Comparison of two treatments
(two-sample t-test)

Assumptions

Treatment X → Popln. is normal with mean μ_X and variance σ_X^2

Treatment Y → Popln. is normal with mean μ_Y and variance σ_Y^2.

All parameters are unknown.

Though σ_X^2 and σ_Y^2 are unknown, it is assumed that $\sigma_X^2 = \sigma_Y^2$.

↓

Assumption of homogeneity of variances.

Goal: To make inference about the mean difference of the two treatments, i.e., to construct CI for $\mu_X - \mu_Y$

or

to test $H_0: \mu_X = \mu_Y$ vs $H_1: \mu_X \neq \mu_Y$

$H_1: \mu_X > \mu_Y$

$H_1: \mu_X < \mu_Y$.

Let $X_1, \ldots, X_m$ be a sample from Treatment X,
$Y_1, \ldots, Y_n$ be a random sample from Treat. Y.

Then, we know the following facts:

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{m}\right),$$

$$\bar{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right)$$

and so

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}\right);$$

$$= \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)$$

under homogeneity.

$$\frac{(m-1)S_X^2}{\sigma_X^2} \sim \chi^2_{m-1},$$

$$\frac{(n-1)S_Y^2}{\sigma_Y^2} \sim \chi^2_{n-1}$$

and so, under **homogeneity of variances**,

$$\frac{1}{\sigma^2}\left\{(m-1)S_X^2 + (n-1)S_Y^2\right\} \sim \chi^2_{m+n-2}.$$

Now, the pooled sample variance is

$$S_p^2 = \frac{(m-1)S_x^2 + (n-1)S_y^2}{m+n-2}.$$

So, we have

$$\frac{(m+n-2)S_p^2}{\sigma^2} \sim \chi^2_{m+n-2}.$$

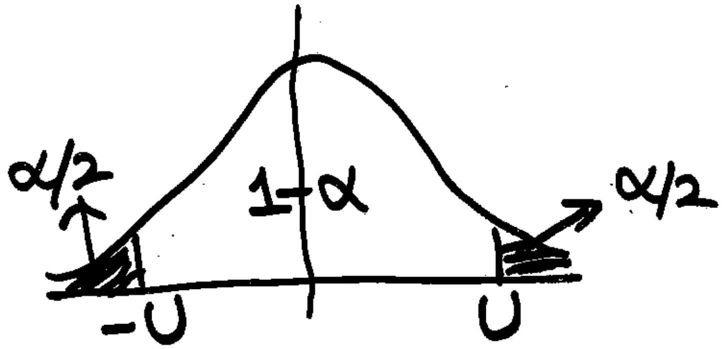
From Student's t-formula, we then have

$$\frac{SNV}{\sqrt{\chi^2_{df}/df}} = \frac{\dfrac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{\sqrt{\sigma^2\left(\frac{1}{m}+\frac{1}{n}\right)}}}{\sqrt{\dfrac{(m+n-2)S_p^2}{\sigma^2}\Big/(m+n-2)}}$$

$$= \frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{S_p\sqrt{\frac{1}{m}+\frac{1}{n}}}$$

has the Student's t-distn. with $m+n-2$ degrees of freedom.

CI for $\mu_x - \mu_y$



The $100(1-\alpha)\%$ CI for $\mu_x - \mu_y$ is

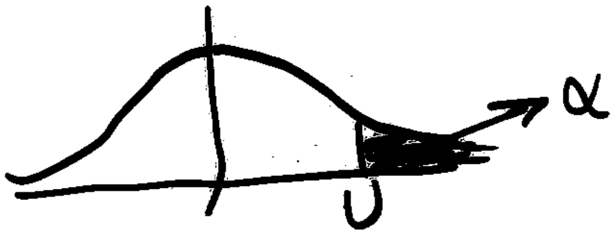
$$\left(\bar{X}-\bar{Y} - U\, S_p\sqrt{\frac{1}{m}+\frac{1}{n}},\ \bar{X}-\bar{Y} + U\, S_p\sqrt{\frac{1}{m}+\frac{1}{n}}\right)$$

