Inference for μ : A&B Ch 7.1 ; Colton Ch 4;

CI(μ): small n: => "Student" 's t distribution

Use when replace σ by s (an estimate of σ) in CI's and tests.

(1) Assume that either
(a) the Y values are either normally distributed or
(b) if not, n is large enough so that the Central Limit Theorem guarantees that the distribution of possible $\bar{y}$'s is well enough approximated by a Gaussian distrn.

(2) Choose the desired degree of confidence [50%, 80%, 90%, 99...] as before.

(3) Proceed as above, except that use t Distribution rather than Z -- find the t value such that xx% of the distribution is between –t and + t. The cutpoints for %-iles of the t distribution vary with the amount of data used to estimate σ.

"Student"'s 't distribution is (conceptual) distribution one gets if...

- take samples (of given size n) from Normal(μ, σ) distribution
- form the quantity $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ from each sample
- compile a histogram of the results

or, in Gossett's own words ...(W.S. Gossett 1908)

"Before I had succeeded in solving my problem analytically, I had endeavoured to do so empirically [i.e. by simulation]. 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 z's 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..."

Student"'s 't distribution (continued)

Distribution (histogram of sampling distribution)

- is symmetric around 0 (just like $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$)
- has a shape like that of the Z distribution, but with SD slightly larger than unity i.e. slightly flatter & more wide-tailed; $Var(t) = \frac{df}{df-2}$
- shape becomes indistinguishable from Z distribution as n -> ∞ (in fact as n goes much beyond 30)
- Instead of $\pm 1.96 \frac{\sigma}{\sqrt{n}}$ to enclose μ with 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
± 1.96	∞	∞

- Test of $\mu = \mu_0$ CI for μ

$$\text{Ratio} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \qquad \bar{x} \pm t\frac{s}{\sqrt{n}}$$

WORKED EXAMPLE : CI and Test of Significance

Response of interest: D: INCREASE (D) IN HOURS OF SLEEP with DRUG

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

Data:

Subject	HOURS of SLEEP† DRUG	PLACEBO	DIFFERENCE Drug - Placebo
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 of 10 differences $SD[d] = 1.77$

Test statistic $= \dfrac{1.78 - [0]}{\frac{1.77}{\sqrt{10}}} = 3.18$ CR:ref$|t_9|$=2.26

Since 3.18 > 2.26, "Reject" H_0

95% CI for μ_D

$= 1.78 \pm t_9 \dfrac{1.77}{\sqrt{10}} = 1.78 \pm 1.26 = 0.5$ to 3.0 hours

NOTE: I deliberately omitted the full data on the drug and placebo conditions: all we need for the analysis are the 10 differences.

What if not sure d's come from a Gaussian Distribution?

[for t: Gaussian data or (via CLT) Gaussian statistic ($\bar{d}$)

WORKED EXAMPLE *C P G Barker* The Lancet Vol 345 . April 22, 1995, p 1047.

Posture, blood flow, and prophylaxis of venous thromboembolism

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 plantegrade 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 overleaf]. 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.

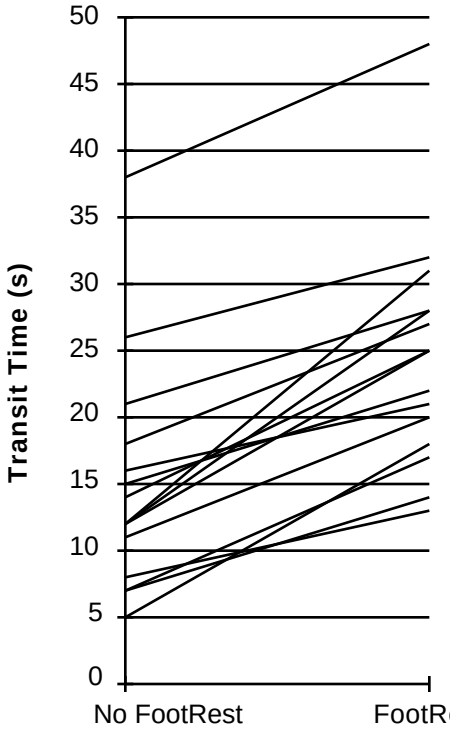
JH's Analysis of raw data [data abstracted by eye, so my calculations won't match exactly with those in text]

$\bar{d}$(SD) = 9.8(4.1); $t = \dfrac{9.8 - [0]}{4.1/\sqrt{15}} = \dfrac{9.8}{1.0} = 9.8 > t_{14,0.05}$ of 2.145

difference is 'off the t-scale'

95% CI for μ_D: $9.8 \pm 2.145[1.0] = 7.7$ to 11.9 s

WORKED EXAMPLE: Leg Elevation (continued)



	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

Remarks:
Whereas mean of 15 differences between 2 conditions 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**...

