Inference for the difference of two population means (σ unknown)

PSLS chapter 18

Two samples situations

When comparing 2 treatments or conditions, ask whether the data sets are **independent samples**: that is, individuals in both samples are selected separately ("independently")

independent random samples

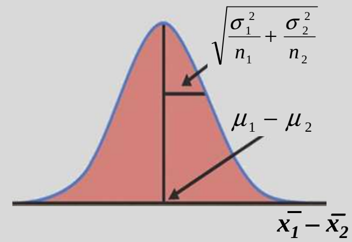
Individual ID	Condition tested	Value recorded
01	A	
02	A	
03	B	
04	A	
05	A	
06	B	
07	B	
08	A	
09	B	
10	B	
…		

matched pairs / repeated measures

Individual or pair ID	Value under condition A	Value under condition B
01		
02		
03		
04		
05		
06		
07		
08		
09		
10		
…		

Sampling distribution of $(\bar{x}_1 - \bar{x}_2)$:

We have **2 independent random samples** of sizes n_1 and n_2 coming from 2 populations. When the both populations are normally distributed, the sampling distribution of $(\bar{x}_1 - \bar{x}_2)$ is also normal.



$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

follows the standard normal distribution N(0,1)

FYI

FYI

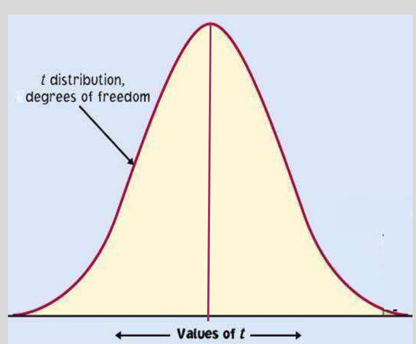
The two-sample *t* statistic follows approximately a *t* distribution with standard error

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The degrees of freedom of the *t* distribution are computed as follows:

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



Two-sample t test

Using 2 independent random samples for a quantitative variable, test

$$H_0\text{: } \mu_1 = \mu_2 \iff \mu_1 - \mu_2 = 0$$

with either a one-sided or a two-sided alternative hypothesis (H_a).

The ***t* statistic**, relevant df, and H_a direction are part of the P-value computation.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \Rightarrow \boxed{t = \frac{\bar{x}_1 - \bar{x}_2}{SE}}$$

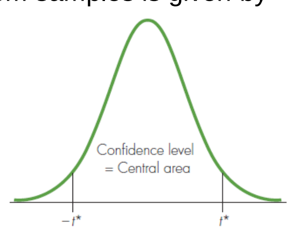
Obtain the test *P*-value (the probability, if H_0 was true, of obtaining a test statistic at least as extreme as that computed from the data). **Conclude in context.**

Two sample t confidence interval

A **level C confidence interval for ($\mu_1 - \mu_2$)** for a quantitative variable in 2 populations based on 2 independent random samples is given by

$$\boxed{(\bar{x}_1 - \bar{x}_2) \pm m}$$

$$m = t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = t^* SE$$



There is area C between -t* and t* under the t distribution with the appropriate degrees of freedom

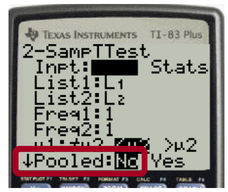
The **margin of error *m*** represents the maximum distance between the estimate and the parameter in C% of all sets of random samples from these populations.

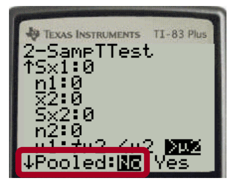
TI calculator : hypothesis test for μ_1- μ_2 (σ_1 and σ_2 unknown)

STAT

2-SampT-Test

From summary statistics
Or from raw data in 2 lists



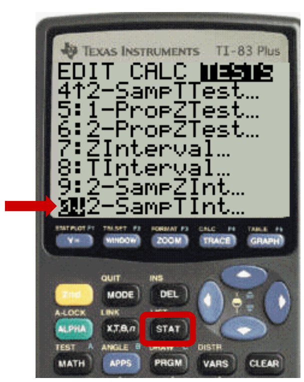


Choose **"Pooled: No"**

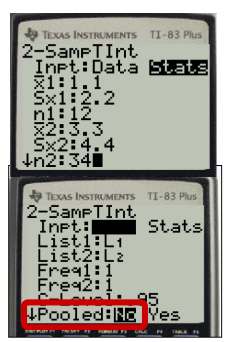
See videos for TI-83 and CrunchIt! on Sapling

