

---

**ΕΠΕΚΤΑΣΗ ΥΠΟΘΕΣΕΩΝ**

---

**Έλεγχος Υποθέσεων:** Είναι μια μέθοδος της Στατιστικής Συμπερασματολογίας (statistical inference) που μας βοηθά να βγάλουμε συμπεράσματα σε σχέση με την τιμή μιας παραμέτρου του πληθυσμού, συγκρίνοντας τα αποτελέσματα του δείγματος με εκείνα που θα περιμέναμε εάν η **υπόθεση** είναι αληθινή.

Η μέθοδος αυτή μας βοηθά να διαπιστώσουμε εάν τα δεδομένα του δείγματος υποστηρίζουν την υπόθεση ότι η παράμετρος του πληθυσμού έχει μια συγκεκριμένη τιμή.

Με άλλα λόγια προσπαθούμε να προσδιορίσουμε την πιθανότητα να προκύψει ένα δείγμα όπου η τιμή της παραμέτρου είναι π.χ.  $\mu$  από έναν πληθυσμό με πραγματική τιμή  $\mu_0$ . Εάν η πιθανότητα είναι μεγάλη τότε η διαφορά μεταξύ του  $\mu$  και  $\mu_0$  οφείλεται στις τυχαίες κυμάνσεις της δειγματοληψίας. Αντίθετα, εάν διαπιστωθεί ότι η πιθανότητα είναι μικρή, τότε η υπόθεση δεν ισχύει.


**Παραδείγματα:** Έλεγχος ποσοστού εκλογικού σώματος υπέρ κάποιου κόμματος, έλεγχος ποιότητας στην βιομηχανία,


---

 **Μηδενική υπόθεση** (Null Hypothesis)  $H_0$ , είναι η υπόθεση που ελέγχουμε.

 **Εναλλακτική υπόθεση** (Alternative Hypothesis)  $H_1$ , είναι η υπόθεση που θα δεχθούμε εάν η μηδενική υπόθεση βρεθεί εσφαλμένη.

 Η **Εναλλακτική υπόθεση** δεν περιέχει το σύμβολο ίσον (=) όσον αφορά την τιμή της παραμέτρου του πληθυσμού αλλά βασίζεται στο  $\neq$ ,  $>$ ,  $<$

 Η αποδοχή ή η απόρριψη της μηδενικής υπόθεσης βασίζεται στη τιμή του **κριτηρίου ελέγχου** (test statistic). Κριτήριο ελέγχου είναι η στατιστική (μεταβλητή) που υπολογίζεται με βάση τις πληροφορίες του δείγματος και ακολουθεί μια γνωστή κατανομή (Z κανονικής κατανομής, t κατανομής student, κτλ). Εφόσον το κριτήριο ακολουθεί μια γνωστή κατανομή, με την χρήση των εμβαδών της αντίστοιχης κατανομής θα προσδιορίσουμε την πιθανότητα με βάση την οποία θα κρίνουμε εάν η  $H_0$  είναι αληθινή ή όχι.

 **Επίπεδο σημαντικότητας** (significant level) και συμβολίζεται με  $\alpha$ , είναι η τιμή της πιθανότητας που μικρότερή της θα θεωρείται απίθανο να αποδοθεί στην τύχη ενώ μεγαλύτερη θα είναι φυσικό να ην αποδώσουμε στην τύχη.

---

# Έλεγχος Υπόθεσης Μέσου (γνωστή $\sigma$ ): κριτήριο $Z$

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

Κριτήριο ελέγχου:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Εφόσον το επίπεδο σημαντικότητας είναι  $\alpha$ , τότε η περιοχή απόρριψης έχει εμβαδόν (πιθανότητα)  $\alpha$ , και η περιοχή αποδοχής εμβαδόν ίσο με  $1-\alpha$

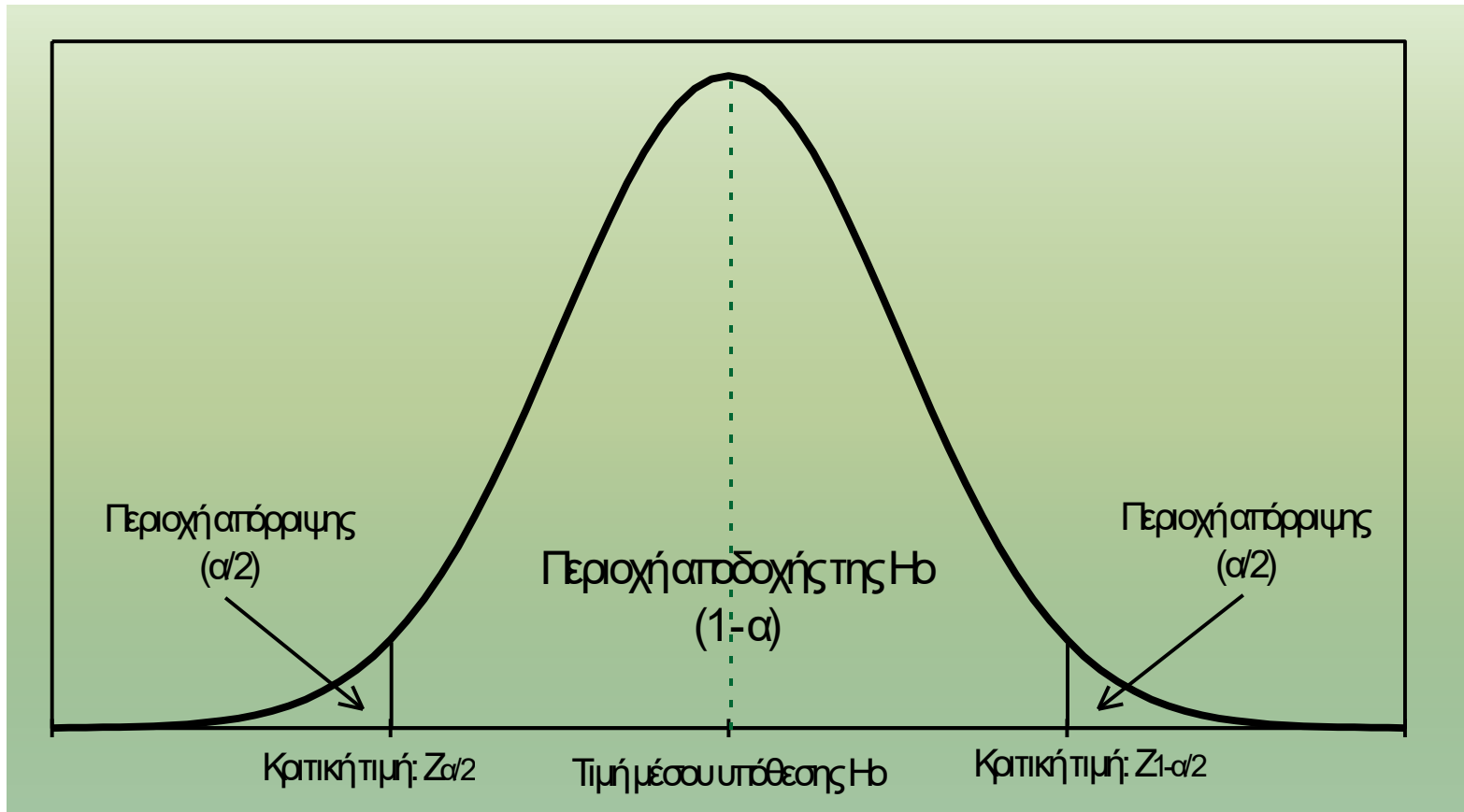
Οι τιμές του κριτηρίου  $Z$  που αντιστοιχούν στα όρια της περιοχής απόρριψης ονομάζονται **κριτικές τιμές** (critical values). Η κάτω τιμή συμβολίζεται με  $Z_{\alpha/2}$  και η άνω τιμή με  $Z_{1-\alpha/2}$

Εάν  $Z$  είναι η τιμή που προκύπτει από τα δεδομένα του δείγματος τότε εάν:

$Z < Z_{\alpha/2}$  ή  $Z > Z_{1-\alpha/2}$  τότε η  $H_0$  απορρίπτεται

$Z_{\alpha/2} < Z < Z_{1-\alpha/2}$  τότε η  $H_0$  γίνεται δεκτή

# Κατανομή Δειγματοληψίας και Περιοχές Απόρριψης και Αποδοχής της $H_0$



## Παράδειγμα:

Μία εταιρεία παράγει συνθετικό δέρμα που χρησιμοποιείται για την υποδηματοποιία. Μία από τις βασικές προδιαγραφές που πρέπει να τηρούνται είναι το πάχος. Από τις προδιαγραφές είναι γνωστό ότι ο συγκεκριμένος τύπος συνθετικού δέρματος παράγεται με μέσο πάχος 4 χιλιοστά και  $\sigma=0,1$  χιλιοστά. Από δείγμα 50 μετρήσεων προέκυψε μέσο πάχος=4,02 χιλιοστά. Το ερώτημα που απασχολεί τον υπεύθυνο παραγωγής είναι εάν η παραγωγή εξελίσσεται ομαλά σύμφωνα με τις προδιαγραφές ή κάποιο πρόβλημα στην παραγωγή προκαλεί μεταβολή στο πάχος του συνθετικού δέρματος.

