Chapter 5

Statistical Inference:

To draw conclusions about population on the basis of sample data. Two procedures (a) estimations (b) Hypothesis Testing.

(I) Estimation

e.g. a. Estimate average starting salary of college graduate

b. Estimate average height/weight of UC students.

In general, estimates based on a sample of data did NOT equal to the TRUE value of the population parameters and will vary from sample to sample. For this reason, the estimates themselves are random variables and have probability distribution.

Def^n: Estimator & estimates

Suppose we have a random sample $x_1, x_2, \ldots x_n$ of observations from some population. An estimator of a population parameter is a rule that tells us how to use the values $x_1, x_2, \ldots x_n$ to estimate the parameter. An estimate is the value obtained after the observations $x_1, x_2 \ldots x_n$ have been substituted into the formula.

e.g. $\bar{x} = \frac{\Sigma x_i}{n}$ is an estimator of μ & denoted by $\hat{\mu}$. Let $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4, x_5 = 5$.

$\bar{x} = \frac{15}{5} = 3$ is the estimate.

Point estimate: a single numerical value calculated from sample data & taken to be indicative of the value of the population parameters.

Ex. How many hours of sleep do college students (college residence) get on a typical weeknight? A r. s. of 22 students living in college residence we asked about their average length of weeknight sleep. The reported weeknight sleep duration had a mean of 7.05 hours and a s.d. of .86 hr.

The Sample mean $\bar{x} = 7.05$ is a point estimate of the mean μ of weeknight sleep durations that would be reported by all students living in college residence. The sample s.d. $s = .86$ is a point estimate of the s.d. σ of weeknight sleep durations that would be reported by all members of the populations. How closely the reported sleep durations corresponds to actual sleep duration is, of course, an open question.

The chief drawback of point estimate is that they provide NO information about their precision.

Interval Estimate

When we try to evaluate the goodness or reliability of an estimator $\hat{\theta}$, we are in general to put some bounds on the possible error of estimation $|\hat{\theta} - \theta|$, where θ represents the true value of the parameter being estimated. The error of estimation $|\hat{\theta} - \theta|$ called the sampling error, measures the distance between the estimated value and the true value of the population parameter.

A systematic method of indicating the precision of an estimator $\hat{\theta}$ exists, provided we know the form of the sampling distribution of θ. We indicate the precision of an estimator by constructing confidence intervals for θ. We use the estimate $\hat{\theta}$ to determine two values $\hat{\theta}_1$, $\hat{\theta}_2$ such that the intervals $(\hat{\theta}_1, \hat{\theta}_2)$ contains the value θ with a specific probabilities. The probability is usually denoted as $1-\alpha$ and the percentage $100(1-\alpha)\%$ is called the confidence level of the confidence interval $(\hat{\theta}_1, \hat{\theta}_2)$

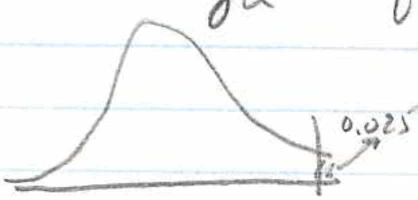
Value of z_α

$Z \sim N(0,1)$ & let α be any number such that $0 < \alpha < 1$

Then z_α denotes the number for which

$$P(Z \geq z_\alpha) = \alpha.$$

e.g. Find z_α if $\alpha = 0.025$



From table $P(0 \leq Z \leq 1.96) = .475$

$\therefore z_\alpha = 1.96$

e.g. Find z_α if $\alpha = .05$, $\alpha = .005$ ...

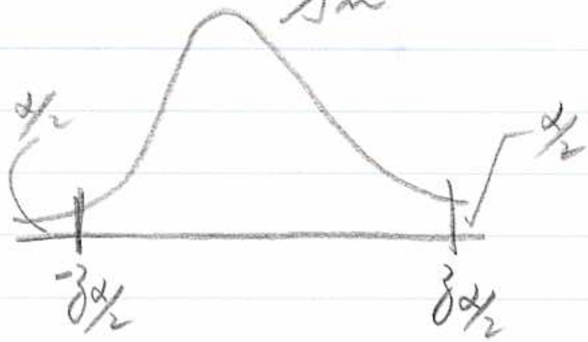
Confidence Interval for Mean with Known Population Variance