Test for $H_0: \mu_x = \mu_y$ vs $H_1: \mu_x > \mu_y$

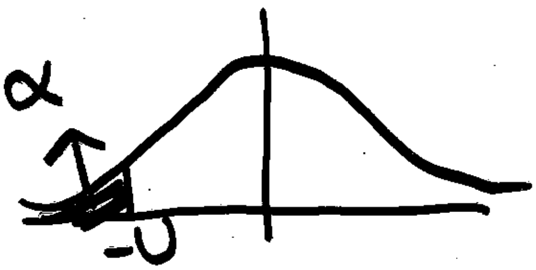
Under $H_0: \mu_x = \mu_y$, $\mu_x - \mu_y = 0$ so that

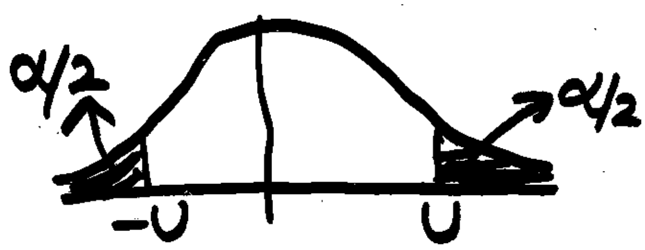
$$\frac{\bar{X}-\bar{Y}}{S_p\sqrt{\frac{1}{m}+\frac{1}{n}}}$$ will have t-distn. with $m+n-2$ df.

CR



At α level of significance, since the critical Region is on the right, find U from t-table. If the calculated test statistic $> U$, reject H_0; otherwise, accept H_0.





Example: Eight 35-39 yeal-old nonpregnant, premenopausal OC users had mean systolic blood pressure of 132.86 mm Hg and a sample standard deviation of 15.34 mm Hg.

Similarly, a sample of twenty one 35-39 year-old nonpregnant, premenopausal non-OC users had mean systolic blood pressure of 127.44 and a standard deviation of 18.23.

$$m = 8,\ \bar{X} = 132.86,\ S_x = 15.34,$$
$$n = 21,\ \bar{Y} = 127.44,\ S_y = 18.23$$
$$S_p^2 = \frac{7(15.34)^2 + 20(18.23)^2}{27} = 307.18$$

So, if we wish to test

$$H_0: \mu_x = \mu_y \quad \text{vs} \quad H_1: \mu_x \neq \mu_y,$$

we find

$$t = \frac{132.86 - 127.44}{\sqrt{307.18\left(\frac{1}{8} + \frac{1}{21}\right)}} = 0.74.$$

At 5% level of significance, CR is $(-\infty, -2.052) \cup (2.052, \infty)$. So, we accept H_0.

95% CI for $\mu_x - \mu_y$ is found to be $(-9.52, 20.36)$. Since 0 is in the interval, we conclude there is no difference in the means.

Example for one-way ANOVA

In an experiment to determine the effect of nutrition on the attention spans of elementary school students, 15 were randomly assigned to three meal plans:

No breakfast, Light and Heavy.

Their attention spans (in mins.) were recorded during a morning reading period as follows:

No	8, 7, 9, 13, 10	47	9.4
Light	14, 16, 12, 17, 11	70	14.0
Heavy	10, 12, 16, 15, 12	65	13.0
	x_{ij}	$\sum_{j=1}^{n_i} x_{ij}$	$\bar{x}_i$

k = No. of treatments $= 3$

$n_1 = n_2 = n_3 = 5$,

$n = n_1 + n_2 + n_3 = 15$,

$\bar{x}_1 = 9.4$, $\bar{x}_2 = 14.0$, $\bar{x}_3 = 13.0$,

$$\bar{x} = \frac{47+70+65}{15} = 12.1333$$

$$\text{Total Sum of Squares} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij}^2 - n\bar{X}^2$$

