1 (Binomial) Model for (Sampling)Variability of Proportion/Count in a Sample

The Binomial Distribution: what it is

- The $n+1$ probabilities $p_0, p_1, ..., p_y, ..., p_n$ of observing $0, 1, 2, \ldots, n$ "positives" in n independent binary trials (*such as in s.r.s of n individuals*)
- Each of the n observed elements is binary (0 or 1)
- There are 2^n possible *sequences* ... but only $n+1$ possible *values*, i.e. $0/n,\ 1/n,\ \ldots,\ n/n$ *(can think of y as sum of n Bernoulli r. v.'s)*
- Apart from sample size (n), the probabilities p_0 to p_n depend on only 1 parameter:
  - the probability that a selected individual will be '+ve" i.e.,
  - the proportion of "+ve" individuals in sampled population
- Usually denote this (un-knowable) proportion by π (sometimes θ)[1]

Author	**Parameter**	**Statistic**
Hanley et al.	π	$p = y/n$
M&M	p	$\hat{p} = y/n$
Miettinen	P	$p = y/n$

- Shorthand: $y \sim$ Binomial(n, π).

How it arises

- Sample Surveys
- Clinical Trials
- Pilot studies
- Genetics
- Epidemiology ...

[1]M & M use p for *population* proportion and $\hat{p}$ or "p-hat" for observed proportion in a *sample*. Others use Greek letter π for population value (parameter) and p for the sample proportion. Greek letters make the distinction clearer, some textbooks do not use them consistently: e.g., for the population proportion and population mean respectively, M & M use the Arabic letter p and the Greek letter μ (mu)! Some authors (e.g., Miettinen) use UPPER-CASE letters, [e.g. P, OR] for PARAMETERS and lower-case letters [e.g., p, or] for statistics (*estimates* of parameters).

Use

- to make inferences about π from observed proportion $p = y/n$.
- to make inferences in more complex situations, e.g. ...
  - Prevalence Difference: $\pi_1 - \pi_0$
  - Risk Difference (RD): $\pi_1 - \pi_0$
  - Risk Ratio, or its synonym Relative Risk (RR): $\pi_1 \ / \ \pi_0$
  - Odds Ratio (OR): $[\ \pi_1/(1-\pi_1)\ ] \ / \ [\ \pi_0 \ / \ (1-\pi_0)\ ]$
  - Trend in several π's

Requirements for y to have a Binomial (n, π) distribution

- Each element in"population" is 0 or 1, but we are only interested in estimating proportion (π) of 1's; we are not interested in individuals.
- Fixed sample size n.
- Elements selected at random and independent of each other; each element in population has same probability of being sampled.
- Denote by y_i the value of the i-th sampled element. Prob$[y_i = 1]$ is constant (it is π) across i. It helps to distinguish the N *population* values Y_1 to Y_N from the n *sampled* values y_1 to y_n. In the 'What proportion of our time do we spend indoors?' example (in Resources), it is the *random* sampling of the temporal and spatial patterns of 0s and 1s that makes y_1 to y_n independent of each other. The Ys, the elements in the population can be related to each other [e.g. spatial distribution of persons] but if elements are chosen at random, the chance that the value of the i-th element chosen is a 1 cannot depend on the value of y_{i-1}: the sampling is 'blind' to the spatial location of the 1's and 0s.

Binomial Tree (overleaf): Even though there are 2^n possible sequences of 0s and 1s, each with its probability, the calculation of the probability that the sequence in the selected sample contains y 1's and $(n-y)$ 0's is greatly simplified by the fact that Prob$[y_1 = 1] = Prob[y_2 = 1] = \cdots = Prob[y_n = 1] = \pi$. Thus we can calculate prob. of any one sequence that contains y 1's and $(n-y)$ 0's. Since all the sequences have same probability, namely $\pi^y(1-\pi)^{n-y}$, we can, in lieu of adding all such probabilities, simply multiply this one probability by the number, ${}^nC_y$, of unique paths to terminal node.

