1 First: Overview of Sampling Distributions

1.1 Examples of Sampling Distributions

Distribution	Statistic whose sampling variability it describes
Binomial(n, π)	no. (or prop'n, p) of 1's in s.r.s. of n from infinite-sized universe containing proportion π of 1's & $(1-\pi)$ of 0's. — can think of p as a *mean* of n 1's and 0's.
Hypergeometric $(N_1\ 1's,\ N_0\ 0's)$	no. (or prop'n, p) of 1's in s.r.s. of size n from universe of 1's and 0's, but where size (N) of universe is finite.
Poisson(μ)	no. of 1's in s.r.s. of n from infinite-sized universe of 0's and 1's, but where π is small, and n is large, so that $np(1-p) \approx np = \mu$ [limiting case of Binomial]. No. of 'events' in a sampled volume of experience (conditions apply ! – see later).
Gaussian	**mean, proportion, count, difference, etc. (n large)**
Student's t	$(\bar{y} - \mu) / (s_y/\sqrt{n}\,)$; $y \sim N(\mu, \sigma)$; $s_y^2 = \sum(y-\bar{y})^2/(n-1)$.
F	ratio of sample variances (used for ANOVA)
???	order statistic as estimate of quantile

1.2 Three ways of calculating sampling variability

1. directly from the relevant discrete distribution, by adding probabilities of the variations in question, e.g. :
   - $0.010 + 0.001 = 0.011$ Binomial prob. of ≥ 9 1's in $n = 10$ if $\pi = 0.5$.
   - 2.5% probability of (Poisson) count ≥ 5 if $\mu = 1.624$
   - 2.5% probability of (Poisson) count ≤ 5 if $\mu = 11.668$
2. from specially-worked out distributions for more complex statistics calculated from continuous or rank data –
   - Student's t, F ratio, χ^2, distribution of Wilcoxon statistic.
3. (**very common**) from the **Gaussian approximation** to the relevant discrete or continuous distribution – by using (an estimate of) the standard deviation of the sampling variation in question and assuming the variation is reasonably symmetric and bell-shaped [every sampling distribution has a standard deviation – its just that it isn't very useful if the distribution is quite skewed or heavy-tailed]. *We give a special name (standard error[1]) to the standard deviation of a sampling distribution in order to distinguish it from the measure of variability of individuals.* Interestingly, we haven't given a special name to the square of the SD of a statistic – we use Variance to denote both SE^2 and SD^2.

1.3 Standard Error (SE) of a sample statistic

What it is

An estimate of the SD of the different values of the sample statistic one would obtain in different random samples of a given size n.

Since we observe only one of the many possible different random samples of a given size, the SD of the sample statistic is not directly measurable.

In this course, in computer simulations, and in mathematical statistics courses, we have the luxury of knowing the relevant information about each element in the population and thus the probabilities of all the possible sample statistics. Thus, for example, we can say that **if** individual Y's are such that $Y \sim N(\mu, \sigma)$, **then** the different possible $\bar{y}$'s will vary from μ in a certain known way. In real life, we don't know the value of μ and are interested in estimating it using the one sample we are allowed to observe. Thus the SE is usually an estimate or a projection of the variation in a *conceptual* distribution i.e. *the SD of all the "might-have-been" statistics.*

Use

If n large enough, the different possible values of the statistic *would* have a Gaussian distribution with a spread of 2-3 SE's on each side of the "true"

[1] Note: Up to Ch 5, M&M use the same notation for the SD of a *mean* or a difference of means as they do for the SD of *individuals* – they use 'SD' for both. Many texts distinguish the two by using SE (Standard Error) when dealing with the SD of a mean or proportion or other *statistic*, and SD when dealing with *individual* variation. M&M in page 500 of Ch 7 say "when the SD of a statistic is *estimated* from the data, the result is called the SE of the statistic." This is a more restricted definition than many authors use. JH's advice: always say what SD or SE one is referring to: the SD or SE of a mean, SD or SE of a median, the SD or SE of a proportion, the SD or SE of a slope, the SD of individual measurements etc. If one sees a SD on its own i.e., without reference to a specific statistic, one would suspect (but cannot be sure) that it is the SD of individuals. *However a SE is never in relation to individuals; it is always in relation to a statistic.*

parameter value [note the "*would have*"]

So, one can calculate the chance of various deviations from the true value.

Can determine under what range of parameter values the observed statistic would/would be an extreme observation.

e.g.

if statistic is $\bar{y}$, we talk of SE of the mean (SEM)

SE($\bar{y}$) describes variation of $\bar{y}$ from μ;

SD(y) describes variation of y from μ (or from $\bar{y}$).

2 Sampling Distribution of $\bar{y}$: Expectation / SD / Shape

- Quantitative variable (characteristic) of interest : Y
- N (effectively) infinite (or sampling with replacement)
- Mean of all Y values in population: μ
- Variance of all Y values in population: σ^2
- **Shape** of distribution of Y's: **Unknown/Unspecified**

- Sample of size n; (i.i.d.) observations $y_1, \ldots, y_n$
- Sample mean: $\bar{y} = (1/n) \sum y_i$

Statistic	**E(Statistic)**	**SD(Statistic)**
$\bar{y}$	μ_y	$\sigma_y/\sqrt{n}$

2.1 ?? Shape of the sampling distribution of $\bar{y}$??

The sampling distribution is the frequency distribution (e.g. in form of histogram or other depiction) we would get if we could observe the mean (or any other calculated statistic) of each of the (infinite number of) different possible random samples of a given size. It quantifies probabilistically how the different possible values of the statistic would vary around some central value. The sampling distribution is strictly conceptual (except, for illustration purposes, in toy classroom exercises where we can actually do the 'what if' exercise for all possible samples from some made-up universe of known values).

Relevance of knowing shape of a sampling distribution:

We will only observe the mean in the one sample we chose; however we can, with certain assumptions, mathematically (beforehand) calculate how far the mean ($\bar{y}$) of a randomly selected sample is likely to be from the mean (μ) of the population. Thus we can say with a specified probability (95% for example) that the $\bar{y}$ that we are about to observe will be no more than Q (some constant, depending on whether we use 90%, 95%, 99%, ...) units from μ. In 'frequentist' inference, we say that in 95% of applications of our procedure, our estimate will come within the stated distance of the target, and so we can have this much 'confidence' in the *procedure*. The probability statement associated with the *confidence interval* for μ is really about the stochastic behaviour of $\bar{y}$ in relation to μ.[2]. We also use the sampling distribution to assess the (probabilistic) distance of a sample mean from some "test" or "Null Hypothesis" value in *statistical tests*.

2.1.1 Example of the distribution of a sample mean:

When summing (or averaging) n 0's and 1's (i.e numbers measured on a 2-point scale), there are only $n+1$ unique possibilities for the result $(0, 1, \ldots, n)$. However, if we were studying a variable, e.g. cholesterol or income, that was measured on a continuous scale, the numbers of possible sample means would be very large and not easy to enumerate. For the sake of illustration, we instead take a simpler variable, that is measured on a *discrete* integer scale

[2] Ideally any description of the CI should involve sentences in which $\bar{y}$ is the subject; μ should not be the subject of the sentence. In the 'frequentist' approach, we are not allowed to say before (or after) the fact that there is a 95% probability that the target will be (is) within the stated distance of where the estimate lands. If one is pretty sure that a particular location is within 15 Km of downtown Montreal, then it is *mathematically* correct to say that one is pretty sure that downtown Montreal is within 15 Km of the location in question. In the frequentist approach, however, it is not '*statistically correct*' to turn this type of statement around and to say that there therefore is a 95% chance that the population mean (μ, the quantity we would like to make inferences about) will not be more than Q units away from the sample mean ($\bar{y}$) we (are about to) observe. The reason has to do with the differeny (asymmetric) logical status of each of the 2 quantities: even though it is unknown, μ is treated as a *fixed* point, while $\bar{y}$ is treated as the *stochastic* element. Thus, for example, if μ were the speed of light, and $\bar{y}$ was a future estimate of it, we cannot speak of μ 'falling' randomly somewhere near $\bar{y}$: instead. In Bayesian inference, it is permitted to speak of the pre-sample and thus the post-sample uncertainty concerning μ.

with a very limited range. However, the principle is the same as for a truly continuous variable.

