Σημειώσεις για τις συναρτήσεις που χρησιμοποιήθηκαν στο εργαστήριο.

Πίνακας περιεχομένων

I) Δημιουργία δήγματος με βάση γνωστή κατανομή:.....1
II) Έλεγχος υποθέσεων.....2
A) T-τεστ : Σύγκριση δύο group δεδομένων.....2
B) Z-τεστ.....4

I) Δημιουργία δήγματος με βάση γνωστή κατανομή:

(στοχαστικής μεταβλητής με γνωστή κατανομή (παράδειγμα κανονικής κατανομής))

Βήματα :

- Δημιουργία στοχαστικής μεταβλητής που να παίρνει ισοπίθανα τιμές στο διάστημα [0.1)

```
=RAND()
```

- Δημιουργία στοχαστικής μεταβλητής με βάση την κανονική κατανομή (με δεδομένη μέση τιμή **μ** και τυπική απόκλιση **σ**)

```
=NORM.INV(RAND();μ;σ)
```

- Παράδειγμα δημιουργίας 9 δηγμάτων με μ =200 και σ= 40

ΙΙ) Έλεγχος υποθέσεων

Α) Τ-τεστ : Σύγκριση δύο group δεδομένων.

Ερώτημα:
Υπάρχουν στατιστικά σημαντικά στοιχεία (statistical significant evidences) για να καταρρίψουν την υπόθεση ότι ο μέσος όρος δύο δηγμάτων είναι

- **ίδιος**
- **ο ένας δεν είναι μεγαλύτερος (ή δέν είναι μικρότερος)**

Συμπληρωματικό Ερώτημα Ι : ένα Τ τεστ δύο δηγμάτων που έχουν δημιουργηθεί με βάση την ίδια κατανομή με τη πιθανότητα θα απορρίπτουν την μηδενική υπόθεση (ενώ αυτή αληθεύει , αφού εμείς επιλέξαμε να τα φτιάξουμε από την ίδια).

Συμπληρωματικό Ερώτημα ΙΙ : ένα Τ τεστ δύο δηγμάτων που έχουν δημιουργηθεί με βάση διαφορετική κατανομή με τη πιθανότητα δέν θα απορρίπτουν την μηδενική υπόθεση (ενώ αυτή αληθεύει , αφού εμείς επιλέξαμε να τα φτιάξουμε από διαφορετική).

Δημιουργία 2 δηγμάτων με βάση την ίδια κατανομή
(κανονική κατανομή με μ =200 και σ= 40)

Εφαρμογή Τ τεστ δύο δηγμάτων

Δυνατές επιλογών που έχουμε σε ένα **Τ τεστ**) :

1.
   - τα δήγματα είναι paired ή όχι
     (π.χ. αντιστοιχούν στο ίδιο "άτομο" υπό δύο διαφορετικές συνθήκες
     ή σε διαφορετικά "άτομα" ?)
   - Αναμένω τα δύο δήγματα να έχουν την ίδια διακύμανση
   - Αναμένω τα δύο δήγματα να έχουν την διαφορετική διακύμανση (αν δεν το ξέρω επιλέγω συνήθως αυτό)
2. το τέστ είναι one or two tail ?

η Μηδενική υπόθεση H_0 αφορά

α) (two tails **)** ισότητα μέσων όρων $\mu_\chi =$ (άρα η εναλλακτική υπόθεση H_1 σημαίνει ότι είτε $\mu_\chi >$ είτε $\mu_\chi <$)

β) (one tail**)** ανισότητα μέσων όρων $\mu_\chi >$, ή $\mu_\chi <$ (ένα από τα δύο όχι είτε το ένα είτε το άλλο)

Σε περίπτωση που πρόκριτε για paired, το libreoffice έχει αυτόματο υπολογισμό μέσω των επιλογών
Δεδομένα → Στατιστικά → Δοκημή Τ με ζεύγη.
Όπου αναφέρονται όχι μόνο το p-value αλλά και οι κρίσιμες τιμές.