$n=50$

$H_0: \mu = 4 \text{ χιλ.}$

$H_1: \mu \neq 4 \text{ χιλ.}$

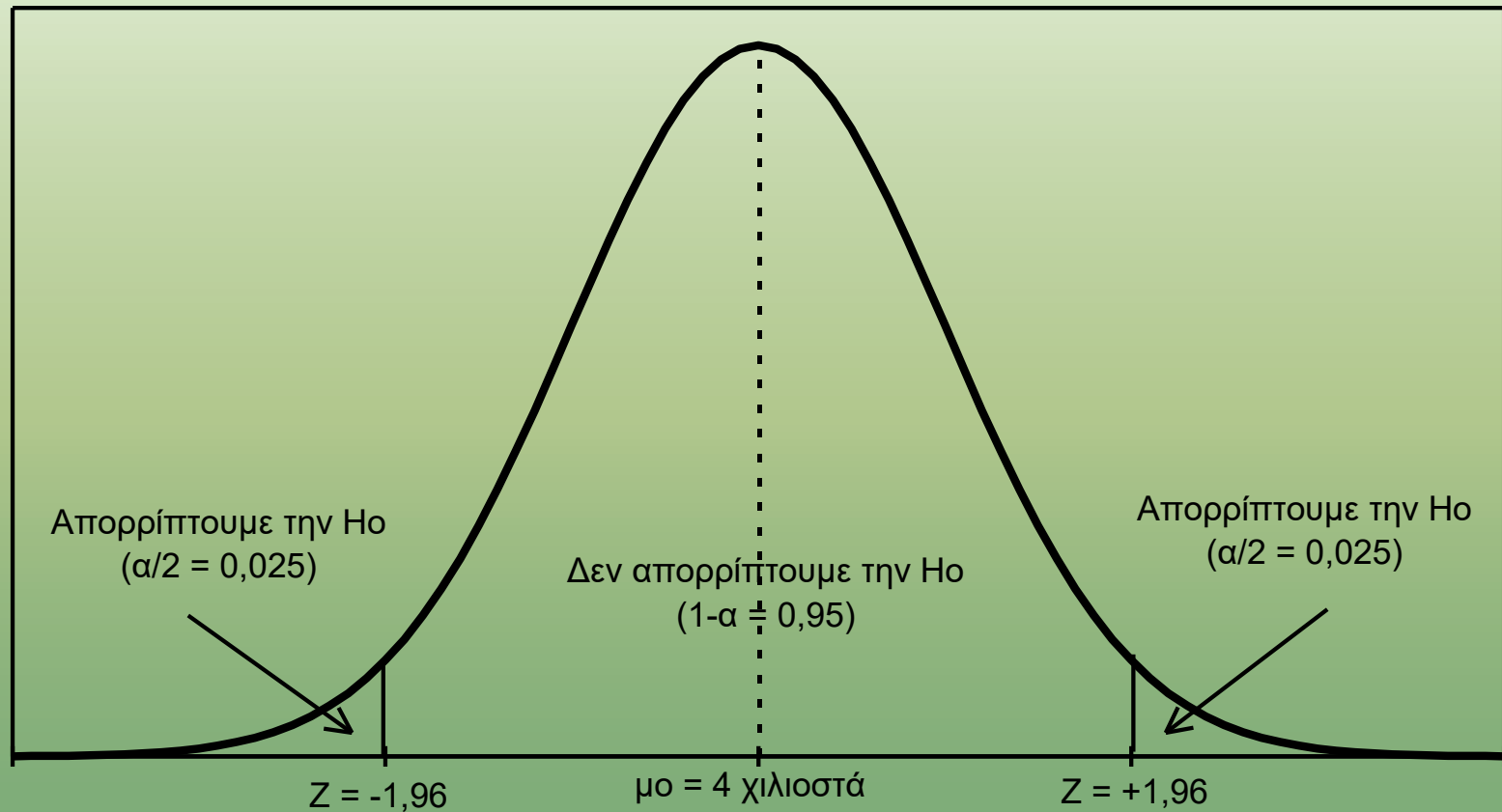
Οι κριτικές τιμές του κριτηρίου  $Z$  σύμφωνα με την τυποποιημένη κανονική κατανομή είναι  $-1,96$  και  $1,96$

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{4,02 - 4,00}{0,1 / \sqrt{50}} = +1,41$$

Εφόσον  $Z=+1,41$ , βλέπουμε ότι  $-1,96 < Z \leq 1,96$  που σημαίνει ότι η τιμή του κριτηρίου  $Z$  βρίσκεται στην περιοχή αποδοχής της  $H_0$

<b>z</b>	<b>0.00</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>
<b>0.0</b>	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
<b>0.1</b>	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
<b>0.2</b>	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
<b>0.3</b>	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
<b>0.4</b>	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
<b>0.5</b>	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
<b>0.6</b>	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
<b>0.7</b>	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
<b>0.8</b>	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
<b>0.9</b>	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
<b>1.0</b>	0.84134	0.84375	0.84614	0.84850	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
<b>1.1</b>	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
<b>1.2</b>	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
<b>1.3</b>	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
<b>1.4</b>	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92786	0.92922	0.93056	0.93189
<b>1.5</b>	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
<b>1.6</b>	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
<b>1.7</b>	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
<b>1.8</b>	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
<b>1.9</b>	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
<b>2.0</b>	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
<b>2.1</b>	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
<b>2.2</b>	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
<b>2.3</b>	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
<b>2.4</b>	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
<b>2.5</b>	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
<b>2.6</b>	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
<b>2.7</b>	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
<b>2.8</b>	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
<b>2.9</b>	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
<b>3.0</b>	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99897	0.99900

# Έλεγχος Υπόθεσης Μέσου (γνωστή $\sigma$ ) με Επίπεδο Σημαντικότητας 0,05





# Έλεγχος Υπόθεσης μέσου του προγράμματος Statgraphics

Sample population=4.02

Population standard deviation=0.1

Sample size=50

95% confidence interval for mean:4.02+/- 0.02772 [3.99228;4.04772]

Null Hypothesis:  $\mu=4,0$

Alternative: Not equal

Computed Z statistic=1.41421

**P-Value=0.1573**

Conclusion: Do not reject the null hypothesis for  $\alpha=0.05$

## The StatAdvisor

Given a sample of 50 observations with a mean of 4.02 and a population standard deviation of 0.1 the computed Z statistic equals 1.41421. since the P-value for the test is grater than or equal to 0.05, the null hypothesis cannot be rejected at the 95.0% confidence level. The confidence interval shows that the values of  $\mu$  supported by the data fall between 3.99228 and 4.04772

★ Απορρίπτουμε την  $H_0$  εάν  $|Z| > |Z_{\alpha/2}|$  (ή  $|Z| > |Z_{1-\alpha/2}|$ )

★ Δεχόμαστε την  $H_0$  εάν  $|Z| < |Z_{\alpha/2}|$  (ή  $|Z| < |Z_{1-\alpha/2}|$ )

## Σχέση μεταξύ του Διαστήματος Εμπιστοσύνης και του ελέγχου της υπόθεσης

$\Delta E: \bar{X} \pm Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$       Μας πληροφορεί για το που μπορεί να βρίσκεται ο άγνωστος  $\mu$  εάν γνωρίζουμε μόνο τον  $\bar{X}$

Ο έλεγχος της  $H_0$  μας λέει εάν ο  $\bar{X}$  μπορεί να προέρχεται από έναν πληθυσμό με μέσο  $\mu_0$ .

Όλες οι τιμές του διαστήματος εμπιστοσύνης μπορούν να γίνουν δεκτές σαν  $\mu_0$  εφόσον απέχουν από τον  $\bar{X}$  μέχρι  $\pm Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$

Με άλλα λόγια αντί να κάνουμε τον κλασικό έλεγχο της  $H_0$ , εκτιμούμαι το διάστημα εμπιστοσύνης και εάν περιλαμβάνεται η τιμή  $\mu_0$  τότε δεχόμαστε την  $H_0$  αλλιώς την απορρίπτουμε.

# Εσφαλμένα Συμπεράσματα από την Διεξαγωγή των Ελέγχων Υποθέσεων

**Σφάλμα Τύπου I:** Η Υπόθεση μηδέν ( $H_0$ ) απορρίπτεται ενώ είναι σωστή

Παράδειγμα: κάποιες ακραίες τιμές στην δειγματοληψία είχαν σαν αποτέλεσμα ο μέσος όρος να ισούται με  $\bar{X} = 4,03$

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{4,03 - 4,00}{0,1 / \sqrt{50}} = 2,12 > 1,96 \quad \text{άρα η } H_0 \text{ απορρίπτεται ενώ είναι σωστή}$$

**Σφάλμα Τύπου II:** Η Υπόθεση μηδέν ( $H_0$ ) δεν απορρίπτεται γίνεται δεκτή ενώ είναι λανθασμένη.

Παράδειγμα:

$$\bar{X} = 4,025 \quad Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{4,025 - 4,00}{0,1 / \sqrt{50}} = 1,77 < 1,96$$

# Εσφαλμένα Συμπεράσματα από την Διεξαγωγή των Ελέγχων Υποθέσεων

Κατά το έλεγχο υποθέσεων υπάρχουν δύο δυνατοί τύποι σφαλμάτων. Το **σφάλμα τύπου I** συμβαίνει όταν απορρίπτουμε μία αληθινή μηδενική υπόθεση, ενώ το **σφάλμα τύπου II** συμβαίνει όταν δεχόμαστε μία ψευδή μηδενική απόφαση.