$$= (8^2 + 7^2 + \ldots + 15^2 + 12^2) - 15(12.1333)^2$$

$$= 2338 - 2208.2667 = 129.7333;$$

Sum of Squares due to Treatments

$$= SST = \sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2$$

$$= 5\{(9.4 - 12.1333)^2 + (14 - 12.1333)^2 + (13 - 12.1333)^2\}$$

$$= 5\{7.4709 + 3.4846 + 0.7512\} = 58.5335;$$

Sum of Squares due to Error

$$= SSE = TSS - SST$$

$$= 129.7333 - 58.5335 = 71.1998$$

ANOVA table

Source	SS	DF	MS	F
Treatment	58.5335	2	29.2668	4.9326
Error	71.1998	12	5.9333	
Total	129.7333	14	*	

At 5% level of significance, $F_{2,12}(0.05) = 3.89$.
Since $4.9326 > 3.89$, we decide to Reject H_0.

<u>CI for μ_1</u>: $\left(\bar{X}_1 - t_{n-p,\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1}},\ \bar{X}_1 + t_{n-p,\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_1}}\right)$

$$t_{n-p,0.025} = t_{12,0.025} = 2.179$$

95% CI for μ_1 is

$$\left(9.4 - 2.179 \times \sqrt{5.9333}\,\frac{1}{\sqrt{5}},\ 9.4 + 2.179 \times \sqrt{5.9333}\,\frac{1}{\sqrt{5}}\right)$$

$$= (9.4 - 2.37, 9.4 + 2.37) = (7.03, 11.77)$$

95% CI for μ_2 is $\left(\bar{X}_2 \pm t_{n-p,\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_2}}\right)$

$$= \left(14 - 2.179 \times \sqrt{5.9333}\,\frac{1}{\sqrt{5}},\ 14 + 2.179 \times \sqrt{5.9333}\,\frac{1}{\sqrt{5}}\right)$$

$$= (14 - 2.37, 14 + 2.37) = (11.63, 16.37)$$

95% CI for μ_3 is $\left(\bar{X}_3 \pm t_{n-p,\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_3}}\right)$

$$= \left(13 - 2.179 \times \sqrt{5.9333}\,\frac{1}{\sqrt{5}},\ 13 + 2.179 \times \sqrt{5.9333}\,\frac{1}{\sqrt{5}}\right)$$

$$= (13 - 2.37, 13 + 2.37) = (10.63, 15.37)$$

<u>CI for $\mu_i - \mu_j$</u>:

$$\left(\bar{X}_i - \bar{X}_j - t_{n-p,\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_i} + \frac{1}{n_j}},\ \bar{X}_i - \bar{X}_j + t_{n-p,\frac{\alpha}{2}} S_p\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}\right)$$

$$t_{n-p,0.025} = t_{12,0.025} = 2.179$$

$$S_p\sqrt{\frac{1}{n_i} + \frac{1}{n_j}} = \sqrt{5.9333}\sqrt{\frac{1}{5} + \frac{1}{5}} = 1.541$$

95% CI for $\mu_1 - \mu_2$ is

$$\left(\bar{X}_1 - \bar{X}_2 - 2.179 \times 1.541,\ \bar{X}_1 - \bar{X}_2 + 2.179 \times 1.541\right)$$

$$= (-4.6 - 3.358,\ -4.6 + 3.358)$$

$$= (-7.958,\ -1.242)$$

95% CI for $\mu_1 - \mu_3$ is

$$\left(\bar{X}_1 - \bar{X}_3 - 3.358,\ \bar{X}_1 - \bar{X}_3 + 3.358\right)$$

$$= (-3.6 - 3.358,\ -3.6 + 3.358)$$

$$= (-9.958,\ -0.242)$$

95% CI for $\mu_2 - \mu_3$ is

$$\left(\bar{X}_2 - \bar{X}_3 - 3.358,\ \bar{X}_2 - \bar{X}_3 + 3.358\right)$$

$$= (1.0 - 3.358,\ 1.0 + 3.358)$$

$$= (-2.358,\ 4.358)$$

- **One-way ANOVA: Span versus Meal**

Source	DF	SS	MS	F	P
Meal	2	58.53	29.27	4.93	0.027
Error	12	71.20	5.93		
Total	14	129.73			

```
S = 2.436   R-Sq = 45.12%   R-Sq(adj) = 35.97%
                               Individual 95% CIs For Mean
                               Based on Pooled StDev
Level  N    Mean   StDev  --+---------+---------+---------+-------
1      5   9.400   2.302  (---------*--------)
2      5  14.000   2.550                   (--------*--------)
3      5  13.000   2.449               (--------*--------)
                          --+---------+---------+---------+-------
                           7.5      10.0      12.5      15.0

Pooled StDev = 2.436
```

