Στατιστικός έλεγχος υποθέσεων (Μέρος 2ο)

Στατιστικός έλεγχος υποθέσεων σε επίπεδο σημαντικότητας **a** για τη διακύμανση **σ^2** ενός κανονικού πληθυσμού με ένα τυχαίο δείγμα μεγέθους **n**

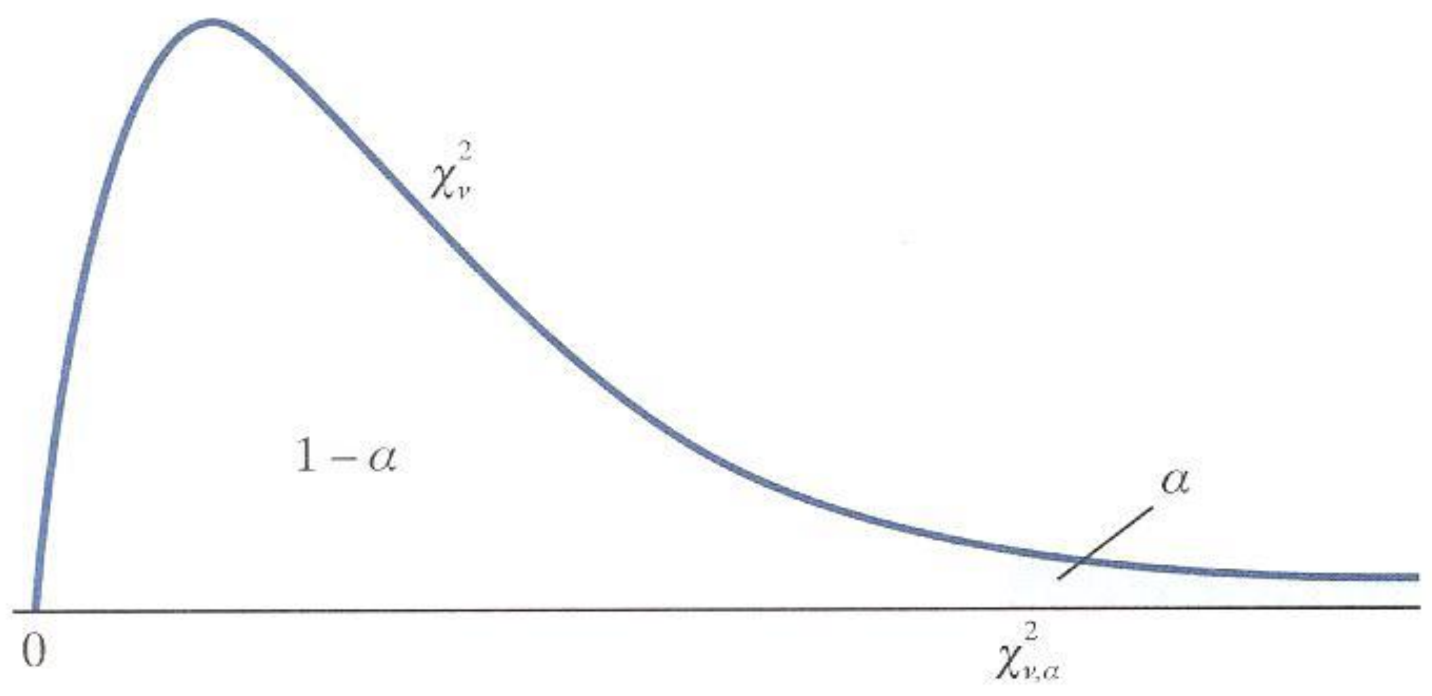
H_0: $\sigma^2 = \sigma_0^2$

Περιοχή απόρριψης της H_0

H_1: $\sigma^2 \neq \sigma_0^2$	H_1: $\sigma^2 > \sigma_0^2$	H_1: $\sigma^2 < \sigma_0^2$
$\frac{(n-1)S^2}{\sigma_0^2} \geq \chi^2_{n-1;\alpha/2}$ ή $\frac{(n-1)S^2}{\sigma_0^2} \leq \chi^2_{n-1;1-\alpha/2}$	$\frac{(n-1)S^2}{\sigma_0^2} \geq \chi^2_{n-1;a}$	$\frac{(n-1)S^2}{\sigma_0^2} \leq \chi^2_{n-1;1-a}$