Πραγματικότητα	Απόφαση	
	Επιλέγω $H_0$	Επιλέγω $H_1$
$H_0$ Αληθινή	Σωστή Απόφαση	Σφάλμα Τύπου I
$H_1$ Αληθινή	Σφάλμα τύπου II	Σωστή Απόφαση

# Παράδειγμα Εσφαλμένων αποφάσεων

Ένα συνηθισμένο παράδειγμα είναι από τον χώρο της δικαιοσύνης. Όταν κάποιος αντιμετωπίζει μια σοβαρή κατηγορία οδηγείται στο δικαστήριο, όπου ο δημόσιος κατήγορο και ο συνήγορος υπεράσπισης παρουσιάζουν αποδείξεις και αναπτύσσουν επιχειρήματα, και τελικά οι ένορκοι αποφασίζουν αν ο κατηγορούμενος είναι αθώος ή ένοχος.

Αυτό που κάνουν οι ένορκοι είναι ένας έλεγχος υποθέσεων.

$H_0$ : ο κατηγορούμενος είναι αθώος

$H_1$ : ο κατηγορούμενος είναι ένοχος



Πραγματικότητα	Απόφαση δικαστηρίου	
	Επιλέγω $H_0$ (Αθώωση)	Επιλέγω $H_1$ (Καταδίκη)
$H_0$ Αληθινή (αθώος)	Σωστή Απόφαση	Σφάλμα Τύπου I
$H_1$ Αληθινή (ένοχος)	Σφάλμα τύπου II	Σωστή Απόφαση

# Ερώτηση ....

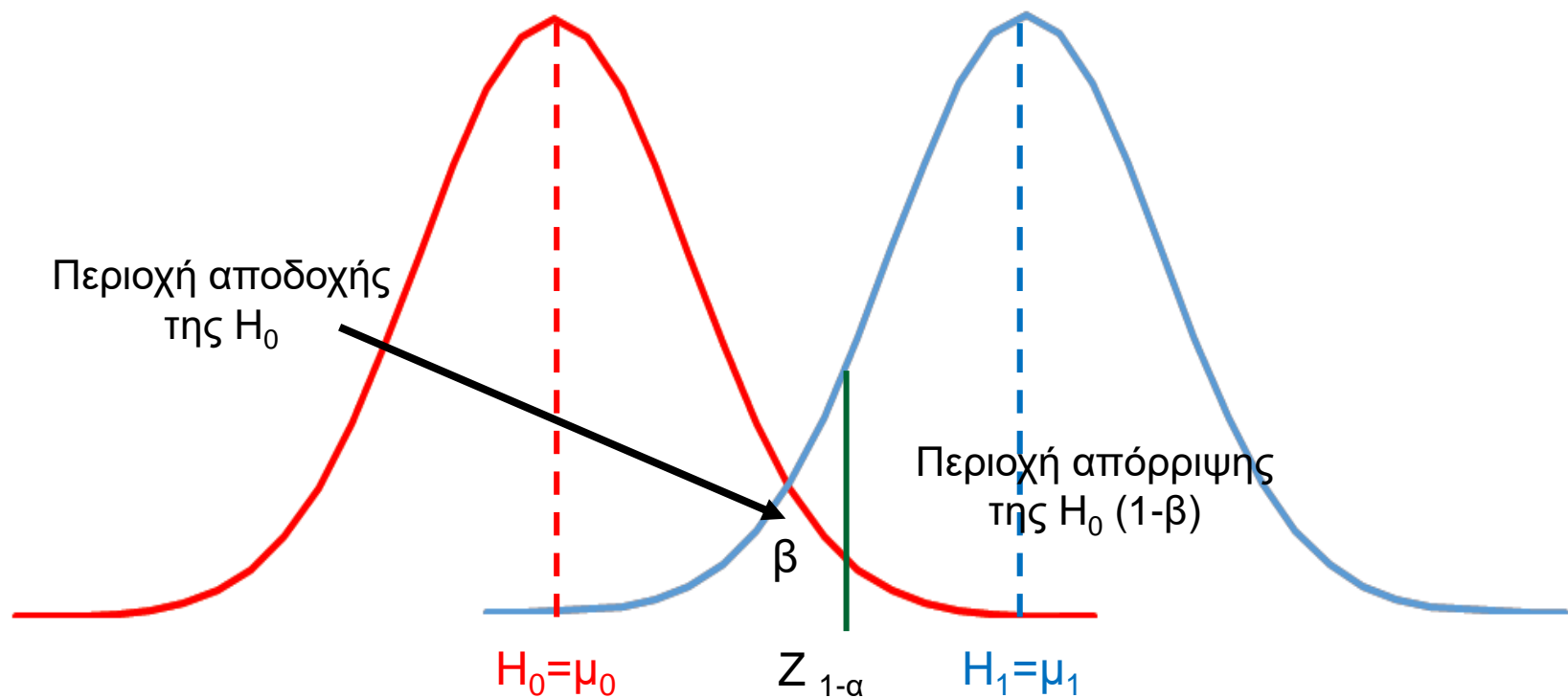
1. Μπορούμε να αποφύγουμε τα σφάλματα κατά τον έλεγχο των υποθέσεων;
2. Μπορούμε να υπολογίσουμε την πιθανότητα να διαπράξουμε κατά τον έλεγχο των υποθέσεων;

$\alpha = P(\text{Σφάλμα Τύπου I}) = P(\text{Απορρίπτω την } H_0 \text{ όταν είναι αληθινή}) =$   
Επίπεδο σημαντικότητας

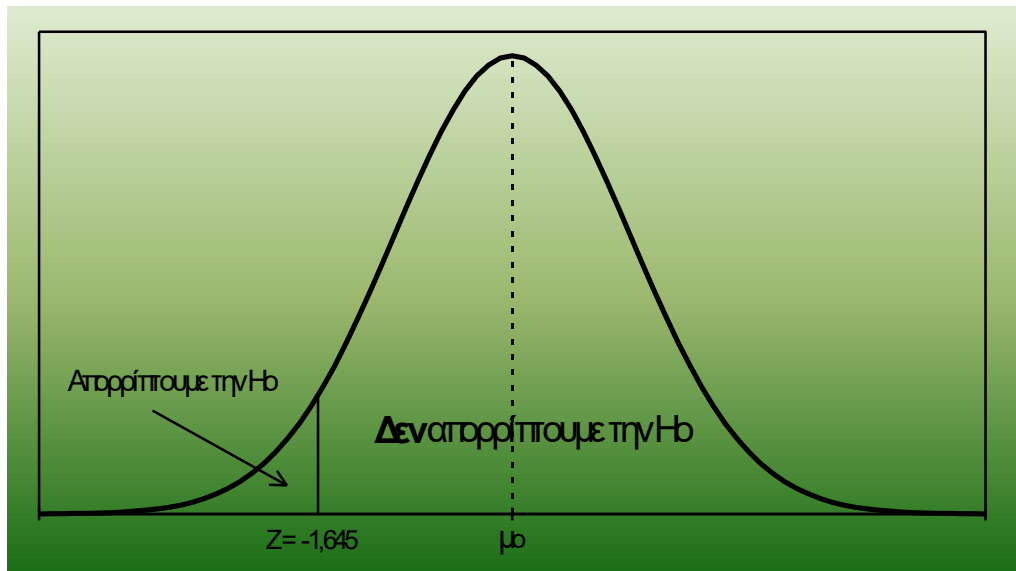
$\beta = P(\text{Σφάλμα Τύπου II}) = P(\text{Δεν απορρίπτω την } H_0 \text{ όταν είναι λάθος})$

**$1-\beta$ : Δύναμη κριτηρίου** ή **Ισχύς ελέγχου** εκφράζεται συνήθως ως συνάρτηση των τιμών  $\mu_1 - \mu_0$  δηλαδή  $1-\beta = f(\mu_1 - \mu_0)$

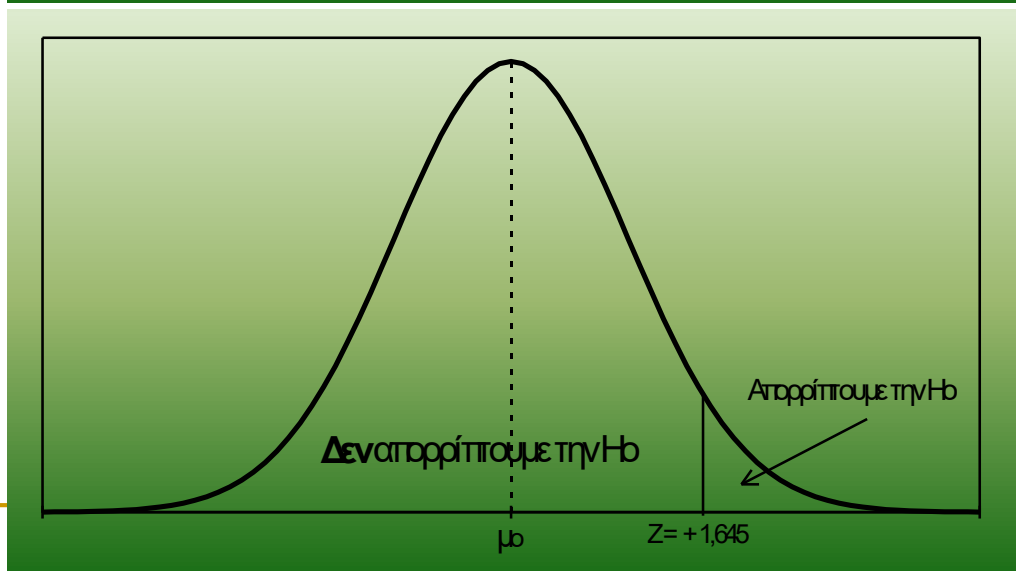
## Κατανομές Δειγματοληψίας και Περιοχές Απόρριψης και Αποδοχής της $H_0$ όταν η $H_0$ είναι Λανθασμένη



## Περιοχές Απόρριψης και Αποδοχής της $H_0$ σε Μονοκατάληκτα Κριτήρια Ελέγχου ( $\alpha = 0.05$ )



$H_0: \mu = \mu_0$   
Έναντι της εναλλακτικής  $H_1:$   
 $\mu < \mu_0$   
Εάν  $Z < Z_\alpha$  απορρίπτουμε την  
υπόθεση μηδέν ( $H_0$ )



$H_0: \mu = \mu_0$   
Έναντι της εναλλακτικής  $H_1:$   
 $\mu > \mu_0$   
Εάν  $Z > Z_\alpha$  απορρίπτουμε την  
υπόθεση μηδέν ( $H_0$ )



# Έλεγχος Υπόθεσης Μέσου (άγνωστη $\sigma$ ): κριτήριο $t$

Τις περισσότερες φορές η τυπική απόκλιση του πληθυσμού είναι άγνωστη και ως εκ τούτου χρησιμοποιούμε την εκτίμηση  $s$  από τα δεδομένα του δείγματος.

