Some Issue in Opinion Polls

Common Approach

- Telephone survey
- Use listed numbers or Random Digit Dialling (RDD)
- Sample sizes usually 1200 to 1600
- One person selected per selected household
- Independent sample each occasion
- Geographic coverage - national, state, specific electorates
- Sample weighting used by age and sex and ?

Sampling Frames Issues

Some surveys use the numbers listed in the electronic white pages (EWP) as the sampling frame.

This is convenient but will not represent households with unlisted numbers and those that have yet to be put on the EWP.

The proportion of households with unlisted telephone numbers is thought to be at least 15 percent in Australia.

For this reason some form of Random Digit Dialling (RDD) can be considered.

ABS data on telephone connections for Queensland for October 2003:

- **95% of households were connected**
- **17% were unlisted**
- **68% percent of people 18 or over had mobile telephones**
  - **this rate varied considerably by age**

Provided people with mobile telephones can be contacted through a landline to the household in which they live, then the popularity of mobile telephones does not reduce the coverage of the EWP or RDD approaches.

The 2003 survey found only 3% people had a mobile telephone but no landline but this rate was almost 7% for people aged 18-29.

Over time the coverage of the population by traditional telephone sampling methods will decrease as more people only have a mobile.

This could happen quite quickly and is likely to affect some parts of the population more than others.

If the people who are not covered are different from those that are, then this can result in a bias in the estimate, which can be called coverage bias.

Margin of Error

Under some assumptions the margin of error due to sampling on an estimated proportion is

$$2\sqrt{\frac{P(1-P)}{n}}$$

Where *P* is the proportion and *n* is the sample size.

Example: ***P*=0.5 and *n*=1600 then we get 0.025 or 2.5%**

***P*=0.5 and *n*=1200 then we get 0.029 or 2.9%**

It is a simple exercise to show that for given *n* the margin of error is maximum when *P*=0.5.

Analysis of Estimated Lead

There is often focus on the ALP lead.

If *P* is the estimated ALP proportion on a 2 party preferred basis, then the estimated lead is

$$\text{Lead} = P-(1-P)=2P-1$$

Example: *P*=0.55 gives a lead of 0.1 or 10%

Because of the multiplication by 2 the margin of error on the estimated lead is twice that on the estimated proportion. So for *n*=1200 the margin of error on the estimated lead will be 5%

Combining Several Polls

One approach to reducing the margin of error is to combine the results of two or more polls. The sample sizes may differ. What is the best way to combine the estimates from two polls?

Consider the combined estimate

$$\hat{P} = w_1 P_1 + w_2 P_2$$

where the weights, w_1**, and** w_2 **have to be chosen.**

Assuming each estimate is unbiased we need

$$w_1 + w_2 = 1 \textbf{ so } w_2 = 1 - w_1$$

Thus we can write

$$\hat{P} = wP_1 + (1 - w\ )P_2$$

We want to minimise the variance of the estimator, which determines the margin of error:

$$Var(\hat{P}) = w^2 Var(P_1) + (1-w)^2 Var(P_2)$$

$$= w^2 \frac{P(1-P)}{n_1} + (1-w)^2 \frac{P(1-P)}{n_2}$$

To find the minimum, differentiate with respect to w and set to zero, giving

$$2P(1-P)\left[\frac{w}{n_1}-\frac{1-w}{n_2}\right]=0$$

Some rearrangement gives

$$w=\frac{n_1}{n_1+n_2}$$

Therefore, we should use

$$\hat{P}=\frac{n_1}{n_1+n_2}P_1+\frac{n_2}{n_1+n_2}P_2$$

So each survey should be given a weight proportional to its sample size. This result be generalised to combining more than two polls.

If there is a trend in the proportions over time then more sophisticated time series methods can be used to combined different polls over a period of time.