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

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

Calculating $\text{Var}(\bar{y}_1 - \bar{y}_2) = \text{Var}(\bar{y}_1) + \text{Var}(\bar{y}_2)$
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. As it is, even this inefficient analysis would have sufficed here because the 'signal' was so much greater than the 'noise'.

See article On Reserve on display of data from pairs.

Sample Size for CI's and test involving μ

n to yield (2-sided) CI with margin of error m at confidence level 1- α (see M&M p 447)

```
                          |<--   margin       |
                          |    of error   -->|
                          |                  |
   (----------------------•------------------)
```

- **large-sample CI:** $\bar{x} \pm Z_{\alpha/2}\,SE(\bar{x}) = \bar{x} \pm m$
- **$SE(\bar{x}) = \sigma / \sqrt{n}$, so...**

$$n = \frac{\sigma^2 \cdot Z_{\alpha/2}^2}{m^2}$$

If n small, replace $Z_{\alpha/2}$ by $t_{\alpha/2}$

Typically, won't know σ so use guesstimate;

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.

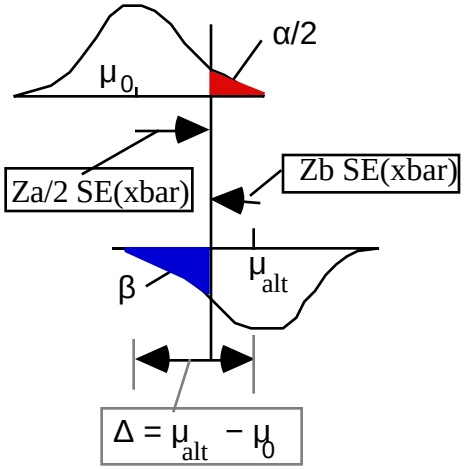
Sample Size for CI's and test involving μ .. cont'd

n for power 1-β if mean μ is Δ units from μ_0 (test value) ; type I error = α (cf Colton p142 or CRC table next)

Need $Z_{\alpha/2}\, SE(\bar{x}) + Z_{\beta} SE(\bar{x}) > \Delta$.

Substitute $(\bar{x}) = \sigma/\sqrt{n}$ and solve for n:

$$\text{so need } n = \frac{\{ Z_{\alpha/2} - Z_{\beta} \}^2 \sigma^2}{\Delta^2}$$



If power is > 0.5, then β < 0.5, and $Z_{\beta} < 0$.

eg. α=0.05 , β=0.2 => $Z_{\alpha/2} = 1.96$ $Z_{\beta} = -0.84$

Technically, if n small, use t-test... see table next page

The question of what Δ to use is not a matter of statistics or samples, or what the last guy found in a study, but rather the difference that makes a difference" i.e it is a clinical judgement, and includes the impact, cost, alternatives, etc...
It is the Δ that IF TRUE would lead to a difference in management or a substantial risk, or whatever...

Sign Test for median

Test:

H_0: $Median_D = 0$

vs

H_{alt}: $Median_D \neq 0$; α =0.05 (2-sided);

DIFFERENCE Drug – Placebo	SIGN of d
0.9	+
-0.9	–
4.3	+
2.9	+
1.2	+
3.0	+
2.7	+
0.6	+
3.6	+
-0.5	–
∑	8+, 2–

Reference: Binomial [n=10; π(+) = 0.5] See Table C (last column of p T9) or Sign Test Table which I have provided in Chapter on Distribution-free Methods.

Upper Tail: Prob(≥ 8+ | π = 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, the fact that all 15/15 had delays makes any formal test unnecesary... 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!".

