, for example, their dimension and sampling techniques. Next are presented the accuracy results, their evolution, a comparison with other countries and relation with election results and some methodological characteristics of polls.

## PRE-ELECTION POLLS IN PORTUGAL

In result of the dictatorship in our country until April of 1974, pre-election polls had no relevance in that period, because they were not allowed. Besides, that, elections were only a deceitful way to make us believe we live in a democracy and, on the other hand, the news covering of electoral campaign was only state propaganda.

Any political inquiring would have always had serious problems in its organization, not only the existing censorship would cause serious obstacles, but also because people would be afraid of answering. In an opinion poll about French elections, carried out in the last days of dictatorship, almost half of people refuse to answer, as the journalist said: “The high percentage of people, who refused to answer, is due to the fear that people have from political police control of answers”<sup>1</sup>.

With the end of dictatorship and the first free elections, there were all the conditions to the beginning of pre-election polls publication. However, according to the government decree that regulate this elections: “it is forbidden to publish as a news report, as a interview, or other, any inquiry whose purpose is to show voter attitude or is vote intention on elections day” (article 62° of electoral law).

This prohibition were not impeditive of the realization of some studies, however, their results were never published directly:

-“according recent polls, the undecided voters percentage is about 71%” *Jornal Novo* (two days before election);

- “Someone has, in an abusive way, used the name of our Porto colleague *Jornal de Notícias* to make telephone polls about party affiliation of people.” *Diário de Notícias*, (29/03/1975);

---

<sup>1</sup> Although this study was carried out before 25 of April 1974, his publication was after the revolution.

- “ The present tendency of Portuguese public opinion is not favourable to the conservative ideologies, as has been quoted by opinion polls, whose origin and technical guarantee we do not know and come to us through foreign newspapers” *Vida Mundial* (13/3/75)

One interesting detail was the publication of the first pre-election poll in Portugal in April of 1974, one day after the election. This poll was carried out by IPOPE (first Portuguese company in public opinion research) about a month before the elections, and has involved 2000 respondents. The results of this poll were:

	PS	MDP	PCP	PPD	CDS	Outros
Poll	47	4	17	21	2	9
Election	38	4	13	26	8	11

Table 1 –Results of the first pre-election poll in Portugal and respective election results

In this poll we have two important results the overestimation of left parties (PS, MDP and PCP) and the underestimation of right parties (PPD and CDS). These two results, together with the high percentage of undecided and non respondents, seem to indicate some difficulty of the right electorate to express their voting intention.

In the following elections occurred the same situation, with almost any pre-election poll. Only in 1991, with the new legislation, this situation changed, according with the new law publication of pre-elections polls results were authorized if is taking place more than one week before Election Day. 16 pre-election polls<sup>2</sup> were carried out in the first election after this new legislation

The electoral legislation was changed again in 2000, allowing the publication of pre-election polls until two days before the Election Day.

---

<sup>2</sup> I consider pre-election poll any poll published in the last month before election.

## MEASURES OF POLL ACCURACY

The study of polls accuracy had developed mainly after the USA polling disaster of 1948. After these elections the Social Science Research Council had set up the Committee on Analysis of Pre-Election Polls and Forecasts with the objective of analysing the causes of the errors in forecasting the results of the 1948 election.

In the report presented to this Committee, Mosteller (1949) present eight different measures for poll accuracy. Seven of these measures are related with the differences between election results and poll estimates, although some of them only involve the winner or the first two candidates. The eighth measure is related with the difference between polls estimation of participation and the real election participation.

- (1) Measure the error in an estimate by the difference in percentage points between the predicted Democratic (Republican) proportion of the total votes cast and the actual Democratic (Republican) proportion.
- (2) Measure the error in an estimate by the difference in percentage points between the predicted Democratic (Republican) proportion of the two-party vote and the actual proportion Democratic (Republican) of the two-party vote.
- (3) Measure the error by averaging the deviation in percentage points between predicted and observed results for each party (without regarding the sign).
- (4) Use the concept of average percentage error, taking the ratio of predicted to actual proportion, and averaging the deviations from 100 percent.
- (5) Use the difference of the oriented differences between predicted and actual results for the two major candidates
- (6) Use the maximum observed difference in percentage points for any party.
- (7) Use the chi-square test.
- (8) Use the electoral vote predicted versus that observed.