Έτσι με άγνωστη τυπική απόκλιση του πληθυσμού, χρησιμοποιούμε το κριτήριο  $t$  *student* με  $n-1$  βαθμούς ελευθερίας για τον έλεγχο της υπόθεσης  $H_0: \mu = \mu_0$

$$t_{n-1} = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$$

Παράδειγμα:

4.7	4.5	4.8	5.0	5.1	5.4	4.9	5.0	5.1	4.3	5.1	5.6	5.6	4.6	5.2
4.5	4.3	4.9	5.2	5.3	5.5	5.2	5.6	4.8	4.5	5.3	5.5	4.9	5.0	5.2

$$H_0: \mu = 5$$

$$H_1: \mu \neq 5$$

$$\bar{X} = \frac{\sum X}{n} = \frac{150.6}{30} = 5.02$$

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} = \sqrt{\frac{4.188}{30-1}} = \sqrt{0.144} = 0.38$$

$$t_{29} = \frac{\bar{X} - \mu_0}{s / \sqrt{n}} = \frac{5.02 - 5}{0.38 / \sqrt{30}} = 0.29 \quad t_{29,0.025} = 2.045 \quad |t_{29}| < |t_{29,0.025}| \quad \text{Άρα ότι η } H_0 \text{ γίνεται δεκτή}$$

# Έλεγχος Υπόθεσης Ποσοστού: κριτήριο Z

$$H_0: \pi = \pi_0$$

με  $H_1: \pi \neq \pi_0$  για δικατάληκτο κριτήριο

ή  $H_1: \pi < \pi_0$  ή  $H_1: \pi > \pi_0$  για μονοκατάληκτο κριτήριο

$$p = X/n$$

$$s_p = \sqrt{\frac{\pi_0(1 - \pi_0)}{n}}$$

κριτήριο Z

$$Z = \frac{p - \pi_0}{s_p}$$

Για δικατάληκτο κριτήριο όταν  $|Z| < |Z_{\alpha/2}|$  ή  $|Z| < |Z_{1-\alpha/2}|$  η  $H_0$  γίνεται δεκτή

Για μονοκατάληκτο κριτήριο προς τα κάτω όταν  $Z > Z_\alpha$  η  $H_0$  γίνεται δεκτή

Για μονοκατάληκτο κριτήριο προς τα πάνω όταν  $Z < Z_{1-\alpha}$  η  $H_0$  γίνεται δεκτή

## Παράδειγμα:

Το Υπουργείο Αγροτικής Ανάπτυξης & Τροφίμων ενόψει της εφαρμογής της νέας ΚΑΠ θέλει να διαπιστώσει πόσοι από τους δικαιούχους άμεσων ενισχύσεων ξεπερνούν το όριο των 5.000 €. Παλαιότερη μελέτη της Δ/σης Αγροτικής Πολιτικής είχε καταλήξει στο συμπέρασμα ότι περίπου το 15% των δικαιούχων απολαμβάνουν ενισχύσεις μεγαλύτερες από 5.000 €. Έμπειρα στελέχη όμως του Υπουργείου εισηγούνται στον Υπουργό ότι το συγκεκριμένο ποσοστό έχει φτάσει στο 25% με τάσεις ανόδου λόγω της επιτυχημένης εφαρμογής των παλαιότερων διορθωτικών προγραμμάτων. Επειδή η βάση δεδομένων των πληρωμών του ΟΠΕΚΕΠΕ δεν είναι κατάλληλα ενημερωμένη, η ηγεσία του Υπουργείου, προκειμένου να ελέγξει τους ισχυρισμούς των στελεχών, αποφασίζει να διεξάγει μία μικρής εμβέλειας έρευνα σε ένα τυχαίο δείγμα 700 δικαιούχων. Σε αυτό το τυχαίο δείγμα, 196 δικαιούχοι απάντησαν ότι απολαμβάνουν άμεσες ενισχύσεις μεγαλύτερες από 5.000 €. α). Να διατυπωθεί η μηδενική και η εναλλακτική υπόθεση β) Η άποψη των στελεχών του υπουργείου είναι σωστή ή λάθος;

A. Διατύπωση ερευνητικών υποθέσεων:

$$H_0: \pi=0.25 \text{ και } H_1: \pi>0.25$$

B.  $p=196/700=0.28$

$$s_p = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{0.25(1-0.25)}{700}} = 0.0164$$

$$Z = \frac{p - \pi_0}{s_p} = \frac{0.28 - 0.25}{0.0164} = 1.83$$

$\alpha=0.05$        $Z_{1-0.05}=Z_{0.95}=+1.645$

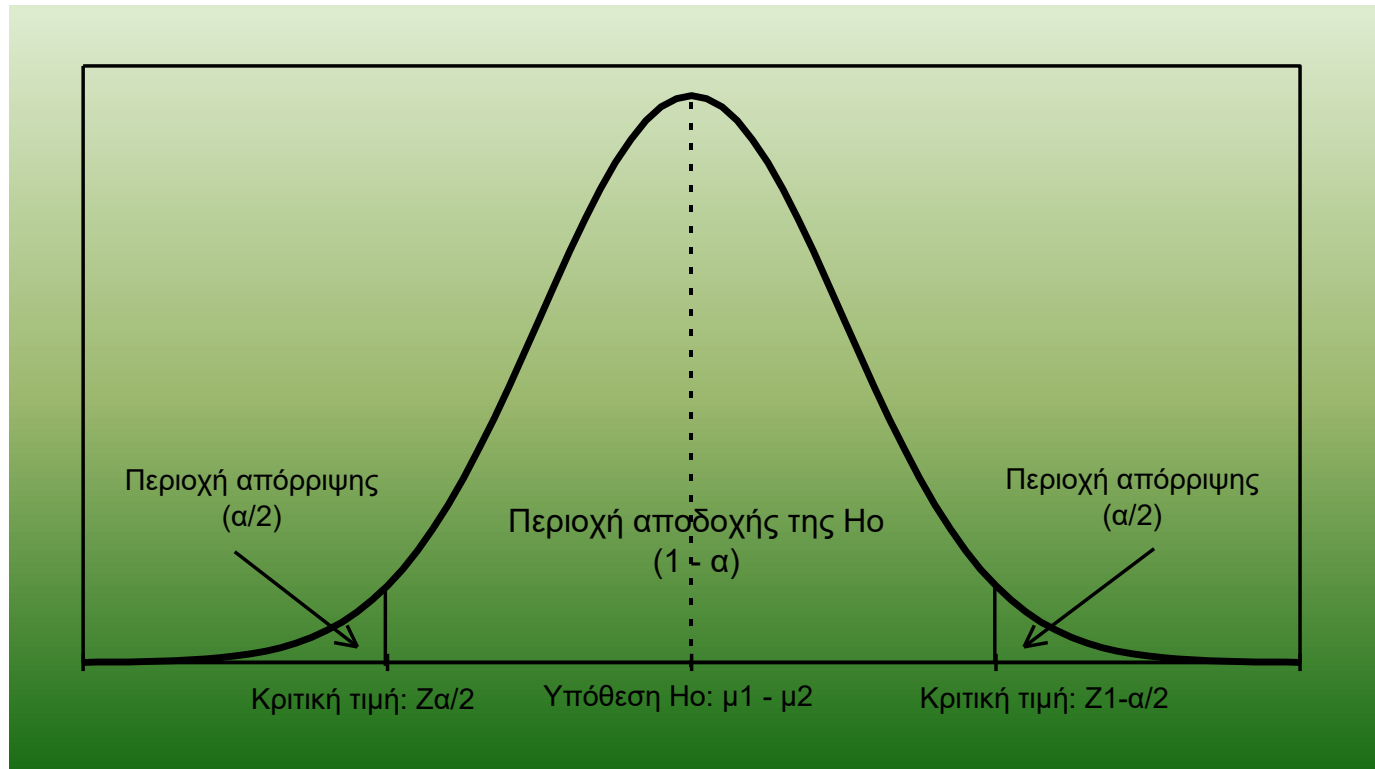
**Επειδή  $Z > Z_{1-\alpha}$  η  $H_0$  απορρίπτεται**

# Έλεγχος Διαφοράς Δύο Μέσων

Είναι έλεγχος υποθέσεων μεταξύ δύο δειγμάτων.