Number of Observations to Ensure Specified Power (1-β) if use 1-sample or paired t-test of Mean

	α = 0.005(1-sided) α = **0.01 (2-sided)**					α = 0.025(1-sided) α = **0.05 (2-sided)**					α = 0.05(1-sided) α = **0.1 (2-sided)**				
β =	**0.01**	**0.05**	**0.1**	**0.2**	**0.5**	**0.01**	**0.05**	**0.1**	**0.2**	**0.5**	**0.01**	**0.05**	**0.1**	**0.2**	**0.5**
Δ/σ															
0.2										99					70
0.3				134	78			119	90	45		122	97	71	32
0.4		115	97	77	45	117	84	68	51	26	101	70	55	40	19
0.5	100	75	63	51	30	76	54	44	34	18	65	45	36	27	13
0.6	71	53	45	36	22	53	38	32	24	13	46	32	26	19	9
0.7	53	40	34	28	17	40	29	24	19	10	34	24	19	15	8
0.8	41	32	27	22	14	31	22	19	15	9	27	19	15	12	6
0.9	34	26	22	18	12	25	19	16	12	7	21	15	13	10	5
1.0	28	22	19	16	10	21	16	13	10	6	18	13	11	8	5
1.2	21	16	14	12	8	15	12	10	8	5	13	10	8	6	
1.4	16	13	12	10	7	12	9	8	7		10	8	7	5	
1.6	13	11	10	8	6	10	8	7	6		8	6	6		
1.8	12	10	9	8	6	8	7	6			7	6			
2.0	10	8	8	7	5	7	6	5			6				
2.5	8	7	6	6		6									
3.0	7	6	6	5		5									

$$\frac{\Delta}{\sigma} = \frac{\mu - \mu_0}{\sigma} = \frac{\text{"Signal"}}{\text{"Noise"}}$$

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 ratios 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-tables, and so slightly larger (and more realistic, when n small) than those given by z-based formula: $n = (z_\alpha + z_\beta)^2 \left(\frac{\sigma}{\Delta}\right)^2$

"Definitive Negative" Studies? Starch blockers--their effect on calorie absorption 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 Journal 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.]

For an **good paper on topic of 'negative' studies**, see Powell-Tuck J "A defence of the small clinical trial: evaluation of three gastroenterological studies." British Medical Journal Clinical Research Ed..292(6520):599-602, 1986 Mar 1. (Resources for Ch 7)

Table 1. Standard Test Meal.

Ingredients

Spaghetti (dry weight)*	100 g
Tomato sauce	112 g
White bread	50 g
Margarine	10 g
Water	250 g
$^{51}CrCl_3$	4 μCi

Dietary constituents†

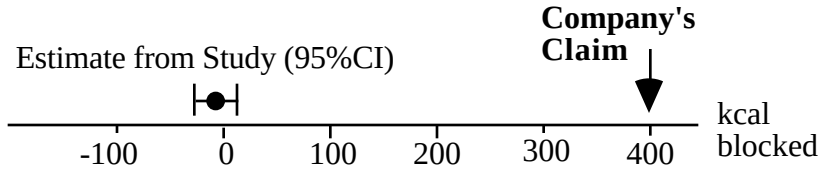
Protein	19 g
Fat	14 g
Carbohydrate (starch)	108 g (97 g)

•Boiled for seven minutes in 1 liter of water.
† Determined by adding food-table contents of each item

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
	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.



Inference for $\mu_1 - \mu_2$: A&B Ch 7.2 ; ColtonCh 4

Somewhat more complex than simply replacing σ_1 and σ_2 by s_1 and s_2 as estimates of σ's in CI's and tests.

Need to distinguish **two theoretical situations** (unfortunately seldom clearly distinguishable in practice) where:

$\sigma_1 = \sigma_2 = \sigma$

use a "pooled" estimate s of σ [see M&M page 550]
[think of s^2_{pooled} as a weighted average of s_1^2 and s_1^2]

t-table is accurate (if Gaussian data)

$\sigma_1 \neq \sigma_2$

use separate estimates of σ_1 and σ_2

adjust d.f. downwards from (n_1+n_2-2) to compensate for inaccuracy

Option 1 (p 549) "software approximation"*

Option 2 (p541) for hand use: df = Minimum[n_1-1, n_2-1]

[M&M are the only ones I know who suggest this option 2; I think they do so because the undergraduates they teach may not be motivated enough to use equation 7.4 page 549 to calculate the reduced degrees of freedom... I agree that the only time one should use option 2 is the 1st time when learning about the t-test]

* The SAS manual says that in its TTEST procedure it uses Satterthwaite's approximation [p. 549 of M&M] for the reduced degrees of freedom unless the user specifies otherwise.

