Inference concerning a single π M&M §8.1 updated Dec 26, 2003

Parameter π: the proportion e.g. ...

- with undiagnosed hypertension / seeing MD during a 1-year span
- responding to a therapy
- still breast-feeding at 6 months
- of pairs where response on treatment > response on placebo
- of US presidential elections where taller candidate expected to win
- of twin pairs where L handed twin dies first
- able to tell imported from domestic beer in a "triangle taste test"
- who get a headache after drinking red wine
- (of all cases, exposed and unexposed) where case was "exposed"

[function of rate ratio & of relative sizes of exposed & unexposed denominators; CONDITIONAL analysis (i.e. "fix" # cases), used for CI for ratio of 2 rates , especially in 'extreme' data configurations...
eg. # seroconversions in RCT of HPV16 Vaccine, NEJM Nov 21, 2002
0/11084.0 W-Y in vaccinated gp. sersus 41/11076.9 W-Y in placebo gp]

Statistic: the proportion p = y/n in a sample of size n. ...

Inferences from y/n to π

FREQUENTIST

via Confidence Intervals and Tests

- **Confidence Interval: where is** π **?**

  supplies a NUMERICAL answer (range)

- **Evidence (P-value) against** H_0**:** $\pi = 0.xx$

- **Test of Hypothesis: Is (P-value) < preset** α **?**

  supplies a YES / NO answer **(uses Pdata |** H_0**)**

BAYESIAN

via posterior probability distribution for, and probabilistic statements concerning, π **itself**

- **point estimate:** median, mode, ...
- **interval estimate:** credible intervals, ...

Software: • "Bayesian Inference for Proportion (Excel)" Resources Ch 8
• First Bayes { http://www.epi.mcgill.ca/Joseph/courses.html }

cf also A&B §4.7; Colton §4. Note that JH's notes use p for statistic, π for parameter.

(FREQUENTIST) Confidence Interval for π from a proportion p = x / n

1. "Exact" (not as "awkward to work with' as M&M p586 say they are)

tables [Documenta Geigy, Biometrika , ...] **nomograms, software**

e.g. what fraction π will return a 4-page questionnaire?
11/20 returns on a pilot test i.e. p= 11/20 =0.55

95% CI (from CI for proportion **table** Ch 8 Resources) 32% to 77%
[To save space, table gives CI's only for p≤0.5, so get CI for π of non-returns: point estimate is 9/20 or 45%, CI is 23% to 68% {1st row, middle column of the X=9 block} Turn this back to 100-68=32% to 100-23=77% returns]

95% CI (Biometrika **nomogram**) 32% to 77%
[uses c for numerator; enter through lower x-axis if p≤0.5; in our case p=0.55 so enter nomogram from the top at c/n = 0..55 near upper right corner; travel downwards until you hit bowed line marked 20 (the 5th line from the top) and exit towards the rightmost border at $\pi_{lower} \approx$ **0.32**; go back and travel downward until hit the companion bowed line marked 20 (the 5th line from bottom) and exit towards the rightmost border at $\pi_{upper} \approx$ **0.77**].

Others may use other names for numerator and statistic, or use symmetry (Binomial[y, n,p] <--> Binomial[n-y, n,1 - p] to save space. Nomogram on next page shows full range, but uses an approxn..