Mitofsky (1998) refers to some lack of consensus about the best way of measure poll accuracy. This author considers Mosteller's measures 3 and 5 as the best choices, although he prefers measure 5 because it is independent of the number of parties in the election, and, also, it is the most used in the publication of results.

The National Council on Public Polls (2002) also chooses measure 5, although divided by 2, in their report about the performance of polls in the USA presidential elections of 2002.

Analysing a sample of polls, carried out between 1979 and 1986, Crespi (1988) used Mosteller's measures 1, 3 e 6, but in the last two he only considered the three first parties or candidates.

Lau (1994) considered another measure, different from Mosteller's measures because it not involves election results, the poll estimate is compared with average result of all the other polls (weighted by sample size). This author justify this measure arguing that he can't use the others measures for judging polls accuracy because it's impossible to assume the stability in the support for the candidates until the election day, specially for polls carried out about two months before the elections.

Reporting the performance of the polls in the 2004 presidential election Traugott (2004) used Mosteller's measures 3 and 5. This author also used another measure, developed by Martin et al (2004), which combine the measure of predictive accuracy with the measure of bias, or the systematic over-or-under-estimate of a given party.

The error calculated by these measures is dependent on how the polls results are presented, with the undecided or with the allocation of the undecided, since considering the undecided has as a consequence, a systematic underestimation of parties and, also, an inadequate comparison with other polls where the allocation of undecided is done.

Mitofsky (1998) considers this problem an important issue in the evaluation of polls accuracy. He refers that although this topic was not analysed in 1949 in the report of the Committee on Analysis of Pre-Election Polls and Forecasts, because at that time it wasn't a major concern, Mosteller's measures become more consistent between themselves when the undecided are allocated proportionally. Martin et al (2003) and Crespi (1988) also considers that the most efficient option is to allocate proportionally the undecided.

In spite of these opinions, this procedure is not unanimous, the National Council on Public Polls (2002), in their report about the performance of polls in the USA presidential elections of 2002, didn't use the undecided allocation.

In this study I consider, in all polls with the undecided voters, the proportional allocation of them. I also consider the proportional allocation of white and non valid votes in the election results, since these votes are rarely presented in polls results.

In the present study I used four of the Mosteller's measures, these measures were chosen mainly because they are also used in other investigations, which allow a comparison with their results. The measures I consider are:

$$\text{Measure 1} = |\text{Party 1 result} - \text{Party 1 poll estimate}|$$

where Party 1 is the election winner

$$\text{Measure 3} = \frac{\sum_{i=1}^4 |\text{Party } i \text{ result} - \text{Party } i \text{ poll estimate}|}{4}$$

$$\text{Measure 5} = |(\text{Party 1 poll estimate} - \text{Party 2 poll estimate}) - (\text{Party 1 result} - \text{Party 2 result})|$$

where Party 1 is the first in poll estimates  
and Party 2 is the second in poll estimates

$$\text{Measure 6} = \text{Máximum}_{i=1}^4 |\text{Party } i \text{ result} - \text{Party } i \text{ poll estimate}|$$

In measures 3 and 6 I only consider the results of the main four parties: PS; PSD; CDU and PP.

## THE DATA

Portuguese legislation enforces that, previously to any publication of every opinion poll about politics, a report of polls methodology and results should be delivered to High Authority for Social Communication<sup>3</sup> (AACS). About 2272 reports of opinion polls, since 1991, as a result of this law are in this institution.

As the focus of this study is pre-election polls, I didn't work on all these reports, I just analysed polls which publication was less than a month before the Election Day. From these polls I excluded the exit polls because they have objectives and methodological characteristics very different from regular pre-election polls.