Suppose we take a r.s. of size n from a normal population having mean μ and variance σ^2

i.e. $X \sim N(\mu, \sigma^2)$.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$



$$1-\alpha = P\left\{-z_{\alpha/2} < Z < z_{\alpha/2}\right\}$$

$$= P\left\{-z_{\alpha/2} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right\}$$

$$= P\left\{\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

which tells us that the RANDOM interval $\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$ will contain the true pop. mean μ with probability of $1-\alpha$, where

$z_{\alpha/2}$ is the number for which $P(Z > z_{\alpha/2}) = \alpha/2$ & $Z \sim N(0,1)$

e.g. Construct a 95% C.I. for the mean:
A student advisor wants to estimate the mean annual income of all college students who graduated last year. The population s.d. is believed to be \$2000. Based on a random sample of 25 college graduates the advisor obtains $\bar{x} = \$19500$. Construct 95% C.I. for the unknown population mean μ.

solution :

$\sigma = \$2000 \quad n = 25, \quad \bar{x} = 19500, \quad 1-\alpha = 0.95$

$$1-\alpha = P\left\{-1.96 \le z \le 1.96\right\}$$

$$= P\left\{-1.96 \le \frac{\bar{x}-\mu}{(\sigma/\sqrt{n})} \le 1.96\right\}$$

$$= P\left\{-1.96\frac{\sigma}{\sqrt{n}} \le \bar{x}-\mu \le 1.96\frac{\sigma}{\sqrt{n}}\right\}$$

$$= P\left\{\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right\}$$

95 % C.I. for μ :

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

Now $\bar{x} = 19500$, $\sigma = 2000$, $n = 25$

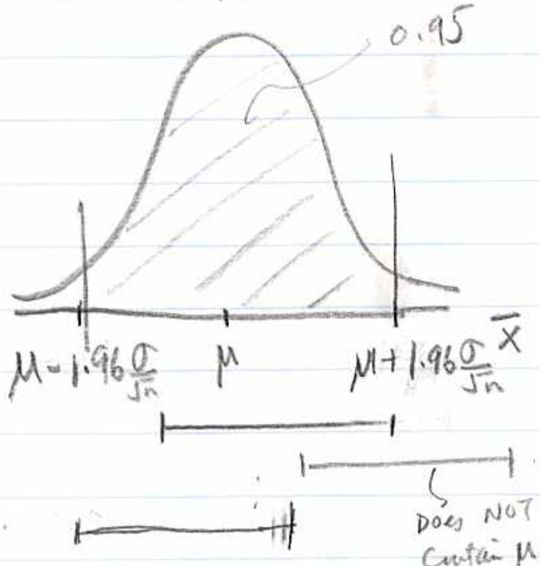
$\therefore$ 95% C.I. is (18716, 20284)

We are 95% <u>confident</u> that the population mean is between \$18716 and \$20284.

We <u>CAN NOT</u> sure <u>this interval</u> contains the population mean, but if we repeat this process a large number of times, 95% of the confidence intervals would contain the population mean.

This confidence interval depends on the specific value $\bar{x} = 19500$ obtained from the sample of 25 observations. Suppose we take another r.s. of 25 observations, calculate a new value of $\bar{x}$, and obtain the 95% C.I.

$\left(\bar{x} - 1.96\frac{\sigma}{5}, \bar{x} + 1.96\frac{\sigma}{5}\right)$. Like the first C.I., this one may or may NOT contain the population mean μ. If we repeated this process, say 1000 times, we would have 1000 different sample means and 1000 different C.I. . A 95% level of confidence means that approximately 95% (or 950) of these C.I. would contain μ and 5% (or 50) would NOT.



$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

It is important that the probability statement

$$P\left\{\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha$$

be interpreted correctly. In this statement, the parameter mean μ is NOT a random variable and does NOT vary from sample to sample, rather the mean μ is an UNKNOWN population parameter.

On the other hand, $\bar{x}$ is a random variable and varies from sample to sample. If we take many samples of size n from the population we get a different value of $\bar{x}$ for each sample. For a particular estimate $\bar{x}$, we can calculate the endpoints of the interval

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$ these endpoints

vary from sample to sample. The probability statement says that $100(1-\alpha)\%$ of these random intervals contain the value μ. We say that we are $100(1-\alpha)\%$ confident our interval contains μ because, essentially our interval is just one of many possible intervals.

Margin of error: $e_{\alpha/2} = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Sample size required for a given margin of error:

$$e_{\alpha/2} = z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \quad n = \left(\frac{z_{\alpha/2}\,\sigma}{e_{\alpha/2}}\right)^2$$

If the resulting value is NOT an integer, the next larger integer should be taken for the required sample size

Eg. Consider light-bulb manufacturing process, given that the true s.d. σ of bulb lifetime is 150 hrs. How large a random sample would the quality control staff need in order to be 95% confident that the sample mean would be within 20 hours of population mean?