Adjustments are not a big issue if sample sizes are large or variances similar.

WORKED EXAMPLE : CI and Test of Significance

Y= Fall in BP over 8 weeks [mm] with Modest Reduction in Dietary Salt Intake in Mild Hypertension (Lancet 25 Feb, 1989)

Test: H_0: $\mu_{Y\ (Normal\ Sodium\ Diet)} = \mu_{Y\ (Low\ Sodium\ Diet)}$
H_{alt}: $\mu_{Y\ (Normal\ Sodium\ Diet)} \neq \mu_{Y\ (Low\ Sodium\ Diet)}$

α =0.05 (2-sided); β =0.20;
==> Power = 80% if $|\mu_{Y(Nl)} - \mu_{Y(Low)}| \geq$ 2mm DBP

Data given: Mean(SEM) Fall in BP

	"Normal" Group (n=53)	"Low" Group (n=50)
SBP	-0.6(1.0)	-6.1(1.1)

Reconstruct s^2's via relation: $s^2 = n\ SEM^2$

Mean(s^2) Fall in BP

	"Normal" (n=53)	"Low" (n=50)
SBP	-0.6(53)	-6.1(60.5)

$$s^2 = \frac{[52]53+[49]60.5}{[52]+[49]} = 56.63;\ s = \sqrt{56.63} = 7.52$$

$$t = \frac{-6.1-\ [-0.6]}{7.52\sqrt{1/53\ +\ 1/50}} = -3.71 \text{ vs } t_{101,05} = 1.98$$

Calculation of t-test using separate variances

Had we used the separate s^2's in each sample we would have calculated

$$t = \frac{-6.1 - [-0.6]}{\sqrt{\frac{53}{53} + \frac{60.5}{50}}} = -3.70$$

This is equivalent to calculating:

$$t = \frac{-6.1 - [-0.6]}{\sqrt{SE_1^2 + SE_2^2}} = \frac{-6.1 - [-0.6]}{\sqrt{1.1^2 + 1.0^2}} = -3.70$$

M&M suggest that the appropriate df for t are

Option 1 (via their eqn. 7.4):	99.5
Option 2 (smaller df):	49

Either way, the t ratio is far beyond the α=0.05 point of the null distribution. Notice that the reduction in df is minimal here because the two sample variances are quite close.

Incidentally, as per their power calculations, the primary response variable was DBP

Mean(SEM) Fall in DBP in the 2 samples:

"Normal" Group (n=53)	"Low" Group (n=50)
-0.9(0.6)	-3.7(0.6)

$$t = \frac{-3.7 - [-0.9]}{\sqrt{0.6^2 + 0.6^2}} = 3.3$$

"Eye test": Judging overlap of two independent CI's

```
|----------x----------|  Overlapping CI's
              |----------x----------|
```

How far apart do two independent $\bar{x}$'s, say $\bar{x}_1$ and $\bar{x}_2$, have to be for a formal statistical test, using an alpha of 0.05 two sided, to be statistically significant?

Need...

$$| \bar{x}_1 - \bar{x}_2 | \geq 1.96 \sqrt{[\{SE[\bar{x}_1]\}^2 + [\{SE[\bar{x}_2]\}^2} \quad \text{if using z-test*}$$

If the 2 SEM's are about the same size (as they would be if the 2 n's, and the per-unit variability, were about the same), then ... [as in exercise X in Chapter 5]

Need... $| \bar{x}_1 - \bar{x}_2 | \geq 1.96 \sqrt{2\ \{SE[\text{each } \bar{x}]\}^2}$

i.e. $| \bar{x}_1 - \bar{x}_2 | \geq 1.96 \sqrt{2}$ SE[each $\bar{x}$] , or... $| \bar{x}_1 - \bar{x}_2 | \geq 2.77$ SE[each $\bar{x}$]