Imagine we are interested in the average number of cars per household μ in a city area with a large number (N) of households. With an estimate of the average number per household and the total number of households we can then estimate the total number of cars $N \times \mu$. It is not easy to get data on every single one of the N, so we draw a random sample, with replacement, of size n. [The sampling with replacement is simply for the sake of simplicity in this example – we would use sampling without replacement in practice].

How much sampling variation can there be in the estimates we might obtain from the sample? What will the degree of "error" or "noise" depend on? Can we anticipate the magnitude of possible error and the *pattern* of the errors in estimation caused by use of a finite sample?

Suppose that:

- 50% have 0 cars,
- 30% have 1 car,
- 20% have 2 cars.

i.e. in all, there are $0.5 \times N$ 0's, $0.30 \times N$ 1's, and $0.20 \times N$ 2's.

You would be correct to object "but how can we know this - this is the point of sampling"; however, this is a purely *conceptual* or "what if" exercise; the relevance will become clear later.

The mean of the entire set of Y's is

$$\mu_Y = 0 \times 0.5 + 1 \times 0.3 + 2 \times 0.2 = 0.7$$

The variance of the Y's is

$$\begin{aligned} \sigma^2_Y &= (0-0.7)^2 \times 0.5 + (1-0.7)^2 \times 0.3 + (2-0.7)^2 \times 0.2 \\ &= 0.49 \times 0.5 + 0.09 \times 0.3 + 1.69 \times 0.2 \\ &= 0.61 \end{aligned}$$

[Thus, the SD, $\sigma = \sqrt{0.61} = 0.78$ is slightly larger than μ].

We take a s.r.s. of $n = 2$ houses, obtain y_1 and y_2, and use $\bar{y} = (y_1 + y_2)/2$ as $\hat{\mu}_Y$. What estimates might we obtain?

The distribution of all possible $\bar{y}$'s when $n = 2$ is:

Probability (frequency)	$\hat{\mu}$ [i.e., $\bar{y}$]	error [$\bar{y} - \mu$]	% error [% of μ]
25%	$\frac{0}{2} = 0.0$	-0.7	-100
30%	$\frac{1}{2} = 0.5$	-0.2	-29
29%	$\frac{2}{2} = 1.0$	+0.3	+43
12%	$\frac{3}{2} = 1.5$	+0.8	+114
4%	$\frac{4}{2} = 2.0$	+1.3	+186

Most of the possible estimates of μ from samples of size 2 will be "off the target" by quite serious amounts. It's not much good saying that "on average, over all possible samples" the sample will produce the correct estimate.

Check:

$$\begin{aligned} &Average[\bar{y}] \\ &\quad = 0 \times 0.25 + 0.5 \times 0.30 + 1.0 \times 0.29 + 1.5 \times 0.12 + 2.0 \times 0.04 \\ &\quad = 0.7 \\ &\quad = \mu \end{aligned}$$

$$\begin{aligned} &Variance[\bar{y}] \\ &\quad = (-0.7)^2 \times 0.25 + \ldots (1.3)^2 \times 0.04 \\ &\quad = 0.305 \\ &\quad = \sigma^2/2 \end{aligned}$$

A sample of size $n = 4$ would give less variable estimates. The distribution of the $3^n = 81$ possible sample configurations, and their corresponding estimates of μ can be enumerated manually as:

Distribution of all possible $\bar{y}$'s when $n = 4$:

Probability (frequency)	$\hat{\mu}$ [i.e., $\bar{y}$]	error $[\bar{y} - \mu]$	% error [% of μ]
6.25%	$\frac{0}{4} = 0.00$	-0.70	-100
15.00%	$\frac{1}{4} = 0.25$	-0.45	-64
23.50%	$\frac{2}{4} = 0.50$	-0.20	-29
23.4%	$\frac{3}{4} = 0.75$	+0.05	+7
17.61%	$\frac{4}{4} = 1.00$	+0.30	+43
9.36%	$\frac{5}{4} = 1.25$	+0.55	+79
3.76%	$\frac{6}{4} = 1.50$	+0.80	+114
0.96%	$\frac{7}{4} = 1.75$	+1.05	+150
0.16%	$\frac{8}{4} = 2.00$	+1.30	+186

Of course, there is still a good chance that the estimate will be a long way from the correct value of $\mu = 0.7$. But the variance or scatter of the possible estimates is less than it would have been had one used $n = 2$.

Check:

$$\begin{aligned}
&Average[\bar{y}] \\
&\quad = 0 \times 0.0625 + 0.25 \times 0.15 + ... + 2.0 \times 0.0016 \\
&\quad = 0.7 \\
&\quad = \mu \\
&Variance[\bar{y}] \\
&\quad = (-0.7)^2 \times 0.0625 + (-0.45)^2 \times 0.15 + \dots \\
&\quad = 0.1525 \\
&\quad = \sigma^2/4
\end{aligned}$$

If we are happy with an estimate that is not more than 50% in error, then the above table says that with a sample of $n = 4$, there is a 23.50 + 23.40 + 17.61 or ≈ 65% chance that our sample will result in an "acceptable" estimate (i.e. within ±50% of μ). In other words, we can be 65% confident that our sample will yield an estimate within 50% of the population parameter μ.

For a given n, we can trade a larger % error for a larger degree of confidence and vice versa e.g. if $n = 4$, we can be 89% confident that our sample will result in an estimate within 80% of or be 25% confident that our sample will result in an estimate within 10% of μ.

If we use a bigger n, we can increase the degree of confidence, or narrow the margin of error (or a mix of the two), since with a larger sample size, the distribution of possible estimates is tighter around μ. With $n = 100$, we can associate a 20% error with a statement of 90% confidence or a 10% error with a statement of 65% confidence.

But one could argue that there are two problems with these calculations: first, **they assumed that we knew both μ and the distribution of the individual** Y's before we start; second, they used manual enumeration of the possible configurations for a small n and Y's with a small number (3) of possible integer values.

2.1.2 What about real situations with a sample of 10 or 100 from an unknown distributions of Y on a continuous scale?

The answer can be seen by examining the sampling distributions as a function of n in the 'cars per household' example, and in other examples dealing with Y's with a more continuous distribution (see Colton p103-108, A&B p80-83 and M&M 403-404). All the examples show the following:

1. As expected, the variation of possible sample means about the (in practice, unknown) target μ is less in larger samples. We can use the variance or SD of $\bar{y}$ to measure this scatter. The SD (scatter) in the possible $\bar{y}$'s from samples of size n is $\sigma/\sqrt{n}$, where σ is the SD of the *individual* Y's.

   This is true no matter what the shape of the distribution of the individual Y's.

2. If the individual Y's **DO HAVE** a Gaussian distribution, then the distribution of all possible $\bar{y}$'s will be Gaussian.