Η τ.μ. $\frac{(n-1)S^2}{\sigma^2}$, για κανονικό πληθυσμό, προσεγγίζεται ικανοποιητικά από την κατανομή χ^2 με n-1 βαθμούς ελευθερίας

Κατανομή χ^2 με *n* βαθμούς ελευθερίας – καμπύλη



$$P(X > \chi^2_{n;a}) = a$$

Στατιστικός έλεγχος υποθέσεων σε επίπεδο σημαντικότητας **α** για τη διαφορά **$\mu_1 - \mu_2$** των μέσων τιμών δύο πληθυσμών με δύο ανεξάρτητα δείγματα μεγέθους **n_1** και **n_2** αντίστοιχα

H_0: $\mu_1 - \mu_2 = \delta$

- Οι διακυμάνσεις **σ_1^2** και **σ_2^2** είναι γνωστές και οι πληθυσμοί είναι κανονικοί
- Οι διακυμάνσεις **σ_1^2** και **σ_2^2** είναι γνωστές και τα μεγέθη των δειγμάτων **n_1** και **n_2** είναι μεγάλα (οτιδήποτε πληθυσμός) [*έλεγχος επιπέδου σημαντικότητας α κατά προσέγγιση*]

Περιοχή απόρριψης της H_0

H_1: $\mu_1 - \mu_2 \neq \delta$	H_1: $\mu_1 - \mu_2 > \delta$	H_1: $\mu_1 - \mu_2 < \delta$
$\lvert Z\rvert = \dfrac{\lvert\bar{X} - \bar{Y} - \delta\rvert}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \geq z_{a/2}$	$Z = \dfrac{\bar{X} - \bar{Y} - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \geq z_a$	$Z = \dfrac{\bar{X} - \bar{Y} - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \leq -z_a$

Στατιστικός έλεγχος υποθέσεων σε επίπεδο σημαντικότητας **α** για τη διαφορά **$\mu_1 - \mu_2$** των μέσων τιμών δύο πληθυσμών με δύο ανεξάρτητα δείγματα μεγέθους **n_1** και **n_2** αντίστοιχα

H_0: $\mu_1 - \mu_2 = \delta$

- Οι διακυμάνσεις **σ_1^2** και **σ_2^2** είναι άγνωστες και τα μεγέθη των δειγμάτων **n_1** και **n_2** είναι μεγάλα (οτιδήποτε πληθυσμός) [*έλεγχος επιπέδου σημαντικότητας α κατά προσέγγιση*]

Περιοχή απόρριψης της H_0

H_1: $\mu_1 - \mu_2 \neq \delta$	H_1: $\mu_1 - \mu_2 > \delta$	H_1: $\mu_1 - \mu_2 < \delta$
$\lvert Z \rvert = \dfrac{\lvert \bar{X} - \bar{Y} - \delta \rvert}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \geq z_{a/2}$	$Z = \dfrac{\bar{X} - \bar{Y} - \delta}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \geq z_a$	$Z = \dfrac{\bar{X} - \bar{Y} - \delta}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \leq -z_a$

Στατιστικός έλεγχος υποθέσεων σε επίπεδο σημαντικότητας **α** για τη διαφορά $\mu_1 - \mu_2$ των μέσων τιμών δύο πληθυσμών με δύο ανεξάρτητα δείγματα μεγέθους n_1 και n_2 αντίστοιχα