TI calculator : C.I. for μ_1- μ_2 (σ_1 and σ_2 unknown)

STAT then **TESTS / 2-SampTInt**



From summary statistics
Or from raw data in 2 lists

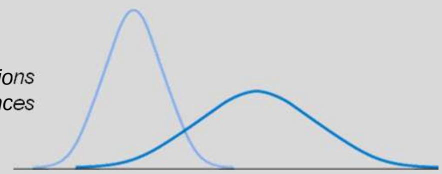


Choose **"Pooled: No"**

See videos for TI-83 and CrunchIt! on Sapling

Two-sample inference: assume equal variance?

Two Normally distributed populations
with unequal variances



There are two versions of the two-sample *t*-test: one **assuming equal population variances ("pooled two-sample test")** and one **not assuming equal population variances**. They have slightly different formulas and df.

In two-sample inference for means, always choose the unpooled ("Pooled:No") *t* procedure.

FYI – NOT required material

Computations for two-sample *t* procedures with pooled variance

When both populations have the *same* variance σ^2, the **pooled estimator of σ^2** is:

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

The pooled sampling distribution for $(\bar{x}_1 - \bar{x}_2)$ has exactly the *t* distribution with ***$(n_1 + n_2 - 2)$* degrees of freedom**.

Confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x}_1 - \bar{x}_2) \pm t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

H_0: $\mu_1 = \mu_2$ against a 1-sided or a 2-sided alternative:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

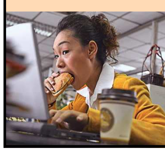
A study assessed the effects of playing a computer game (solitaire) during lunch on behavioral and physiological variables. A sample of 44 healthy adults was randomly assigned to eat lunch either with the computer game distraction or without distraction. All participants ate the same breakfast and lunch.

Thirty minutes after lunch, participants were offered cookies ("biscuits") as a snack and were instructed to eat as many or as few cookies as they liked. The published findings were shown as mean ± standard deviation. Do they provide evidence that, on average, distraction during meals leads to greater snack intake later?

Measure	Distraction (n = 22)	No distraction (n = 22)
Biscuit intake (g)	52.1 ± 45.1*	27.1 ± 26.4

H_0: $\mu_{\text{distraction}} = \mu_{\text{no}}$ ⇔ $(\mu_{\text{distraction}} - \mu_{\text{no}}) = 0$

H_a: $\mu_{\text{distraction}} > \mu_{\text{no}}$ ⇔ $(\mu_{\text{distraction}} - \mu_{\text{no}}) > 0$ (one-sided)



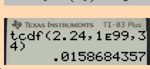
Measure	Distraction (n = 22)	No distraction (n = 22)
Biscuit intake (g)	52.1 ± 45.1	27.1 ± 26.4

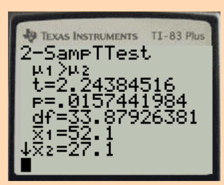
H_0: $\mu_{\text{distraction}} = \mu_{\text{no}}$

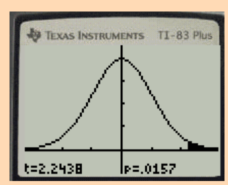
H_a: $\mu_{\text{distraction}} > \mu_{\text{no}}$

$$t = \frac{\bar{x}_{distr} - \bar{x}_{no}}{\sqrt{\frac{s^2_{distr}}{n_{distr}} + \frac{s^2_{no}}{n_{no}}}} = \frac{52.1 - 27.1}{\sqrt{\frac{45.1^2}{22} + \frac{26.4^2}{22}}} = \frac{25}{11.14} \approx 2.24$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} = \frac{\left(\frac{45.1^2}{22} + \frac{26.4^2}{22}\right)^2}{\frac{1}{21}\left(\frac{45.1^2}{22}\right)^2 + \frac{1}{21}\left(\frac{26.4^2}{22}\right)^2} \approx 34$$

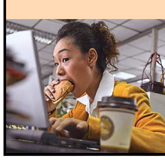






P = 0.016, statistically significant

➔ We reject H_0

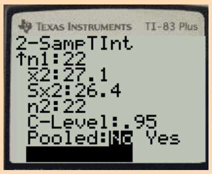


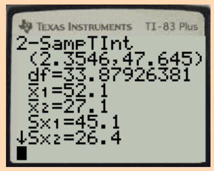
-The study provides good evidence (P = 0.016) that, on average, a computer distraction during meals leads to greater snack/biscuit intake later, compared with eating without distraction.