   BUT...

   even if the individual Y's DO NOT a Gaussian distribution...

   the larger the n [and the more symmetric and unimodal the distribution of the individual Y's], the more the distribution of possible $\bar{y}$'s resembles a Gaussian distribution. And for many distributions, this approximation is already quite good for samples of $n = 30$ or fewer.

The sampling distribution of $\bar{y}$ [or of a sample proportion, or slope or correlation, or other statistic created by aggregation of individual observations ..] is, for a large enough n [and under other conditions[3]], close to Gaussian in shape no matter what the shape of the distribution of individual Y values. This phenomenon is referred to as the CENTRAL LIMIT THEOREM.

We use the notation $Y \sim Distribution(\mu_y, \sigma_y)$ as shorthand to say that "Y has a certain type of distribution with mean μ_y and standard deviation σ_y".

In this notation, the Central Limit Theorem says that

$$\boxed{\begin{array}{c}\text{if } Y \sim ???(\mu_Y,\ \sigma_Y),\text{ then}\\[1em] \bar{y} \sim N(\mu_Y, \sigma_Y/\sqrt{n}),\text{ if } n \text{ is large enough and ...}\end{array}}$$

The Gaussian approximation to certain Binomial distributions is an example of the Central Limit Theorem in action: Individual (Bernoulli) Y's have a 2-point distribution: a proportion (1 - π) have the value $Y = 0$ and the remaining proportion π have $Y = 1$.

The mean (μ) of all (0, 1) Y values in population is π.

The variance (σ^2) of all Y values in population

$$\sigma^2 = (0-\pi)^2 \times (1-\pi) + (1-\pi)^2 \times \pi = \pi(1-\pi)$$

From a sample of size n:

observations $y_1, y_2, ..., y_n$ (sequence of n 0's and 1's)

sample mean $\bar{y} = \frac{\sum y_i}{n} = \frac{number\ of\ 1's}{n} = p.$

CLT ...

If $Y \sim Bernoulli(\mu = \pi,\ \sigma = \sqrt{\pi[1-\pi]})$, then

$p = \bar{y} \sim N(\pi, \sqrt{\pi[1-\pi]}/\sqrt{n})$ if n is sufficiently 'large' and π is not extreme.[4]

Returning to example on estimating $\mu_{cars/household}$.

If $n = 100$, then the SD of possible $\bar{y}$'s from samples of size $n = 100$ is $\sigma/\sqrt{100} = 0.78/10 = 0.078$. Thus, we can approximate the distribution of possible $\bar{y}$'s by a Gaussian distribution with a mean of 0.7 and a standard deviation of 0.078, to get ...

		Interval	**Prob.**	**% Error**
$\mu \pm 1.00SD(\bar{y})$	0.7 ± 0.078	0.62 to 0.77	68%	±11%
$\mu \pm 1.50SD(\bar{y})$	0.7 ± 0.117	0.58 to 0.81	87%	±17%
$\mu \pm 1.96SD(\bar{y})$	0.7 ± 0.143	0.55 to 0.84	95%	±20%
$\mu \pm 3.00SD(\bar{y})$	0.7 ± 0.234	0.46 to 0.93	99.7%	±33%

[The Gaussian-based intervals are only slightly different from the results of a computer simulation in which we drew samples of size 100 from the above Y distribution]

If this variability in the possible estimates is still not acceptable and we use a sample size of $n = 200$, the standard deviation of the possible $\bar{y}$'s is not

[3] On the degree of symmetry and dispersion of the distribution of the individual Y's.

[4] E[no. 'positive' = numerator = $\sum y_i$] needs to be sufficiently far 'inland' from 0 and from 1, and n needs to be large enough that Binomial(n, π) distribution does not have much probability mass on 0 or n,, i.e., so that the Gaussian approximation to it does not spill over onto, and thus place substantial probability mass on, sillyvalues such as $\cdots -3, -2, -1$ or on $n+1, n+2, \ldots$. One Rule of Thumb for when the Gaussian approximation provides a reasobable accurate approximation is that both $n \times \pi \geq 5$ and $n \times (1-\pi) \geq 5$, i.e. the expected number of 'positives' should be 'inland' by at least 5 from both boundaries.

halved (divided by 2) but rather divided by $\sqrt{2} = 1.4$. We would need to go to $n = 400$ to cut the s.d. down to half of what it is with $n = 100$.

[Notice that in all of this (as long as we sample with replacement, so that the n members are drawn independently of each other), the size of the population (N) didn't enter into the calculations at all. The errors of our estimates (i.e. how different we are from μ on randomly selected samples) vary directly with σ and inversely with $\sqrt{n}$. However, if we were interested in estimating $N\mu$ rather than μ, the absolute error would be N times larger, *although the relative error would be the same in the two scales.*]

Message from diagram opposite:

The variation of means is closer to Gaussian than the variation of the individual observations (the panel where we have a mean of $n = 1$ values can be taken as the distribution of individual Y's), and the bigger the sample size, the closer to Gaussian: with large enough n, you could not tell from the sampling distribution of the means what the shape of the distribution of the individual 'parent' observations was. Averages of $n = 16$ are "effectively" Gaussian in this example. How 'fast' the CLT will 'kick in' is a function of how asymmetric the distribution of Y is.

2.1.3 Annother example of central limit theorem at work: word lengths

The distribution of the lengths of words has a long right tail (see '$n = 1'$ panel in Fig 2), but the (sampling) distribution of the possible values of the sample mean when $n = 2$ has less of a long right tail, and the distribution of $\bar{y}_{n=4}$ is less asymmetric and closer to Gaussian, and that of $\bar{y}_{n=16}$ even more so.

You can think of the effects of increasing n as two-fold:

- It makes for a 'finer' measuring scale (just as with a ruler with finer gradations). For example, if the Y's are recorded with a 'bin-width' of δY (integers in our two examples), then the sample mean has a 'bin-width' of $\delta Y \div n$.

- Extreme sums, and thus extreme means, are less likely: with large enough n, there are enough extremes from each end of the distribution that they will tend to cancel each other.

