

- Ορισμός 1: Στατιστική υπόθεση είναι μία πρόταση για έναν πληθυσμό, που συνήθως εκφράζεται σε μία σχέση:
 - για κάποια παράμετρο του πληθυσμού (π.χ. μέση τιμή)
 - σχετικά με την κατανομή που ακολουθεί ένας πληθυσμός (π.χ. αν ακολουθεί κανονική κατανομή)
 - για σύγκριση δειγμάτων (π.χ. αν διαφέρουν ως προς τη μέση τιμή ή αν συσχετίζονται τα δείγματα)
 - κ.α.
- Ορισμός 2: Στατιστικό test TS είναι μία στατιστική συνάρτηση, η τιμή της οποίας προσδιορίζεται από ένα δείγμα (ή και περισσότερα) και με βάση την τιμή της δεχόμαστε την μηδενική υπόθεση H_0 ή την απορρίπτουμε.

• Είδη σφαλμάτων:

	Απόδοχή της H_0	Απορρίψη της H_0
Αληθής η H_0	Ορθή απόφαση (πιθαν. $1-\alpha$)	Σφάλμα τύπου I (πιθαν. α)
Αληθής η H_1	Σφάλμα τύπου II (πιθαν. β)	Ορθή απόφαση (πιθαν. $1-\beta$)

Η ποσότητα α ονομάζεται επίπεδο σημαντικότητας (significance level)

* ΠΟΛΥ ΣΗΜΑΝΤΙΚΟ ΣΧΟΛΙΟ

Όταν επιλέγουμε μικρή τιμή για το α , δεχόμαστε "μεγάλες" αποκλίσεις της τιμής του στατιστικού test, συνεπώς η πιθανότητα να αποδεχθούμε την H_0 είναι μεγάλη. Ωστόσο, αν για μικρό α απορρίψουμε την H_0 , είναι αρκετά βέβαιο ότι η H_1 είναι αληθής.

Λόγω των παραπάνω, αν για μικρό α απορρίψουμε την H_0 , αλλά παρόλα αυτά η H_0 είναι αληθής, τότε οδηγούμαστε στο Σφάλμα τύπου I.

Δηλαδή, έχουμε απορρίψει την H_0 και είμαστε πολύ βέβαιοι ότι ισχύει η H_1 , ενώ στην πραγματικότητα δεν ισχύει.

Από την άλλη, αν για μικρό α αποδεχθούμε την H_0 , αλλά παρόλα αυτά

Από την άλλη, αν για μικρό α αποδεχθούμε την H_0 , αλλά πορτοκάλι αυτό η H_1 είναι αληθής, τότε οδηγούμαστε στο Σφάλμα τύπου II. Δηλαδή, αποδεχόμαστε την H_0 με μικρή βεβαιότητα, λόγω της μικρής τιμής του α , δίδει δεχόμαστε μεχάδες αποδείξεις ότι τιμή του στατιστικού test, ενώ στην πραγματικότητα ισχύει η H_1 .

Με βάση τα παραπάνω:

- Για μικρό α (που είναι το σύνθημα) το Σφάλμα Τύπου I είναι πιο σημαντικό από το Σφάλμα Τύπου II
- $\alpha \downarrow \Rightarrow \text{ΣΦ. I} \downarrow$ και $\text{ΣΦ. II} \uparrow$
 $\alpha \uparrow \Rightarrow \text{ΣΦ. I} \uparrow$ και $\text{ΣΦ. II} \downarrow$

Άρα, η επιλογή κατάλληλης τιμής για το α είναι πολύ κρίσιμη!