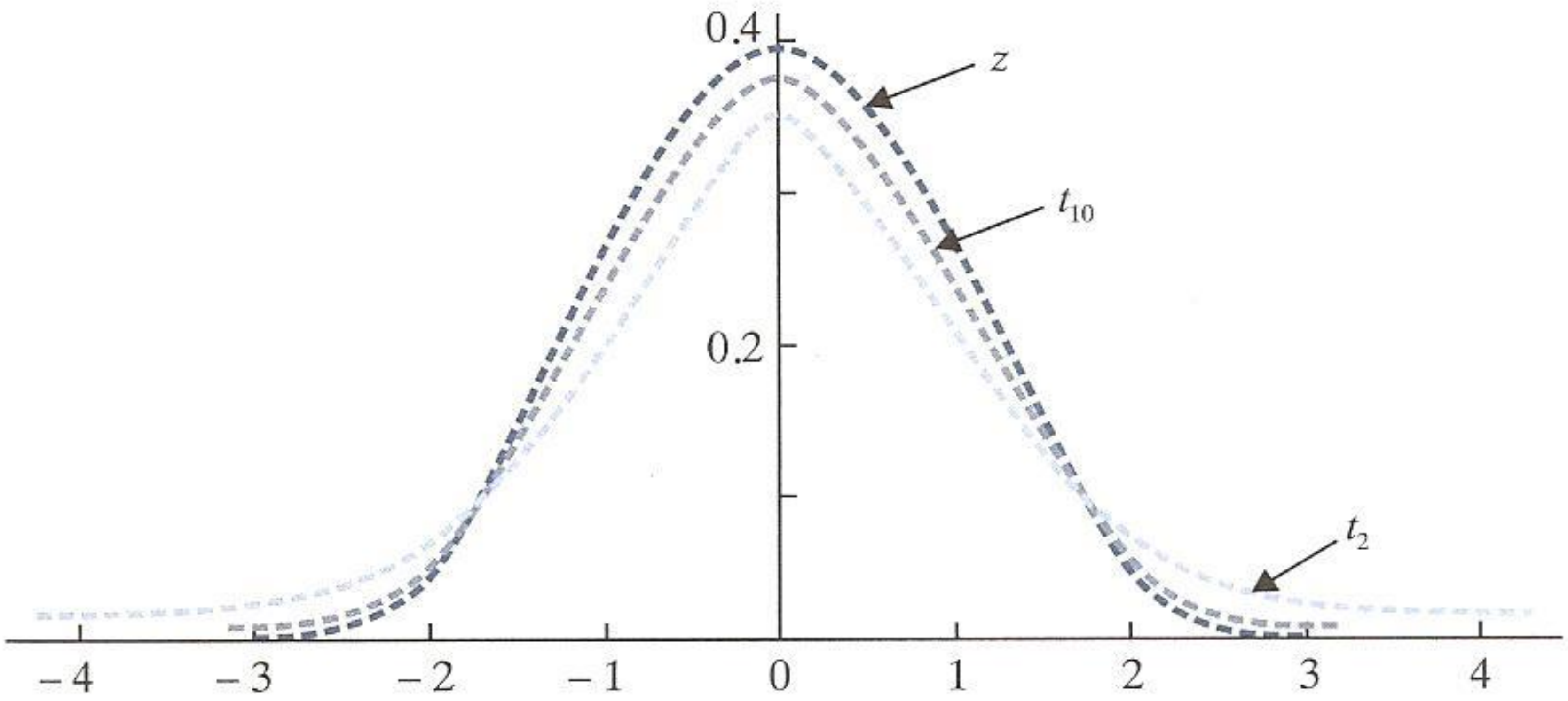
H_0: $\mu_1 - \mu_2 = \delta$

- Οι διακυμάνσεις σ_1^2 και σ_2^2 είναι άγνωστες και ίσες, οι πληθυσμοί είναι κανονικοί και τα μεγέθη των δειγμάτων τους n_1 και n_2 είναι οτιδήποτε
- $n = n_1 + n_2 - 2, \qquad S^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

Περιοχή απόρριψης της H_0

H_1: $\mu_1 - \mu_2 \neq \delta$	H_1: $\mu_1 - \mu_2 > \delta$	H_1: $\mu_1 - \mu_2 < \delta$
$\lvert T\rvert = \frac{\lvert\bar{X} - \bar{Y} - \delta\rvert}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \geq t_{n;a/2}$	$T = \frac{\bar{X} - \bar{Y} - \delta}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \geq t_{n;a}$	$T = \frac{\bar{X} - \bar{Y} - \delta}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \leq -t_{n;a}$

Η κατανομή *t* (ή Student) με *n* βαθμούς ελευθερίας – καμπύλη



Η συνάρτηση πυκνότητας της t_ν *για* $\nu = 2$ *και* $\nu = 10$ *και της* $N(0, 1)$.