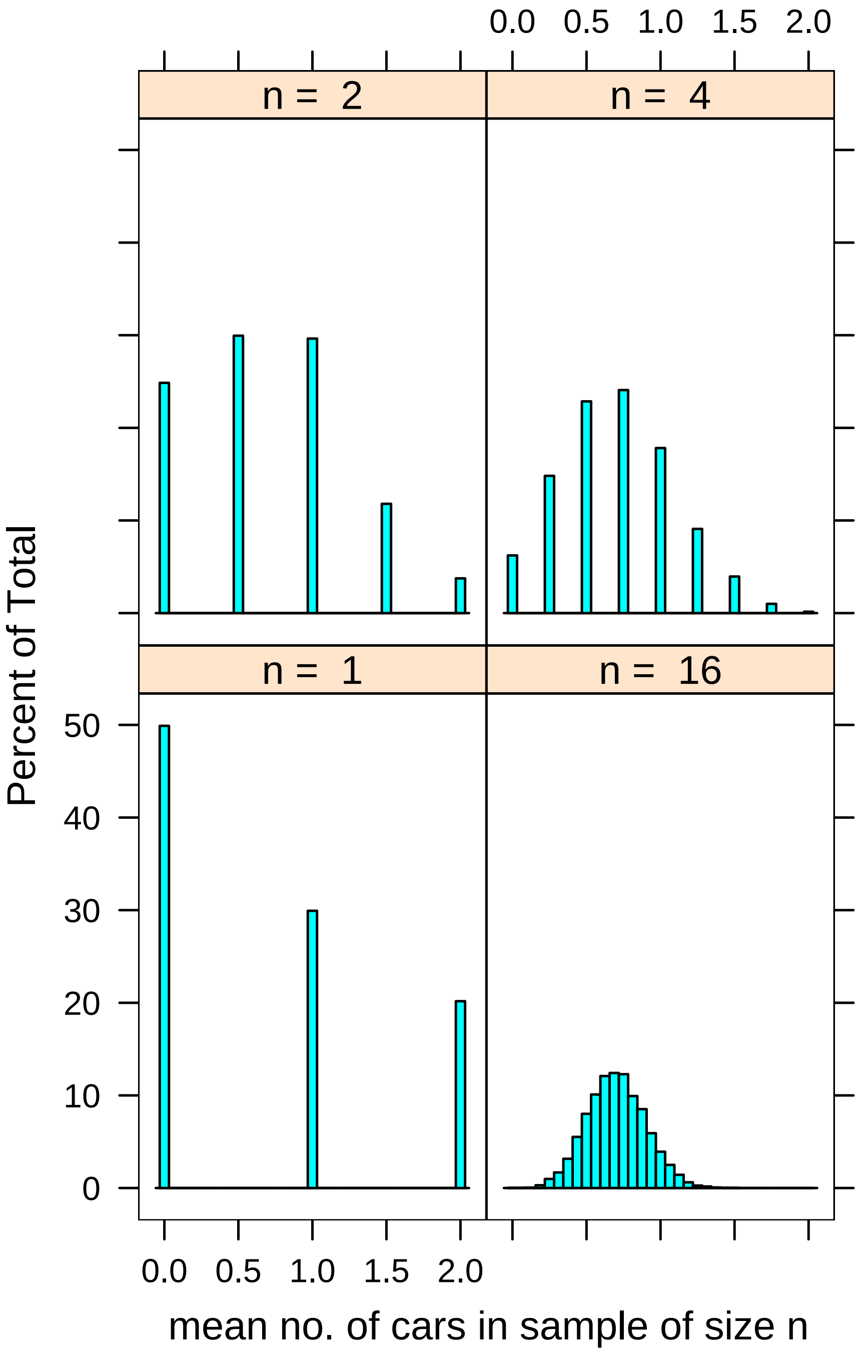


Figure 1: Illustration of Central Limit Theorem

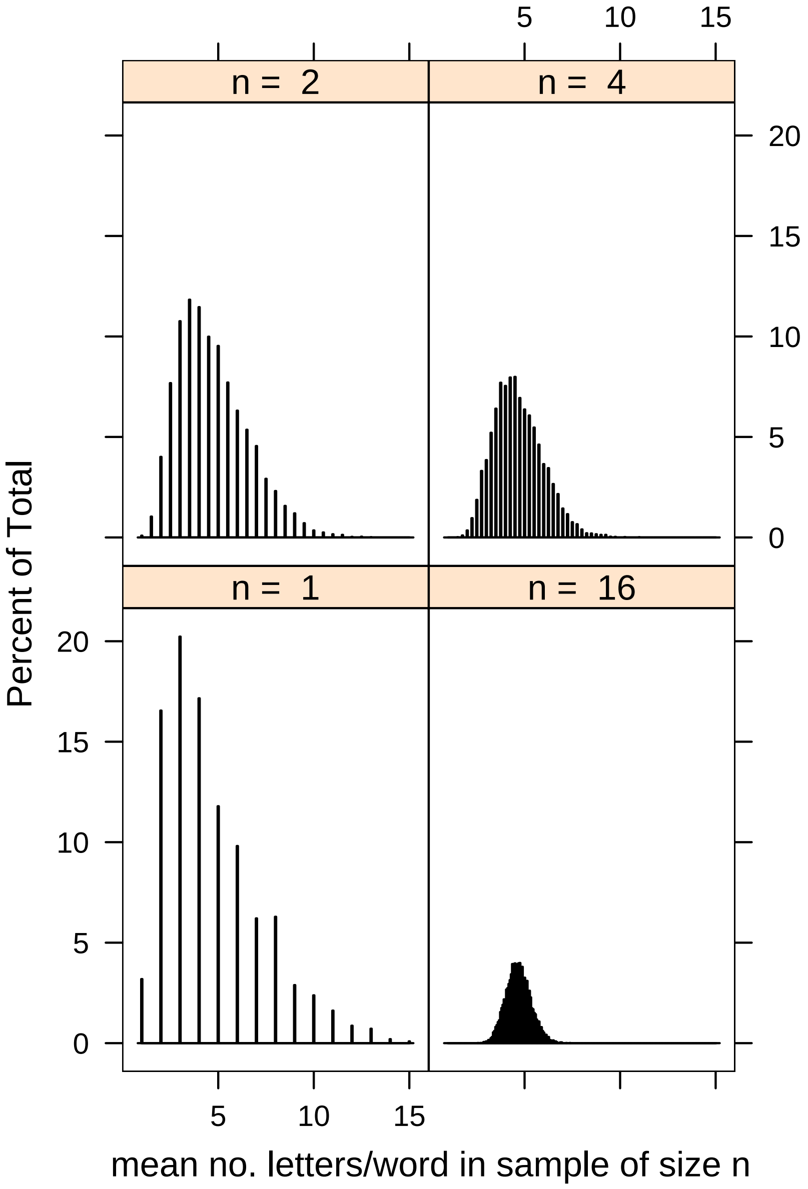


Figure 2: Illustration of Central Limit Theorem

2.1.4 Other examples of central limit theorem in action work:

- Lengths of n-th generation copies of the 1-metre bar:

  Suppose we use a piece of string (or a large photocopier) to make 2 copies of the 1-metre prototype bar `http://en.wikipedia.org/wiki/Metre`. But suppose that in doing so, we make independent errors of either +1cm or -1cm. From each of these 2, we we 2 second generation copies, again with independent +/- errors of 1mm, and from these 8 third generation copies, etc.. What would be the distribution of the lengths of the 2^{16} 16-th generation copies? They will have a binomial-shape distribution, ranging from 84cm to 116cm, and centered on 100cm. A plot this (using
  `plot(100+seq(0,16),dbinom(seq(0,16),16,0.5),type="h")`
  say, in R) you will see that it has the shape of what Gauss called the Law or Errors. If you make the errors smaller, but have more of them, the variation will be effectively on a continuous scale. One way to establish the Normal density $\phi(y, \mu, \sigma)$ is to apply Stirling's formula (`http://en.wikipedia.org/wiki/Stirling's_approximation`) to the Binomial probabilities in the case of a large n and "success" probability $\pi = 0.5$.