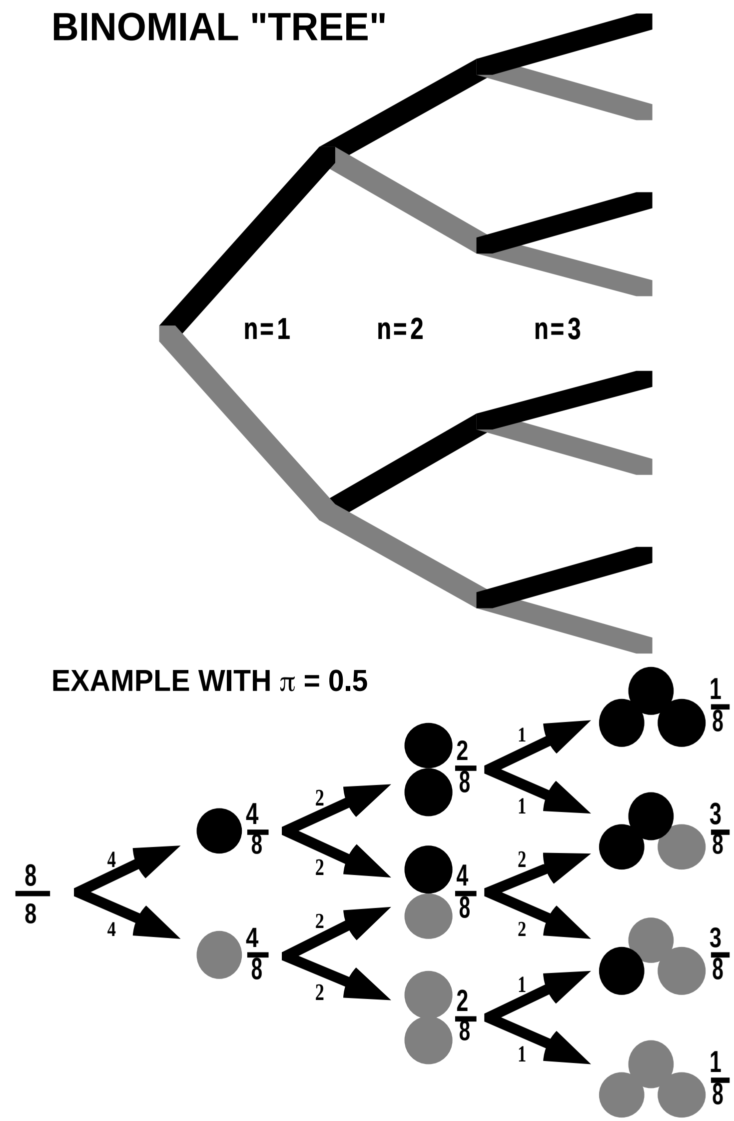

Figure 1: Binomial($n = 3,\ \pi = 0.5$)

1.1 Does the Binomial Distribution Apply if... ?

Interested in	π	the proportion of 16 year old girls in Québec protected against rubella
Choose	$n = 100$	girls: 20 at random from each of 5 randomly selected schools [‘cluster’ sample]
Count	y	how many of the $n = 100$ are protected

• Is $y \sim$ Binomial($n = 100, \pi$)?

“SMAC”	π	Prob[“abn’l” | Healthy] =0.03 for each chemistry in Auto-analyzer with $n = 18$ channels
Count	y	How many of $n = 18$ give abnormal result.

• Is $y \sim$ Binomial($n = 18, \pi = 0.03$)? (cf. Ingelfinger: Clin. Biostatistics)

Interested in	π_u	proportion in ‘usual’ exercise classes and in
	π_e	expt’l. exercise classes who ’stay the course’
Randomly	4	classes of
Allocate	$\underline{25}$	students each to usual course
	$n_u = 100$	
	4	classes of
	$\underline{25}$	students each to experimental course
	$n_e = 100$	
Count	y_u	how many of the $n_u = 100$ complete course
	y_e	how many of the $n_e = 100$ complete course

• Is $y_u \sim$ Binomial($n_u = 100, \pi_u$) ? Is $y_e \sim$ Binomial($n_e = 100, \pi_e$) ?

Sex Ratio	$n = 4$	children in each family
	y	number of girls in family

• Is variation of y across families Binomial (n = 4, π = 0.49)?

Pilot Study		To estimate proportion π of population that is eligible & willing to participate in long-term research study, keep recruiting until obtain
	$y = 5$	who are. Have to approach n to get y.

• Can we treat $y \sim$ Binomial(n, π)?

1.2 Calculating Binomial probabilities:

Exactly

- pdf: formula Prob[y] = ${}^nC_y\,\pi^y\,(1-\pi)^{n-y}$.
- cdf:
  - direct summation of terms in pdf
  - Using link between this sum and the cdf of the F distribution[2].
- Tables: CRC; Fisher and Yates; Biometrika Tables; Documenta Geigy
- Spreadsheet — Excel function `BINOMDIST`(y, n, π, cumulative)
- Statistical Packages:
  - `SAS PROBBNML(p, n, y)` function
  - `Stata` function `Binomial(n,k,p)`
  - `R` functions `dbinom(), pbinom(), qbinom()`:
    probability, distri'n, and quantile functions.

Using an approximation

- Poisson Distribution (n large; small π)
- Normal (Gaussian) Distribution (n large or midrange π)
  - Have to specify *scale* i.e., if say $n = 10$, whether summary is a

	r.v.	e.g.	E	SD
count:	y	2	$n \times \pi$	$\{n \times \pi \times (1-\pi)\}^{1/2}$
				$n^{1/2} \times \sigma_{Bernoulli}$
proportion:	$p = y/n$	0.2	π	$\{\pi \times (1-\pi)/n\}^{1/2}$
				$\sigma_{Bernoulli}/n^{1/2}$
percentage:	$100p\%$	20%	$100 \times \pi$	$100 \times SD[p]$

  - same core calculation for all 3 [only the *scale* changes]

[2] Fisher 1935: see Resources

2 Inference concerning a proportion π, based on s.r.s. of size n

The **Parameter** π of interest: the proportion, e.g., ...