Solution:

$$1-\alpha = P\left\{ -z_{\alpha/2} \le \frac{\bar{X}-\mu}{(\sigma/\sqrt{n})} \le z_{\alpha/2} \right\}$$

$$= P\left\{ -\frac{\sigma}{\sqrt{n}} z_{\alpha/2} \le \bar{X}-\mu \le \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \right\}$$

$$\therefore \quad z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 20 \Rightarrow \sqrt{n} = \frac{z_{\alpha/2}\,\sigma}{20} = \left(\frac{1.96 \times 150}{20}\right)^2$$

$n = 216.09$, Sample size required is 217

How about 99% C.I.

replace $z_{\alpha/2} = 1.96$ by $z_{\alpha/2} = 2.58$

We have

$$n = \frac{(2.58)^2(150)^2}{400} = 374.4225$$

$\therefore$ Sample size required is 375

Formula for commonly constructed Confidence Interval - Known Variance

Level of Confidence $1-\alpha$	α	$\alpha/2$	$z_{\alpha/2}$	C.I. $\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$
.9	.1	.05	1.645	$\left(\bar{x} - 1.645\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.645\frac{\sigma}{\sqrt{n}}\right)$
.95	.05	.025	1.96	$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$
.99	.01	.005	2.58	$\left(\bar{x} - 2.58\frac{\sigma}{\sqrt{n}},\ \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}\right)$

Desirable Properties of C.I.

1. The interval should have a high level of confidence $1-\alpha$.
2. C.I. should have a narrow width

In most cases we would like the probability that our C.I. contains the mean to be very high, say 90% or more. We also would like the C.I. to be very narrow, so that an estimate is precise.

Defn Width of a C.I. for μ

The width W of a C.I. for the pop. mean is $W = 2\, z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Width of C.I for the mean depends on:

1. level of confidence of the C.I. $1-\alpha$
2. s.d. of population: σ
3. Sample size n

Properties of C.I.

1. If σ, n are constants, $1-\alpha \uparrow$, $W \uparrow$.
   e.g. a 99% C.I. will be wider than 95% C.I.
2. $\sigma \downarrow$, $W \downarrow$. If the population is highly concentrated our estimate of the mean is very reliable. Because the estimate is very reliable, only a narrow confidence interval is required.
3. $n \uparrow$ $W \downarrow$! As we obtain more information, our estimate should become better as reflected by a narrower C.I. Note that in order to CUT the width of a C.I. into half, it is necessary to multiply the sample size by a factor of 4.

σ fixed, n_1 for W_1

$$W_1 = \frac{W}{2} \qquad \text{let size } n \text{ corresponds to } W.$$

$$\frac{\sigma}{\sqrt{n_1}} z_{\alpha/2} = \frac{1}{2} z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \sqrt{n_1} = 2\sqrt{n}$$

$$\Rightarrow n_1 = 4n$$

To decrease the width of C.I. we must either use a smaller level of confidence $(1-\alpha)$ which decreases $z_{\alpha/2}$ or increases the sample size n. By making the sample size larger & larger one can make the C.I. (for any value α) as narrow as desired, but at an <u>increased</u> cost of sampling.

<u>Confidence Interval for mean with unknown population Variance</u>

If the observed sample mean is $\bar{x}$ and the observed sample s.d. s then the confidence interval for the mean having level of confidence $100(1-\alpha)\%$ is given by

$$\left(\bar{x} - t_{(\frac{\alpha}{2}, n-1)} \frac{s}{\sqrt{n}}, \ \bar{x} + t_{(\frac{\alpha}{2}, n-1)} \frac{s}{\sqrt{n}}\right)$$

where $t_{\alpha/2, n-1}$ is the number such that $P(t > t_{\alpha/2, n-1}) = \frac{\alpha}{2}$

e.g. A newly hired employee of the U.S. Postal Service wants to estimate the mean annual income of first-year mail carriers. Assume the population of incomes is approximately normal and the population

variance is unknown. The employee takes a r.s. of 25 1^{st}-year carriers & obtains $\bar{x} = \$19500$, & $s = \$2000$
Construct a 95% C.I. for the unknown population mean μ.