- Generating random numbers from a Gaussian distribution:

  Since Φ^{-1}, the inverse of the cdf of a N(0,1) random variable does not have a closed form, the inverse cdf method of obtaining Gaussian random numbers has to rely on an approximation involving powers.[5] Another way to produce values that have close to a N(0,1) distribution is by summing $n = 12$ realizations from a Uniform(0,1) distribution and subtracting 6 from the sum.

  ```
  # sum of 12 random numbers from U(0,1)

  r = function(dummy) sum(runif(12))-6 ;
  sum.12.uniforms = sapply(1:50000,r);
  hist(sum.12.uniforms,breaks=50)
  ```

- There is also a CTL that applies to sums of *independent* but *not identically distributed* random variables. The key element is the independence. See the cartoon "The Central Limit Theorem in Action (courtesy Lawrence Joseph)" in the Resources page. If the components were

[5] For an exact method, see `http://en.wikipedia.org/wiki/Box-Muller_transform`

correlated, say because of weather, then it would impede the cancellation of extremes.

```
days=2000;

walk.to.bus = rnorm(days,mean=4,sd=1);
wait.for.bus = runif(days,4,16);
bus.ride = rnorm(days,mean=20,sd=2);
walk.up.hill = rgamma(days,scale=2,shape=3/2);

hist(walk.to.bus); summary(walk.to.bus);
hist(wait.for.bus); summary(wait.for.bus);
hist(bus.ride); summary(bus.ride);
hist(walk.up.hill); summary(walk.up.hill);

total.time = walk.to.bus + wait.for.bus + bus.ride +
walk.up.hill;

summary(total.time);
c(mean(total.time),sd(total.time),var(total.time))
hist(total.time)
boxplot(total.time)
```

3 Standard Error (SE) of combination or weighted average of estimates

$$
\begin{aligned}
SE(\sum estimates) &= \sqrt{\sum([SE\ of\ each\ estimate]^2)} \\
SE(constant \times estimate) &= constant \times SE(estimate) \\
SE(constant + estimate) &= SE(estimate) \\
SE(\sum w_i \times estimate_i) &= \sqrt{\sum(w_i^2 \times [SE\ estimate_i]^2)} \qquad (1)
\end{aligned}
$$

This last one is important for combining estimates from stratified samples, and for meta-analysis:

In an estimate for the overall population, derived from a stratified sample, the weights are chosen so that the overall estimate is unbiased for the weighted average of the stratum-specific parameters i.e. the w's are the relative sizes of the segments (strata) of the overall population (see "combining estimates ... entire population" below). The parameter values usually differ between strata: this is why stratified sampling helps. The *estimate* for this weighted avearge of the stratum-specific parameters is formed as a weighted average of the age-specific parameter estimates, and so one has no choice in the choise of weights: they must reflect the proportions of population in the various strata.

If instead, one had several estimates of a single parameter value (a big assumption in the 'usual' approach to meta-analyses), but each estimate had a different uncertainty (precision), one should take a weighted average of them, but with the weights inversely proportional to the amount of uncertainty in each. From the formula above one can verify by algebra or trial and error that the smallest variance for the weighted average is obtained by using weights proportional to the inverse of the variance (squared standard error) of each estimate. If there is variation in the parameter value, a 'fixed effects' SE is too small. The 'random effects' approach to meta-analyses weights each estimate in inverse relation to an amalgam of (i) each SE and (ii) the 'greater-than-random' variation between estimates [it allows for the possibility that the parameter estimates from each study would not be the same, even if each study used huge n's). The SE of this weighted average is larger than that using the simpler (called fixed effects) model; as a result, CI's are also wider.

3.1 Combining Estimates from Subpopulations to form an Estimate for the Entire Population

Suppose several (say k) sub-populations or "strata" of sizes N_1, N_2, ... N_k, form one entire population of size $\sum N_k = N$. Suppose we are interested in the average level of a quantitative characteristic, or the prevalence of a qualitative characteristic in the entire population. Denote this numerical or binary characteristic in each individual by Y, and an average or proportion (or total) across all individuals in the population by θ. It could stand for a mean (μ), a total ($T_{amount} = N \times \mu$), a proportion ($\pi$), a percentage ($\% = 100\pi$) or a total count ($T_c = N \times \pi$).

Examples:

If Y is a measured variable (i.e. "numerical")

μ:	the annual (per capita) consumption of cigarettes
T_{amount}:	the total undeclared yearly income ($T_{amount} = N \times \mu$ and conversely $\mu = T_{amount}/N$)

If Y is a binary variable (i.e. "yes / no")

π:	the proportion of persons who exercise regularly
$100\pi\%$:	the percentage of children who have been fully vaccinated
$N\pi$:	the total number of persons who need R_x for hypertension ($T_c = N\pi$; $\pi = T_c/N$)

The sub-populations might be age groups, the 2 sexes, occupations, provinces, etc. There is a corresponding θ_i for the i-th of the k sub-populations. Rather than study every individual each each stratum, one might instead measure Y in a sample from each stratum.

3.2 Estimate of overall μ, π, *or* $\pi\%$, by combining estimates:

Sub Popln	Size	Relative Size $W_i = N_i/N$	Sample Size	Estimate of θ_i	SE of estimate
1	N_1	W_1	n_1	e_1	$SE(e_1)$
...	...	...	...	...	...
...	...	...	...	...	...
k	N_k	W_k	n_k	e_k	$SE(e_k)$
Total	$\sum N_i = N$	$\sum W_i = 1$	$\sum n_i = n$	$\sum W_i e_i$	$\sum W_i^2 [SE(e_i)]^2$

Note 1 To estimate T_{amount} or T_c, use weights $W_i = N_i$;

Note 2 If any sampling fraction $f_i = n_i/N_i$ is substantial, the SE of the e_i should be scaled down i.e. it should be multiplied by $\sqrt{(1 - f_i)}$.

Note 3 If variability in Y within a stratum is smaller than across strata, the smaller SE obtained from the SE's of the individual stratum specific estimates more accurately reflects the uncertainty in the overall estimate. Largest gain over SRS is when large inter-stratum variability.

4 (FREQUENTIST) Inference for μ – small n: Student's t distribution

Use: when we replace σ by s (an estimate of σ) when forming CI's, or carrying out statistics tests, using the sample mean and the standard error of the mean.[6] We proceed in the usual way – expressing the distance of $\bar{y}$ from μ in terms of multiples of $\text{SE}_{\bar{y}} = s/\sqrt{n}$ – except that we use a different 'reference' distribution than the usual Z (Gaussian)one. The percentiles of this new distribution are further from 0 than the familiar 0.84, 1.28, 1.645, 1.96, etc,of the Z distribution: how much further depends on the amount of data (i.e., the $(n-1)$ used to estimate σ^2.

To paraphrase, and quote from, Student's 1908 paper... (*italics* by JH)

> (Until now) "the usual method of calculating the probability that "μ is within a given distance of $\bar{x}$ [7] is to assume $\mu \sim N(\bar{x}, s/\sqrt{n})$, where s is the standard deviation of the sample, and to use the tables of the (Normal) probability integral." *But, with smaller n, the value of s "becomes itself subject to increasing error."* In some instances, we can use a more reliable value of s from earlier experiments, but "in some chemical, many biological, and most agricultural and large scale experiments," we are forced to "judge of the uncertainty of the results from a small sample, which itself affords the only indication of the variability." Inferential methods for such small-scale experiments had "hitherto been outside the range of statistical enquiry."

Rather than merely complain, Gosset did something about it.

> Although it is well known that the method of using *the normal curve is only trustworthy when the sample is "large", no one has yet told us very clearly where the limit between "large" and "small" samples is to be drawn.* The *aim* of the present paper is to *determine the point at which we may use the tables of the (Normal) probability integral* in judging of the significance of the *mean* of a series of experiments, and to furnish *alternative tables for use when the number of experiments is too few.*

[6] it is also used in a wider context, where we have a ratio of a Gaussian random variable, and the square root of an independent random variable that has a chi-squared distribution.

[7] This way of writing, i.e., of making μ the subject of the sentence, was commonplace in 1908; it is not politically or statistically correct today, unless one adopts a Bayesian viewpoint, where the focus is directly on the pre- and post-data uncertainty concerning μ. [JH]

Student assumed that the Y values are normally distributed, so that $\bar{y}$ has a Gaussian sampling distribution.[8]

“Student's” t distribution is the (conceptual) distribution one would get if one...