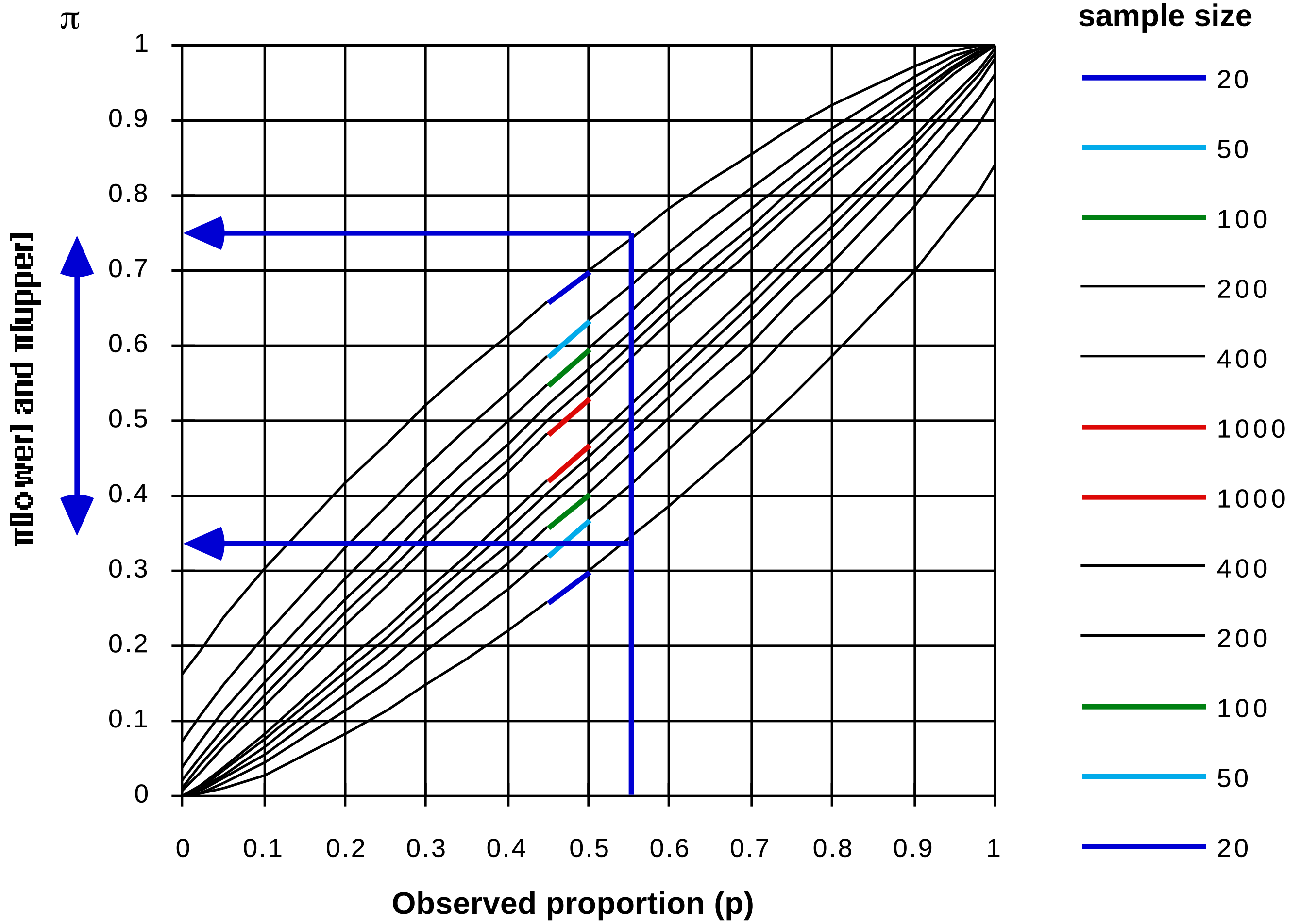
Notice link between 100(1 - α)% CI and two-sided test of significance with a preset α. If true π were < π_{lower}, there would only be less than a 2.5% probability of obtaining, in a sample of 20, this many (11) or more respondents; likewise, if true π were > π_{lower}, there would be less than a 2.5% probability of obtaining, in a sample of 20, this many (11) or fewer respondents. The 100(1 - α)% CI for π includes all those parameter values such that if the oberved data were tested against them, the p-value (2-sided) would not be < α.

e.g. Experimental drug gives p = $\frac{\text{0 successes}}{\text{14 patients}}$ => π = ??

95% CI for π (from table) 0% to 23%
CI "rules out" (with 95% confidence) possibility that π>23%
[might use a 1-sided CI if one is interested in putting just an upper bound on risk: e.g. what is upper bound on π = probability of getting HIV from HIV-infected dentist? see JAMA article on "zero numerators" by Hanley and Lippman-Hand (in Resources for Chapter 8) .

CI for π -- using nomogram (many books of statistical tables have fuller versions)

95% CI for π



(Asymmetric) CI in above Nomogram: approx. formula $\pi = \dfrac{1 - \dfrac{n}{n+z^2} + \dfrac{2np}{n+z^2} \pm \dfrac{z\sqrt{4np - 4np^2 + z^2}}{n+z^2}}{2}$ (cf later page)

See Biometrika Tables for Statisticians for the "exact" Clopper-Pearson version.
calculated so that Binomial Prob [≥ p | π_{lower}] = Prob[≤ p | π_{upper}] = 0.025 exactly.

Inference concerning a single π M&M §8.1

(FREQUENTIST) **Confidence Interval for** π from a proportion p = x / n

1. Exactly, but by *trial and error*, via SOFTWARE with Binomial probability function

Exactly, *and directly,* using table of (or SOFTWARE function that gives) the percentiles of the F distribution

See spreadsheet "**CI for a Proportion (Excel spreadsheet, based on exact Binomial model**) " under Resources for Chapter 8. In this sheet one can obtain the direct solution, or get there by trial and error. Inputs in bold may be changed.

The general "Clopper-Pearson" method for obtaining a Binomial-based CI for a proportion is explained in 607 Notes for Chapter 6.