Τα Τ-τεστ μπορούν να υπολογιστούν με την χρήση μίας συνάρτησης.
=TTEST(**Array1** ;**Array2** ; **Tails** ; **Type**)

όπου
Array1 ; τα κελιά με τα δεδομένα του πρώτου group
Array2 ; τα κελιά με τα δεδομένα του δεύτερου group

Tails ; 1 ή 2 (one tail ή two tails)

Type :

1 τα δήγματα είναι paired

2 τα δήγματα αναμένετε να έχουν την ίδια διακύμανση

3 δήγματα δεν αναμένετε να έχουν την ίδια διακύμανση (ή απαλά δεν ξέρω)

Παράδειγμα

p value :

paired , one tail

θεωρώντας ίδια διακήμανση, one tail

μή θεωρώντας ίδια διακήμανση, one tail

(η διαφορά του να χρησημοποιήσεις 2 tails αντί για 1 είναι ότι διπλασιάζετε η τιμή του p δηλαδή :

=TTEST(**Array1** ;**Array2** ; **2** ; **Type**) = 2* TTEST(**Array1** ;**Array2** ; **1** ; **Type**)

έτσι ώστε το p να συγκρίνετε πάντα με το επίπεδο σημαντικότητας α (π.χ. το 0.05) και να μήν χριάζετε να χωρίσεις το α σε αριστερό και δεξιό μέρος.)

Link για διαφορές του ttest

B) Z-τεστ

A

Δεδομένα : ένας μέσος όρος **<X>**, και η διακύμανση της κατανομής σ_{χ} .
Σύγκριση του <X>με την υπόθεση μας για μ_{χ}
(μ_{χ} η μέση κατανομή του πληθυσμού π.χ. τι θα παίρναμε εάν μπορούσαμε να κάναμε άπειρα πειράματα ελέγχοντας όλα τα ενδεχόμενα)

Ερώτημα : Ο μέσος όρος που παρατήρησα <X> στις μετρήσεις μου υποστηρίζουν ή εναντιώνετε στην υπόθεση ότι η μέση τιμή του πληθυσμού είναι μ_{χ} (είτε γνωρίζοντάς το σ_{χ} είτε εκτιμώντας το με βάση την διακύμανση του δήγματος (link).
Παραδείγματα στο αρχείο Hypothesis_testing_examples_1.xlsx στο eclass.

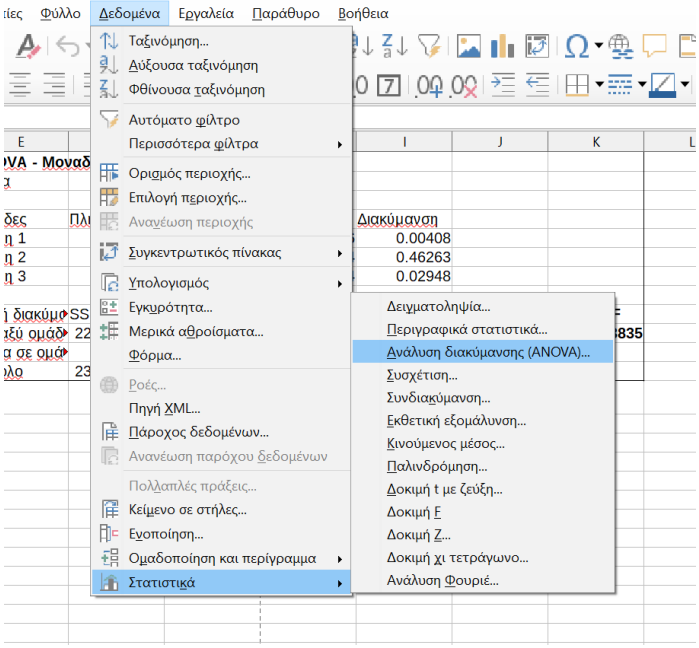
B

Δεδομένα : Ένα σέτ μετρήσεων, (και όχι υποχρεωτικά η διακύμανση της κατανομής $\sigma_{\chi\ ,\ \mu\iota\alpha\varsigma}$ και μπορεί να εκτιμηθεί από το σετ των μετρήσεων)
Ερώτημα : Το σέτ μετρήσεων μας προσφέρουν στατιστικά σημαντικά στηχεία για να απορίψουμε την υπόθεση ότι η μέση τιμή του πληθυσμού είναι μ_{χ} (το όπου το σ_{χ} μπορεί να εκτιμάτε το με βάση την διακύμανση του δήγματος ή να δοθεί σαν δεδομέν(link).
=ztest(**Array1,** $\mu_{\chi,}$ σ_{χ})
Ζτεστ θεωρεί κανονική κατανομή και κατά συνέπεια θα πρέπει να εφαρμόζετε για μέγεθος δήγματος μεγαλύτερου του 30.
Παραδείγματα στο αρχείο Hypothesis_testing_examples_1.xlsx στο eclass.