- with undiagnosed hypertension / seeing MD during a 1-year span
- who would respond to a specific 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 left-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 who would become HPV-infected, where seroconversion was in a vaccinated subject:

  e.g. in RCT of HPV16 Vaccine, NEJM Nov 21, 2002: 0 seroconversions in 11084.0 W-Y in vaccinated group vs. 41 in 11076.9 W-Y in placebo group.

  [this proportion is a function of the parameter of interest, the efficacy of the vaccination]

Inference via **Statistic**: the number (y) or proportion $p = y/n$ 'positive' in an s.r.s. of size n.

Frequentist	**Bayesian**
- based on prob[data $\vert\theta$], i.e. - probability statements about data	- based on prob[$\theta\vert$data], i.e., - probability statements about π
Evidence (P-value) against H_0: $\pi = \pi_0$ Test of H$_0$: Is P-value < (preset) α? CI: interval estmate	- point estimate: e.g. mean / mode of posterior distribution of π - (credible) interval

See "Bayesian Inference for a Proportion (Excel)" under Resources Ch 8
cf also A&B §4.7; Colton §4.

2.1 (Frequentist) Confidence Interval for π, based on an observed proportion $p = y/n$

2.1.1 "Exact", first-principles, Confidence Interval

This is not as "awkward to work with" as M & M p. 586 imply.

Example 1

Q: What fraction π of a population would return a 4-page questionnaire?

Data: In a pilot test, $(y =)$ 11 of $(n =)$ 20 return it, i.e. $p = \hat{\pi} = 11/20 = 0.55$

Logic behind the exact Clopper-Pearson $100(1-\alpha)\%$ **CI:**

Limits are calculated so that Binomial Prob $[\geq y \mid \pi_{lower}]$ = Prob $[\leq y \mid \pi_{upper}] = \alpha/2$, exactly. [See Biometrika Tables for Statisticians]

How one can obtain exact CI:

- Tables [Homegrown table[3], Documenta Geigy, Biometrika Tables]
- Nomograms[4]

  Read **horizontally**, nomogram shows the variability of proportions from s.r.s samples of size n. Read **vertically**, it shows: (i) CI $\rightarrow$ symmetry as $p \rightarrow 0.5$ or $n \rightarrow \infty$ [in fact, as $n \times p$ & $n(1-p) \rightarrow \infty$] (ii) widest uncertainty at $\pi = 0.5 \Rightarrow$ can use as a 'worst case scenario'

- software
  - by trial and error, via the Binomial cdf in Excel, R, SAS, ...

[3]To save space, table gives CI's only for $p \leq 0.5$, 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]

[4]95% CI (Biometrika nomogram) 32% to 77%. Nomogram uses c for numerator; enter through lower x-axis if $p \leq 0.5$; in our case $p = 0.55$ so enter nomogram from the top at $\frac{c}{n} = 0.55$ 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 $\pi_{upper} \approx 0.77$.
Notice link between $100(1-\alpha)\%$ CI and two-sided test of significance with a preset α. If true π were $< \pi_{lower}$, 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 $> \pi_{upper}$, there would be less than a 2.5% probability of obtaining, in a sample of 20, this many (11) or fewer respondents. The $100(1-\alpha)\%$ 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 $< \alpha$.