Can obtain these limits by trial and error (e.g. in spreadsheet) or directly using the link between the Binomial and F tail areas (also implemented in spreadsheet). The basis for the latter is explained by Liddell (method, and reference, given at bottom of Table of 95% CI's).

Spreadsheet opens with example of an observed proportion p = 11/20.

(FREQUENTIST) **Confidence Interval for** π

NOTES

- Turn spreadsheet of Binomial Probabilities (Table C) 'on its side' to get CI(π) .. simply find those columns (π values) for which the probability of the observed proportion is small
- Read horizontally, Nomogram [previous page] shows the variability of proportions from SRS samples of size n. [very close in style to table of Binomial Probabilities, except that only the central 95% range of variation shown , & all n's on same diagram]

  Read vertically, it shows:

  - CI -> symmetry as p -> 0.5 or n - > ∞ [in fact, as np & n(1-p) - > ∞]
  - widest uncertainty at p=0.5 => can use as a 'worst case scenario'
  - cf. the ± 4 % points in the 'blurb' with Gallup polls of size n ≈ 1000. Polls are usually *cluster* (rather than SR) samples, and so have bigger margins of error [wider CI's] than predicted from the Binomial.

95%CI? IC? ... Comment dit on... ?

[La Presse, Montréal, 1993] L'Institut Gallup a demandé récemment à un échantillon représentatif de la population canadienne d'évaluer la manière dont le gouvernement fédéral faisait face à divers problèmes économiques et général. Pour 59 pour cent des répondants, les libéraux n'accomplissent pas un travail efficace dans ce domaine, tandis que 30 pour cent se déclarent de l'avis contraire et que onze pour cent ne formulent aucune opinion.

La même question a été posée par Gallup à 16 reprises entre 1973 et 1990, et ne n'est qu'une seule fois, en 1973, que la proportion des Canadiens qui se disaient insatisfaits de la façon dont le gouvernement gérait l'économie a été inférieure à 50 pour cent.

Les conclusions du sondage se fondent sur **1009** interviews effectuées entre le 2 et le 9 mai 1994 auprès de Canadiens âgés de 18 ans et plus. **Un échantillon de cette ampleur donne des résultats exacts à 3,1 p.c., près dans 19 cas sur 20.** La marge d'erreur est plus forte pour les régions, par suite de l'importance moidre de l'échantillonnage; par exemple, les **272** interviews effectuées au Québec ont engendré une marge d'erreur de **6 p.c**. dans **19 cas sur 20**.

2. CI for π based on "large-n" behaviour of p, or fn. of p

CI: $p \pm z\,SE(p) = p \pm z\sqrt{\frac{p[1-p]}{n}}$

e.g. p = 0.3, n=1000

95%CI for π

$= 0.3 \pm 1.96\sqrt{\frac{0.3[0.7]}{1000}}$

$= 0.30 \pm 1.96(0.015)$

$= 0.30 \pm 0.03$

$= 30\% \pm 3\%$

<u>Note</u>: the ± 3% is pronounced and written as "± 3 <u>percentage points</u>" to avoid giving the impression that it is 3% *of* 30%

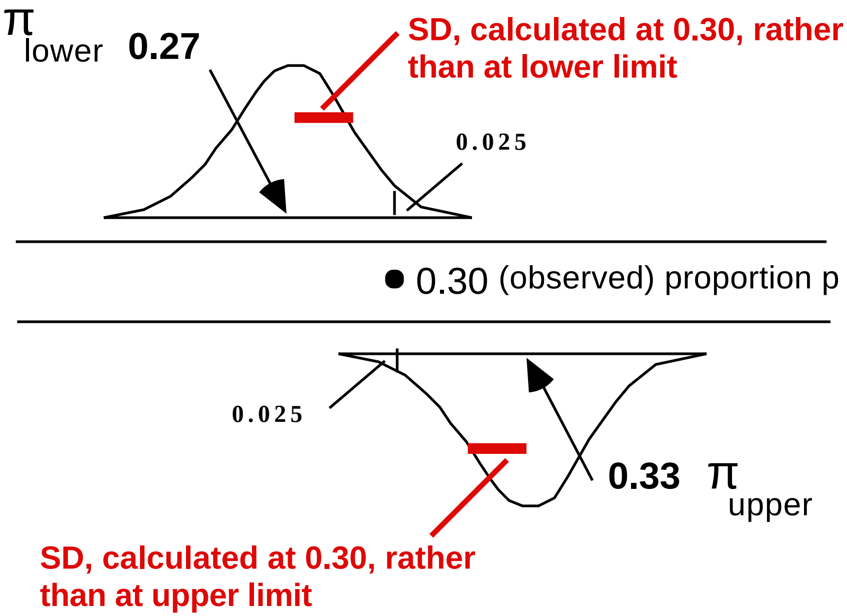
<u>**"Large-sample n": How large is large?**</u>

- A rule of thumb: when the expected no. of positives, np, and the expected no. of negatives, n(1-p), are both bigger than 5 (or 10 if you read M&M).