Άλλα παραδείγματα χρήσης στατιστικών τέστ στο LibreOffice .

Anova ANalisis Of Varience

Εφαρμογή χρησιμοποιώντας τις επιλογές :

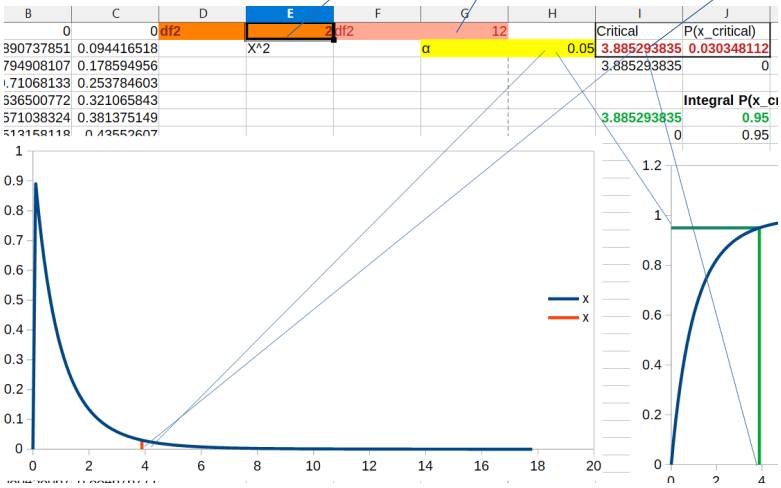


Αποτελέσματα :
P value , το F είναι μεγαλίτερο ή μικρότερο από το Fcrit
Το Fcrit εχαρτ'ατε από 2 βαθμούς ελευθερίας
τον αριθμό των γκρούπς – 1
και τον αριθμό των μετρίσεων - τον αριθμό των γκρούπς.
Αναλυτικό παράδειγμα υπολογισμού anova στο αρχείο :anova_pH (2019_04_12 12_59_09 UTC)

Μέτρηση 1	Μέτρηση 2	Μέτρηση 3
1.16	3.57	1.16
1.24	3.57	1.15
1.06	4.04	1.06
1.16	4.74	1.3
1.16	2.9	1.5

ANOVA - Μοναδικός παράγοντας

Άλφα	0.05					
Ομάδες	Πλήθος	Άθροισμα	Μέσος	Διακύμανση		
Στήλη 1	5	5.78	1.156	0.00408		
Στήλη 2	5	18.82	3.764	0.46263		
Στήλη 3	5	6.17	1.234	0.02948		
Πηγή διακύμ	SS	df	MS	F	Τιμή-P	Κρίσιμο F
Μεταξύ ομάδ	22.01441333	2	11.00720667	66.5503537	3.19941E-07	3.885293835
Μέσα σε ομά	1.98476	12	0.165396667			
Σύνολο	23.99917333	14				



Τέστ χ^2

Παράδειγμα από Γενετική.
1) Σύγκριση προβλεπόμενης αναλογίας φαινοτίπων με βάση τον δεύτερο νόμο του Mentel για πεπερασμένο δήγμα.

Παράδειγμα εφαρμογής του ελέγχου χ^2 *στον* δεύτερο νόμο του Μέντελ με βάση την συχνότητα εμφάνισης φαινοτύπων στην F2 γενιά

$$\chi^2 = \frac{(E_1 - O_1)^2}{E_1} + \frac{(E_2 - O_2)^2}{E_2} + \frac{(E_3 - O_3)^2}{E_3} + \frac{(E$$

Μέγεθος δήγματος	600		
Συχνότητα Υπόθεσης	Expected Values	Observed values	(E-O)^2/E
9/16	337.5	330	0.166667
3/16	112.5	115	0.055556
3/16	112.5	111	0.02
1/16	37.5	44	1.126667
κ (βαθ.Ελευθ)		X^2	**1.368889**
3	71%	Παραμάτων θα έδιναν μεγαλήτερο X^2	
p	0.05	κρήσιμη τιμή Xcr^2	7.814728

p το εμβαδόν που αντιστοιχεί στα πειράματα με $\chi^2 > \chi^2_{cr}$

Επισήμανση της διαφοράς σε περίπτωση που χρησημοποιήτε το τέστ χ^2 για τον έλενχο γεννετικής ισοροπίας : οι βαθμού ελευθερίας δεν είναι ο αριθμός των κατηγοριών-1 όπως σε αυτό το παράδειγμα αλλά ίσος με τον αριθμώ των παραμέτρων του μοντέλου γεννετικής ισοροπίας (λυγότεροι)

Κατανομές που χρησιμοποιούνται στα τεστ υποθέσεων και η σημασία των βαθμών ελευθερίας.