Σκοπός του κριτηρίου είναι να ελέγξουμε την διαφορά μεταξύ των μέσων δύο πληθυσμών με βάση τις παρατηρήσεις δύο ανεξαρτήτων δειγμάτων που έχουν επιλεγεί από τους δύο πληθυσμούς.

## Κατανομή Δειγματοληψίας της Διαφοράς δύο Μέσων και Περιοχές Απόρριψης και Αποδοχής της $H_0$



	<u>Δικατάληκτο κριτήριο</u>	<u>Μονοκατάληκτο κριτήριο</u>	<u>Μονοκατάληκτο κριτήριο</u>
$H_0: \mu_1 - \mu_2 = \theta_0$	$H_0: \mu_1 - \mu_2 = 0$	$H_0: \mu_1 - \mu_2 = 0$	$H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 - \mu_2 \neq \theta_0$	$H_1: \mu_1 - \mu_2 \neq 0$	$H_1: \mu_1 - \mu_2 > 0$	$H_1: \mu_1 - \mu_2 < 0$

Για μεγάλα δείγματα ( $n_1, n_2 > 30$ )

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad Z = \frac{(\bar{X}_1 - \bar{X}_2) - \theta_0}{s_{\bar{X}_1 - \bar{X}_2}}$$

Για μικρά δείγματα ( $n_1, n_2 < 30$ )



Χρησιμοποιούμε το κριτήριο *t student*

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

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

$$t_{n_1 + n_2 - 2} = \frac{(\bar{X}_1 - \bar{X}_2) - \theta_0}{s_{\bar{X}_1 - \bar{X}_2}}$$

## Παράδειγμα

Το ΕΘΙΑΓΕ θέλει να ελέγξει εάν το γεωργικό οικογενειακό εισόδημα (ΓΟΕ) των γεωργών με υψηλό μορφωτικό επίπεδο διαφέρει από το ΓΟΕ των γεωργών με χαμηλό μορφωτικό επίπεδο. Δύο τυχαία δείγματα 150 γεωργών με υψηλό ΓΟΕ και 200 γεωργών με χαμηλό ΓΟΕ έδωσαν τα εξής αποτελέσματα:

	<u>Υψηλό Μορφωτικό</u> <u>Επίπεδο</u>	<u>Χαμηλό Μορφωτικό</u> <u>Επίπεδο</u>
Μέγεθος δείγματος	n=150	n=200
Μέσος	105 χιλ €	115 χιλ €
Τυπική Απόκλιση	23 χιλ €	35 χιλ. €

$$H_0: \mu_1 - \mu_2 = 0 \quad H_1: \mu_1 - \mu_2 \neq 0$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{23^2}{150} + \frac{35^2}{115}} = 3.107 \quad Z = \frac{(\bar{X}_1 - \bar{X}_2)}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{(105 - 115)}{3,107} = -3.22$$

Για επίπεδο σημαντικότητας  $\alpha=0.05$   $Z_{\alpha/2}=Z_{0.025}=-1.96$

$Z < Z_{0.025}$  άρα η  $H_0$  απορρίπτεται και γίνεται δεκτή η εναλλακτική υπόθεση. Δηλαδή το ΓΟΕ διαφέρει ανάλογα με το μορφωτικό επίπεδο του γεωργού.

## Παράδειγμα

Ένας γεωπόνος μελετά την απόδοση βαμβακιού σε δύο περιοχές κατά την περασμένη καλλιεργητική περίοδο. Το πρόβλημα που τον απασχολεί είναι εάν η καλλιέργεια βαμβακιού στην πρώτη περιοχή απέδωσε όσο και στην δεύτερη. Ο μελετητής συλλέγει ένα τυχαίο δείγμα 18 εκμεταλλεύσεων στην μία περιοχή και ένα δείγμα 20 εκμεταλλεύσεων στην δεύτερη περιοχή. Τα δεδομένα έχουν ως εξής:

	<u>Περιοχή I</u>	<u>Περιοχή II</u>		
Μέγεθος δείγματος	n=18	n=20		
Μέσος	3.5 ton/ha	2.7 ton/ha	$H_0: \mu_1 - \mu_2 = 0$	$H_1: \mu_1 - \mu_2 \neq 0$
Τυπική Απόκλιση	1.3 ton/ha	1.5 ton/ha		

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{(18 - 1)1.3^2 + (20 - 1)1.5^2}{(18 - 1) + (20 - 1)} = 1.9856$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{1.9856 \left( \frac{1}{18} + \frac{1}{20} \right)} = 0.4578$$

$$t_{36} = \frac{(3.5 - 2.7)}{0.4578} = 1.747$$

Για επίπεδο σημαντικότητας  $\alpha=0.05$   $t_{36, \alpha/2} = Z_{36, 0.0255} = 2.028$

άρα  $|t_{36}| < |t_{v, \alpha/2}|$  όπου σημαίνει ότι δεχόμαστε την  $H_0$  δηλαδή δεν υπάρχει διαφορά στην απόδοση βαμβακιού μεταξύ των δύο περιοχών.

# Παρατηρούμενα και Πειραματικά δεδομένα

- Ερευνητική υπόθεση: Η κατανάλωση δημητριακών με υψηλή περιεκτικότητα σε φυτικές ίνες για πρωινό, καταναλώνουν λιγότερες θερμίδες για μεσημεριανό.
- Δείγμα 150 ατόμων
- Έλεγχος διαφοράς μέσω των πληθυσμών.
- $H_0: \mu_1 - \mu_2 = 0$        $H_1: \mu_1 - \mu_2 < 0$
- Ο έλεγχος για ίσες διασπορές έδωσε  $F=0.3845$  και τιμή  $p=0.0008$



# ...αποτελέσματα Excel

	A	B	C
1	t-Test: Two-Sample Assuming Unequal Variances		
2			
3		<i>Consumers</i>	<i>Nonconsumers</i>
4	Mean	604.02	633.23
5	Variance	4103	10670
6	Observations	43	107
7	Hypothesized Mean Difference	0	
8	df	123	
9	t Stat	-2.09	
10	P(T<=t) one-tail	0.0193	
11	t Critical one-tail	1.6573	
12	P(T<=t) two-tail	0.0386	
13	t Critical two-tail	1.9794	

Βάση των αποτελεσμάτων συμπεραίνουμε ότι αυτοί που καταναλώνουν για πρωινό δημητριακά με φυτικές ίνες, καταναλώνουν λιγότερες θερμίδες το μεσημεριανό.

**Είναι αυτή η μοναδική ερμηνεία;**

# Πειραματικά δεδομένα

- Πρέπει να εξασφαλίσουμε ότι οι δύο πληθυσμοί έχουν παρόμοια χαρακτηριστικά.
- Όσο μεγαλύτερο είναι το δείγμα τόσο πιθανότερη είναι η ομοιομορφία των πληθυσμών.
- Τα πειραματικά δεδομένα απαιτούν περισσότερη προετοιμασία και σημαντικό μεγαλύτερο κόστος.
- Πολλές φορές δεν είναι δυνατό να προκύψει ελεγχόμενο περιβάλλον.
- Η διαφορά μεταξύ παρατηρούμενων και πειραματικών δεδομένων επηρεάζει την ερμηνεία κάθε στατιστικού αποτελέσματος και όχι μόνο τη διαφορά των μέσων δύο πληθυσμών.

---

# Διαφορά Δύο Μέσων: Σύγκριση κατά Ζεύγη

Προηγουμένως, όταν συγκρίναμε δύο πληθυσμούς τα δύο δείγματα που εξετάζαμε ήταν ανεξάρτητα.

Εάν παρόλο αυτά, κάποιο πείραμα σχεδιαστεί με τέτοιο τρόπο όπου κάθε παρατήρηση του ενός δείγματος να αντιστοιχεί σε μία παρατήρηση του άλλου δείγματος, αυτό λέγεται *πείραμα σύγκρισης κατά ζεύγη*

---

---

# Παράδειγμα

Ο Διευθυντής μιας εταιρείας προσφοράς καριέρας ενδιαφέρεται να ερευνήσει τις προσφορές που δέχονται απόφοιτοι MBA.

Ιδιαίτερα, θέλει να δει εάν κάτοχοι MBA που προέρχονται από το Οικονομικά Τμήματα αμείβονται περισσότερο από κατόχους MBA που προέρχονται από το Πολυτεχνείο .

Σε μία προκαταρκτική έρευνα επέλεξε ένα δείγμα 50 κατόχων MBA όπου οι μισοί προέρχονταν από Οικονομικά Τμήματα και οι υπόλοιποι από Πολυτεχνικές σχολές.

Μπορούμε από τα δεδομένα να συμπεράνουμε ότι οι κάτοχοι MBA που προέρχονται από οικονομικές σπουδές δέχονται καλύτερες προσφορές σε σύγκριση με τους κατόχους MBA που προέρχονται από τεχνολογικές σπουδές;

---

# Παράδειγμα

## Αναγνωρίζω την παράμετρο

Η παράμετρος είναι η διαφορά μεταξύ δύο μέσων ( $\mu_1 =$  μέσος μηνιαίος προσφερόμενος μισθός για απόφοιτους οικονομικών σχολών και  $\mu_2 =$  μέσος μηνιαίος προσφερόμενος μισθός για αποφοίτους Τμημάτων Marketing) .