- took (an infinite number of) samples, of a given size n, from a Normal(μ, σ) distribution
- formed the ratio $t = (\bar{y} - \mu) \ / \ (s/\sqrt{n})$ from each sample
- compiled a histogram of the ratios.

In fact, to check that his derivation was correct, Gosset[9] actually performed a simulation in which he followed the above process:

> Before I had succeeded in solving my problem analytically, I had endeavored to do so empirically. The material I used was a ... table containing the height and left middle finger measurements of 3000 criminals.... The measurements were written out on 3000 pieces of cardboard, which were then very thoroughly shuffled and drawn at random... each consecutive set of 4 was taken as a sample... [i.e. $n = 4$ above]... and the mean [and] standard deviation of each sample determined.... This provides us with two sets of... 750 (ratios) on which to test the theoretical results arrived at. The height and left middle finger... table was chosen because the distribution of both was approximately normal...”

Sampling distribution of t

- is symmetric around 0 (just like Z distribution)
- has shape like that of the Z distribution, but with SD slightly larger than unity i.e. slightly flatter & more wide-tailed; $Var[t] = df/(df-2)$.
- its shape becomes indistinguishable from that of Z distribution as $n \to \infty$ (in fact as n goes much beyond 30.)

- Instead of $\pm 1.96\sigma/\sqrt{n}$ for 95% confidence, we need

Multiple	n	Degrees of freedom (‘df’)
±3.182	4	3
...	...	...
±2.228	11	10
...	...	...
±2.086	21	20
...	...	...
±2.042	31	30
...	...	...
±1.980	121	120
...	...	...
±1.960	∞	∞

-

Test of $\mu = \mu_0$	Confidence Interval (CI) for μ
$t\,ratio = (\bar{y} - \mu_0/(s/\sqrt{n})$	$\bar{y} \pm t \times s/\sqrt{n}$

4.1 WORKED EXAMPLE: CI and Test of Significance

Response of interest: D: Increase (D) in hours of sleep with a test medication.

Test:

$$\begin{array}{cccl} \mu_D & = & 0 & H_0 \\ & \neq & 0 & H_{alt} \\ \\ \alpha & = & 0.05 & \text{2 sided} \end{array}$$

[8] even if the Y's were not normally distributed, but n was sufficiently large, the Central Limit Theorem would guarantee that the distribution of all possible $\bar{y}$'s is close to a Gaussian distribution – but with large enough n, one would have sufficient degrees of freedom to estimate σ quite precisely, and so the problem would disappear.

[9] Student. The probable error of a mean, *Biometrika* 1908.

Data:[10]

	Hours of Sleep †		Difference:
Subject	Drug	Placebo	Drug minus Placebo
			d
1	6.1	5.2	0.9
2	7.0	7.9	-0.9
3	8.2	3.9	4.3
4	•	•	2.9
5	•	•	1.2
6	•	•	3.0
7	•	•	2.7
8	•	•	0.6
9	•	•	3.6
10	•	•	-0.5
			$\bar{d} = 1.78$ $SD[d] = 1.77$

Test statistic: $t = (1.78 - 0)/(1.77/\sqrt{10} = 3.18.$

Critical Value: $|t_9| = 2.26$

Since $3.18 > |t_9|$, we "reject" H_0.

95% CI for μ_D:

$$\begin{array}{l} 1.78 \pm t_9 \times \text{ SE}_{\bar{d}} \\ 1.78 \pm 2.26 \times (1.77/\sqrt{10}) \\ 1.78 \pm 1.26 \\ 0.5 \text{ to } 3.0 \text{ hours} \end{array}$$

4.2 Another worked Example, with graphic:

Posture, blood flow, and prophylaxis of venous thromboembolism. *CPG Barker, The Lancet Vol 345. Aprill 22, 1995, p. 1047.*

Sir–Ashby and colleagues (Feb 18, p 419) report adverse effects of posture on femoral venous blood flow. They noted a moderate reduction velocity when a patient was sitting propped up at 35° in a hospital bed posture and a further pronounced reduction when the patient was sitting with legs dependent. Patients recovering from operations are often asked to sit in a chair with their feet elevated on a footrest. The footrests used in most hospitals, while raising the feet, compress the posterior aspect of the calf. Such compression may be important in the aetiology of venous thrombo-embolism. We investigated the effect of a footrest on blood flow in the deep veins of the calf by dynamic radionuclide venography.

Calf venous blood flow was measured in fifteen young (18-31 years) healthy male volunteers. 88 MBq technetium-99m-labelled pertechnetate in 1 mL saline was injected into the lateral dorsal vein of each foot, with ankle tourniquets inflated to 40 mm Hg, and the time the bolus took to reach the lower border of the patella was measured (Sophy DSX Rectangular Gamma Camera). Each subject had one foot elevated with the calf resting on the footrest and the other plantigrade on the floor as a control. The mean transit time of the bolus to the knee was 24.6 s (SE 2.2) for elevated feet and 14.8 s (SE 2.2) for control feet [see figure 3]. The mean delay was 9.9 s (95% CI 7.8-12.0).

Simple leg elevation without hip flexion increases leg venous drainage and femoral venous blood flow. The footrest used in this study raises the foot by extension at the knee with no change in the hip position. Ashby and colleagues' findings suggest that such elevation without calf compression would produce an increase in blood flow. Direct pressure of the posterior aspect of the calf therefore seems to be the most likely reason for the reduction in flow we observed. Sitting cross-legged also reduced calf venous blood flow, probably by a similar mechanism. If venous stasis is important in the aetiology of venous thrombosis, the practice of nursing patients with their feet elevated on footrests may need to be reviewed.

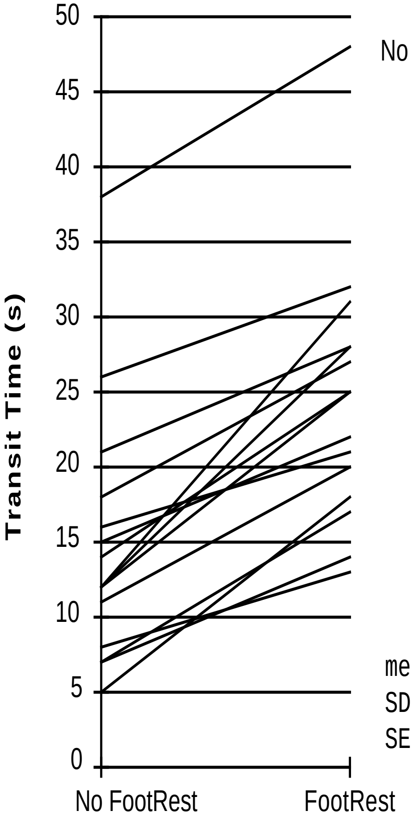
[*Data abstracted from diagram; calculations won't match exactly those in text*]

$\bar{d}(SD) = 9.8(4.1)$; $t = (9.8 - 0)/(4.1/\sqrt{15}) = 9.8/1.0 = 9.8$

Critical ratio: $t_{14,0.05} = 2.145$. So, the observed difference is 'off the t-scale'. This corroborates the impression gained from visual display of the data.

95% CI for μ_D : $9.8 \pm 2.145 \times 1.0$ i.e., $7.7s$ to $11.9s$.