In these conditions I found in AACS 394 reports, however, 46 of them were incomplete not having a copy of the newspaper report or even the results. I also exclude other 2 polls that, in spite of having a complete report and being published less than a month before the Election Day, their work in the field was about two months before this day. So, as a result, I work with a sample of 346 polls which represent about 88% of all pre-election polls published in Portugal.

Analysing the distribution of polls by the nature of elections related to them, I would highlight the fact that more than a half of these polls are about Town Hall elections. This fact has the consequence of strongly increase the weight of local polls in this sample, about 70%.

	N	%
Parliament	106	31
Town Hall	191	55
Regional Parliament	7	2
European Parliament	27	8
Presidential	15	4

Table 2 – Distribution of polls by the nature of elections

---

<sup>3</sup> AACS is an institution under supervision of national parliament.

Considering the sampling method of the polls in this sample, more than a half is based on a random procedure. This distribution is similar to the one find by Crespi (1988) in their study, where 54% of polls have also a random sampling method.

	N	%
Random	195	56
Quotas	151	44

Table 3 - Distribution of polls by sampling method

It is, however, interesting to say that this distribution is not homogeneous regarding the scope of the poll. Polls with random sampling and quota sampling are equally distributed among national polls, while at local polls random sampling, are the majority, about 60%.

Almost half of polls have carried out their inquiring using telephone interview, while the face to face interview was the least technique used.

Face to face Interview	65	19
Telephone Interview	169	49
Ballot simulation	90	26
Not mentioned	22	6

Table 4 - Distribution of polls by inquiring technique

About half of the polls have samples with less than 800 respondents. Also in this case the distribution is different among national and local polls, for example, all polls with samples less than 400 respondents are local and the great majority of polls with samples larger than 1000 respondents are national.

	N	%
400 or less	70	20
401 to 800	92	27
801 to 1000	89	26
More than 1000	86	25
Not mentioned	9	3

Table - 5 Distribution of polls by sample dimension

## RESULTS

One of the most relevant result is the strong correlation between all these error measures. These results are similar with the ones obtained by Crespi (1988), who also find higher correlations between Mosteller's measures 1, 3 and 6 (this author didn't use measure 5).

Pearson Correlation Coefficient			
	Measure 3	Measure 5	Measure 6
Measure 1	0,81	0,82	0,84
Measure 3		0,88	0,93
Measure 5			0,89

Table 5 - Pearson Correlation Coefficient between error measures

Considering the mean of the four measures, presented in table 6, we may say that the accuracy of polls in Portugal is somehow low, for example, the winner is estimated, in average, with an error larger than 5% (Measure 1).

Also interesting to highlight is the extremely high error that some polls present, for example, in one of them the error in the estimate of the difference between the two first candidates reach almost 36%.

	Minimum	Maximum	Mean	Standard Deviation
Measure 1	0,02	18,47	5,27	3,94
Measure 3	0,21	11,95	3,71	2,29
Measure 5	0,00	35,76	9,03	7,16
Measure 6	0,21	21,99	6,60	3,95

Table 6 – Descriptive statistics of error measures

The comparison with the results in other studies highlight the poor performance of Portuguese polls in term of accuracy, although Portuguese results being similar with the ones obtained by Crespi (1988), they are significantly lower than results in other

studies. The English polls in 1992 also present similar results, but those elections were characterized by a large failure of polls industry, all polls estimating a Labors winning, but the elections winner was the Conservative Party.

	Portugal 1981-2005	EUA (sample)	EUA 1956-2004	EUA 2002	EUA 2004	GB 1992	GB 1997
Winning error (measure 1)	5,27	5,67					
Mean Error (measure 3)	3,71	5,68	1,8		1,2	2,7	2,2
Difference error (measure 5)	9,03	-	3,3	4,8	1,9	9,0	4,8
Maximum error (measure 6)	6,60	6,75					
FONTE		Crespi (1988)	Traugott (2004)	NCPP (2002)	Traugott (2004)	Moon (2005)	Moon (2005)

Table 7 – Average error measure in several studies

A possible explanation for the lower accuracy of polls in Portugal may be related with the gap of time between their work field and the elections day, only after 2000 was allowed to publish polls results in the last week before elections day, while in USA and Great Britain there is no such restriction.