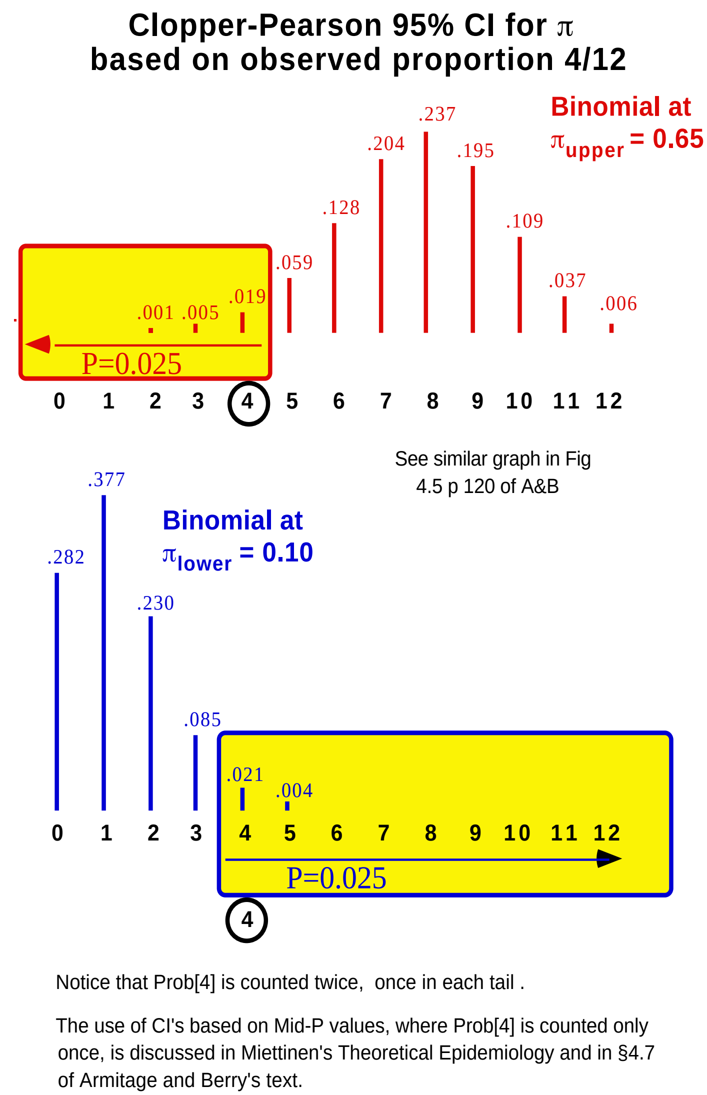

Figure 2: Logic behind Clopper-Pearson CI for proportion π

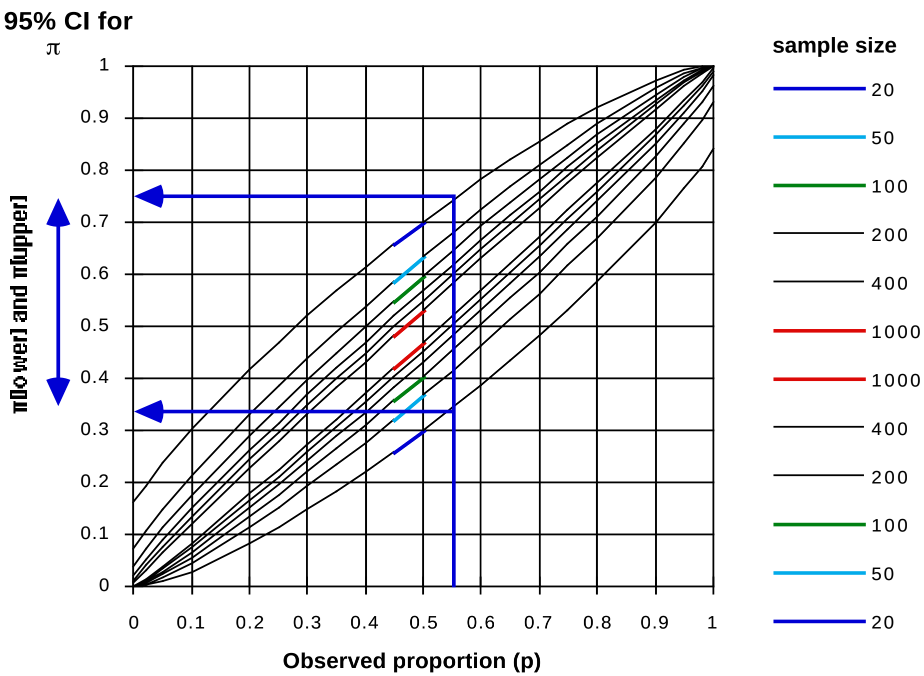

Figure 3: Logic behind Clopper-Pearson CI for proportion π

– using function that gives the inverse of the cdf of the F distribution.[5]

– directly in R, Stata etc

Example 2

Experimental drug gives $y = 0$ successes in $n = 14$ patients[6]; $\Rightarrow \pi =??$

2.1.2 "Approximate", first-principles, Confidence Interval

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

This asymmetric CI (see later) was used to create CI's in nomogram below.

2.1.3 CI based on Gaussian approximation to sampling distribution of p, or function of p

- **CI:** $p \pm z \times \mathrm{SE}[p] = p \pm z \times \{p(1-p)\}^{1/2}$
- e.g. $p = y/n = 300/1000 = 0.3$
- 95% CI: $0.30 \pm 1.96 \times \{0.3 \times 0.7/1000\}^{1/2} = 0.30 \pm 0.03 = 30\% \pm 3.0\%$
- **NB:** The $\pm 3.0\%$ is pronounced and written as "$\pm$ 3 percentage points" to avoid giving the impression that it is 3% of 30%.
- **"Large-n": How Large is large?**
  – A rule of thumb: when the expected no. of positives, $n \times \pi$, and the expected no. of negatives, $n \times (1-\pi)$, 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), i.e., if, with the symmetric Gaussian approximation, neither of the tails of the distribution spills over a boundary (0 or 1 if proportions,

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

[6]95% CI for π (from table) 0% to 23. CI "rules out" (with 95% confidence) possibility that $\pi > 23\%$. [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).

or beyond 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.