- JH's rule: when you can't find the CI tabulated anywhere!

- if the distribution is not 'crowded' into one corner (cf. the shapes of binomial distributions in the Binomial spreadsheet -- in Resources for Ch 5), i.e., if, with the symmetric Gaussian approximation, neither of the tails of the distribution spills over a boundary (0 or 1 if proportions, or 0 or n if on the count scale),

See M&M p383 and A&B §2.7 on Gaussian approximation to Binomial.



SE-based (sometimes referrred to in texts and software output as "Wald" CI's) use the same SE for the upper and lower limits -- they calculate **one SE at the point estimate**, rather than two separate SE's, calculated at each of the two limits.

<u>**From SAS**</u>

```
DATA CI_propn;
INPUT n_pos  n ;
LINES;
       300 1000
;
PROC genmod data = CI_propn; model n_pos/n = /
    dist = binomial link = identity waldci ;  RUN;
```

<u>**From Stata**</u> **immediate command:** `cii 1000 300`

```
clear      * Using datafile
input n_pos  n
     140  500  * glm doesn't like file with 1 'observation'
     160  500  * so..........split across 2 'observations'
end
glm n_pos , family(binomial n) link(identity)
```

2. CI for π based on "large-n" behaviour of p... continued

Other, more accurate and more theoretically correct, large-sample (Gaussian-based) constructions

The "usual" approach is to form a symmetric CI as

point estimate ± a multiple of the SE.

This is technically incorrect in the case of a distribution, such as the binomial, with a variance that changes with the parameter being measured. In construction of CI's [see diagram on page 1 of material on Ch 6.1] there are two distributions involved: the binomial at π_{upper} and the binomial at π_{lower}. They have different shapes and different SD's in general. Approaches i and ii (below) take this into account.

i Based on Gaussian approximation to binomial distribution, **but with SD's calculated at limits** SD = $\sqrt{\pi[1-\pi] / n}$ **rather than at the point estimate itself** {"usual" CI uses SD = $\sqrt{p[1-p] / n}$ }

If define CI for π as (π_L, π_U),

where Prob[sample proportion ≥ p | π_L] = $\alpha/2$
Prob[sample proportion ≤ p | π_U] = $\alpha/2$

and if use Gaussian approximations to Binomial(n,π_L) **and** Binomial(n,π_U), and solve

$$p = \pi_L + z_{\alpha/2}\sqrt{\frac{\pi_L [1-\pi_L]}{n}}$$

and

$$p = \pi_U - z_{\alpha/2}\sqrt{\frac{\pi_U [1-\pi_U]}{n}}$$

for π_L and π_U,

This leads to asymmetric $100(1-\alpha)\%$ limits of the form:

$$\frac{1 - \frac{n}{n + z^2} + \frac{2np}{n + z^2} \pm \frac{z\sqrt{4np - 4np^2 + z^2}}{n + z^2}}{2}$$

Rothman(2002-p132) attributes this method **i** to Wilson 1927.

ii Based on Gaussian distribution of a **variance-stabilizing transformation of the binomial**, again **with SD's calculated at the limits** rather than at the point estimate itself

$$[\sin[\sin^{-1}[\sqrt{p}] - \frac{z}{2\sqrt{n}}]]^2 , [\sin[\sin^{-1}[\sqrt{p}] + \frac{z}{2\sqrt{n}}]]^2$$

as in most calculators, $\sin^{-1}$ & the * in sin[*] measured in radians.

E.g. with α = 0.05, so that z=1.96, we get: * from Mainland

Method	n=10 p=0.0	n=10 p=0.3	n=20 p=0.15	n=40 p=0.075
1.	[0.00, 0.28]	[0.11, 0.60]	[0.05, 0.36]	[0.03, 0.20]
2.	[0.09, 0.09]	[0.07, 0.60]	[0.03, 0.33]	[0.01, 0.18]
"usual"	[0.00, 0.00]	[0.02, 0.58]	[–0.01, 0.31]	[–0.01, 0.16]
Binomial*	[0.00, 0.31]	[0.07, 0.65]	[0.03, 0.38]	[0.02, 0.20]

* from Mainland

References:
- Fleiss, Statistical Methods for Rates and Proportions
- Miettinen, Theoretical Epidemiology, Chapter 10.