The great majority of the polls in this study ended their work in the field more than 12 days away from Election Day, while in Crespi (1988) study only 31% of polls were on these conditions.

Comparison of results between polls that ended their work in the field more than a week before the Election Day and the ones that ended in the last week, confirms this hypothesis, since all measures considered indicate lower errors among this last ones.

Crespi (1988) reach similar results, concluding, also, that the number of days between the work in the field and the Election Day is affecter that influences the estimate error, too.

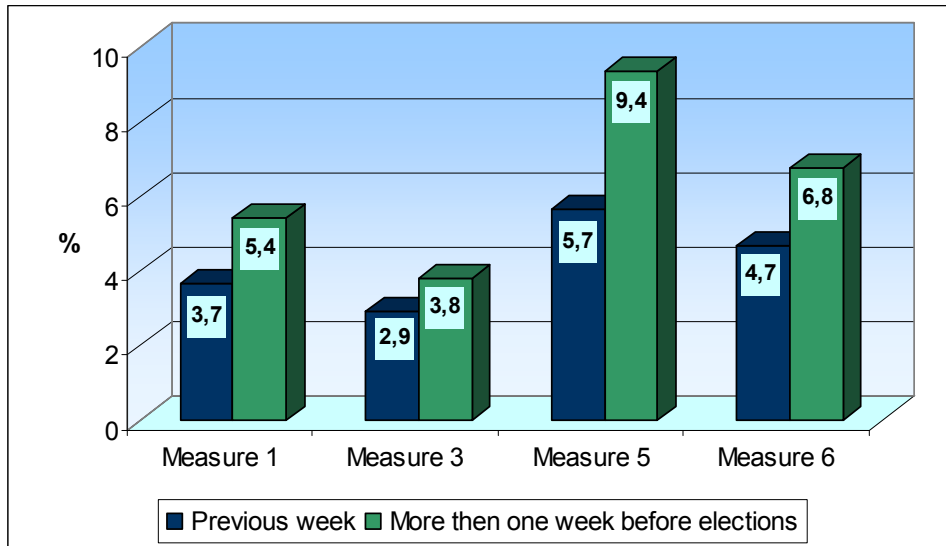


Figure 1 – Average error measure and gap of time between polls inquiring and elections day

### Accuracy Evolution

Analyzing measure 5<sup>4</sup> evolution since 1991, first year of pre-election poll publications, we can see a great irregularity, with constant variations of average error. The worst year happens in 1993 when the average error exceeded 12%, on the other hand, this year, for the first time, the average error is lower than 4%.

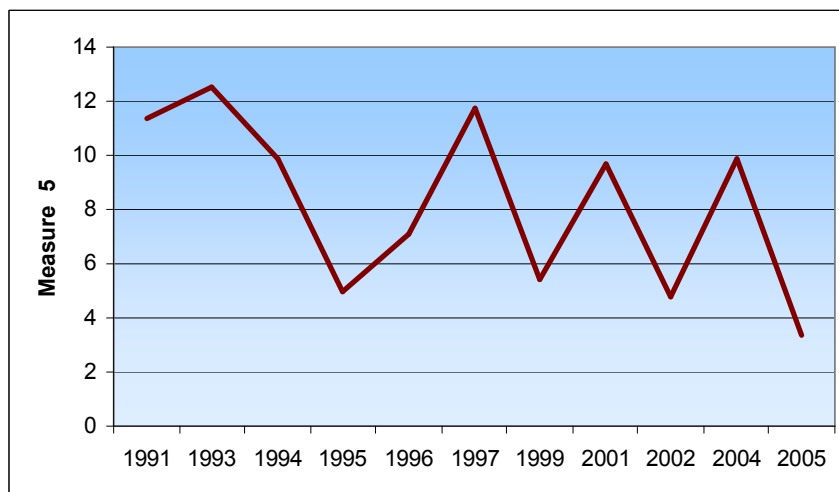


Figure 2 – Evolution of measure 5 average.