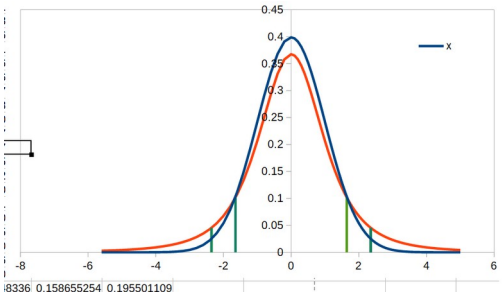
Στο Αρχείο distriiutions_df.xlsx
θα βρείτε παραδείγματα όπου δείχνουν την Κατανομές που χρισιμοποιούντε στα τέστ που αναφέραμε (μαζί με τα διαγράμματα των αθριστικών κατανομών που μας δίνουν τα εμβαδά που εμπλέκονται στους υπολογισμούς του επιπέδου σημαντικότητας σε σχέση με τις κρισημες τημές και το p value):

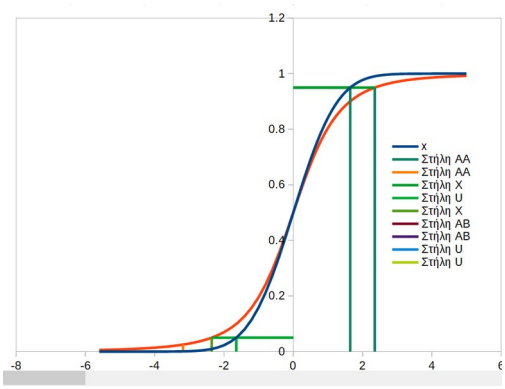
a) κανονική κατανομή (με $\mu_\chi=0$, $\sigma_\chi=1$)
b) κατανομή t (σαν συνάρτηση των βαθμών ελευθερίας)
c) κατανομή χ^2 (σαν συνάρτηση των βαθμών ελευθερίας)
d) κατανομή F που χρησιμοποιείτε στο ANOVA (σαν συνάρτηση 2 βαθμών ελευθερίας)

Ερωτήματα που μπορούν να κατανοιθούν εξετάζοντας τα διαγράμματα του αρχείου.

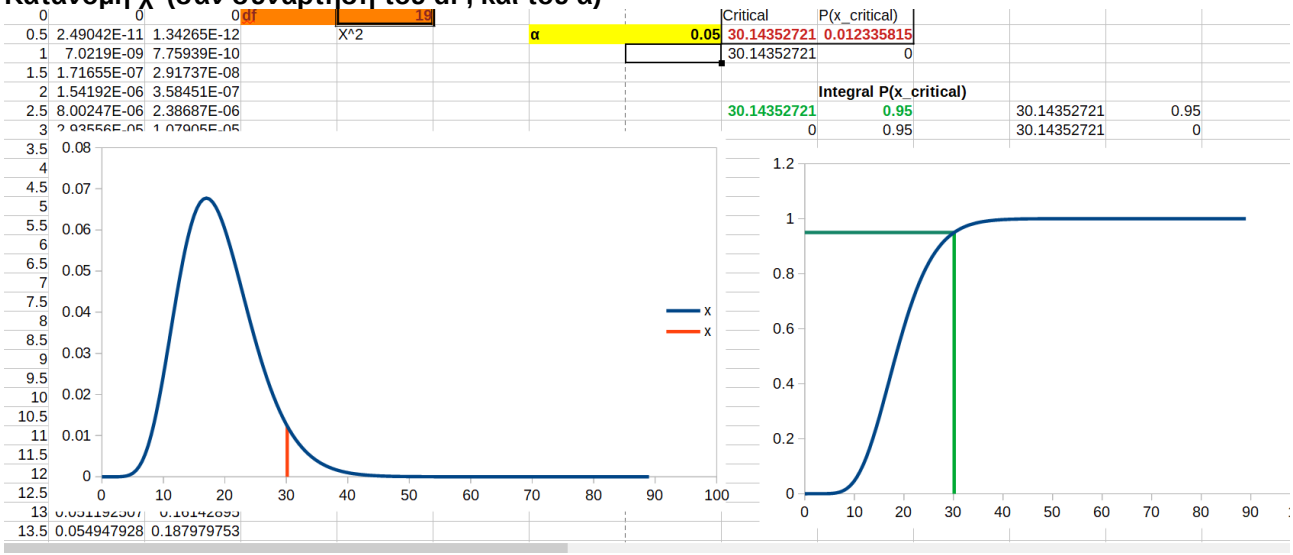
1) Πόσο διαφέρει η κατανομή t από την κανονική, και πως όταν εχουμε μεγαλιτερο δήγμα (περισσότερους βαθμούς ελευθερίας) οι δύο κατανομές πλησιάζουν)
2) Τι αλλάζει όταν αλλάζουμε το διάστημα εμπιστοσήνης.
3)Πως αλλάζουν οι κρίσιμες τιμές των τέστ όταν αλλάζουν οι βαθμοί ελευθερίας.