*If using t rather than z, multiple would be somewhat higher than 1.96, so that when multiplied by $\sqrt{2}$ it might be higher than 2.77, closer to 3. Thus **a rough answer to the question could b**e

$$| \bar{x}_1 - \bar{x}_2 | \geq 3 \text{ SE[each } \bar{x}]$$

This means that even when two 100(1-α)% CI's overlap slightly, as above, the difference between the two means could be statistically significant at the α level. This is why Moses, in his article on graphical displays (see reserve material) advocates plotting the 2 CI's formed by

$$\bar{x}_1 \pm 1.5 \text{ SE}[\bar{x}_1] \text{ and } \bar{x}_2 \pm 1.5 \text{ SE}[\bar{x}_2]$$

Thus, we can be reasonably sure that if the CI's do not overlap (i.e. if $\bar{x}_1$ and $\bar{x}_2$ are more than 3 SE[each $\bar{x}$] apart) the difference between them is statistically significant at the alpha=0.05 level.

[estimate ±1.5 SE(estimate) corresponds to an 86% CI if using Z distribution].

Note: above logic applies for other symmetric CI's too.

Inferences regarding means --- Summary

Situation	Object	σ **known** (or large n's)	σ **unknown**
1 Popln. μ , σ_x	**CI** for μ	$\bar{x} \pm z \frac{\sigma_x}{\sqrt{n}}$	$\bar{x} \pm t_{n-1} \frac{s_x}{\sqrt{n}}$
(sample of n)	**Test** μ_0	$z = \frac{\bar{x} - \mu_0}{\frac{\sigma_x}{\sqrt{n}}}$	$t_{n-1} = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}}$
1 Popln. under 2 condns.	**CI** for $\Delta = \mu_{(d)}$	$\bar{d} \pm z \frac{\sigma_d}{\sqrt{n}}$	$\bar{d} \pm t_{n-1} \frac{s_d}{\sqrt{n}}$
(sample of n within-pair differences {d=x_1-x_2})	**Test** Δ_0	$z = \frac{\bar{d} - \Delta_0}{\frac{\sigma_d}{\sqrt{n}}}$	$t_{n-1} = \frac{\bar{d} - \Delta_0}{\frac{s_d}{\sqrt{n}}}$
2 Poplns.	**CI** for $\Delta = \mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2 \pm z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 \pm t_{df}\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}$
(independent samples of n_1 and n_2)	**Test** Δ_0	$z = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$t_{df} = \frac{\bar{x}_1 - \bar{x}_2 - \Delta_0}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}}$

Notes:

•Pooled $s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$ (weighted average of the two s^2 's) •df = $(n_1-1) + (n_2-1) = n_1 + n_2 - 2$

•If it appears that σ_1^2 is very different from σ_2^2, then a "separate variances" t-test is used with df reduced to account for the differing σ^2 's

Sample Size for CI for $\mu_1 - \mu_2$

CI($\mu_1 - \mu_2$)

n's to produce CI for $\mu_1 - \mu_2$ with prespecified margin of error

• large-sample CI:

$$\bar{x}_1 - \bar{x}_2 \pm Z\, SE(\bar{x}_1 - \bar{x}_2) = \bar{x}_1 - \bar{x}_2 \pm \text{margin of error}$$

• $SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}$

• if use equal n's, then ...

$$\text{n per group} = \frac{2\sigma^2 Z^2}{[\text{margin of error}]^2}$$

example:

* 95% CI for difference in mean Length of Stay (LOS);
* desired Margin of Error for difference: 0.5 days,
* anticipate SD of individual LOS's, in each situation, of 5 days.

95% -> α=0.05 -> $Z_{\alpha/2} = 1.96$

$$\text{n per group} = \frac{2 \cdot 5^2 \cdot 1.96^2}{[0.5]^2} \quad 800$$

Contrast formula for test and formula for CI:
CI: no null and al. values for comparative parameter; notice also absence of beta.

See reference to Greenland [bottom of next column].