Στατιστικός έλεγχος υποθέσεων σε επίπεδο σημαντικότητας **α** για τη διαφορά $\boldsymbol{\mu_1 - \mu_2}$ των μέσων τιμών δύο πληθυσμών με δύο ανεξάρτητα δείγματα μεγέθους $\mathbf{n_1}$ και $\mathbf{n_2}$ αντίστοιχα

H_0: $\mu_1 - \mu_2 = \delta$

- Όταν τα $\mathbf{n_1}$ και $\mathbf{n_2}$ είναι μικρά, ο πληθυσμοί όχι κανονικοί και οι διακυμάνσεις $\boldsymbol{\sigma_1^2}$ και $\boldsymbol{\sigma_2^2}$ είτε γνωστές είτε άγνωστες (ίσες ή όχι) δεν μπορούμε να πραγματοποιήσουμε στατιστικό έλεγχο υποθέσεων σε επίπεδο σημαντικότητας α

Στατιστικός έλεγχος υποθέσεων σε επίπεδο σημαντικότητας **a** για τη διαφορά δύο διωνυμικών ποσοστών **p_1** και **p_2** με δύο ανεξάρτητα δείγματα μεγέθους **n_1** και **n_2** αντίστοιχα

H_0: $p_1 = p_2$

- $n_i\widehat{p_i} \geq 5$ και $n_i(1-\widehat{p_i}) \geq 5$, i =1, 2
- $\widehat{p_i}$ είναι το ποσοστό επιτυχιών στο δείγμα n_i, i =1, 2
- $\hat{p} = \frac{n_1\widehat{p_1}+n_2\widehat{p_2}}{n_1+n_2}$

Περιοχή απόρριψης της H_0

H_1: $p_1 \neq p_2$	H_1: $p_1 > p_2$	H_1: $p_1 < p_2$
$\frac{\lvert\widehat{p_1}-\widehat{p_2}\rvert}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \geq z_{a/2}$	$\frac{(\widehat{p_1}-\widehat{p_2})}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \geq z_a$	$\frac{(\widehat{p_1}-\widehat{p_2})}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \leq -z_a$

Στατιστικός έλεγχος υποθέσεων σε επίπεδο σημαντικότητας **α** για την ισότητα των διακυμάνσεων $\boldsymbol{\sigma_1^2}$ και $\boldsymbol{\sigma_2^2}$ δύο κανονικών πληθυσμών με δύο ανεξάρτητα δείγματα μεγέθους $\mathbf{n_1}$ και $\mathbf{n_2}$ αντίστοιχα