Solution: $n = 25$, $\bar{x} = 19500$, $s = 2000$, $1-\alpha = .95$
Thus we have $\alpha = 0.05$, $\frac{\alpha}{2} = .025$. $d.f. = 24$

$$t_{.025, 24} = 2.064.$$

C.I. is $\left(\bar{x} - t_{\alpha/2, n-1} \frac{s}{\sqrt{n}},\ \bar{x} + t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}\right)$

or $\left(19500 - 2.064 \frac{(2000)}{\sqrt{25}},\ 19500 + 2.064 \frac{2000}{\sqrt{25}}\right)$

or $(18674.40, 20325.60)$

Thus we have 95% confident that the mean income of the population is between \$18674.40 & \$20325.60

The C.I. based on t-score is always WIDER than the C.I. based on Z-score because the later uses more information (the population variance is known).

Sample size n	d.f. $n-1$	95% C.I.
5	4	$\bar{x} \pm 2.78 \frac{s}{\sqrt{n}}$
10	9	$\bar{x} \pm 2.26 \frac{s}{\sqrt{n}}$
20	19	$\bar{x} \pm 2.09 \frac{s}{\sqrt{n}}$
30	29	$\bar{x} \pm 2.05 \frac{s}{\sqrt{n}}$
∞	∞	$\bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$

Note that when d.f. exceeds 30, C.I. using t is approximately the same as CI obtained using standard normal distribution.

Confidence Intervals for Proportions

At the level of confidence $1-\alpha$, a confidence interval for p is given by

$$\left(\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}},\quad \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)$$

Margin of error: $e_{\alpha/2} = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}\hat{q}}{n}}$

$$\left(e_{\alpha/2}\right)^2 = \left(z_{\alpha/2}\right)^2 \frac{\hat{p}\hat{q}}{n}$$

$$\Rightarrow n = \frac{\hat{p}\hat{q}\left(z_{\alpha/2}\right)^2}{\left(e_{\alpha/2}\right)^2}$$

This equation can not be used directly because it involves the sample proportion $\hat{p}$, which will not be known at the outset of the investigation. If there is no such estimate available, substitute 0.5 for $\hat{p}$ in the formula. The product $\hat{p}\,\hat{q}$ cannot exceed 0.25 (the value when $\hat{p}=0.5$), so the largest possible value for n is

$$n = \frac{0.25\left(z_{\alpha/2}\right)^2}{\left(e_{\frac{\alpha}{2}}\right)^2}$$

This equation shows the largest sample needed so that the probability is $1-\alpha$

5.13A

FROM A1

C. H. Jan 28 2001

KLEIN: Support high for Alberta Tories

"But when you consider his ability to hang onto that popularity, that's the remarkable finding here."

The poll showed Albertans' passion for Klein is slightly surpassed by their support for the Tory party. The provincial government got a 68-per-cent approval rating.

The survey was done by Kanji and Barry Cooper of the University of Calgary. In half-hour telephone sessions, 1,003 people were interviewed from Oct. 22 through Nov. 5. Results are accurate to within plus or minus three per cent, 19 times out of 20.

m.e. = 0.0309

QUOTABLE

"If you ask any Albertan what the election is all about, they know"

LIBERAL LEADER NANCY MACBETH

The exercise took place before Canadian Alliance Leader Stockwell Day's $792,000 legal bill to Alberta taxpayers became public, and prior to the full implications to the pocketbook of surging power and natural gas prices.

However, Kanji said he didn't expect those issues to greatly hurt the Tories — the government survived the kerfuffle over Bill 11 virtually unscathed, and the Alberta Liberals haven't offered a credible alternative to anyone sour on Klein.

"These results come after the whole Bill 11 ordeal, so I think it's remarkable to find after all the discourse and debate on Bill 11, and all the protesting, the polls still show the Tories heading into the next election with majority support."

Liberal Leader Nancy MacBeth said the big issue will be deregulation and the government is vulnerable.

"If you ask any Albertan what the election is all about, they know — having opened their electricity bill in the last couple of weeks — exactly what this election is going to be about," said MacBeth.

The Liberals have been criticized in recent days for being slow off the mark when it was foolproof.

"The policy will be based on lower prices to consumers, more supply and suppliers, and insuring the economic health of Alberta in the future and we're working on it now."

She wouldn't comment on elements of the poll which showed her personal popularity dipping, even in Edmonton, a traditional Liberal stronghold.

"I'm really not going to get in to commenting on individual parts of the poll because it doesn't go anywhere. I mean, there's a poll, you've got the results, and the one that's going to matter to me is the one on election day," she said. "You're aware of it, you know it, but we're working and fighting hard to do a better job which I know we can do for Alberta families."