<sup>4</sup> Since the four measures I considered were strongly correlated, there is no advantage in working with all of them at the same time, so I choose measure 5, because it was used in several studies, which allows the comparison with them. Mosteller (1949), referring to this measure, said that their only problem is their complexity of explaining.

All this variation seems to point out that different kind of elections have different performances of polls accuracy. If we consider the type of elections in the evolution of polls accuracy, major differences can be found between the various kind of elections, especially Parliament and Town Hall, with this last ones having nearly twice the average error of the first ones.

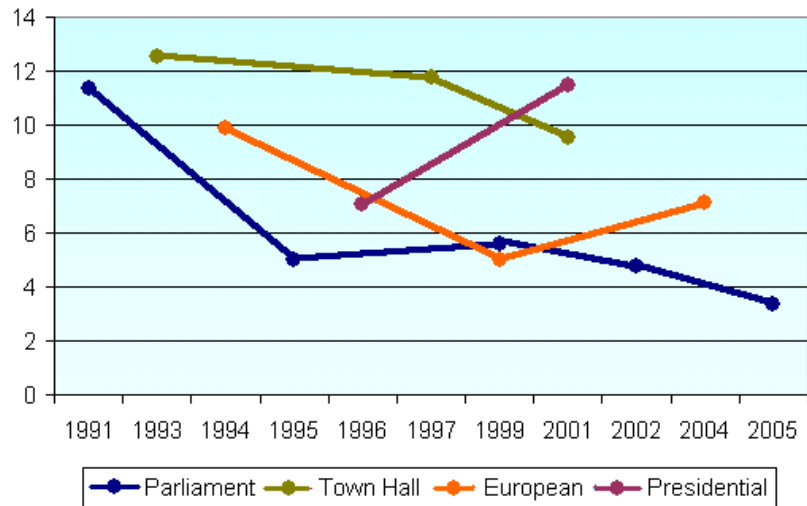


Figure 3 – Evolution of measure 5 average and type of election.

These results raise the question of possible differences between local and national polls in terms of their accuracy. Figure 4 show that local polls have, also, an average error twice larger than national polls. This result is not only due to the fact that local polls are majorly about Town Hall elections, even in local polls about Parliament elections, local polls have an average error of 6,9%, while in national polls the average error was 4,8%.

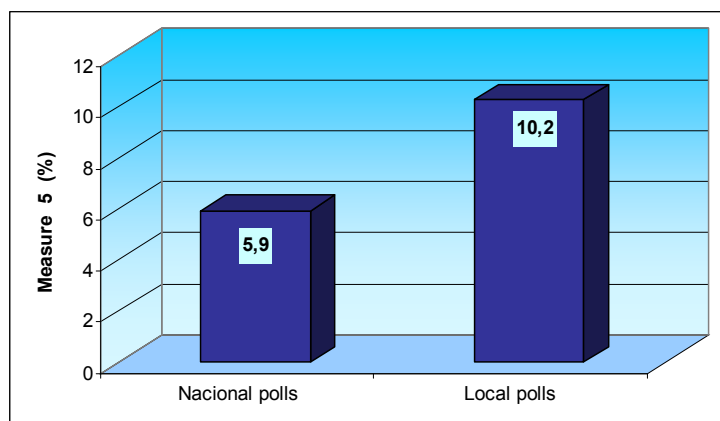


Figure 4 – Measure's 5 average in local and national polls.

The lower accuracy of local polls is certainly related with logistic and methodological issues. Local polls are normally less important, which means less investment and in consequence fewer resources for their organization.

### Accuracy and Polls Methodology

Analyzing the relation between methodological issues and polls accuracy, and beginning by sampling method, we can see that there are no significant differences on accuracy between polls with random sampling and polls with quota sampling. Although polls with quota sampling have an average error slightly lower. Crespi (1988) also find no relation between polls accuracy and sampling method.