- p-value: Καλούμε p-value ενός test την μικρότερη τιμή για την οποία απορρίπτουμε την H_0 .
 - Μικρή p-value $\Rightarrow$ απόρριψη της H_0
 - Μεγάλη p-value $\Rightarrow$ όχι απόρριψη της H_0
 - Στην πράξη, πρώτα υπολογίζουμε την p-value και μετά επιλέγουμε κατάλληλο επίπεδο σημαντικότητας.
 - Αν $p\text{-value} < \alpha \Rightarrow$ απόρριψη H_0
 - Αν $p\text{-value} \geq \alpha \Rightarrow$ όχι απόρριψη της H_0
- Κατηγορίες στατιστικών ελέγχων (μέση τιμή)

	H_0	H_1	TS	p-value	Έλεγχος με TS και α	Έλεγχος με p-value
Γνωστή σ	$\mu = \mu_0$	$\mu \neq \mu_0$	$\sqrt{n} \cdot \frac{\bar{X} - \mu_0}{\sigma}$	$2P(Z \geq TS)$	Απόρριψη H_0 , αν $ TS \geq z_{\frac{\alpha}{2}}$ Αλλιώς, όχι απόρριψη H_0	Απόρριψη H_0 , αν $p\text{-value} < \alpha$ Αλλιώς, όχι απόρριψη
	$\mu \leq \mu_0$	$\mu > \mu_0$		$P(Z \geq TS)$	Απόρριψη H_0 , αν $TS \geq z_\alpha$ Αλλιώς, όχι απόρριψη H_0	
	$\mu \geq \mu_0$	$\mu < \mu_0$		$P(Z \leq TS)$	Απόρριψη H_0 , αν $TS \leq -z_\alpha$ Αλλιώς, όχι απόρριψη H_0	
Άγνωστη σ	$\mu = \mu_0$	$\mu \neq \mu_0$	$\sqrt{n} \cdot \frac{\bar{X} - \mu_0}{s}$	$2P(T_{n-1} \geq TS)$	Απόρριψη H_0 , αν $ TS \geq t_{n-1, \frac{\alpha}{2}}$ Αλλιώς, όχι απόρριψη H_0	
	$\mu \leq \mu_0$	$\mu > \mu_0$		$P(T_{n-1} \geq TS)$	Απόρριψη H_0 , αν $TS \geq t_{n-1, \alpha}$ Αλλιώς, όχι απόρριψη H_0	
	$\mu \geq \mu_0$	$\mu < \mu_0$		$P(T_{n-1} \leq TS)$	Απόρριψη H_0 , αν $TS \leq -t_{n-1, \alpha}$ Αλλιώς, όχι απόρριψη H_0	

Άσκηση 1

Tuesday, January 14, 2025 2:09 PM

Ένα παλαιότερο δείγμα από ψάρια στη λίμνη Michigan υποδείκνυε ότι η μέση τιμή της συγκέντρωσης μιας τοξικής ουσίας ήταν 11.2 ppm ανά ψάρι, με τυπική απόκλιση 2 ppm.

Έστω τώρα ένα νέο δείγμα 10 ψαριών με τις εξής συγκεντρώσεις στην ουσία αυτή:

11.5, 12, 11.6, 11.8, 10.4, 10.8, 12.2, 11.9, 12.4, 12.6

Αν η τυπική απόκλιση του δείγματος παραμένει στα 2 ppm, να ελεγχθεί η υπόθεση ότι η μέση συγκέντρωση στην τοξική αυτή ουσία συνεχίζει να έχει την τιμή 11.2 ppm. Να χρησιμοποιηθεί επίπεδο σημαντικότητας 5%.

$$\alpha \cdot 100\% = 5\% \Rightarrow \boxed{\alpha = 0.05}$$

$$\sigma = 2$$

$$\mu_0 = 11.2$$

Γνωστή τυπική απόκλιση, αίρα!

$$H_0: \mu = 11.2 \quad \text{vs} \quad H_1: \mu \neq 11.2$$

$$\bar{X} = \frac{11.5 + 12 + 11.6 + 11.8 + 10.4 + 10.8 + 12.2 + 11.9 + 12.4 + 12.6}{10} = 11.72$$

$$TS = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} = \sqrt{10} \cdot \frac{11.72 - 11.2}{2} \Rightarrow \boxed{TS \approx 0.822}$$

$$|TS| = 0.822$$

$$|TS| < z_{\frac{\alpha}{2}} \Rightarrow \text{Δεν απορρίπτουμε την } H_0$$

$$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$$

Άσκηση 2

Tuesday, January 14, 2025

2:09 PM

Έστω ένας στατιστικός έλεγχος με $H_0: \mu \leq 100$ και $H_1: \mu > 100$.