Two-way ANOVA (Randomized Block Design)

The cellular phone industry is involved in a battle for customers, with each company devising its own price plan to lure customers, with the cost of a cell phone minute varying drastically depending on the no. of minutes per month used by the customer.

The monthly costs (in dollars) computed by 4 companies for peak-time callers at low (20 mins. per month), middle (150 mins. per month), and high (1000 mins. per month) usage levels are as follows:

	Companies					
Usage	A	B	C	D	Total	Means
Low	27	24	31	23	105	26.25
Middle	68	76	65	67	276	69.00
High	308	326	312	300	1246	311.50
Total	403	426	408	390	1627	
Means	134.33	142.0	136.0	130.0		

Here, we have:

$\bar{X}_{i.}$ $\qquad$ $\bar{X}_{.j}$

Sum of Squares due to Error

$= SSE = TSS - SST - SSB$

$= 189{,}798.9167 - 222.25 - 189{,}335.1667$

$= 241.5.$

<u>DF</u> Total SS $= n-1 = 12-1 = 11$,

SST $= k-1 = 4-1 = 3$,

SSB $= b-1 = 3-1 = 2$,

SSE $= 11-3-2 = 6 = (k-1)(b-1)$.

ANOVA Table

Source	SS	DF	MS	F
Treatments	222.25	3	74.083	1.84
Blocks	189335.17	2	94667.585	2351.99
Error	241.5	6	40.25	–
Total	189798.92	11	–	–

We have $F_{3,6}(0.05) = 4.76$, $F_{2,6}(0.05) = 5.14$.

Since $F = 1.84 < 4.76$, we <u>accept</u> H_0: Treatment effects are all equal (i.e., no difference in mean cost among the 4 companies) at 5% los.

Since $F = 2351.99 > 5.14$, we <u>reject</u> H_0: Block effects are all equal at 5% los, and conclude that the block means (between the 3 blocks) are unequal.

$k = 4$ treatments,

$b = 3$ blocks,

$n = kb = 4 \times 3 = 12$ obsns,

$$\bar{X} = \frac{1}{12} \sum_{i=1}^{k} \sum_{j=1}^{b} X_{ij} = \frac{1627}{12} = 135.583$$

Total Sum of Squares = TSS

$$= \sum_{i=1}^{k} \sum_{j=1}^{b} X_{ij}^2 - n\bar{X}^2$$

$$= (27^2 + \cdots + 300^2) - 12(135.583)^2$$

$$= 189{,}798.9167;$$

Sum of Squares due to Treatments

$$= SST = b \sum_{i=1}^{k} (\bar{X}_{i.} - \bar{X})^2$$

$$= 3\{(134.33 - 135.583)^2 + \cdots + (130.0 - 135.583)^2\}$$

$$= 222.25;$$

Sum of Squares due to [redacted] Blocks

$$= SSB = k \sum_{j=1}^{b} (\bar{X}_{.j} - \bar{X})^2$$

$$= 4\{(26.25 - 135.583)^2 + \cdots + (311.5 - 135.583)^2\}$$

$$= 189{,}335.1667;$$

Two-way ANOVA: Dollars versus Usage, Company

Source	DF	SS	MS	F	P
Usage	2	189335	94667.6	2351.99	0.000
Company	3	222	74.1	1.84	0.240
Error	6	242	40.3		
Total	11	189799			

S = 6.344   R-Sq = 99.87%   R-Sq(adj) = 99.77%