2. CI for π based on "large-n" behaviour of <u>logit</u> transformation of proportion

iii Based on Gaussian distribution of the **logit transformation** of the estimate (p, the observed proportion) of the parameter π

<u>PARAMETER</u>: LOGIT[π] = log [ODDS] = log [π / (1 - π)]

= log ["Proportion POSITIVE" / "Proportion NEGATIVE"]

<u>STATISTIC</u>: logit[p] = log [odds] = log [p / (1 - p)]

(Here, log = 'natural' log, i.e. to base e, which some write as *ln*)
(UPPER CASE/Greek = parameter; lower case/Roman = statistic)

Reverse transformation (to get back from LOGIT to π)...

$\pi = \dfrac{\text{ODDS}}{1 + \text{ODDS}} = \dfrac{\exp[\text{LOGIT}]}{1 + \exp[\text{LOGIT}]}$; and likewise p <-- logit

$\pi_{\text{LOWER}} = \dfrac{\exp[\text{LOWER limit of LOGIT}]}{1+\exp[\text{ LOWER limit of LOGIT }]}$; π_{UPPER} likewise

SE[logit] = Sqrt[1 / #positive + 1 / #negative]

e.g. p = 3/10 => estimated odds = 3/7 => logit = log[3/7] = -0.85

SE[logit] = Sqrt[1/3 + 1/7] = 0.69

CI in LOGIT scale: -0.85 ± 1.96×0.69 = { -2.2, 0.5}

CI in π scale: $\left\{ \dfrac{\exp[-2.2]}{1+\exp[-2.2]}, \dfrac{\exp[0.5]}{1+\exp[0.5]} \right\} = \{ 0.10, 0.67\}$

From SAS

```
DATA CI_propn; INPUT n_pos  n ;
LINES;
    3 10
;
PROC genmod data = CI_propn;
model n_pos/n = /
      dist = binomial
      link = logit waldci ;
```

From Stata

```
clear
input   n_pos   n
             1   5
             2   5
end
glm n_pos, family(binomial n) link(logit)
```

anti-logit [logit] = $\dfrac{\exp[\text{logit}]}{1 + \exp[\text{logit}]}$ Greenland calls it the "expit" function

2. CI for π based on "large-n" behaviour of <u>log</u> transformation of proportion

iv Based on Gaussian distribution of estimate of **log**[π]

<u>PARAMETER</u>: log[π]

<u>STATISTIC</u>: log[p]

Reverse transformation (to get back from log[π] to π)...

π = antilog[log[π]] = exp[log[π]] ; and likewise p <-- log[p]

π_{LOWER} = exp[LOWER limit of log[π]] ; π_{UPPER} likewise

SE[log[p]] = Sqrt[1 / #positive - 1 / #total]

Limits for π from p = 3/10 : exp[log[3/10] ± z × Sqrt[1/3 - 1/10]]

<u>Exercises</u>:

1 Verify that you get same answer by calculator and by software

2 Even with these logiy and log transformations, the Gausian distribution is not accurate at such small sample sizes as 3/10. Compare their preformance (against the exact methods) for various sample sizes and numbers positive.

From SAS

```
DATA CI_propn; INPUT n_pos  n ;
LINES;
    3 10
;
PROC genmod data = CI_propn;
model n_pos/n = /
      dist = binomial
      link = log waldci ;
```

From Stata

```
clear
input   n_pos   n
             1   5
             2   5
end
glm n_pos, family(binomial n) link(log)
```

1200 are hardly representative of 80 million homes /220 million people!!

The Nielsen system for TV ratings in U.S.A.
(Excerpt from article on "Pollsters" from an airline magazine)

"...Nielsen uses a device that, at one minute intervals, checks to see if the TV set is on or off and to which channel it is tuned. That information is periodically retrieved via a special telephone line and fed into the Nielsen computer center in Dunedin, Florida.

With these two samplings, Nielsen can provide a statistical estimate of the number of homes tuned in to a given program. A rating of 20, for instance, means that 20 percent, or 16 million of the 80 million households, were tuned in.
To answer the criticism that 1,200 or 1,500 are hardly representative of 80 million homes or 220 million people, Nielsen offers this analogy:

Mix together 70,000 white beans and 30,000 red beans and then scoop out a sample of 1000. the mathematical odds are that the number of red beans will be between 270 and 330 or 27 to 33 percent of the sample, which translates to a "rating" of 30, plus or minus three, with a 20-to-1 assurance of statistical reliability. The basic statistical law wouldn't change even if the sampling came from 80 million beans rather than just 100,000." ...

Why, if the U.S. has a 10 times bigger population than Canada, do pollsters use the same size samples of approximately 1,000 in both countries?