Τ-κατανομή με 3 βαθμούς ελευθερίας (κελί H2 στο αρχείο distriiutions_df.xlsx) και επίπεδα εμπιστοσύνης α = 0.05 (κελί K2 στο αρχείο distriiutions_df.xlsx)

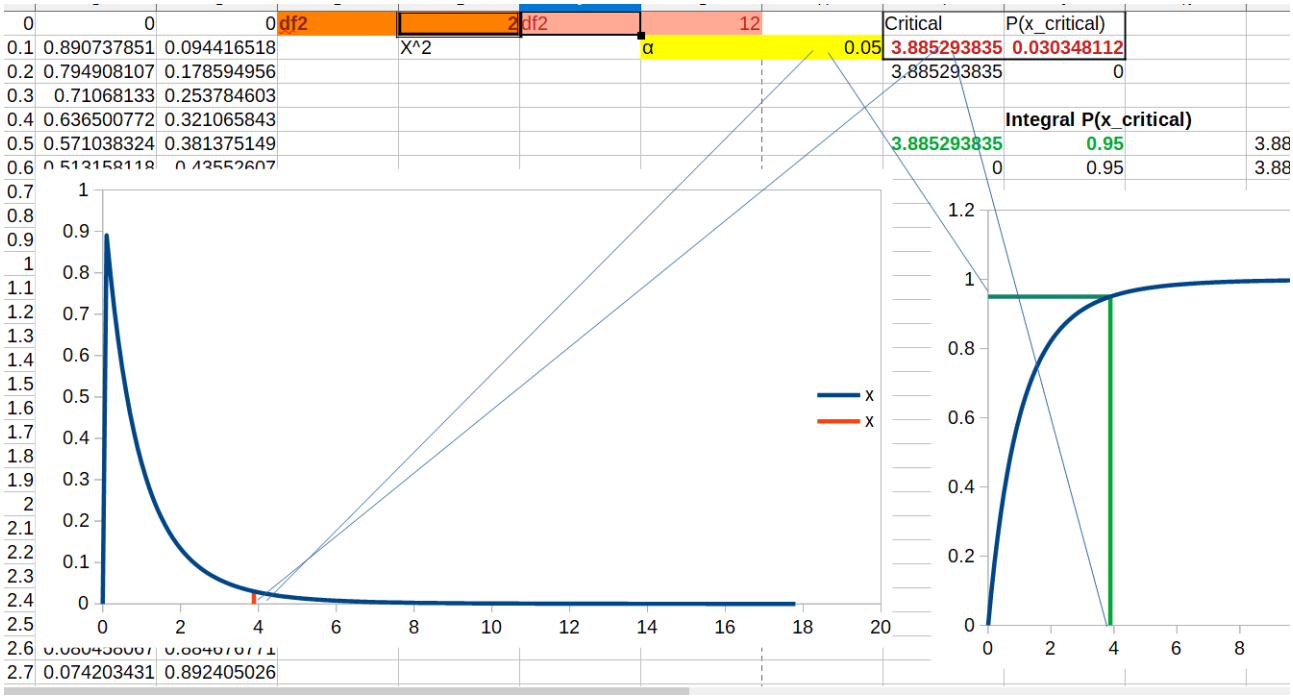




Κατανομή χ^2 (σαν συνάρτηση του df , και του α)



Κατανομή F.