- **In SAS**

```
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;
```

- **In Stata**

```
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)
```

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

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

$$\textit{point estimate} \ \pm \ \textit{a multiple of SE}$$

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 A andB (below) take this into account.

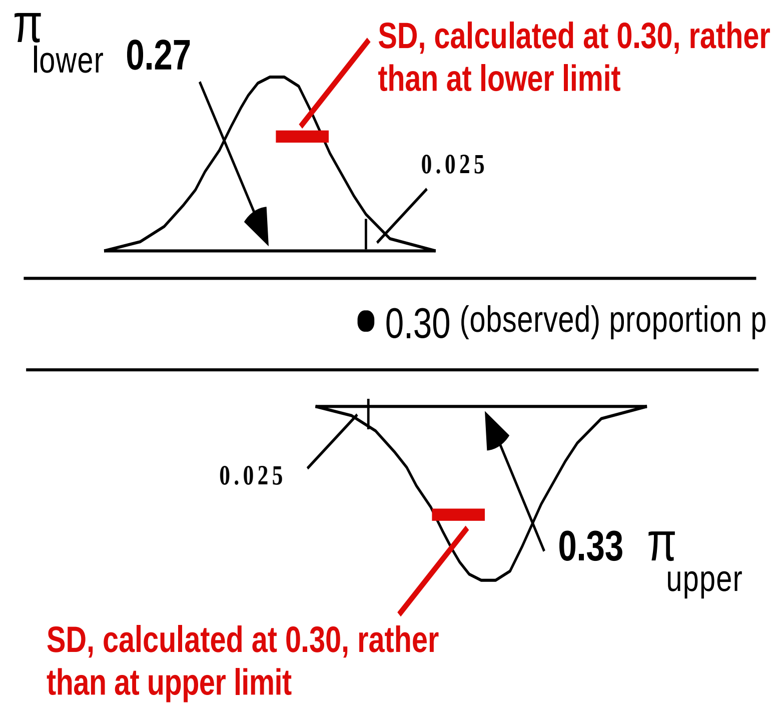

Figure 4: Closer to a 'first-principles' CI

Method A. Based on Gaussian approximation to binomial distribution, but with **SD's calculated separately at the lower limit** π_{LOWER} **and upper limit** π_{UPPER} of π.

The "usual" (not-a-first-principles) CI uses a *common* $SD = \{p[1-p]/n\}^{1/2}$, i.e. the one calculated at the point estimate, i.e., at $\pi = p$.

Define CI for π as the pair of π values $\{\pi_{LOWER}, \pi_{UPPER}\}$ such that,

$$\text{Prob}[p \geq p_{obs} \mid \pi_{LOWER}] = \alpha/2 \ \ \& \ \ \text{Prob}[p \leq p - obs \mid \pi_{UPPER}] = \alpha/2.$$

Use Gaussian approximations to Binomial(n, π_{LOWER}) and

Binomial(n, π_{UPPER}), and solve

$$p = \pi_{LOWER} + z_{\alpha/2}\{\pi_{LOWER} \times [1 - \pi_{LOWER}]/n\}^{1/2}$$

for π_{LOWER} and

$$p = \pi_{UPPER} + z_{\alpha/2}\{\pi_{UPPER} \times [1 - \pi_{UPPER}]/n\}^{1/2}$$

for π_{UPPER}.

This leads to **asymmetric** $100(1-\alpha)\%$ limits of the form[7]

$$\{\pi_{LOWER}, \pi_{UPPER}\} = \frac{1 - \frac{n}{n+z^2} + \frac{2np}{n+z^2} \mp \frac{z\sqrt{4np-4np^2+z^2}}{n+z^2}}{2}$$

Rothman (2002 - p.132) attributes this method ?? to Wilson 1907

Method B: 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 $\alpha = 0.05$, so that $z = 1.96$, we get:

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

[7] References: Fleiss, Statistical Methods for Rates & Proportions; Miettinen, Theoretical Epidemiology, Ch 10. See also `binconf` function in (downloadable) `Hmisc` library in `R`.

Method C: Based on **Gaussian distribution of the logit transformation** of the estimate (p, the observed proportion) and of the parameter π.

Parameter: [8]

$\text{logit}\{\pi\} = \log\{ODDS\}$ [9] $= \log\{\frac{\pi}{(1-\pi)}\} = \log\{\frac{\text{PROPORTION "Positive"}}{\text{PROPORTION "Negative"}}\}$

Statistic: $\text{logit}\{p\} = \log\{odds\} = \log\{\frac{\text{proportion "Positive"}}{\text{proportion "Negative"}}\}$.