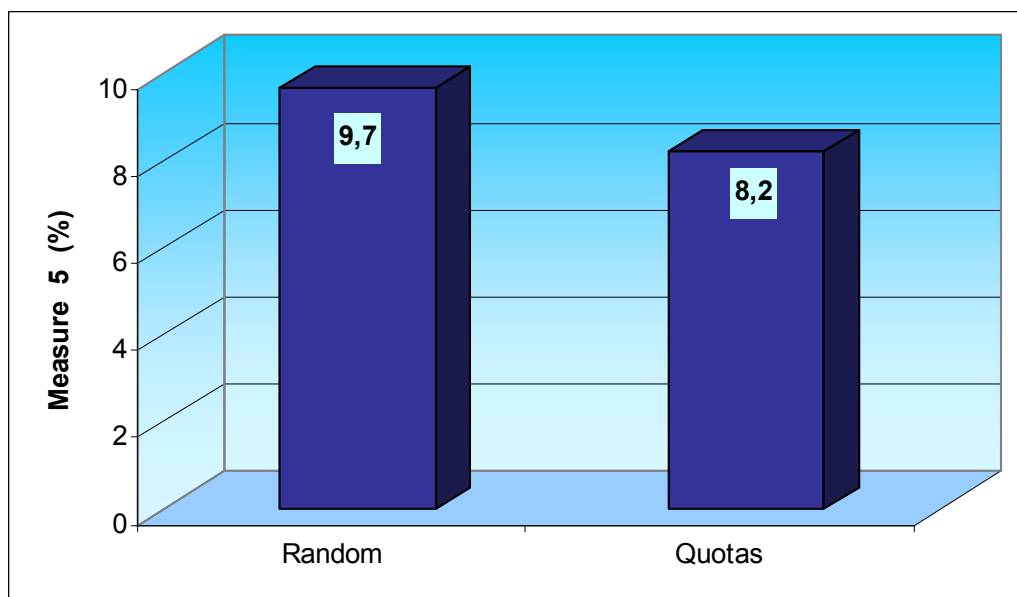


Figure 5 – Measure’s 5 average by sampling method.

The sample dimension is significantly correlated in a negative way with polls error ( $r = -0,127; \alpha \leq 0,02$ ). As an example, polls with samples smaller than 400 respondents have an average error of 12%. These results are similar to the ones obtained by DeSart and Holbrook (2003), who concluded, in their study about USA 1996 e 2000 election polls, that polls with smaller samples tend to be less accurate.

These results are one of the possible explanations for the local polls lower accuracy, since the dimension of their samples is normally smaller than in national polls.

Crespi (1988) and Lau (1994), however, don't reach the same conclusion; both of them have considered that polls accuracy and sample dimension are not correlated.

Another issue related with sampling procedure is the number of days of work in the field. Both Crespi (1988) and Lau (1994) have concluded that this variable is one of the most important in the explanation of polls accuracy, according to these authors, polls with more days in the field are more accurate. Lau (1988) had considered that each day more of work in the field increases polls accuracy in a half percentage point.

The great majority of Portuguese pre-election polls, about 80%, spend one to five days of work in the field, however, I didn't find any significant differences between these polls and polls with more days of work in the field.

Crespi (1988) and Lau (1994) have also studied other issues related with sampling procedures, for instance, Lau considers that avoiding weekend days improves polls performance. Crespi concluded that polls that conduct call-backs when handling with "not-at-homes" are more accurate. Concluded also that polls in which the respondent is chosen as the oldest or youngest man or woman in the household were the most accurate. On the other hand, this author refers that polls with no sample stratification don't have significant differences of accuracy with polls that uses stratification.

In the present study I didn't analyze these issues because they aren't available in polls report. However and also related with sampling procedures, I find that polls in which the inquiring is made by ballot simulation are the most accurate. On the other hand, face to face interviews are the least accurate.