On Saturday, the Liberals announced they would be airing provincewide ads Monday "to give television viewers a more personal look at Nancy MacBeth."

"We're on the verge of an election," said Kieran Leblanc, the party's communications vice-president, in a news release. "It's time to re-introduce Albertans to Nancy MacBeth, and let them see for themselves that there's a strong, articulate, experienced leader ready to do a job."

Kanji said one of the problems facing the Liberals is the perception they are largely platform-less with an election just around the corner. On top of that, Albertans don't feel a personal warmth towards MacBeth. When asked to rate their warmth factor, 62 per cent toasted the premier, with 26 per cent saying he gave them a definite chill.

With MacBeth, however, nearly half of those who responded — about 500 people — said they felt cold towards her. Just over 20 per cent said they felt a glow towards the Liberal leader.

"Were the Liberals to change their leader and come up with a new, more dynamic person, and actually make some attempt to put forward policy and alternatives, that might make a difference," said Kanji.

He identified another trend that emerged during the research.

In past surveys, respondents pointed to Klein's success in restoring order to Alberta's finances as a key reason for support. That's waning, however, as people now view that as an accomplishment from the past.

"There is a solid close to a third of Albertans who say he's popular because he has no real competition. Earlier on,

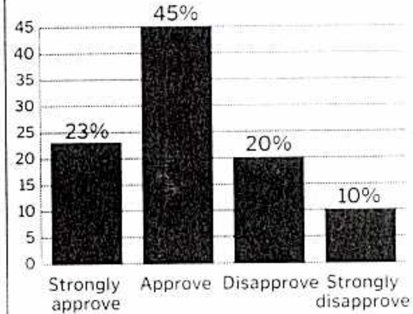
Question: For each of the followi[ng] indicate if you strongly approve, ap[prove] strongly disappr[ove]

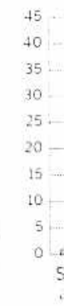
Beginning with your views on the Klein government in Alberta, do you say you strongly approve, approve, disapprove, or strongly disapprove of the Klein government's performance?

Wh[at] perf[ormance] Mac[Beth]

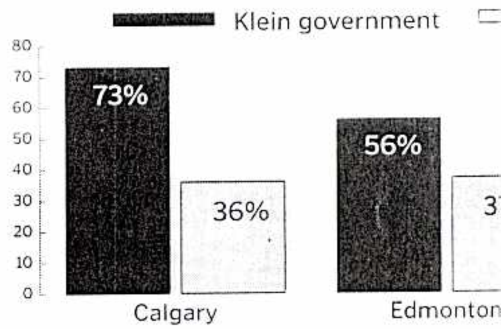
Klein government's performance rating from 1999 Alberta Advantage Survey: 64%

Libe[ral] rati[ng] 199[9] Sur[vey]

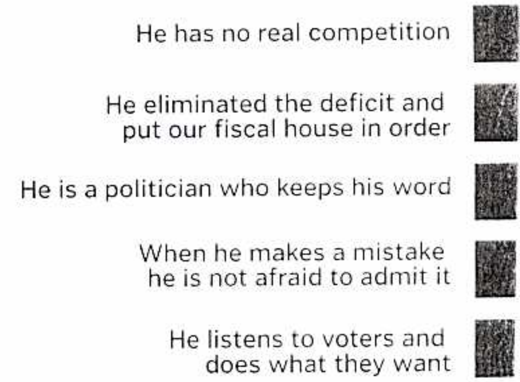




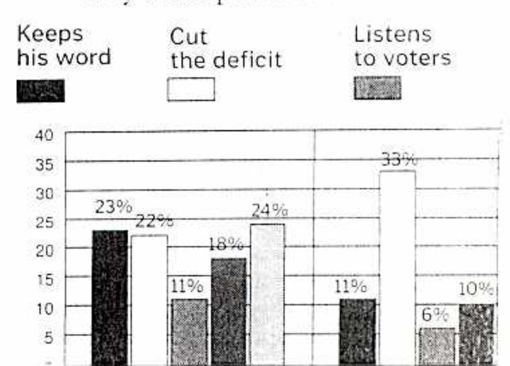
Performance ratings (by soc[...]



Question: It seems that no matter wh[...] remains remarkably popular among Al[...] the premier is so well-liked by s[...]