[10] table deliberately omits the full data on the drug and placebo conditions: this is to emphasize that all we need for the analysis are the 10 *differences*.



	No FootRest	FootRest	Delay
	38	48	10
	26	32	6
	21	28	7
	18	27	9
	16	21	5
	15	22	7
	14	25	11
	12	28	16
	12	31	19
	12	25	13
	11	20	9
	8	13	5
	7	17	10
	7	14	7
	5	18	13
mean	14.8	24.6	9.8
SD	8.5	8.7	4.1
SEM	2.2	2.2	1.0

Figure 3: Raw data, and summary statistics

Remarks:

Whereas the mean, 9.8, of the 15 within-person between-conditionc differences is arithmetically equal to the difference of the 2 means of 15, the SE of the mean of these 15 differences is not the same as the SE of the difference of two independent means. In general...

$$Var(\bar{y}_1 - \bar{y}_2) = Var(\bar{y}_1) + Var(\bar{y}_2) - 2 \times Covariance(\bar{y}_1, \bar{y}_2)$$

Double-check that one can arrive at the SE of 1.1 for the mean delay by subtracting twice the covariance from the sum of the two variances, and then taking the square root of this.

Indeed, the effect of pairing is to remove the intrinsic between-person variance, and focus the within-person differences. **Applying an inefficient statistical analysis to data collected by an efficient statistical design is a common 'Type III' error!**

Authors continue to report the SE of each of the 2 means, but the 2 separate SEs are of little use here, since we are not interested in the difference of means, but in the mean difference.

Calculating

$$Var(\bar{y}_1 - \bar{y}_2) = Var(\bar{y}_1) + Var(\bar{y}_2) = 2.2^2 + 2.2^2 = 9.7$$

so that the SE$_{diff.\,in\,means}$ is $\sqrt{9.7} = \sqrt{2} \times 2.2 = 3.1$ assumes that we used *one set of 15 subjects* for the No FootRest condition, and *a different set of 15* for the FootRest condition, a much noisier contrast.

Fortunately, it turned out that in this study the signal is much greater than the 'noise'. Thus, even the inefficient (2-independent samples) analysis, based on a $SE_{\bar{y}_1 - \bar{y}_0} = 3.1$, would have produced a statistically significant 2-sample t-ratio of $9.8/3.1 = 3.2$.

See article (in jh's catalogued collection) on display of data from pairs.

4.3 Sample Size for CI's and test involving μ

4.3.1 n required for a (2 sided) CI with margin of error (ME) at confidence level $1 - \alpha$

```
<-- Margin of Error(ME) --   •  -- Margin of Error(ME) -->
<-------------------------   •  ------------------------->
```

- large-sample CI: $\bullet \pm \text{ME} = \bar{y} \pm Z_{\alpha/2} SE(\bar{y})$
- $\text{SE}(\bar{y}) = \sigma/\sqrt{n}$, so solving for n...
- $\boxed{n = (\sigma^2 \times Z^2_{\alpha/2}) \,/\, \text{ME}^2.}$
- If n small, replace $Z_{\alpha/2}$ by $t_{\alpha/2}$

Typically we do not know σ, so we use use a pre-study estimate of it.

In planning n for example just discussed, authors might have had pilot data on inter leg differences in transit time – with both legs in the No FootRest position. Sometimes, one has to 'ask around' as to what the SD of the d's will be. Always safer to assume a higher SD than might turn out to be the case.

4.3.2 n required to have power $1 - \beta$ when testing $H_0 : \mu = \mu_0$, if unknown mean, μ, is Δ units from μ_0, i.e., if $\mu_{alt} - \mu_0 = \Delta$, and if test is carried out with Probability[type I error] $= \alpha$.

[cf. Fig 4, as well as Colton p. 142, and CRC table on next page.]

- Assume that the 'unit variability', σ_Y, is the same under H_0 and H_{alt}, so that
$$\text{SE}_0[\bar{y}] = \text{SE}_{alt}[\bar{y}] = \sigma_Y/\sqrt{n}.$$
- Need
$$Z_{\alpha/2} \times \text{SE}_0[\bar{y}] + Z_\beta \times \text{SE}_{a;lt}[\bar{y}] \geq \Delta.$$
- Substitute $\text{SE}[\bar{y}] = \sigma_Y/\sqrt{n}$.
- Solve for n:

$\boxed{n \geq [Z_{\alpha/2} + Z_\beta]^2 \times [\sigma_Y/\Delta]^2}$ σ_Y/Δ is the "*noise-to-signal*" ratio.

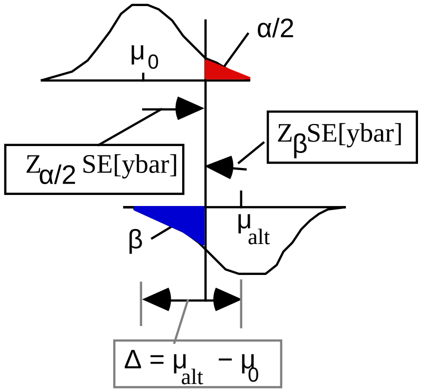


Figure 4: Link between test size (α), sample size, n, power $(1 - \beta)$ and Δ.

Notes:

- To make life simpler, JH has made the diagram and formula in terms of the *absolute* values of $Z_{\alpha/2}$ and Z_β. Thus, *be careful* with the *sign* of Z_β: If $\mu_{alt} > \mu$, then the alternative distribution of $\bar{y}$ is to the right of μ_0 (as in diagram), so that a power of more than 50% means that technically, Z_β is negative. e.g. :
$$\alpha = 0.05 \,\&\, \beta = 0.2 \;\; \Rightarrow Z_{\alpha/2} = 1.96 \,\&\, Z_\beta = -0.84.$$
If back-solving for Z_β (and thus β) in terms of n, Δ and σ_Y, be especially careful as to the sign of Z_β: **always draw a diagram.**
- While it can be α *or* $\alpha/2$, its *always* $1 - \beta$, *never* $1 - \beta/2$!
- *Technically*, if n is small, should use the more conservative $t_{\alpha/2}$ and t_β: see table on the following page. Since the required n is a function of $t_{\alpha/2}$ and t_β and vice versa, arriving at this table takes some iteration.
- The question of what Δ to use is not a matter of statistics or samples, or what the last researcher found in a study, but rather the "difference that would make a difference" i.e., it is a clinical judgement, and includes the impact, cost, alternatives, etc... JH thinks of it as the Δ that IF IT WERE TRUE would lead to a difference in management or a substantial risk, or ...

4.4 Sign Test for median

Test:

$$\begin{array}{lcll} \text{Median}_D & = & 0 & H_0 \\ & \neq & 0 & H_{alt} \\ & & & \\ \alpha & = & 0.05 & \text{2 sided} \end{array}$$

Reference: Binomial [$n = 10$; $\pi(+) = 0.5$]. See also Sign Test Table which I have provided in Chapter on Distribution-free Methods.

Data:

DIFFERENCE	SIGN
Drug-Placebo	
0.9	+
-0.9	-
4.3	+
2.9	+
1.2	+
3.0	+
2.7	+
0.6	+
3.6	+
-0.5	-
$\sum$ 8+, 2-	

Upper-tail: Prob[$\geq 8+ \mid \pi = 0.5$] $= 0.0439 + 0.0098 + 0.0010 = 0.0547$.
2-tails: $P = 0.0547 + 0.0547 = 0.1094$. P > 0.05 (2-sided) ...less powerful than t-test.