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

$$\pi = \frac{\text{ODDS}}{1 + \text{ODDS}} = \frac{\exp[LOGIT]}{1 + \exp[LOGIT]}.$$

likewise...

$$p = \frac{odds}{1 \ + \ odds} = \frac{\exp[logit]}{1 \ + \ \exp[logit]}.$$

$\pi_{LOWER} = \frac{\exp\{\text{LOWER limit of LOGIT}\}}{1+\exp\{\text{LOWER limit of LOGIT}\}} = \frac{\exp\{logit - z_{\alpha/2}SE[logit]\}}{1+\exp\{logit - z_{\alpha/2}SE[logit]\}}$

π_{UPPER} likewise.

$SE[logit] = \{\frac{1}{\#\text{ positive}} + \frac{1}{\#\text{ negative}}\}^{1/2}$

e.g. $p = 3/10 \Rightarrow odds = 3/7 \Rightarrow logit = \log[3/7] = -0.85$.

$SE[logit] = \{1/3 + 1/7\}^{1/2} = 0.69$

$\Rightarrow$ 95% CI in LOGIT[π] scale: $-0.85 \pm 1.96 \times 0.69 = \{-2.2, 0.5\}$

$\Rightarrow$ CI in π scale: $\{\exp(-2.2)/(1 + \exp(), \exp(0.5)/(1 + \exp(0.5)\}$

SAS	Strata
`DATA CI_propn;`	`clear`
`INPUT n_pos n;`	`input n_pos n`
`LINES;`	`1 5`
`3 10`	`2 5`
`;`	`end`
`PROC genmod;`	`glm n_pos,`
`model n_pos/n = /`	`family (binomial n) link (logit)`
`dist = binomial`	
`link = logit waldci;`	

Anti-logit[logit] $= \frac{\exp[logit]}{1 \ + \ \exp[logit]}$ Greenland calls it the "expit function".

Method D: Based on Gaussian distribution of estimate of **log** $[\pi]$

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

Parameter: $\log[\pi]$

Statistic: $\log[p]$

Reverse transformation $\pi = antilog[log[\pi]] = \exp[log[\pi]]$

Likewise $p \leftarrow log[p]$

$\pi_{LOWER/UPPER} = \exp\{\text{LOWER/UPPER limit of } \log[\pi]\}$.

$SE[logit] = \{\frac{1}{\# \text{ positive}} - \frac{1}{n}\}^{1/2}$

Limits for π from p = 3/10 : $\exp[\log[3/10] \pm z \times \{1/3 - 1/10\}^{1/2}]$

Exercises:

1. Verify that you get same answer by calculator and by software
2. even with these logit and log transformations, the Gaussian distribution is not accurate at such small sample sizes as $\frac{3}{10}$. Compare their performance (against the exact methods) for various sample sizes and numbers positive.

SAS	Stata
`DATA CI_propn;`	`input n_pos n`
`INPUT n_pos n;`	`1 5`
`LINES;`	`2 5`
`3 10;`	`end`
`;`	
`PROC genmod;`	`glm n_pos,`
`model n_pos/n = /`	`family (binomial n) link (logit)`
`dist = binomial`	
`link = log waldci;`	

3 Applications, and notes

3.1 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 faon 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.

3.2 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.	$80,000,000\,[0.3 \pm 0.03]$	=	24, 000, 000	±	2, 400, 000
Canada	$8,000,000[0.3 \pm 0.03]$	=	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.

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

3.3.1 Number of Smokers rises by Four Points: Gallup Poll

The Gazette, Montreal, August 8, 1981

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.**

Had the percentage *in the population* really risen? Without a SE (or margin of Error, ME) for each percentage, we are unable to judge whether the 'jump' from 41% to 45% is real or maybe just sampling variation. By 1985, margins of error in the reporting of polls had became mandatory...

3.3.2 39% of Canadians Smoked in Past Week: Gallup Poll

The Gazette, Montreal, Thursday, June 27, 1985

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%
1981	....	45%
1980	....	41
1979	....	44%
1978	....	47%
1977	....	45%
1975	....	47%

(* Smoked regularly or occasionally)

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.**

4 Test of $H_0 : \pi = \pi_{NULL}$

4.1 n small enough → Use Exact Binomial probabilities

- Testing H_0: $\pi = \pi_0$ vs H_a: $\pi \neq \pi_0$ [or H_a: $\pi > \pi_0$]
- Observe $p = y/n$.
- Calculate Prob[observed y, or a y that is more extreme $\mid \pi_0$] using H_{alt} to specify which y's are more extreme i.e. provide even more evidence for H_a and against H_0.
  or...
  use correspondence between a $100(1-\alpha)\%$ CI and a test which uses a 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]

Examples

1. A common question is whether there is evidence against the proposition that a proportion $\pi = 1/2$ [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 $\pi = 1/2$, 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 $\pi = 0.5$ vs $\pi > 0.5$.