Επειδή ενδιαφερόμαστε να διαπιστώσουμε εάν κάτοχοι MBA από οικονομικές σχολές έχουν προσφορές με υψηλότερους μέσους μισθούς, η εναλλακτική υπόθεση θα είναι **μεγαλύτερη από...**

Υπολογισμός της στατιστικής F για έλεγχο ίσων διακυμάνσεων καταδεικνύει ότι ο έλεγχος των μέσων θα γίνει για ίσες διακυμάνσεις.

$$H_0 : (\mu_1 - \mu_2) = 0$$

$$H_1 : (\mu_1 - \mu_2) > 0$$

# Παράδειγμα Αποτελέσματα Excel

Υπολογισμός

t-Test: Two-Sample Assuming Equal Variances		
	<i>Finance</i>	<i>Marketing</i>
Mean	65,624	60,423
Variance	360,433,294	262,228,559
Observations	25	25
Pooled Variance	311,330,926	
Hypothesized Mean Difference	0	
df	48	
t Stat	1.04	
P(T<=t) one-tail	0.1513	
t Critical one-tail	1.6772	
P(T<=t) two-tail	0.3026	
t Critical two-tail	2.0106	

Η τιμή του ελέγχου είναι ( $t = 1.04$ ) και η τιμή  $p$  0.1513.

Συνεπώς τα δεδομένα μας δεν μπορούν να στηρίξουν την υπόθεση ότι οι μέσες προσφορές που δέχονται οι κάτοχοι MBA με οικονομικές σπουδές είναι υψηλότερες από αυτές των κατόχων MBA με τεχνολογικές σπουδές.

# Παράδειγμα

Επαναλαμβάνουμε το πείραμα με τον εξής τρόπο

Αποκτούμε πρόσβαση στις βαθμολογίες των κατόχων MBA και βρίσκουμε τον μέσο της βαθμολογίας του πτυχίου (οικονομικής και τεχνολογικής κατεύθυνσης).

Οι βαθμοί έχουν μέγιστο 10 και ελάχιστο 5 συνεπώς χωρίζουμε το διάστημα 5 σε 25 ζώνες βαθμολογίας 0.02.

Στη συνέχεια επιλέγουμε τυχαία έναν απόφοιτο κάθε κατεύθυνσης από κάθε ζώνη βαθμολογίας.

Σε αυτό το παράδειγμα το πείραμα έχει σχεδιαστεί με τέτοιο τρόπο ώστε κάθε παρατήρηση του ενός δείγματος να αντιστοιχεί με μία παρατήρηση από το άλλο δείγμα. Η αντιστοιχία έχει γίνει με κριτήριο την βαθμολογία των αποφοίτων.

Για το έλεγχο της υπόθεσης αρχικά υπολογίζουμε την διαφορά μεταξύ των παρατηρήσεων κάθε ζεύγους.

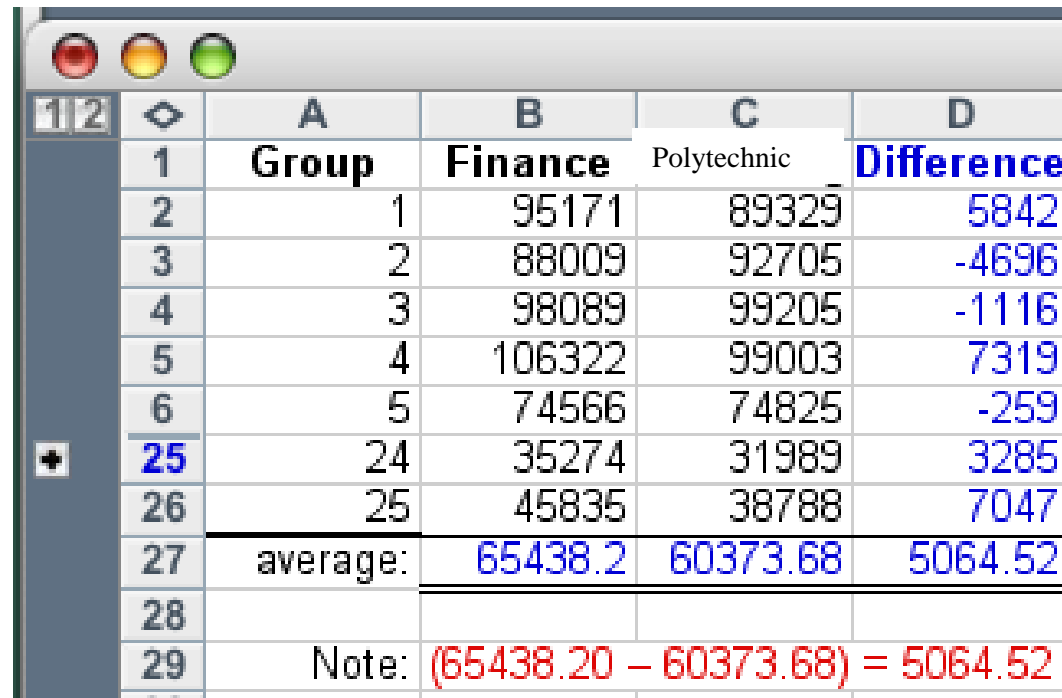
---



# Παράδειγμα

## Αναγνωρίζω την παράμετρο

Οι αριθμοί με την μαύρη γραφή είναι τα δεδομένα των προσφορών για αρχικό μισθό ενώ αυτά με την μπλε έχουν υπολογιστεί.



	A	B	C	D
1	<b>Group</b>	<b>Finance</b>	Polytechnic	<b>Difference</b>
2	1	95171	89329	5842
3	2	88009	92705	-4696
4	3	98089	99205	-1116
5	4	106322	99003	7319
6	5	74566	74825	-259
25	24	35274	31989	3285
26	25	45835	38788	7047
27	average:	65438.2	60373.68	5064.52
28				
29	Note:	(65438.20 - 60373.68) = 5064.52		

Η διαφορά των μέσων είναι ίση με τον μέσο των διαφορών, και ως εκ τούτου θα θεωρήσουμε τον «μέσο του πληθυσμού των διαφορών» ως την παράμετρο που πρέπει να ελέγξουμε :

$$\mu_1 - \mu_2 = \mu_D$$

# Έλεγχος των διαφορών

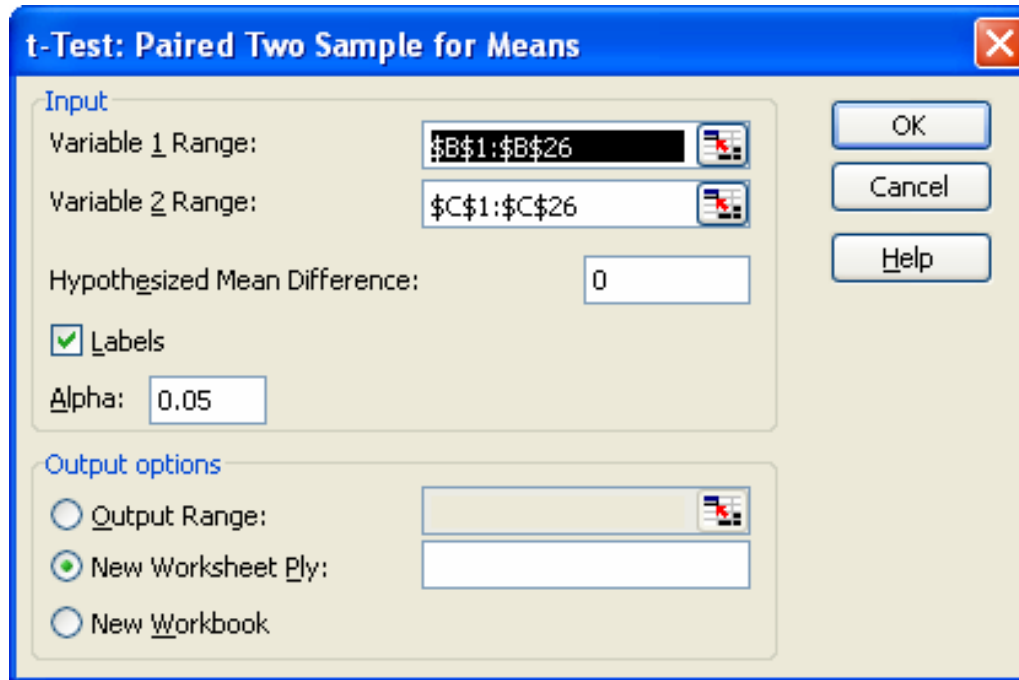
**Αναγνωρίζω την παράμετρο**

Ο έλεγχος για τον μέσο  $\mu_D$  των διαφορών δύο πληθυσμών υπολογίζεται από τον τύπο:

$$t = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n_D}}$$

Που ακολουθεί κατανομή t-Student με  $n_D - 1$  βαθμούς ελευθερίας, με την προϋπόθεση ότι ο πληθυσμός των διαφορών έχει κανονική κατανομή..

### Data, Data Analysis, t-Test: Paired Two- Sample for Means



The image shows the 't-Test: Paired Two Sample for Means' dialog box in Microsoft Excel. The dialog box has a blue title bar and a close button (X) in the top right corner. It is divided into two main sections: 'Input' and 'Output options'. In the 'Input' section, 'Variable 1 Range' is set to '\$B\$1:\$B\$26', 'Variable 2 Range' is set to '\$C\$1:\$C\$26', and 'Hypothesized Mean Difference' is set to '0'. The 'Labels' checkbox is checked, and 'Alpha' is set to '0.05'. In the 'Output options' section, the 'New Worksheet Ply' radio button is selected. On the right side of the dialog box, there are three buttons: 'OK', 'Cancel', and 'Help'.