Why is the premier so well liked? (



Jan 27, 20

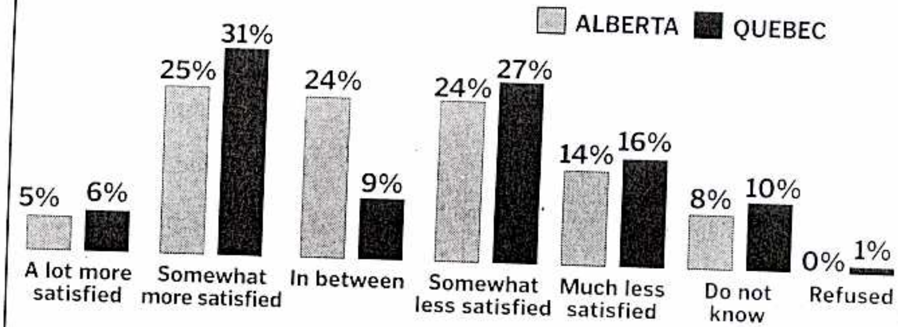
5.13B

FROM A1

POLL: Ottawa 'dictatorial'

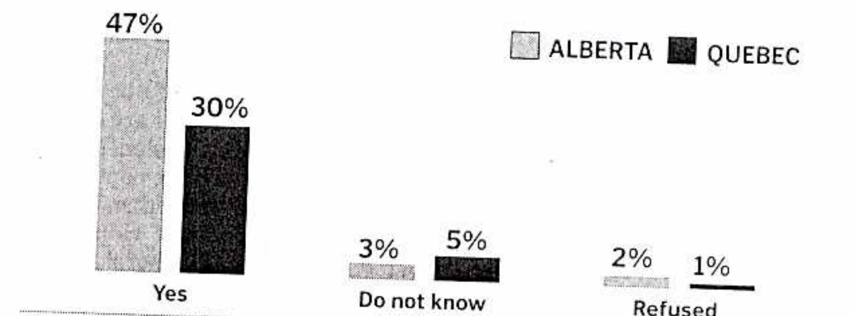
Though Albertans are on balance more satisfied, their rate of increasing alienation matches that in Quebec. In Quebec, 43% say that they have become less satisfied in the last five years compared to 38% in Alberta.

Compared to five years ago, would you say you are a lot more satisfied, somewhat more, somewhat less, or much less satisfied?

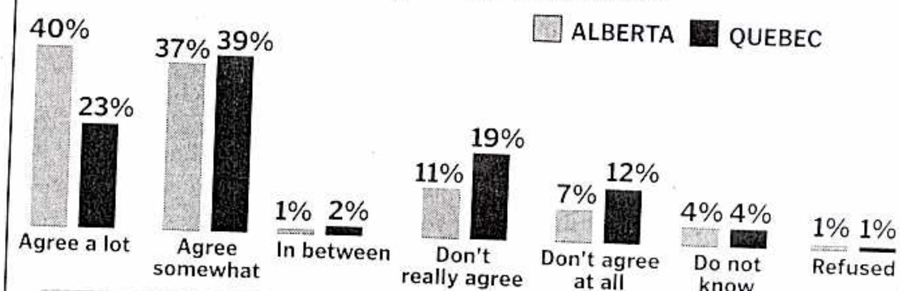


Do you agree that ....

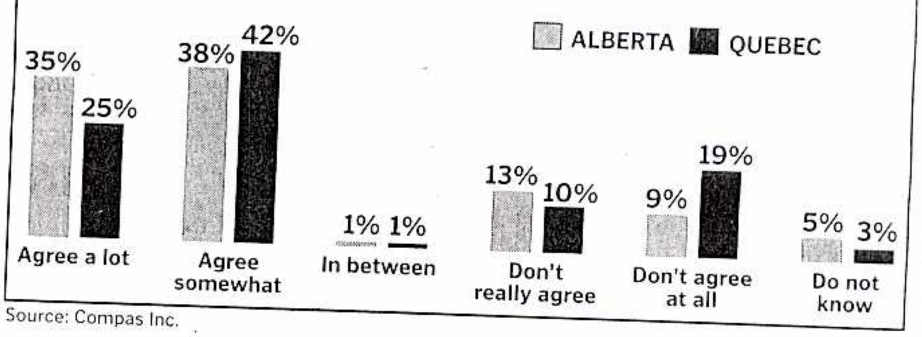
Albertans/Quebecers should insist on a major constitutional revision that would set strict limits on Ottawa's ability to control Alberta government policy and also how much subsidy Ottawa would be allowed to take from Alberta taxpayers.