Answer : it depends on WHAT IS IT THAT IS BEING ESTIMATED. With n=1,000, the SE or uncertainty of an estimated PROPORTION 0.30 is indeed 0.03 or 3 percentage points. However, if interested in the NUMBER of households tuned in to a given program, the best estimate is 0.3N, where N is the number of units in the population (N=80 million in the U.S. or N=8 million in Canada). The uncertainty in the 'blown up' estimate of the TOTAL NUMBER tuned in is blown up accordingly, so that e.g. the estimated NUMBER of households is

```
U.S.A. 80,000,000[0.3 ± 0.03] = 24,000,000 ± 2,400,000
Canada. 8,000,000[0.3 ± 0.03] =  2,400,000 ±   240,000
```

2.4 million is a 10 times bigger absolute uncertainty than 240,000. Our intuition about needing a bigger sample for a bigger universe probably stems from absolute errors rather than relative ones (which in our case remain at 0.03 in 0.3 or 240,000 in 2.4 million or 2.4million in 24 million i.e. at 10% irrespective of the size of the universe. It may help to think of why we do not take bigger blood samples from bigger persons: the reason is that we are usually interested in concentrations rather than in absolute amounts and that concentrations are like proportions.

The "Margin of Error blurb" introduced (legislated) in the mid 1980's

Montreal Gazette August 8, **1981**
NUMBER OF SMOKERS RISES BY FOUR POINTS: GALLUP POLL

Compared with a year ago, there appears to be an increase in the number of Canadians who smoked cigarettes in the past week - up from 41% in 1980 to 45% today. The question asked over the past few years was: **"Have you yourself smoked any cigarettes in the past week?"** Here is the national trend:

Smoked cigarettes in the past week
Today..........................45%
1980............................41
1979............................44
1978............................47
1977............................45
1976...........................Not asked
1975............................47
1974............................52

Men (50% vs. 40% for women), young people (54% vs. 37% for those > 50) and Canadians of French origin (57% vs. 42% for English) are the most likely smokers. ***Today's results are based on 1,054 personal in-home interviews with adults, 18 years and over, conducted in June.***

The Gazette, Montreal, Thursday, June 27, **1985**
39% OF CANADIANS SMOKED IN PAST WEEK: GALLUP POLL

Almost two in every five Canadian adults (39 per cent) smoked at least one cigarette in the past week - down significantly from the 47 percent who reported this 10 years ago, but at the same level found a year ago. Here is the question asked fairly regularly over the past decade: "**Have you yourself smoked any cigarettes in the past week?"** The national trend shows:

Smoked cigarettes in the past week
1985............................39%
1984............................39
1983............................41
1982*..........................42 (* Smoked regularly or occasionally)
1981............................45
1980............................41
1979............................44
1978............................47
1977............................45
1975............................47

Those < 50 are more likely to smoke cigarettes (43%) than are those 50 years or over (33%). Men (43%) are more likely to be smokers than women (36%).
Results are based on 1,047 personal, in-home interviews with adults, 18 years and over, conducted between May 9 and 11. ***A sample of this size is accurate within a 4-percentage-point margin, 19 in 20 times.***

Test of Hypothesis that ≠= some test value

1. n small enough -> Binomial Tables/Spreadsheet

ie if testing H_0: $\pi = \pi_0$ vs H_a: $\pi \neq \pi_0$ [or H_a: $\pi > \pi_0$]

and if observe x / n ,

then calculate

Prob(observed x, or an x that is more extreme | π_0)

using H_a to specify which x's are more extreme i.e. provide even more evidence for H_a and against H_0.

or use correspondence between a 100(1-α)% CI and a test which uses an alpha level of α i.e. check if CI obtained from CI table or nomogram includes π value being tested

[there may be slight discrepancies between test and CI: the methods used to construct CI's don't always correspond exactly to those used for tests]

e.g. 1 A common question is whether there is evidence against the proposition that a proportion π=1/2 [Testing preferences and discrimination in psychophysical matters e.g., therapeutic touch, McNemar's test for discordant pairs when comparing proportions in a paired-matched study, the non-parametric' Sign Test for assessing intra-pair differences in measured quantities, ...]. Because of the special place of the Binomial at π=1/2, the tail probabilities have been calculated and tabulated. See the table entitled "Sign Test" in the chapter on Distribution-Free Methods.

M&M (2nd paragraph p 592) say that "we do not often use significance tests for a single proportion, because it is uncommon to have a situation where there is a precise proportion that we want to test". But they forget paired studies, and even the sign test for matched pairs, which they themselves cover in section 7.1, page 521. They give just 1 exercise (8.18) where they ask you to test π=0.5vs $\pi > 0.5$.

e.g. 2 Another example, dealing with responses in a setup where the "null" is π=1/3, the "Triangle Taste Test" is described in the next page.