-Snack intake is statistically significantly greater, on average, after eating lunch with a computer distraction than without (P = 0.016).

With 95% confidence, how much more do people snack, on average, when distracted during meals?

Measure	Distraction (n = 22)	No distraction (n = 22)
Biscuit intake (g)	52.1 ± 45.1	27.1 ± 26.4

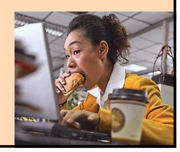




$$SE = \sqrt{\frac{s_{distr}^2}{n_{distr}} + \frac{s_{no}^2}{n_{no}}} = \sqrt{\frac{45.1^2}{22} + \frac{26.4^2}{22}} \approx 11.14$$

$$(\bar{x}_1 - \bar{x}_2) \pm m$$
$$(52.1 - 27.1) \pm (2.032 * 11.14)$$
$$25 \pm 22.6$$
$$(2.4, 47.6)$$

We are 95% confident that, under conditions similar to those of this experiment, healthy adults eat somewhere between 2.4 g and 47.6 g more of snack, on average, after eating lunch while distracted compared with no distraction.



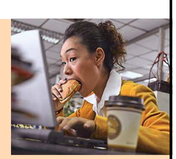
Conditions for inference on means

1. The data are **2 independent random samples** representing 2 much larger populations.

2. The data are Normally distributed OR the combined sample size ($n_1 + n_2$) is large enough to overcome lack of Normality.

   Two-sample t procedures are especially **robust** when both sample sizes are equal and both sample distributions are similar.

Measure	Distraction (n = 22)	No distraction (n = 22)
Biscuit intake (g)	52.1 ± 45.1*	27.1 ± 26.4



- Randomized experiment with independent random samples
- Skewed data but large enough total sample size in 2 equal samples

➔ The t procedures are appropriate here.

In 2015 the internet was hotly debating whether the dress in a photo was white and gold (WG) or black and blue (BB). Researchers recruited a random sample of adults, some who perceived the dress WG and some who perceived it BB.

Here are their relative brain activity in the middle frontal gyrus (MFG) using fMRI brain imaging.

group	N	Mean	StDev	SE Mean
BB	14	-0.275	0.249	0.066
WG	14	0.352	0.296	0.079

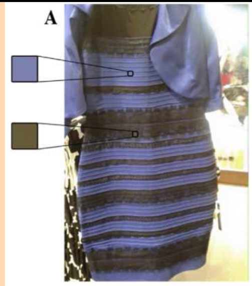
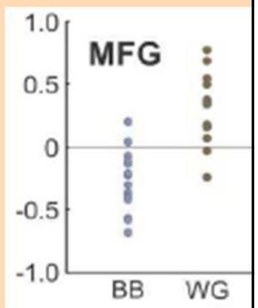


Fig. 1 – The Dress. a) The original image

doi:10.1016/j.cortex.2015.08.017

Is there evidence that, on average, brain activity in the MFG differs based on how the dress is perceived? The *P*-value is

A) not appropriate here
B) greater than 5%
C) between 1% and 5%
D) between 0.1% and 1%
E) less than 0.1%



Conclude in context.

The picture of the dress was posted on Tumblr in 2015 by Caitlin McNeill (Scotland) after noticing her friends saw different colors in the photograph.

The hashtag #TheDress started trending worldwide on Twitter as the debate when global.

Patients with persistent shoulder pain are often told to try shoulder surgery, but is it an effective treatment? A 2017 study randomly assigned patients with persistent shoulder pain to either shoulder surgery (1) or sham surgery (2: incision only). Shoulder condition was assessed with the Oxford Shoulder Score (ranging from 0 for worst to 48 for best) a year later.

```
Two-Sample T-Test and CI

Sample   N   Mean  StDev  SE Mean
1       72  39.40   9.30      1.1
2       65  38.10   9.60      1.2

Difference = μ (1) - μ (2)
Estimate for difference:  1.30
95% CI for difference:  (-1.90, 4.50)
```

What can we conclude, on average, about shoulder condition a year after treatment in the population of patients with persistent shoulder pain?

A. Surgery results in significantly better condition than a placebo.
B. Surgery provides somewhat better condition than a placebo.
C. The study is inconclusive on the relative effectiveness of surgery versus placebo.