2. Another example, dealing with responses in a setup where the "null" is $\pi_0 = 1/3$, example 4.3 is described below.

3. The First Recorded P-Value??? (by a physician no less!) [10]

[10] related by Stigler in his History of Statistics

"AN ARGUMENT FOR DIVINE PROVIDENCE, TAKEN FROM THE CONSTANT REGULARITY OBSERVED IN THE BIRTHS OF BOTH SEXES."

John Arbuthnot, 1667-1735 physician to Queen Anne

Arbuthnot claimed to demonstrate that divine providence, not chance, governed the sex ratio at birth.

To prove this point he represented a birth governed by chance as being like the throw of a two-sided die, and he presented data on the christenings in London for the 82-year period 1629-1710.

Under Arbuthnot's hypothesis of chance, for any one year male births will exceed female births with a probability slightly less than one-half. (It would be less than one-half by just half the very small probability that the two numbers are exactly equal.)

But even when taking it as one-half Arbuthnot found that a unit bet that male births would exceed female births for eighty-two years running to be worth only $(1/2)^{82}$ units in expectation, or

$$\frac{1}{4\ 8360\ 0000\ 0000\ 0000\ 0000\ 0000}$$

a **vanishingly small number**.

"From whence it follows, that it is Art, not Chance, that governs."

4.2 Large n: Gaussian Approximation

Test: $\pi = \pi_0$

Test Statistic: $(p - \pi_0)/SE[p] = (p - \pi_0)/\{\pi_0 \times (1 - \pi_0)/n\}^{1/2}$

Note:
- The *test* uses the *NULL* SE, based on the (specified) π_0.
- The "usual" *CI* uses an SE based on the *observed* p.

4.2.1 (Dis)Continuity Correction

Because we approximate a discrete distribution [where p takes on the values $\frac{0}{n}$, $\frac{1}{n}$, $\frac{2}{n}$, ... $\frac{n}{n}$ 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.

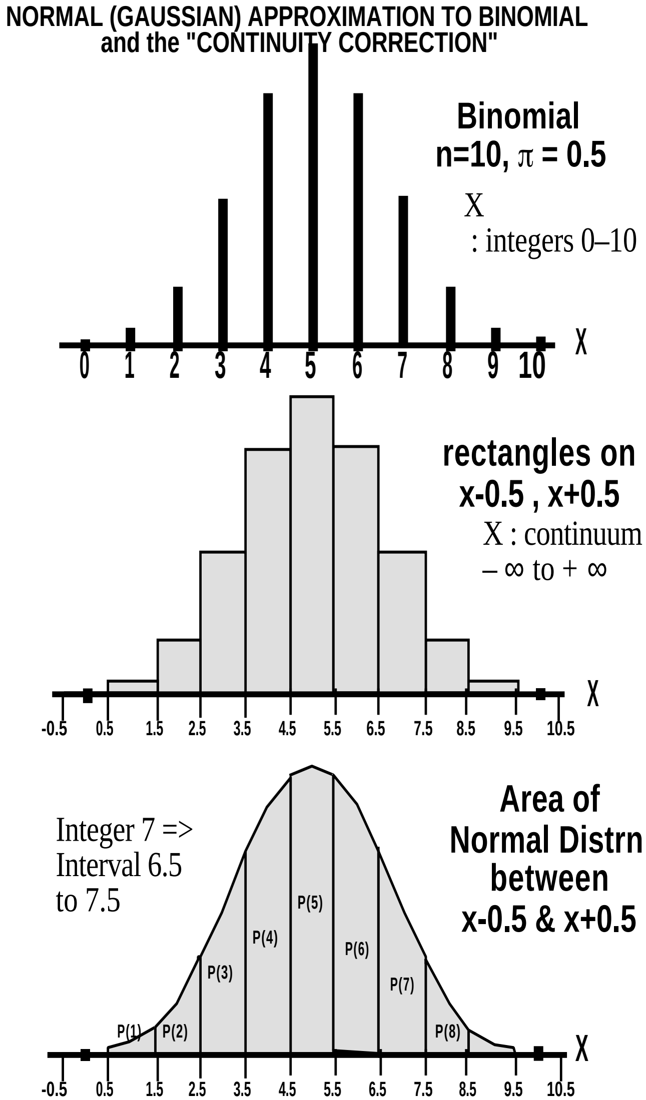

Figure 5: From discrete to continuous

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 $y - n \times \pi_0$ by 0.5 . If we work in the *proportion scale*, the absolute deviation is reduced by $\frac{0.5}{n}$ 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

$$z_c = \frac{|p - n\pi_0| - \frac{0.5}{n}}{SE[p]} = \frac{|p - n\pi_0| - \frac{0.5}{n}}{\pi_0[1 - \pi_0]/n}^{1/2}$$

†Colton [who has a typo in the formula on p $\cdots$] 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.