2. Large n : Gaussian Approximation

$$\textbf{Test } \pi = \pi_0: \; z = \frac{p - \pi_0}{SE[p]} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0[1-\pi_0]}{n}}}$$

Note that the test uses a $\sqrt{\text{variance}}$ based on the (specified) π_0. The "usual" CI uses a $\sqrt{\text{variance}}$ based on the (observed) p.

(Dis)Continuity Correction†

Because we approximate a discrete distribution [where p takes on the values 0/n, 1/n, 2/n, ... n/n corresponding to the integer values (0,1,2, ..., n) in the numerator of p] by a continuous Gaussian distribution, authors have suggested a 'continuity correction' (or if you are more precise in your language, a 'discontinuity' correction). This is the same concept as we saw back in §5.1, where we said that a binomial count of 8 became the interval (7.5, 8.5) in the interval scale. Thus, e.g., if we want to calculate the probability that proportion out of 10 is ≥ 8, we need probability of ≥ 7.5 on the continuous scale.

If we work with the count itself in the numerator, this amounts to reducing the absolute deviation $y - n\pi_0$ by 0.5 . If we work in the proportion scale, the absolute deviation is reduced by 0.5/n viz.

$$z_c = \frac{|y - n\pi_0| - 0.5}{SE[y]} = \frac{|y - n\pi_0| - 0.5}{\sqrt{n\pi_0[1-\pi_0]}}$$

or

$$z_c = \frac{|p - \pi_0| - 0.5/n}{SE[p]} = \frac{|p - \pi_0| - 0.5/n}{\sqrt{\frac{\pi_0[1-\pi_0]}{n}}}$$

†Colton [who has a typo in the formula on p ___] and A&B deal with this; M&M do not, except to say on p386-7 "because most statistical purposes do not require extremely accurate probability calculations, we do not emphasize use of the continuity correction". There are some 'fundamental' problems here that statisticians disagree on. The "Mid-P" material (below) gives some of the flavour of the debate.

EXAMPLE of Testing π: THE TRIANGLE TASTE TEST

As part of preparation for a double blind RCT of lactase-reduced infant formula on infant crying behaviour, the experimental formulation was tested for its similarity in taste to the regular infant formula . n mothers in the waiting room at MCH were given the TRIANGLE TASTE TEST i.e. they were each given 3 coded formula samples -- 2 containing the regular formula and 1 the experimental one. Told that "2 of these samples are the same and one sample is different", p = y/n correctly identify the odd sample. Should the researcher be worried that the experimental formula does not taste the same? (assume infants are no more or less taste-discriminating than their mothers) [study by Ron Barr, Montreal Children's Hospital]

The null hypothesis being tested is

H_0: π(correctly identified samples) = 0.33 against H_a: π() > 0.33

[here, for once, it is difficult to imagine a 2-sided alternative -- unless mothers were very taste-discriminating but wished to confuse the investigator]

We consider two situations (the real study with n=12, and a hypothetical larger sample of n=120 for illustration)

<u>• 5 of n = 12 mothers correctly identified the odd sample.</u>

i.e. p = 5/12 = 0.42

Degree of evidence against H_0

= Prob(5 or more correct | π=0.33)... - a Σ of 8 probabilities

= 1 - Prob(4 or fewer correct | π=0.33) ...- a shorter Σ of only 5

= 1 - [P(0) + P(1) + P(2) + P(3) + P(4)] = 0.37*

Using n=12, and p=0.30 in Table C gives 0.28; using p=0.35 gives 0.42. Interpolation gives 0.37 approx.

Can also obtain this probability directly via Excel, using the function

1 - BINOMDIST(4, 12, 0.33333, TRUE)

So, by conventional criteria (Prob < 0.05 is considered a cutoff for evidence against H_0) there is <u>not a lot of evidence to contradict</u> the H_0 of taste similarity of the regular and experimental formulae.

With a sample size of only n=12, <u>we cannot rule out</u> the possibility that a sizable fraction of mothers could truly distinguish the two.

Our observed proportion of 5/12 projects to a one-sided 95% CI of "as many as 65% in the population get it right". In this worst-case scenario, assuming that the percentage of right answers in the population is a mix of a proportion π_{can} who can really tell and one third of the remaining (1-π_{can}) who get it right by guessing, we equate

$$0.65 = \pi_{can} + (1-\pi_{can}) / 3$$

giving us an upper bound π_{can} = (0.65-0.33) / (2/3) = 0.48 or 48%.