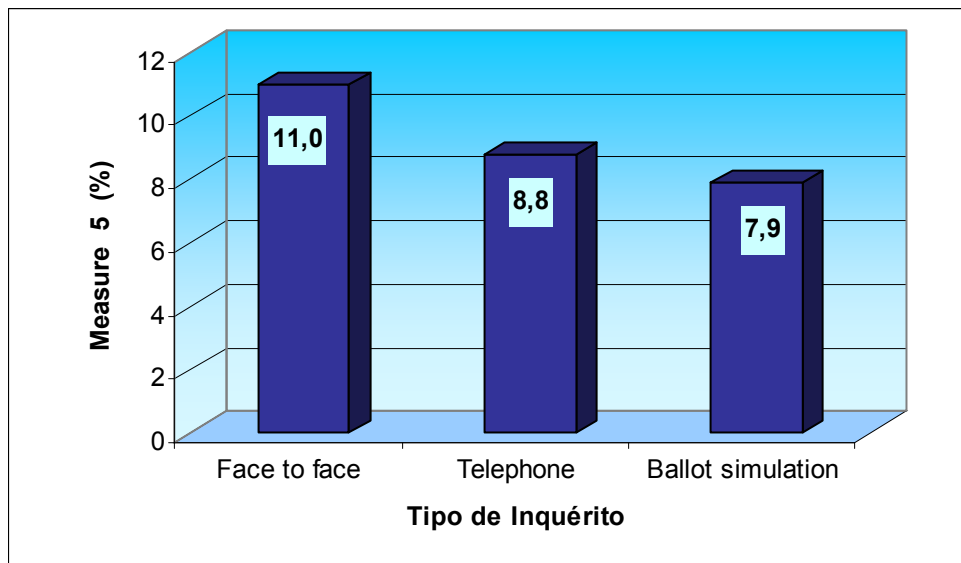


Figure 6 – Measure’s 5 average by inquiring procedure.

Another aspect of polls analysed is the refusal rate, however its correlation with polls accuracy is almost zero. On the other hand, the number of undecided respondents is positive and significantly correlated with polls error ( $r = 0,127; \alpha \leq 0,02$ ), which means that polls with less undecided are more accurate.

### Accuracy and Election Results

Analyzing the election results, I find that the variable less correlated with polls accuracy is the turnout rate, Pearson correlation coefficient between turnout and polls error is nearly zero, which means the absence of any relation. Traugott, quoted by Crespi (1988), also considered that polls methodology works equally in elections with high or low turnout. Crespi (1988), however, find significant negative correlation between turnout and polls accuracy, concluded that elections with larger turnout are more accurate.

The election result more correlated with polls accuracy is the difference between the two first parties, the correlation of this difference with polls error is significantly positive ( $r = 0,21; \alpha \leq 0,00$ ), which means that elections with a previous “known winner” are less accurate. Crespi (1988) also reached the same conclusion.

Another election result studied was the changes between election results and previous election results. Considering polls related to elections with a different winner from previous election and polls related to elections with the same winner, we can see that these last polls are less accurate.