Albertans/Quebecers have little say in how the federal government spends money that comes substantially from Alberta/Quebec.



Alberta/Quebec taxpayers subsidize the rest of the country too much.



Source: Compas Inc.

When asked if they'd like to join the United States, six per cent of Albertans said yes, but wanted to remain neighbourly with the other provinces. Four per cent of people in Quebec want to become Americans.

Sixty-two per cent of Canadian Alliance voters in Alberta are not satisfied with the province's position in Confederation and favour a major constitutional overhaul, compared with 21 per cent of Liberals.

In Quebec, 65 per cent of Bloc voters want the same thing, compared with 22 per cent of Liberals.

Albertans, more than Quebecers, tend to view Ottawa as "dictatorial and disrespectful."

But each province said its situation is unique.

"French-speaking Quebecers may feel that they have a special history of grievance and they may feel that they have a greater foundation for independent nationhood than do Albertans. But Albertans have a greater sense of being frozen out of federal politics and federal government decision-making, and they also have a stronger perception of the federal government as autocratic," concluded COMPAS.

As many as 67 per cent of Albertans said the federal government treats the province dictatorially, compared with 34 per cent of people in Quebec.

As well, 69 per cent of Albertans — as opposed to 47 per cent of Quebecers — complained the country doesn't back the politicians they prefer.

In Alberta, 77 per cent said they feel they have little say in how the federal government spends their money. Sixty-two per cent of Quebecers agreed.

Asked if they feel they subsidize the rest of the country too much, 35 per cent of Albertans and 25 per cent of Quebecers said yes.

According to Conrad Winn, director of COMPAS, the poll showed an acceleration of western discontent.

"I was startled by the fact that half of Albertans want constitutional change. The poll results show the dissatisfaction is authentic; it's genuine," he said.

"The only irony is that for at least a generation, federal leaders and national media have been preoccupied with Quebec ... taking for granted that the rest of the country is loyal.

"However, they've been asleep at the wheel. Albertans have been turbulent in their attitude towards Confederation," said Winn.

Liberal Senator Dan Hays, recently named Speaker of the Senate, said more needs to be done on the part of all elected officials to carry Alberta's voice into Ottawa.

"I'm going to continue to do what I can to make sure Alberta is looked after well," said Hays.

The poll surveyed 400 Albertans and 400 Quebecers and was conducted Jan. 25 to Jan. 27. It is deemed accurate to within five percentage points 19 times out of 20.

m.e. = 0.04899

Poll shows scandal eroding Liberal support

Last Updated Tue, 17 Feb 2004 7:59:19

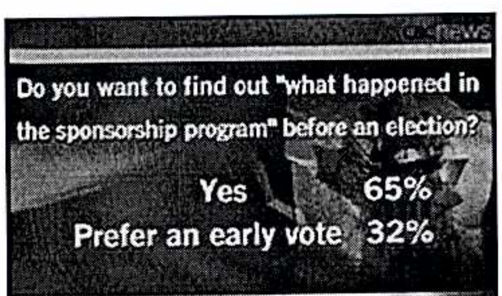
OTTAWA - The furor over the federal government sponsorship scandal continues to erode Liberal support, with only 35 per cent of Canadians saying they would vote for the party in an election, according to a new poll.

- INDEPTH: Sponsorship Scandal

The Ipsos-Reid survey, published in *The Globe and Mail*, found support for the Liberals dropped four per cent from 39 per cent just three days ago and 48 per cent four weeks ago – before the auditor general's report was made public.

The Conservative party was up to 27 per cent, compared with 19 per cent, while the NDP remained stable at 17 per cent.

The poll also revealed that nearly two-thirds of Canadians want more information about the scandal before an election is held.



Prime Minister Paul Martin has refused to say whether he would delay an election call in the wake of mounting criticism over the government's sponsorship program.

He said Canadians are "entitled to more information," but they are also "entitled to say a new government

YOUR TURN

WRITE TO US:
Send your comments to lette

JOIN THE DISCUSSION:
Share your thoughts on this a news stories!

Email This Story
Printable Version

5.13c

5.13d

with a new agenda ought to seek a mandate...and a new prime minister ought to seek a mandate, so there really is a balance that has to be done here."

The poll also found that 29 per cent blame former prime minister Jean Chrétien for the scandal, compared to 22 per cent for Martin.

