ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΣΤΑΤΙΣΤΙΚΗ

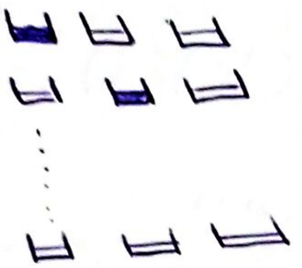
Διάλεξη 9η

22/03/2018

Έλεγχος Στατιστικών Υποθέσεων

Στατιστικό Τεστ

Δοκιμή κρασιών - Τριπλό τεστ:



p = πιθανότητα σωστής διάκρισης

$p = \frac{1}{3}$: δεν είναι

$p > \frac{1}{3}$ είναι.

Μηδενική Υπόθεση (Null Hypothesis): $H_0: p = \frac{1}{3}$ v $H_a: p > \frac{1}{3}$ εναλλακτική (Alternative)

n δοκιμές, p σταθερό

X = αριθμός των σωστών αναγνωρίσεων

$X \sim Bin(n,p) \equiv \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0,1,\dots,n$

$n = 10$ δοκιμές, $x = 0,1,\dots,10$

Η περιοχή απόρριψης της H_0 είναι γνωστή ως κρίσιμη περιοχή (κ.π.) ή C. (critical region).

$C_5 = \{5,6,7,8,9,10\}$, θα απορρίψω την μηδενική H_0, αν $x \geq 5$

$C_8 = \{8,9,10\}$, θα απορρίψω την H_0, αν $x \geq 8$.

Στατιστικό Συμπέρασμα:

Πραγματικότητα	Αποδοχή H_0	Απόρριψη H_0
Αληθεύει H_0	0 πιθ = $1-\alpha$	Σφάλμα Ι πιθ = α
Αληθεύει H_1	Σφάλμα ΙΙ πιθ = β	0 πιθ = $1-\beta$

$\alpha = P(\text{απορ } H_0 \mid \text{αληθεύει } H_0)$
$= P(\text{σφάλμα τύπου I})$

$\beta = P(\text{δεχ } H_0 \mid \text{αληθεύει } H_1)$
$= P(\text{σφάλμα τύπου II})$

$$\alpha(C_5) = P\left(X \geq 5 \mid X \sim Bin\left(n=10, p=\tfrac{1}{3}\right)\right) = \sum_{x=5}^{10} \binom{10}{x} \left(\tfrac{1}{3}\right)^x \left(\tfrac{2}{3}\right)^{10-x}$$

$$= 0{,}2131$$

$$\alpha(C_8) = P\left(X \geq 8 \mid X \sim Bin\left(n=10, p=\tfrac{1}{3}\right)\right) = \sum_{x=8}^{10} \binom{10}{x} \left(\tfrac{1}{3}\right)^x \left(\tfrac{2}{3}\right)^{10-x}$$

$$= 0{,}0034$$

$\Rightarrow$ άρα καλύτερη είναι η C_8

$$\beta(C_5) = P\left(\underset{(X \leq 4)}{X < 5} \mid X \sim Bin(n=10, p=0{,}7)\right) = \sum_{x=0}^{4} \binom{10}{x} (0{,}7)^x \cdot (0{,}3)^{10-x} =$$

$$= 0{,}0473.$$

$$\beta(C_8) = P\left(X \leq 7 \mid X \sim Bin(n=10, p=0{,}7)\right) = \sum_{x=0}^{7} \binom{10}{x} (0{,}7)^x (0{,}3)^{10-x}$$

$$= 0{,}6171.$$

Τα στοιχεία του τεστ

1. Την μηδενική H_0 και την εναλλακτική H_a.
2. Το προκαθορισμένο επίπ. σημαντικότητας α.
3. Την σ.σ. του τεστ και κρ. περ. C.
4. Την τιμή της σ.σ. που προκύπτει απ' τα δεδομένα.
5. Το συμπέρασμα., αν η τιμή πέσει στην κρίσιμη περιοχή, απορρ. H_0.

(όπου σ.σ. = στατιστική συνάρτηση)

~o~o~o~o~o~o~o~o

απλή: $p = \frac{1}{3}$, σύνθετη: $p > \frac{1}{3}$ ← δεξιόπλευρα

α: 0.01 , 0.05 (προκαθορίζεται) : $p < \frac{1}{3}$ ← αριστερόπλευρα

$p \neq \frac{1}{3}$ δίπλευρα

→ για δίπλευρη εναλλακτική : $C: (-\infty, c_1] \cup [c_2, \infty)$

→ για δεξιόπλευρη -//- : $C: [c_1, \infty)$

→ για αριστερόπλευρη -//- : $C: (-\infty, c_2]$

Ορισμός: P-value (τιμή p) : Η πιθανότητα η στατιστική συνάρτηση να πάρει τιμές πιο ακραίες από αυτήν που παρατηρήθηκε όταν αληθεύει η H_0.

αν p-τιμή $\leq \alpha$ τότε απορρ. η H_0.

π.χ

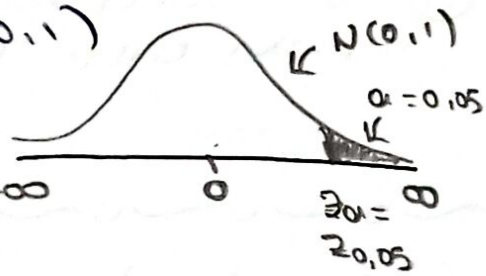
72, 69, 89, 75, 103, 121, 114, 100, 85, 99, $n=10$

$X_1,\dots,X_{10}$ τ.δ. από $N(\mu,\sigma^2=49)$, $a=0.05$

$H_0: \mu=88$ v $H_a: \mu>88$

$\hat{\mu}=\bar{x}$, $\quad \bar{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right) \to \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

$$Z=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\overset{H_0}{\sim}N(0,1)$$



Κρ. περ: $Z\geq z_a\ (=z_{0.05}=1.645)$

Αφού ο δειγματικός μέσος από τα δεδομένα είναι: $\bar{x}=92$ και $z_r=1.81>1.645$, απορρ. H_0.

$P(Z\geq 1.81 \mid Z\overset{H_0}{\sim}N(0,1))=0.0351<a$, απορρ. H_0