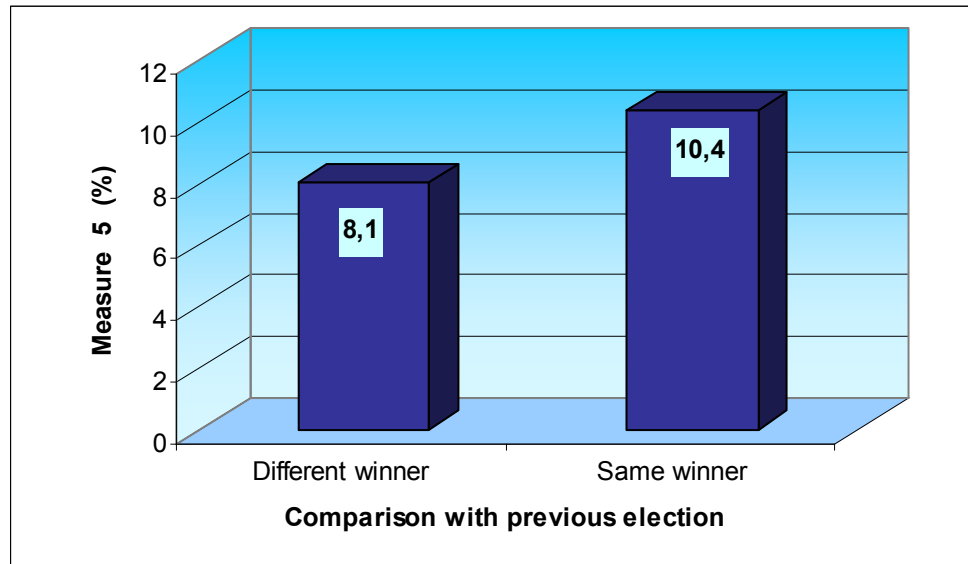


Figure 7 – Measure 5 average and comparison with previous election.

This situation, however, is related to the previous situation, since the difference between the first parties is larger in elections with the same winner (average of 18%) than in elections with different winner (average of 9%).

Another way to analyse the relation of polls accuracy with the changes between actual election and the previous one, is considering an index of the difference between both elections, based on qui-square statistic. However I don't find any significant correlation of this index with polls accuracy.

## CONCLUSIONS

All the error measures considered in this study are strongly correlated between themselves. Also, they have similar results in terms of relations with other variables, although measure 3 offers less significant differences.

On the other hand, all these measures stressed the differences between polls estimates and elections results, but don't reflect any bias, because they don't consider the possible existence of a systematic under or over estimation of a certain party, since they work with absolute value of errors.

It will be interesting to enlarge this study, in order to investigate bias in Portuguese polls. This extension of the study could be based on the Predictive Accuracy Measure, proposed by Martin et al (2003). This measure not only evaluates the estimates error, but also considers which party the error favours.

Polls accuracy in Portugal have been increasing, after a period in the first years of the 1990 decade where polls produce large errors, last pre-election polls reach good levels of accuracy. This accuracy improvement has been clearer in polls related with parliament elections.

Sample dimension was the methodological issue more correlated with polls accuracy; polls with larger samples tend to be more accurate. Also polls with ballot simulation are the most accurate, especially in comparison with polls with face to face interview. On the other hand the sampling method, probability or nonprobability sampling, doesn't reveal any relation with polls accuracy.

While other studies refer to a relation between elections turnout and polls accuracy, I didn't find any relation between these two variables. The stronger relation between election results and polls accuracy is the proximity between the first two parties, polls related to elections with a "previously" known winner tend to be the least accurate.

## BIBLIOGRAPHY

Crespi, Irving, Pre-election Polling: Sources of Accuracy and Error, Russel Sage Foundation, New York, 1988.

DeSart, J. and Holbrook, T., Campaigns, Polls and States: Assessing the Accuracy of Statewide Presidential Trial-Heat Polls, *Political Science Quarterly*, vol. 56: 431-439.

Lau, R. R., An Analysis of the Accuracy of “Trial Heat” Polls During the 1992 Presidential Election, *Public Opinion Quarterly*, Vol. 58: 2-20, 1994

Martin, E.A.; Traugott, M.W. and Kennedy, C., A Review and Proposal for a New Measure of Poll Accuracy, AAPOR 58<sup>th</sup> Annual Conference, Nashville, 2003.

Mitofsky, W., The polls Review: Was 1996 a Worse Year for Polls than 1948?, *Public Opinion Quarterly*, Vol. 62: 230-249, 1998.

Mosteller, F.; Hyman, H.; McCarthy, P.J.; Marks, E.S. and Truman D.B., The Pre-election Polls of 1948: Report to the Committee on Analysis of Pre-election Polls and Forecasts, Social Science Research Council, New York, 1949.

National Council on Public Polls Polling Review Board, Analysis of the 2002 Election Polls, Press release: 2002/12/19, available in <http://www.ncpp.org/2002SenGovPoll/2002ElectionPolls.html>

Traugott, M. W., The performance of the Polls in the 2004 Presidential Election, WAPOR regional conference on “Elections, News Media and Public Opinion”, Pamplona-Spain, 2004.