4.3 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: $\pi() > 0.33$

[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)

Data: $y = 5$ of $n = 12$ mothers correctly identified the odd sample., i.e. $p = 0.42$.

Degree of evidence against H_0 :

$$
\begin{aligned}
&= Prob(5\ or\ more\ correct |\ \pi = 0.33)\ ...\ (a \sum\ of\ 8\ probabilities)\\
&= 1\ -\ Prob(4\ or\ fewer correct |\ \pi = 0.33)\ ...\ (a\ shorter \sum\ of\ only\ 5)\\
&= 1\ -\ [P(0)\ +\ P(1)\ +\ P(2)\ +\ P(3)\ +\ P(4)]\\
&= 0.37*
\end{aligned}
$$

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 the exact probability (0.03762) directly via Excel, using the function `BINOMDIST(4, 12, 0.333, TRUE)`, or using `1-sum(dbinom(1:4,12,1/3))` in `R`.

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

With a sample size of only $n = 12$, we cannot rule out 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 $(1 - \pi_{can})$ who get it right by guessing, we equate $0.65 = \pi_{can} + (1 - \pi_{can})/3$, giving us an upper bound $\pi_{can} = (0.65 - 0.33)/(2/3) = 0.48$, or 48%.

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

P(0) =	0.67^{12}	= 0.008
P(1) =	$\frac{12 \times 0.33 \times P(0)}{[1 \times 0.67]}$	= 0.048
P(2) =	$\frac{11 \times 0.33 \times P(1)}{[2 \times 0.67]}$	= 0.131
P(3) =	$\frac{10 \times 0.33 \times P(2)}{[3 \times 0.67]}$	= 0.215
P(4) =	$\frac{9 \times 0.33 \times P(3)}{[4 \times 0.67]}$	= 0.238
$\sum$		= 0.640

so Prob[5 or more correct $\mid \pi = 0.33] = 1 - 0.64 = 0.36$.

%said that = 0.32 which is mathematically incorrect

What if 50 of 120 mothers identified odd sample?

Test $\pi = 0.33$: $z = (0.42^* - 0.33)/\{0.33 \times (1 - 0.33)/20\}^{1/2} = 2.1$.

So $P = \text{Prob}[\geq 50 \mid \pi = 0.33] = \text{Prob}[Z \geq 2.1] = 0.018$

* 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, and we are dealing with the probability. of obtaining 49.5/120 or more. With $n = 120$, the continuity correction does not make a large difference; however, with smaller n, and its coarser grain, the continuity correction [which makes differences smaller] is more substantial.

5 Planning: Sample Size for CI's and Tests

5.1 n to yield (2-sided) CI with margin of error ME at confidence level $1-\alpha$

(see M&M p. 593, Colton p. 161)

| — margin of error — |

(· • ·) CI

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

$$SE(p) = \{p[1-p]/n\}^{1/2},$$

so ...

$$n = \frac{p[1-p] \times Z^2_{\alpha/2}}{ME^2}$$

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

$$n = \frac{0.25 \times Z^2_{\alpha/2}}{ME^2} \qquad (1)$$

5.2 n for power $1-\beta$ to "detect" a population proportion π that is Δ units from π_0; type I error = α.

(Colton, p. 161)

$$\begin{aligned} n &= \frac{\left\{Z_{\alpha/2}\sqrt{\pi_0[1-\pi_0]} - Z_\beta\sqrt{\pi_1[1-\pi_1]}\right\}^2}{\Delta^2} \\ &\approx \left\{Z_{\alpha/2}\right\}^2 \left\{\frac{\sqrt{\pi[1-\pi]}}{\Delta}\right\}^2 [*] \\ &= \left\{Z_{\alpha/2} - Z_\beta\right\}^2 \left\{\frac{\sigma_{0,1}}{\Delta}\right\}^2 \qquad (2) \end{aligned}$$

* where π is average of π_0 and π_1.

<u>*Notes*</u>: Z_β will be negative; formula is same as for testing μ

5.2.1 Worked Example 1: sample size to test for preferences $\pi = 0.5$ 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 $1-\beta$ against: H_{alt}: $\pi(+) = 0.65$ say.

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

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

$\alpha = 0.05$ (2-sided) & $\beta = 0.2 \Rightarrow Z_\alpha = 1.96$; $Z_\beta = -0.84$

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

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

5.2.2 Worked Example 2: sample size for Δ Taste Test: $\pi_{correct} = 1/3$ vs. $\pi > 1/3$

If set $\alpha = 0.05$ (hardliners might allow 1-sided test here), then $Z_\alpha = 1.645$; If want 90% power, then $Z_\beta = -1.28$; Then using equation 2 above...

$\pi_{correct}$:	0.4	0.5	0.6	0.7	0.8
n for 90 Power against this π	400	69	27	14	8