ΤΥΠΟΛΟΓΙΟ ΜΑΣ061

Α. Περιγραφική Στατιστική

Δειγματικός μέσος: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, Δειγματική διασπορά: $S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$,

Δειγματική τυπική απόκλιση: $S = \sqrt{S^2}$

Β. Πιθανότητες

- **Αθροιστικός Νόμος:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **Πολλαπλασιαστικός Νόμος:** $P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$
- **Θεώρημα Ολικής Πιθανότητας:** $P(B) = P(B|A) \cdot P(A) + P(B|A') \cdot P(A')$
- **Θεώρημα Bayes:** $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$

Γ. Τυχαίες Μεταβλητές και Κατανομές

- **Μέση τιμή τυχαίας μεταβλητής:** $\mu = E(X) = \sum x_i f(x_i)$
- **Διασπορά τυχαίας μεταβλητής:** $\sigma^2 = V(X) = \sum x_i^2 f(x_i) - \mu^2$

Στοιχεία γνωστών κατανομών

Κατανομή		Συνάρτηση πιθανότητας	Μέση τιμή	Διασπορά
Bernoulli	$Bern(p)$	$p^x(1-p)^{1-x},\ x = 0,1$	p	$1-p$
Διωνυμική	$Bin(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x},\ x = 0,1,\dots n$	np	$np(1-p)$
Γεωμετρική	$Geo(p)$	$p(1-p)^{x-1},\ x = 1,2,\dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Αρνητική διωνυμική	$NB(r,p)$	$\binom{x-1}{r-1}p^r(1-p)^{x-r},\ x = r, r+1,\dots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Υπεργεωμετρική	$H(N,r,n)$	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}},\ x = 0,1,\dots r$	$\frac{nr}{N}$	$\frac{nr}{N} \cdot \frac{N-r}{N} \cdot \frac{N-n}{N-1}$
Poisson	$P(\lambda)$	$e^{-\lambda}\frac{\lambda^x}{x!},\ x \geq 0$	λ	λ
Κανονική	$N(\mu,\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),\ x \in \mathbb{R}$	μ	σ^2

Δ. Κανονική κατανομή και Κατανομές Δειγματοληψίας

Έστω ότι η τ.μ. $X \sim N(\mu,\sigma^2)$. Τότε η τυποποιημένη τυχαία μεταβλητή $Z = (X-\mu)/\sigma \sim N(0,1)$. Επίσης, η τυχαία μεταβλητή $\overline{X}$ που δηλώνει τη μέση τιμή n παρατηρήσεων από το δείγμα, ακολουθεί την κανονική κατανομή $N(\mu, \sigma^2/n)$. Αν το τυχαίο δείγμα προέρχεται από διωνυμική κατανομή $Bin(n,p)$ και το μέγεθος του δείγματος είναι μεγαλύτερο του 30, τότε $X \sim N(np, np(1-p))$.

Ε. Συμπερασματολογία για μικρά δείγματα

Ε1. Έλεγχος υποθέσεων σχετικά με το ποσοστό p διωνυμικής κατανομής με επίπεδο σημαντικότητας α, όταν το μέγεθος του δείγματος είναι μικρότερο του 30.

Υπόθεση	Χωρίο Απόρριψης της H_0
$H_0 : p \geq p_0$ με $H_1 : p < p_0$	$A = \{0,1,\dots,c\}$, όπου c λύση της $P(X \leq c) \leq \alpha$
$H_0 : p \leq p_0$ με $H_1 : p > p_0$	$A = \{c, c+1,\dots,n\}$, όπου c λύση της $P(X \leq c-1) \geq 1-\alpha$
$H_0 : p = p_0$ με $H_1 : p \neq p_0$	$A = \{0,1,\dots,c_1\} \cup \{c_2,\dots,n\}$, όπου c_1, c_2 λύσεις των $P(X \leq c_1) \leq \alpha/2$, $P(X \leq c_2 - 1) \geq 1-\alpha/2$

Ε2. Στατιστική συμπερασματολογία σχετικά με τη μέση τιμή μ ενός κανονικού πληθυσμού.

Γνωστή διασπορά σ^2

Υπόθεση	Ελεγχοσυνάρτηση	Χωρίο Απόρριψης
$H_0 : \mu \geq \mu_0$ με $H_1 : \mu < \mu_0$	$Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$	$K : Z \leq -z_\alpha$
$H_0 : \mu \leq \mu_0$ με $H_1 : \mu > \mu_0$	$Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$	$K : Z \geq z_\alpha$
$H_0 : \mu = \mu_0$ με $H_1 : \mu \neq \mu_0$	$Z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$	$K : \lvert Z \rvert \geq z_{\alpha/2}$

$100(1-\alpha)\%$ Δ.Ε. για τη μ

$$\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

Άγνωστη διασπορά σ^2

Υπόθεση	Ελεγχοσυνάρτηση	Χωρίο Απόρριψης
$H_0 : \mu \geq \mu_0$ με $H_1 : \mu < \mu_0$	$T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$	$K : T \leq -t_{n-1,\alpha}$
$H_0 : \mu \leq \mu_0$ με $H_1 : \mu > \mu_0$	$T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$	$K : T \geq t_{n-1,\alpha}$
$H_0 : \mu = \mu_0$ με $H_1 : \mu \neq \mu_0$	$T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$	$K : \lvert T \rvert \geq t_{n-1,\alpha/2}$

$100(1-\alpha)\%$ Δ.Ε. για τη μ

$$\left(\overline{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}},\ \overline{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right)$$

ΣΤ. Συμπερασματολογία για μεγάλα δείγματα

ΣΤ1. Στατιστική συμπερασματολογία σχετικά με τη μέση τιμή μ (πληθυσμιακή διαπορά σ^2 άγνωστη) για δείγματα μεγαλύτερα του 30.