Έστω ένα δείγμα μεγέθους 20 με δειγματικό μέσο $\bar{X}=105$.

Να βρεθεί το p value του ελέγχου, αν η τυπική απόκλιση είναι γνωστή και ίση με

(α) 5

(β) 10

$$n=20$$

$$\bar{X}=105$$

$$H_0: \mu \leq 100 \quad \text{vs} \quad H_1: \mu > 100$$

$$a) \sigma=5$$

$$TS = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} = \sqrt{20} \cdot \frac{105 - 100}{5} = \sqrt{20} \approx 4.47$$

$$\begin{aligned} p\text{-value} &= P(Z \geq TS) = P(Z \geq 4.47) = 1 - P(Z < 4.47) = 1 - \Phi(4.47) = \\ &\approx 1 - 1 = 0 \end{aligned}$$

$$b) \sigma=10$$

$$TS = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma} = \sqrt{20} \cdot \frac{105 - 100}{10} = \frac{1}{2} \cdot \sqrt{20} \approx 2.235$$

$$\begin{aligned} p\text{-value} &= P(Z \geq TS) = P(Z \geq 2.235) = 1 - P(Z < 2.235) = 1 - \Phi(2.235) \approx \\ &\approx 1 - 0.9875 = 0.0125 \end{aligned}$$

$$\ast \underline{\Sigma \alpha = 0.01}$$

$$a) p\text{-value} \approx 0 < 0.01 = \alpha \Rightarrow \text{Απορρίπτουμε την } H_0$$

$$b) p\text{-value} = 0.0125 > 0.01 = \alpha \Rightarrow \text{Δεν απορρίπτουμε την } H_0$$

Άσκηση 3

Tuesday, January 14, 2025 2:09 PM

Μια επιχείρηση έχει έσοδα 2000 ευρώ ανά μέρα. Για να ελέγξουν οι διοικούντες αν τα έσοδα της επιχείρησης έχουν επηρεαστεί θετικά από την αλλαγή στην οικονομία, αποφάσισαν να μελετήσουν τα έσοδα των επόμενων 8 ημερών. Αν αυτά ήταν τα παρακάτω:

2050, 2212, 1880, 2121, 2205, 2018, 1980, 2188

(α) ποια είναι η μηδενική και η εναλλακτική υπόθεση;

(β) είναι τα δεδομένα αρκετά, με επίπεδο σημαντικότητας 5%, για να αποδείξουν ότι έχει επηρεαστεί η επιχείρηση;

$$\mu_0 = 2000$$

$$n = 8 \text{ ημέρες}$$

α) Θέλουμε να ελέγξουμε αν έχουν αυξηθεί τα έσοδα αυστηρά, ορατά λαμβάνουμε υπόψη του ισότιμα. Συνεπώς:

$$H_0: \mu \leq \mu_0 \quad \text{vs} \quad H_1: \mu > \mu_0$$

$$\beta) \alpha = 0.05$$

$$\bar{x} = \frac{2050 + 2212 + 1880 + 2121 + 2205 + 2018 + 1980 + 2188}{8} \approx 2081.75$$

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \approx 1444.79 \Rightarrow \boxed{s \approx 120.19}$$

$$TS = \sqrt{n} \cdot \frac{\bar{x} - \mu_0}{s} = \sqrt{8} \cdot \frac{2081.75 - 2000}{120.19} \Rightarrow \boxed{TS \approx 1.924}$$

$$t_{n-1, \alpha} = t_{7, 0.05} \approx 1.8946 < 1.924 = TS \Rightarrow$$

Απορρίπτουμε την $H_0 \Rightarrow$ Ισχύει η $H_1 \Rightarrow$ Τα έσοδα έχουν αυξηθεί