**t-Test: Paired Two Sample for Means**

**Input**

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

Labels

Alpha:

**Output options**

Output Range:

New Worksheet Ply:

New Workbook

OK  
Cancel  
Help

# Παράδειγμα

## Υπολογισμός

	A	B	C
1	t-Test: Paired Two Sample for Means		
2			
3		<i>Finance</i>	<i>Marketing</i>
4	Mean	65,438	60,374
5	Variance	444,981,810	469,441,785
6	Observations	25	25
7	Pearson Correlation	0.9520	
8	Hypothesized Mean Difference	0	
9	df	24	
10	t Stat	3.81	
11	P(T<=t) one-tail	0.0004	
12	t Critical one-tail	1.7109	
13	P(T<=t) two-tail	0.0009	
14	t Critical two-tail	2.0639	

Η τιμή έλεγχου είναι ίση με  $t=3.81$  ενώ η κριτική τιμή  $t_{0.05, 24}=1.7109$  και η τιμή  $p$  ίση με 0.0004. Βρισκόμαστε σε περιοχή απόρριψης άρα έχουμε επαρκή στατιστικά στοιχεία για να απορρίψουμε την μηδενική υπόθεση.

Διαπιστώνουμε ότι με τη σύγκριση κατά ζεύγη μπορέσαμε να εξάγουμε από τα δεδομένα μας την πληροφορία ότι οι απόφοιτοι MBA που προέρχονται από οικονομικά Τμήματα δέχονται υψηλότερες προσφορές από αποφοίτους MBA που προέρχονται από Τμήματα. Marketing

# Έλεγχος Διαφοράς Δύο Ποσοστών: κριτήριο Z

Ελέγχουμε την διαφορά μεταξύ των ποσοστών  $\pi_1$  και  $\pi_2$  δύο πληθυσμών. Ο έλεγχος θα γίνει βάση των ποσοστών  $p_1$  και  $p_2$  που έχουν προκύψει από δύο ανεξάρτητα δείγματα

Η μηδενική υπόθεση είναι ότι η διαφορά των ποσοστών μεταξύ δύο πληθυσμών έχει μια σταθερή τιμή  $\theta_0$ . Εάν το  $\theta_0$  είναι ίσο με μηδέν οι πιθανοί έλεγχοι είναι:

Δικατάληκτο κριτήριο

$$H_0: \pi_1 - \pi_2 = 0$$

$$H_1: \pi_1 - \pi_2 \neq 0$$

Μονοκατάληκτο κριτήριο

$$H_0: \pi_1 - \pi_2 = 0$$

$$H_1: \pi_1 - \pi_2 > 0$$

Μονοκατάληκτο κριτήριο

$$H_0: \pi_1 - \pi_2 = 0$$

$$H_1: \pi_1 - \pi_2 < 0$$

Κοινό ή συνδυασμένο (combined) ποσοστό  $p_c$ : 
$$p_c = \frac{X_1 + X_2}{n_1 + n_2} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$s_{p_1 - p_2} = \sqrt{p_c(1 - p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$Z = \frac{(p_1 - p_2)}{s_{p_1 - p_2}}$$

Εάν  $|Z| < |Z_{\alpha/2}|$  τότε αποδεχόμαστε την  $H_0$

## Παράδειγμα

Το τμήμα έκδοσης πιστωτικών καρτών της τράπεζας θέλει να διερευνήσει εάν η οικογενειακή κατάσταση των κατόχων πιστωτικών καρτών επιδρά στο ποσοστό των κατόχων που έχουν ανεξόφλητες ληξιπρόθεσμες υποχρεώσεις. Τα δύο τυχαία δείγματα 150 παντρεμένων και 200 ανύπαντρων κατόχων πιστωτικής κάρτας έδωσαν επίσης τα εξής αποτελέσματα:

	<u>Παντρεμένοι</u>	<u>Ανύπαντροι</u>
Μέγεθος δείγματος	$n_1=150$	$n_2=200$
Αριθμός κατόχων με ανεξόφλητες ληξιπρόθεσμες υποχρεώσεις	$X_1=4$	$X_2=9$
Εκτίμηση ποσοστού	$p_1=4/150=0.027$	$p_2=9/200=0.045$

$$H_0: \pi_1 - \pi_2 = 0$$

$$H_1: \pi_1 - \pi_2 \neq 0$$

$$p_c = \frac{X_1 + X_2}{n_1 + n_2} = \frac{4 + 9}{150 + 200} = 0.037$$

$$s_{p_1 - p_2} = \sqrt{p_c(1 - p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = \sqrt{0.037(1 - 0.037)\left(\frac{1}{150} + \frac{1}{200}\right)} = 0.02$$

$$Z = \frac{(p_1 - p_2)}{s_{p_1 - p_2}} = \frac{(0.027 - 0.045)}{0.02} = -0.90$$

Για επίπεδο σημαντικότητας  $\alpha=0.05$   
 $Z_{\alpha/2}=Z_{0.025}=-1.96$  δηλαδή  $|Z| > |Z_{0.025}|$  που σημαίνει ότι **δεχόμαστε την  $H_0$**

# Έλεγχος Διαφοράς δύο Διακυμάνσεων: κριτήριο F

Ο Έλεγχος ισότητας των διακυμάνσεων δύο πληθυσμών βασίζεται στο λόγο των δύο διακυμάνσεων και όχι στην διαφορά τους. Όταν τα δύο δείγματα προέρχονται από κανονικούς πληθυσμούς, τότε ο λόγος των διακυμάνσεων ακολουθεί μία γνωστή θεωρητική κατανομή που ονομάζεται **κατανομή F**.

$$F_{\nu_1 \nu_2} = \frac{S_1^2}{S_2^2}$$

Όπου  $\nu_1 = \eta_1 - 1$  και  $\nu_2 = \eta_2 - 1$

## Δικατάληκτο κριτήριο

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ ή } \sigma_1^2 / \sigma_2^2 = 1$$

$$H_1: \sigma_1^2 \neq \sigma_2^2 \text{ ή } \sigma_1^2 / \sigma_2^2 \neq 1$$

## Μονοκατάληκτο κριτήριο

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ ή } \sigma_1^2 / \sigma_2^2 = 1$$

$$H_1: \sigma_1^2 > \sigma_2^2 \text{ ή } \sigma_1^2 / \sigma_2^2 > 1$$

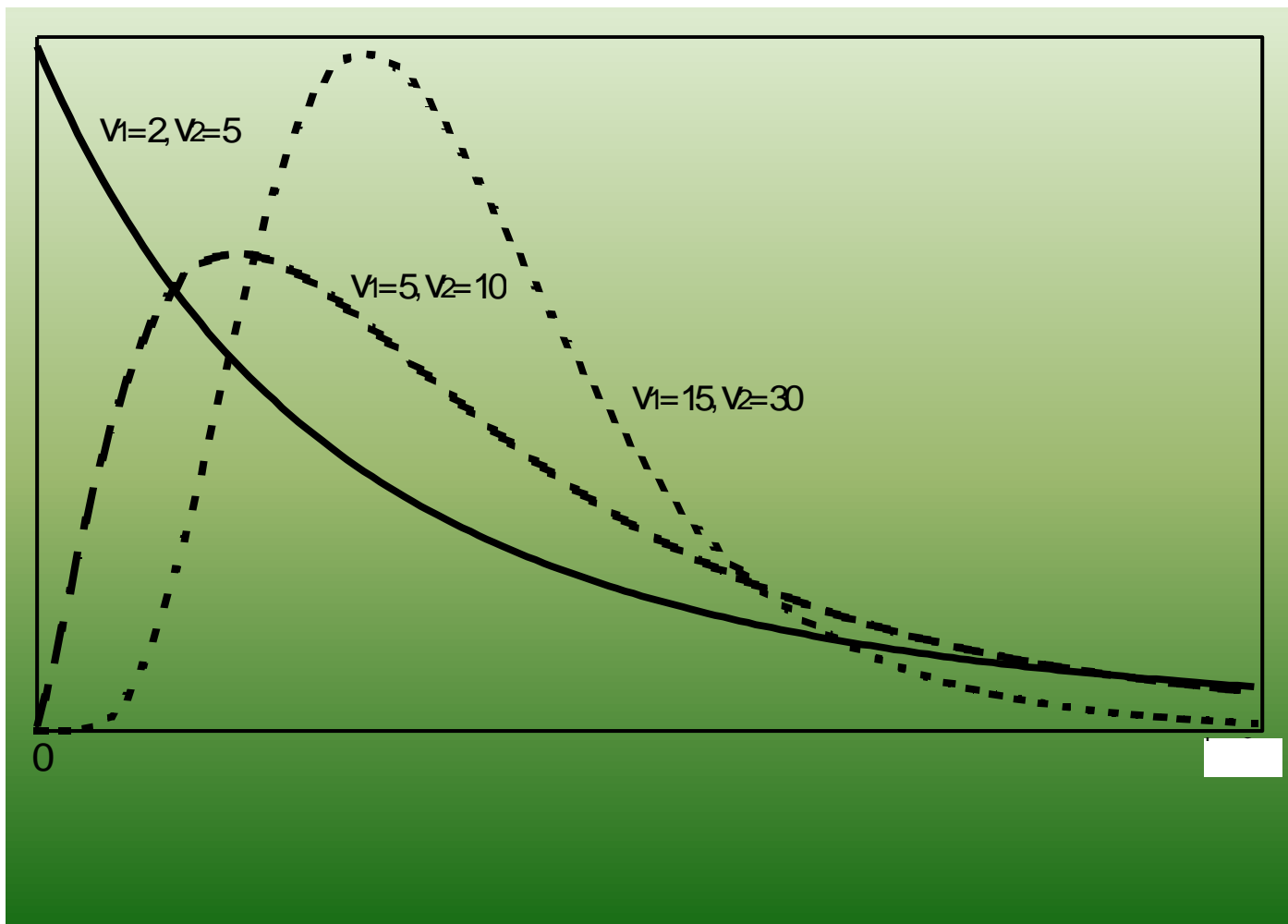
## Μονοκατάληκτο κριτήριο

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ ή } \sigma_1^2 / \sigma_2^2 = 1$$