*These calculations can be done easily even on a calculator or spreadsheet without any combinatorials:

$P(0) = 0.67^{12}$	= 0.008
P(1) = 12 x 0.33 x P(0) / [1 x 0.67]	= 0.048
P(2) = 11 x 0.33 x P(1) / [2 x 0.67]	= 0.131
P(3) = 10 x 0.33 x P(2) / [3 x 0.67	= 0.215
<u>P(4)</u> = 9 x 0.33 x P(3) / [4 x 0.67]	= <u>0.238</u>
Σ	= 0.640

so Prob(5 or more correct | π = 0.33) = 1 - 0.64 = <u>0.32</u>

<u>• 50 of 120 (p=0.42) mothers identified odd sample.</u>

$$\textbf{Test } \pi = 0.33: \quad z = \frac{0.42^* - 0.33}{\sqrt{\frac{0.33[1-0.33]}{120}}} = 2.1$$

So P = Prob[≥ 50 | π = 0.33] = Prob[Z ≥ 2.1] = 0.018

* We treat the proportion 50/120 as a contimuous measurement; in fact it is based on an integer numerator 50; we should treat 50 as 49.5 to 50.5 so ≥50 is really > 49.5 .

The Prob. of obtaining 49.5/120 or more is te Prob. of Z = $\frac{0.413 - 0.33}{\sqrt{\frac{0.33[1-0.33]}{120}}}$ or more. With n=120, the continuity correction does not make a large difference; however, with smaller n, and its coarser grain, the <u>continuity correction</u> [which makes differences smaller] is more substantial.

Sample Size for CI's and Tests involving π

n to yield (2-sided) CI with margin of error m at confidence level 1- α (see M&M p 593, Colton p161)

|--- margin of error --- >|
(----------------------•------------------------) CI

- see CI's as function of n in tables and nomograms
- (or) large-sample CI: $p \pm Z_{\alpha/2}\, SE(p) = p \pm m$

$$SE(p) = \sqrt{\frac{p[1-p]}{n}}, \text{ so... } n = \frac{p[1-p] \bullet Z_{\alpha/2}^2}{m^2}$$

If unsure, use largest SE i.e. when p=0.5 i.e.

$$n = \frac{0.25 \bullet Z_{\alpha/2}^2}{m^2} \qquad [1.c]$$

n for power 1-β to "detect" a population proportion π_1 that is Δ units from π_0 (test value) ; type I error = α (Colton p 161)

$$n = \frac{\left\{ Z_{\alpha/2}\sqrt{\pi_0[1-\pi_0]} - Z_\beta\sqrt{\pi_1[1-\pi_1]} \right\}^2}{\Delta^2} \qquad [1.t]$$

$$\approx \left\{ Z_{\alpha/2} - Z_\beta \right\}^2 \left\{ \frac{\sqrt{\pi[1-\pi]}}{\Delta} \right\}^2 \qquad [1.t]\approx$$

where π is average of π_0 and π_1

$$= \left\{ Z_{\alpha/2} - Z_\beta \right\}^2 \left\{ \frac{\sigma_{0/1}}{\Delta} \right\}^2$$

Notes: Z_β will be negative; formula is same as for testing μ

Worked example 1: sample size for
Test that π(preferences) = 0.5 vs. $\pi \neq 0.5$
or
Sign Test that median difference = 0

Test: H_0: $Median_D = 0$ vs H_{alt}: $Median_D \neq 0$
$\alpha = 0.05$ (2-sided);

or H_0: $\pi(+) = 0.5$ vs H_{alt}: $\pi(+) > 0.5$

For Power **1**-β against: H_{alt}: $\pi(+) = 0.65$ say

[at π=ave of 0.5 & 0.65, $\sqrt{\pi[1-\pi]} = = 0.494$]

$$n \approx \left\{ Z_{\alpha/2} - Z_\beta \right\}^2 \left\{ \frac{0.494}{0.15} \right\}^2$$

α=0.05 (2-sided) & β=0.2 ...

$Z_\alpha = 1.96$; $Z_\beta = -0.84$,

$(Z_{\alpha/2} - Z_\beta)^2 = \{1.96 - (-0.84)\}^2 \approx 8$, i.e.

$$n \approx 8 \left\{ \frac{0.494}{0.15} \right\}^2 = 87$$

Worked example 2: sample size for Δ Taste Test

π(correct) = 1/3 vs. π >1/3

If set α =0.05 (hardliners might allow **1**-sided test here), then Z_α = 1.645; If want 90% pwer, then Z_β = -1.28; Then using eqn [1.t] above...

n's for 90% Power against...

π(correct)=	0.4	0.5	0.6	0.7	0.8
	400	69	27	14	8

(See also homegrown exercise # ___)