Μαθηματικό υπόβαθρο στην διαφορά του υπολογισμού της διακύμανσης, και των βαθμών ελευθερίας στα είδη των τεστ.

$\mu_x = \mu_y \rightarrow t \rightarrow \in N\left(0, \sigma_t\right)$

Type :

1 τα δήγματα είναι <u>paired</u>

$\mu_x = \mu_y \rightarrow \bar{x} \approx \bar{y} \rightarrow \overline{x-y} \rightarrow \in Norm\left(0, \sigma_{\overline{x-y}}\right)$

$\sigma^2_{\overline{x-y}} \rightarrow \sigma^2_{\bar{x}} + \sigma^2_{\bar{y}}, \alpha\nu\, x\, \kappa\alpha\iota\, y$ είναι ασυσχέτιστα

$\sigma^2_{\bar{y}} \rightarrow \sigma^2_y / \sqrt{(M)}$

$\sigma^2_{\bar{x}} \rightarrow \sigma^2_x / \sqrt{(N)},$

$$\sigma^2_x / \sqrt{(N)} \rightarrow S^2_x = \frac{\sum \left(x_i - \bar{x}\right)^2}{(N-1)} \sigma^2_y / \sqrt{(M)} \rightarrow S^2_y = \frac{\sum \left(y_i - \bar{y}\right)^2}{(M-1)}$$

,

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s^2_x}{N} + \frac{s^2_x}{M}}} \rightarrow \bar{x} - \bar{y} \rightarrow \in N\left(\mu_x - \mu_y, \sqrt{\frac{s^2_x}{N} + \frac{s^2_x}{M}}\right)$$

$$df = \frac{2}{\left(\frac{1}{N}+\frac{1}{M}\right)}$$

Βαθμοί ελευθερίας : (N+M)/2-1 ή

2 τα δήγματα αναμένετε να έχουν την ίδια διακύμανση

$$\mu_x = \mu_y \rightarrow \bar{x} \approx \bar{y} \rightarrow \overline{x-y} \rightarrow \in Norm\left(0, \sigma_p\right)$$

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{N}+\frac{s_x^2}{M}}} \rightarrow \frac{\bar{x}-\bar{y}}{s_p\sqrt{\frac{1}{N}+\frac{1}{M}}} \rightarrow \bar{x}-\bar{y} \rightarrow \in N\left(\mu_x-\mu_y, s_p\sqrt{\frac{1}{N}+\frac{1}{M}}\right)$$

$$s_p^2 = \frac{(N-1)*s_{\bar{x}}^2+(M-1)*s_{\bar{y}}^2}{N+M-2}$$

DF = N+M-2

3 unpaired και τα δήγματα δεν αναμένετε να έχουν την ίδια διακύμανση

$$\mu_x = \mu_y \rightarrow \bar{x} \approx \bar{y} \rightarrow \bar{x}-\bar{y} \rightarrow \in Norm\left(\mu_x-\mu_y, \sigma_{\bar{x}-\bar{y}}\right)$$

$$\sigma_{\bar{y}}^2 \rightarrow \sigma_y^2/\sqrt{(M)}$$
$$\sigma_{\bar{x}}^2 \rightarrow \sigma_x^2/\sqrt{(N)},$$

$$\sigma_x^2/\sqrt{(N)} \rightarrow S_x^2 = \frac{\sum\left(x_i-\bar{x}\right)^2}{(N-1)}\sigma_y^2/\sqrt{(M)} \rightarrow S_y^2 = \frac{\sum\left(y_i-\bar{y}\right)^2}{(M-1)}$$

,

$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{N}+\frac{s_x^2}{M}}} \rightarrow \frac{\bar{x}-\bar{y}}{s_p\sqrt{\frac{1}{N}+\frac{1}{M}}} \rightarrow \bar{x}-\bar{y} \rightarrow \in Norm\left(\mu_x-\mu_y, s_p\sqrt{\frac{1}{N}+\frac{1}{M}}\right)$$

$$s_p^2 = \frac{(N-1)*s_{\bar{x}}^2+(M-1)*s_{\bar{y}}^2}{N+M-2}$$

DF = N+M-2