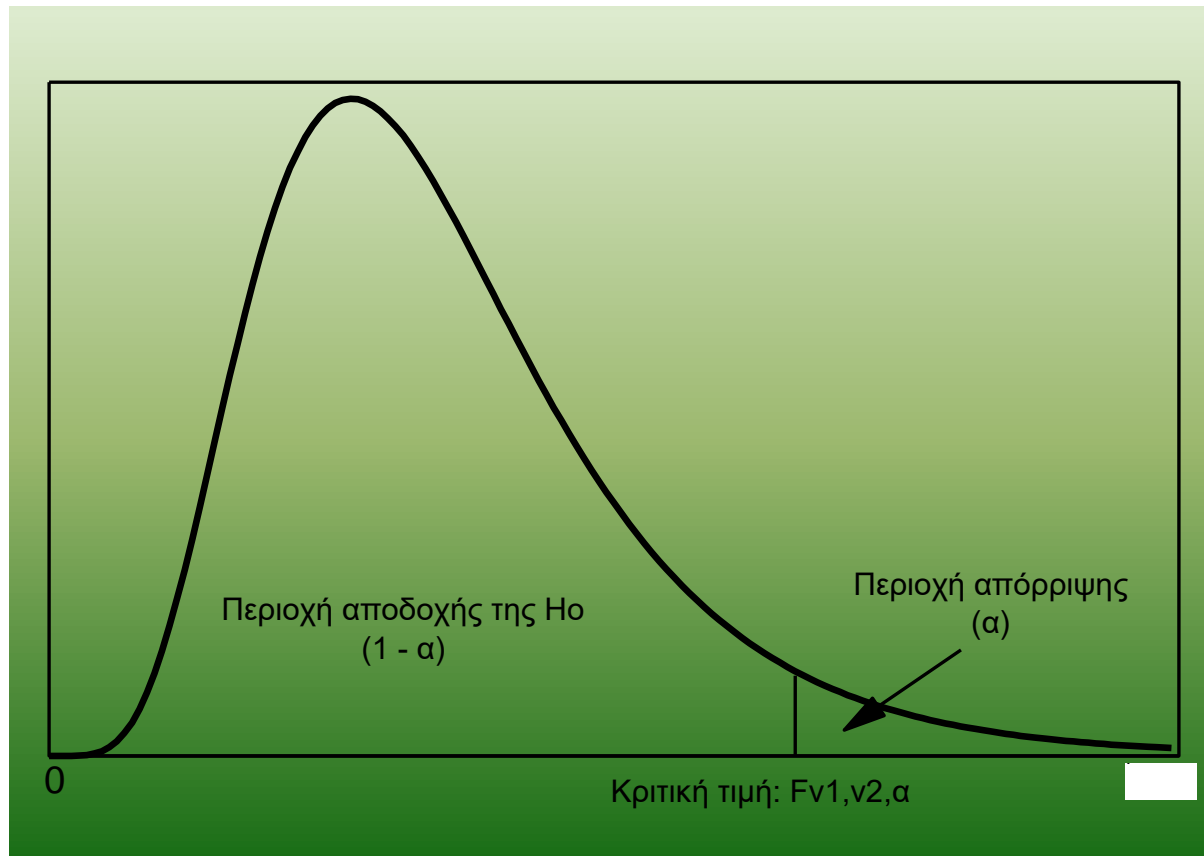
$$H_1: \sigma_1^2 < \sigma_2^2 \text{ ή } \sigma_1^2 / \sigma_2^2 < 1$$



## Κατανομές του Κριτηρίου $F_{v_1, v_2}$ για Διάφορες Τιμές των $v_1$ και $v_2$



## Κατανομή Δειγματοληψίας του Κριτηρίου $F=s_1^2/s_2^2$ και Περιοχές Απόρριψης και Αποδοχής της $H_0$



Εάν η τιμή του  $F$  είναι μεγαλύτερη από την κριτική τιμή  $F_{v_1, v_2, \alpha}$  δηλαδή  $F_{v_1, v_2} > F_{v_1, v_2, \alpha}$  τότε απορρίπτουμε την  $H_0$  και γίνεται δεκτή η  $H_1$

# Έλεγχος Ανεξαρτησίας Ιδιοτήτων: κριτήριο $\chi^2$

$X_1$ : ο αριθμός των παρατηρήσεων που έχουν το υπο εξέταση χαρακτηριστικό

$p_1 = X_1/n_1$ : το ποσοστό των παρατηρήσεων

$n_1 - X_1$ : ο αριθμός των παρατηρήσεων που δεν έχουν το συγκεκριμένο χαρακτηριστικό

Το κριτήριο  $\chi^2$  ελέγχει την ίδια υπόθεση με το κριτήριο Z αλλά βασίζεται στις συχνότητες εμφάνισης ή μή του χαρακτηριστικού δηλ. Τα  $X_1$ ,  $X_2$ ,  $n_1 - X_1$  και  $n_2 - X_2$  και όχι τις τιμές των ποσοστών  $p_1$  και  $p_2$ . Επίσης ο συγκεκριμένο κριτήριο μας επιτρέπει να επεκτείνουμε τον έλεγχο διαφοράς μεταξύ των ποσοστών για περισσότερους από δύο πληθυσμούς

$H_0: \pi_1 = \pi_2 = \dots = \pi_c$

Ελέγχουμε εάν οι πραγματικές ή παρατηρούμενες συχνότητες ( $f_o$ ) διαφέρουν στατιστικά σημαντικά από τις αναμενόμενες ( $f_e$ ). Εάν οι διαφορές είναι μικρές (στατιστικά ασήμαντες), τότε γίνεται δεκτή η  $H_0$  δηλ. Τα δύο χαρακτηριστικά είναι ανεξάρτητα.. Εάν η  $H_0$  απορριφθεί, τότε οι πραγματικές συχνότητες δεν συμβιβάζονται με την υπόθεση της ανεξαρτησίας.

$$\chi^2_v = \sum \frac{(f_o - f_e)^2}{f_e}$$

Όπου:

$f_o$  = πραγματικές ή παρατηρούμενες συχνότητες

$f_e$  = αναμενόμενες ή θεωρητικές συχνότητες

$v$  = βαθμοί ελευθερίας: (αριθμός γραμμών-1)(αριθμός στηλών-1)

Η κατανομή του κριτηρίου  $\chi^2$  έχει σχήμα ανάλογο με εκείνο της κατανομής F και τείνει προς την κανονική κατανομή για  $v > 30$

## Παράδειγμα

Το τμήμα έκδοσης πιστωτικών καρτών της τράπεζας θέλει να διερευνήσει εάν η οικογενειακή κατάσταση των κατόχων πιστωτικών καρτών επιδρά στο ποσοστό των κατόχων που έχουν ανεξόφλητες ληξιπρόθεσμες υποχρεώσεις. Τα δύο τυχαία δείγματα 150 παντρεμένων και 200 ανύπαντρων κατόχων πιστωτικής κάρτας έδωσαν επίσης τα εξής αποτελέσματα:

Πίνακας $f_o$		Οικογενειακή κατάσταση		
		Παντρεμένοι	Ανύπαντροι	Σύνολα
Ανεξόφλητες Υποχρεώσεις	Ναι	$X_1=4$	$X_2=9$	$X=13$
	Όχι	$\frac{n_1-X_1=146}{n_1=150}$	$\frac{n_2-X_2=191}{n_2=200}$	$\frac{n-X=337}{n=350}$
Σύνολα		$n_1=150$	$n_2=200$	$n=350$

Πίνακας $f_e$		Οικογενειακή κατάσταση		
		Παντρεμένοι	Ανύπαντροι	Σύνολα
Ανεξόφλητες Υποχρεώσεις	Ναι	$\frac{150 \cdot 13}{350} = 5,6$	$\frac{200 \cdot 13}{350} = 7,4$	$X=13$
	Όχι	$\frac{150 \cdot 337}{350} = 144,4$	$\frac{200 \cdot 337}{350} = 192,6$	$n-X=337$
Σύνολα		$n_1=150$	$n_2=200$	$n=350$

$$x_v^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$f_o$	$f_e$	$(f_o - f_e)$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
4	5.6	-1.6	2.56	0.457
9	7.4	1.6	2.56	0.346
146	144.4	1.6	2.56	0.018
191	192.6	-1.6	2.56	0.013

$$x^2 = 0.834$$

$$X^2_{(2-1)(2-1)} = 0.834$$

$$X^2_{0.5, 1} = 3.841$$

$X^2_{0.5, 1} > X^2_{(2-1)(2-1)}$  άρα η  $H_0$  γίνεται δεκτή που σημαίνει ότι τα δύο χαρακτηριστικά (ή ιδιότητες) είναι ανεξάρτητα