The poll of 1,055 Canadians was taken from Wednesday – the day after the auditor general released her report – to Sunday. The survey is considered accurate to plus or minus 3.1 percentage points, 19 times out of 20. The rate of undecided voters was not known.

lecture note page 5.12

$$m.e. = 1.96 \times \sqrt{\frac{0.5 \times 0.5}{1055}}$$

$$= 0.03017$$

Written by CBC News Online staff

Headlines:Canada

- Data bank for adverse drug reactions flawed
- Fraser to name names in sponsorship controversy
- Poll shows scandal eroding Liberal support
- Nunavut awaits choice of new premier
- Martin will appear at sponsorship inquiry
- Chrétien ducks queries on sponsorship scandal
- Wayne leaving federal politics
- Former Chrétien aide creates another Liberal nomination battle
- Grenade shuts down Canada-U.S. border crossing

Jobs | Contact Us | Help
Terms of Use | Privacy | Copyright | Other Policies
Copyright © CBC 2004

$100(1-\alpha)\%$ C.I. for $\mu_1-\mu_2$ is

$$\left(\bar{x}_1-\bar{x}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\ ,\ \bar{x}_1-\bar{x}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

If $n_1, n_2 \geq 30$, σ_1^2 : unknown, σ_2^2 : unknown, replace σ_1^2 by s_1^2, σ_2^2 by s_2^2.

e.g. In a gender discrimination case, an employee alleged that a large corporation paid men more than women for comparable work.

$n_1 = 100$
$\bar{x}_1 = 20600$
$s_1 = 3000$

$n_2 = 100$
$\bar{x}_2 = 19700$
$s_2 = 2500$

Construct a 95 % C.I. for $\mu_1-\mu_2$

Sln: Sample sizes are large, we can use sample variance in place of population variance & obtain the C.I. by using the formula:

$$\left(\bar{x}_1-\bar{x}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\ ,\ \bar{x}_1-\bar{x}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right)$$

$$\Rightarrow \left(900 - 1.96\sqrt{\frac{(3000)^2}{100}+\frac{(2500)^2}{100}}\ ,\ 900 + 1.96\sqrt{\frac{3000^2}{100}+\frac{2500^2}{100}}\right.$$

or ($ 134.60, $ 1665.40)

Because this interval contains only +ve values, we can be quite confident that $\mu_1 - \mu_2 > 0$. Thus, it is reasonable to assume that the mean salary for males exceeds the mean salary for females.

C. I. for $\mu_1 - \mu_2$ when variances are unknown & sample sizes are small (pop. normal)

The problem of finding a C. I. has been solved for the special case when the unknown variances are equal. A general solution has not been found for the case where the unknown variances are unequal.

$\bar{x}_1, \bar{x}_2, \quad s_1^2, \quad s_2^2$

$$S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

100 $(1-\alpha\%)$ C. I $\quad \mu_1 - \mu_2$

C. I.

$$\left(\bar{x}_1 - \bar{x}_2 - t_{\frac{\alpha}{2}, \nu} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}, \ \bar{x}_1 - \bar{x}_2 + t_{\frac{\alpha}{2}, \nu} \cdot \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}\right.$$

$\nu = n_1 + n_2 - 2$ & S_p^2 is the pooled variance estimate of the common variance!

e.g. Two manufacturing companies produce carbide drills tips that are used to cut holes in steel sheets. A customer wishing to know which drill tips have the longer life. He purchase independent samples of $n_1 = 20$ drill tips from company 1 and $n_2 = 15$ drill tips from company 2. The mean lifes of the drill tips are $\bar{x}_1 = 78$ minutes & $\bar{x}_2 = 84$ minutes. The population variance are unknown but assumed to be equal. The sample variances are $s_1^2 = 41$, $s_2^2 = 36$

Construct a 95% C.I. for $\mu_1 - \mu_2$

Soln:

$$s_p^2 = \frac{(20-1)(41) + (15-1)36}{20+15-2} = \frac{1283}{33} = 38.88$$

95% C.I. $\frac{\alpha}{2} = 0.025$, d.f. $= n_1 + n_2 - 2 = 33$

$$t_{\frac{\alpha}{2}, 33} = 2.04$$

$\therefore$ 95% C.I. for $\mu_1 - \mu_2$,

$$\left(78 - 84 - 2.04\sqrt{\frac{38.88}{20} + \frac{38.88}{15}},\ 78 - 88 + 2.04\sqrt{\frac{38.88}{20} + \frac{38.88}{15}}\right)$$

$$= (-10.34, -1.66)$$

95% C.I. contain only −ve values, $\Rightarrow$ drill tips of company #1 do NOT last as long, on the average by company 2.