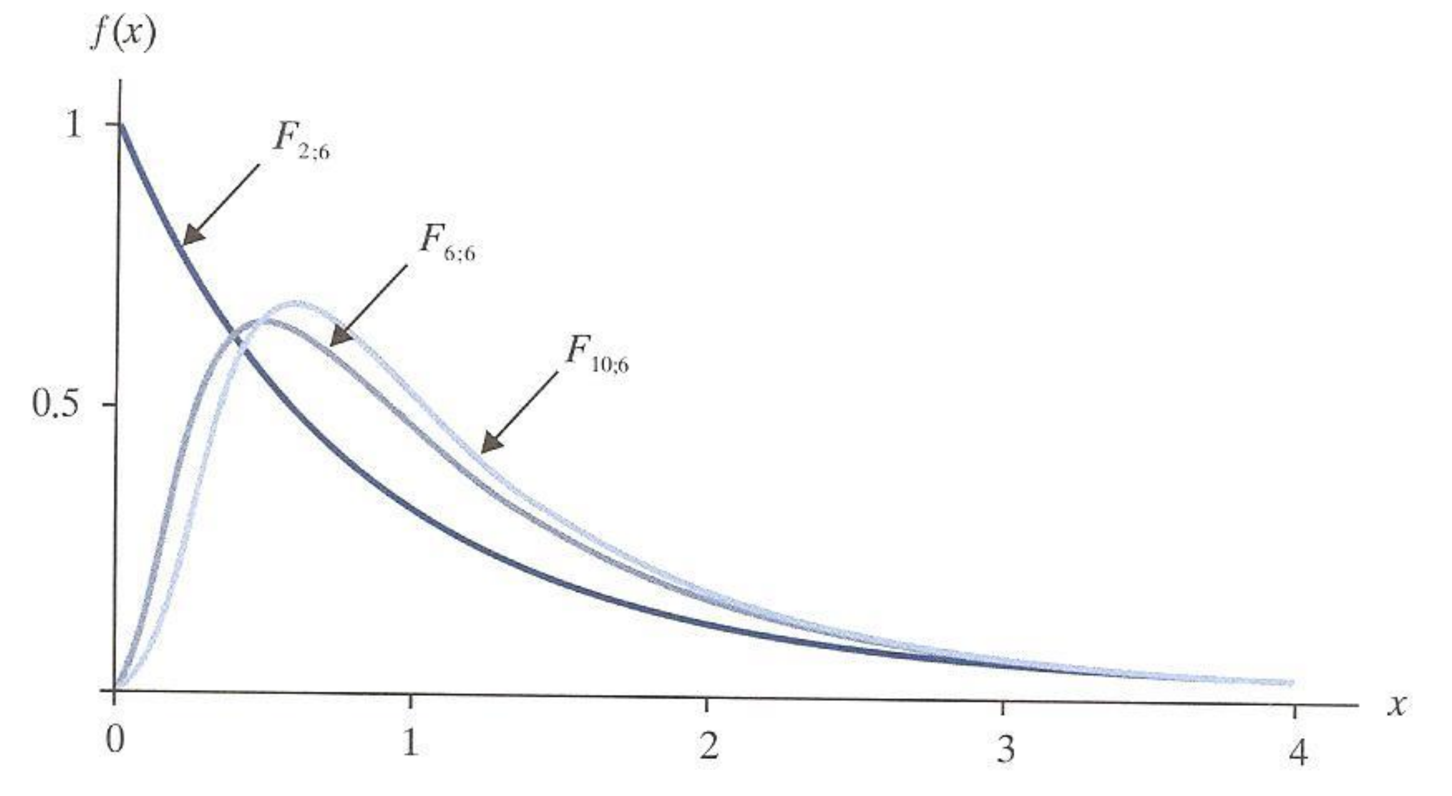
$\mathbf{H_0: \sigma_1^2 = \sigma_2^2}$

Περιοχή απόρριψης της $\mathbf{H_0}$

$H_1: \sigma_1^2 \neq \sigma_2^2$	$H_1: \sigma_1^2 > \sigma_2^2$	$H_1: \sigma_1^2 < \sigma_2^2$
$\frac{s_1^2}{s_2^2} \geq F_{n_1-1;n_2-1;a/2}$ ή $\frac{s_1^2}{s_2^2} \leq F_{n_1-1;n_2-1;1-a/2}$	$\frac{s_1^2}{s_2^2} \geq F_{n_1-1;n_2-1;\alpha}$	$\frac{s_1^2}{s_2^2} \leq F_{n_1-1;n_2-1;\alpha}$

Η τ.μ. $\frac{s_1^2}{s_2^2}$, για κανονικούς πληθυσμούς, προσεγγίζεται ικανοποιητικά από την κατανομή F με n_1-1 και n_2-1 βαθμούς ελευθερίας

Η κατανομή *F* με *n* και *m* βαθμούς ελευθερίας – καμπύλη



Η συνάρτηση πυκνότητας της $F_{v;m}$ για $v = 2,6,10$ και $m = 6$.

Παράδειγμα 3

- Επιθυμούμε να ελέγξουμε αν υπάρχει διαφορά στη μέση τιμή πώλησης ενός είδους καυσίμου μεταξύ της Αθήνας και της Θεσσαλονίκης (μια δεδομένη ημέρα)
- Για το σκοπό αυτό καταγράφονται οι τιμές πώλησης (σε λεπτά του ευρώ) ενός λίτρου του συγκεκριμένου καυσίμου σε 15 και 10 τυχαία επιλεγμένα πρατήρια βενζίνης από την Αθήνα και Θεσσαλονίκη αντίστοιχα:

Αθήνα	74,6	75,4	83,0	82,2	81,7	80,0	77,3	84,5	79,2	79,9	84,0	79,3	79,0	82,0	77,0
Θεσσαλ.	82,9	86,1	78,8	82,1	80,6	82,0	79,5	86,4	87,0	79,3					

- Να ελέγξετε αν υπάρχει διαφορά στις μέσες τιμές πώλησης μεταξύ των δυο πόλεων
- Ο έλεγχος να γίνει σε επίπεδο σημαντικότητας 5% υποθέτοντας ότι οι τιμές κατανέμονται κανονικά