In above example on Blood Flow, fact that all 15/15 had delays makes any formal test unnecessary... the "**Intra-Ocular Traumatic Test**" says it all.

[Q: could it be that always raised the left leg, and blood flow is less in left leg? Doubt it but ask the question just to point out that just because we find a numerical difference doesn't necessarily mean that we know what caused the difference

Famous scientist, begins by removing one leg from an insect and, in an accent I cannot reproduce on paper, says "quick march". The insect walks briskly. The scientist removes another leg, and again on being told "quick march" the insect walks along... This continues until the last leg has been removed, and the insect no longer walks. Whereupon the Scientist, again in an accent I cannot convey here, pronounces "There! it goes to prove my theory: when you remove the legs from an insect, it cannot hear you anymore!".

4.4.1 Number of Observations to ensure specified power β if use 1-sample or paired t-test concerning μ_Y or μ_d

Required n for test where $\alpha = 0.005$ (1-sided) or $\alpha = 0.01$ (2-sided)

	β	0.01	0.05	0.10	0.20	0.50
	Power	99%	95%	90%	80%	50%
Δ/σ						
0.2						
0.3					134	78
0.4			115	97	77	45
0.5		100	75	63	51	30
0.6		71	53	45	36	22
0.7		53	40	34	28	17
0.8		41	32	27	22	14
0.9		34	26	22	18	12
1.0		28	22	19	16	10
1.2		21	16	14	12	8
1.4		16	13	12	10	7
1.6		13	11	10	8	6
1.8		12	10	9	8	6
2.0		10	8	8	7	5
2.5		8	7	6	6	
3.0		7	6	6	5	

Notes:

- $\Delta/\sigma = (\mu - \mu_0)/\sigma =$ "Signal" / "Noise"
- Table entries transcribed from Table IV.3 of CRC Tables of Probability and Statistics. Table IV.3 tabulates the n's for the Signal/Noise ratio increments of 0.1, and also includes entries for $\alpha = 0.01$ (1sided) / 0.02 (2-sided). See also Colton, page 142.
- Sample sizes based on t-distribution, and so slightly larger (and more realistic, when n small) than those given by Z-based formula: $n = (Z_\alpha + Z_\beta)^2 \times (\sigma/\Delta)^2$.

Required n for test where $\alpha = 0.025$ (1-sided) or $\alpha = 0.05$ (2-sided)

	β	0.01	0.05	0.10	0.20	0.50
	Power	99%	95%	90%	80%	50%
Δ/σ						
0.2						99
0.3				119	90	45
0.4		117	84	68	51	26
0.5		76	54	44	34	18
0.6		53	38	32	24	13
0.7		40	29	24	19	10
0.8		31	22	19	15	9
0.9		25	19	16	12	7
1.0		21	16	13	10	6
1.2		15	12	10	8	5
1.4		12	9	8	7	
1.6		10	8	7	6	
1.8		8	7	6		
2.0		7	6	5		
2.5		6				
3.0		5				

Required n for test where $\alpha = 0.05$ (1-sided) or $\alpha = 0.1$ (2-sided)

	β	0.01	0.05	0.10	0.20	0.50
	Power	99%	95%	90%	80%	50%
Δ/σ						
0.2						70
0.3			122	97	71	32
0.4		101	70	55	40	19
0.5		65	45	36	27	13
0.6		46	32	26	19	9
0.7		34	24	19	15	8
0.8		27	19	15	12	6
0.9		21	15	13	10	5
1.0		18	13	11	8	5
1.2		13	10	8	6	
1.4		10	8	7	5	
1.6		8	6	6		
1.8		7	6			
2.0		6				
2.5						
3.0						

4.5 “Definitive Negative” Studies: Starch Blockers – their effect on calorie absorbtion from a high-starch meal.

Abstract: It has been known for more than 25 years that certain plant foods, such as kidney beans and wheat, contain a substance that inhibits the activity of salivary and pancreatic amylase. More recently, this antiamylase has been purified and marketed for use in weight control under the generic name “starch blockers.” Although this approach to weight control is highly popular, it has never been shown whether starch-blocker tablets actually reduce the absorption of calories from starch. Using a one-day calorie-balance technique and a high-starch (100 g) meal (spaghetti, tomato sauce, and bread), we measured the excretion of fecal calories after normal subjects had taken either placebo or starch-blocker tablets. If the starch-blocker tablets had prevented the digestion of starch, fecal calorie excretion should have increased by 400 kcal. However, fecal reduce the absorption of calories from starch. Using a one-day calorie-balance technique and a high-starch (100 g) meal (spaghetti, tomato sauce, and bread), we measured the excretion of fecal calories after normal subjects had taken either placebo or starch-blocker tablets. **If the starch-blocker tablets had prevented the digestion of starch, fecal calorie excretion should have increased by 400 kcal. However, fecal calorie excretion was the same on the two test days (mean ± S.E.M., 80 ± 4 as compared with 78 ± 2). We conclude that starch-blocker tablets do not inhibit the digestion and absorption of starch calories in human beings.**
Bo-Linn GW. et al New England J of Medicine. 307(23):1413-6, 1982 Dec 2.

Overview of Methods: The one-day calorie-balance technique begins with a preparatory washout in which the entire gastrointestinal tract is cleansed of all food and fecal material by lavage with a special calorie-free, electrolyte-containing solution. The subject then eats the test meal, which includes $^{51}CrCl_3$ as a non absorbable marker. After 14 hours, the intestine is cleansed again by a final washout. The rectal effluent is combined with any stool (usually none) that has been excreted since the meal was eaten. The energy content of the ingested meal and of the rectal effluent is determined by bomb calorimetry. The completeness of stool collection is evaluated by recovery of the non absorbable marker.]

——-

See Powell-Tuck J. “A defence of the small clinical trial: evaluation of three gastroenterological studies.” Br Med J Clinical Research Ed..292(6520): 599-602, 1986 Mar 1. (under Resources on webpage). for a good paper on ‘negative’ studies,

Table 1: Standard Test Meal

Ingredients		**Dietary constituents***	
Spaghetti (dry weight)**	100 g	Protein	19 g
Tomato sauce	112 g	Fat	14 g
White bread	50 g	Carbohydrate (starch)	108 g (97 g)
Margarine	10 g		
Water	250 g		
$^{51}CrCl_3$	$4\mu Ci$		

*Determined by adding food-table contents of each item.

**Boiled for seven minutes in 1 liter of water.

Table 2. Results in Five Normal Subjects on Days of Placebo and Starch-Blocker Tests.

	Placebo Test Day			Starch-Blocker Test Day		
	Duplicate Test Meal*	Rectal Effluent	Marker Recovery	Duplicate Test Meal*	Rectal Effluent	Marker Recovery
subject	kcal	kcal	%	kcal	kcal	%
1	664	81	97.8	665	76	96.6
2	675	84	95.2	672	84	98.3
3	682	80	97.4	681	73	94.4
4	686	67	95.5	675	75	103.6
5	676	89	96.3	687	83	106.9
Means	677	80	96.4	676	78	100
± S.E.M.	±4	±4	±0.5	±4	±2	±2

Does not include calories contained in three placebo tablets (each tablet, 1.2±0.1 kcal) or in three Carbo-Lite tablets (each tablet, 2.8±0.1 kcal) that were ingested with each test meal.

Is this a Definitive Negative Study?

```
---0-----100-----200-----300----- | <-  Company's Claim: 400 kcal
--***----100-----200-----300----- |

---0-----100-----200-----300-----400-- kcal blocked
  ***  95% CI estimate from study
```