Sample Size for test of μ_1 versus μ_2

Test H_0: $\mu_1 = \mu_2$ vs H_a: $\mu_1 \neq \mu_2$

n's for power of 100(1 – β)% if $\mu_1 - \mu_2 = \Delta$; Prob(type I error) = α

(cf. Colton p 145 or CRC tables)

$$\text{n per group} = 2\,\frac{\{Z_{\alpha/2} - Z_\beta\}^2 \sigma^2}{\Delta^2} = 2(Z_{\alpha/2} - Z_\beta)^2 \left\{\frac{\sigma}{\Delta}\right\}^2$$

Note that if $\beta < 0.5$, $Z_\beta < 0$ (also, Z_β always 1-sided).

example α=0.05 (2-sided) and β=0.2 ...

$Z_{\alpha/2} = 1.96$; $Z_\beta = -0.84$,

$2(Z_{\alpha/2} - Z_\beta)^2 = 2\{1.96 - (-0.84)\}^2 \approx 16$, i.e.

n per group $\approx 16 \cdot \{\text{noise/signal ratio}\}^2$

These formulae are easily programmed in a spreadsheet. There are also specialized software packages for sample size and statistical power See web page under Resources for Chapter 7.

Greenland S. "On sample-size and power calculations for studies using confidence intervals". American Journal of Epidemiology. 128(1):231-7, 1988 Jul. Abstract: A recent trend in epidemiologic analysis has been away from significance tests and toward confidence intervals. In accord with this trend, several authors have proposed the use of expected confidence intervals in the design of epidemiologic studies. This paper discusses how expected confidence intervals, if not properly centered, can be misleading indicators of the discriminatory power of a study. To rectify such problems, the study must be designed so that the confidence interval has a high probability of not containing at least one plausible but incorrect parameter value. To achieve this end, conventional formulas for power and sample size may be used. Expected intervals, if properly centered, can be used to design uniformly powerful studies but will yield sample-size requirements far in excess of previously proposed methods.

Number of Observations PER GROUP to Ensure Specified Power (1 - β) if use 2-sample t-test of 2 Means

	α = 0.005(1-sided) α = **0.01 (2-sided)**					α = 0.025(1-sided) α = **0.05 (2-sided)**					α = 0.05(1-sided) α = **0.1 (2-sided)**				
β =	**0.01**	**0.05**	**0.1**	**0.2**	**0.5**	**0.01**	**0.05**	**0.1**	**0.2**	**0.5**	**0.01**	**0.05**	**0.1**	**0.2**	**0.5**
$\frac{\Delta}{\sigma}$															
0.2															137
0.3										87					61
0.4					85				100	50			108	78	35
0.5				96	55		106	86	64	32		88	70	51	23
0.6		101	85	67	39	104	74	60	45	23	89	61	49	36	16
0.7	100	75	63	50	29	76	55	44	34	17	66	45	36	26	12
0.8	77	56	49	39	23	57	42	34	26	14	50	35	28	21	10
0.9	62	46	39	31	19	47	34	27	21	11	40	28	22	16	8
1.0	50	38	32	26	15	38	27	23	17	9	33	23	18	14	7
1.2	36	27	23	18	11	27	20	16	12	7	23	16	13	10	5
1.4	27	20	17	14	9	20	15	12	10	6	17	12	10	8	4
1.6	21	16	14	11	7	16	12	10	8	5	14	10	8	6	4
1.8	17	13	11	10	6	13	10	8	6	4	11	8	7	5	
2.0	14	11	10	8	6	11	8	7	6	4	9	7	6	4	
2.5	10	8	7	6	4	8	6	5	4		6	5	4	3	
3.0	8	6	6	5	4	6	5	4	4		5	4	3		

$$\frac{\Delta}{\sigma} = \frac{\mu_1 - \mu_2}{\sigma} = \frac{\text{"Signal"}}{\text{"Noise"}}$$

Table entries transcribed from Table IV.4 of CRC Tables of Probability and Statistics. Table IV.4 tabulates the n's for the Signal/Noise ratios increments of 0.1, and also includes entries for alpha=0.01(1-sided)/0.02(2-sided).

See also Colton, page 145

Sample sizes based on t-tables, and so slightly larger (and more realistic) than those given by z-based formula: n/group = $2(z_{\alpha/2} + z_{\beta})^2 \left(\frac{\sigma}{\Delta}\right)^2$

See later (in Chapter 8) for unequal sample sizes i.e. $n_1 \neq n_2$