Υπόθεση	Ελεγχοσυνάρτηση	Χωρίο Απόρριψης
$H_0 : \mu \geq \mu_0$ με $H_1 : \mu < \mu_0$	$Z = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim N(0,1)$	$K : Z \leq -z_\alpha$
$H_0 : \mu \leq \mu_0$ με $H_1 : \mu > \mu_0$	$Z = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim N(0,1)$	$K : Z \geq z_\alpha$
$H_0 : \mu = \mu_0$ με $H_1 : \mu \neq \mu_0$	$Z = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim N(0,1)$	$K : \lvert Z \rvert \geq z_{\alpha/2}$

$100(1-\alpha)\%$ Δ.Ε. για τη μ

$$\left(\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}},\ \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

Σημείωση: Αν η διασπορά σ^2 του πληθυσμού είναι γνωστή, τότε η ελεγχοσυνάρτηση που πρέπει να χρησιμοποιηθεί είναι η

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

ΣΤ2. Στατιστική συμπερασματολογία σχετικά με το ποσοστό p για δείγματα μεγαλύτερα του 30.

Υπόθεση	Ελεγχοσυνάρτηση	Χωρίο Απόρριψης
$H_0 : p \geq p_0$ με $H_1 : p < p_0$	$Z = \dfrac{X/n - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)$	$K : Z \leq -z_\alpha$
$H_0 : p \leq p_0$ με $H_1 : p > p_0$	$Z = \dfrac{X/n - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)$	$K : Z \geq z_\alpha$
$H_0 : p = p_0$ με $H_1 : p \neq p_0$	$Z = \dfrac{X/n - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)$	$K : \lvert Z \rvert \geq z_{\alpha/2}$

$100(1-\alpha)\%$ Δ.Ε. για το p

$$\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

Ζ. Συμπερασματολογία για δύο πληθυσμούς, μεγάλα δείγματα

Ζ1. Στατιστική συμπερασματολογία σχετικά με την παράμετρο $\mu_1 - \mu_2$ δύο πληθυσμών, όταν τα δείγματα που λαμβάνουμε είναι μεγαλύτερα του 30.

Υπόθεση	Ελεγχοσυνάρτηση	Χωρίο Απόρριψης
$H_0 : \mu_1 - \mu_2 \geq \delta_0$ με $H_1 : \mu_1 - \mu_2 < \delta_0$	$Z = \dfrac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n}}} \sim N(0,1)$	$K : Z \leq -z_\alpha$
$H_0 : \mu_1 - \mu_2 \leq \delta_0$ με $H_1 : \mu_1 - \mu_2 > \delta_0$	$Z = \dfrac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n}}} \sim N(0,1)$	$K : Z \geq z_\alpha$
$H_0 : \mu_1 - \mu_2 = \delta_0$ με $H_1 : \mu_1 - \mu_2 \neq \delta_0$	$Z = \dfrac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n}}} \sim N(0,1)$	$K : \lvert Z \rvert \geq z_{\alpha/2}$

$100(1-\alpha)\%$ Δ.Ε. για την $\mu_1 - \mu_2$

$$\left(\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},\ \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right)$$

Σημείωση: Αν η διασπορά σ^2 του πληθυσμού είναι γνωστή, τότε η ελεγχοσυνάρτηση που πρέπει να χρησιμοποιηθεί είναι η

$$Z = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Ζ2. Στατιστική συμπερασματολογία σχετικά με τη διαφορά ποσοστών $p_1 - p_2$ δύο πληθυσμών. Ορίζουμε την ποσότητα

$$\hat{p} = \frac{n_1\hat{p_1} + n_2\hat{p_2}}{n_1 + n_2}$$

Υπόθεση	Ελεγχοσυνάρτηση	Χωρίο Απόρριψης
$H_0 : p_1 \geq p_2$ με $H_1 : p_1 < p_2$	$Z = \dfrac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim N(0,1)$	$K : Z \leq -z_\alpha$
$H_0 : p_1 \leq p_2$ με $H_1 : p_1 > p_2$	$Z = \dfrac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim N(0,1)$	$K : Z \geq z_\alpha$
$H_0 : p_1 = p_2$ με $H_1 : p_1 \neq p_2$	$Z = \dfrac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim N(0,1)$	$K : \lvert Z \rvert \geq z_{\alpha/2}$

$100(1-\alpha)\%$ Δ.Ε. για $p_1 - p_2$

$$\left(\hat{p_1} - \hat{p_2} - z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}},\ \hat{p_1} - \hat{p_2} + z_{\alpha/2}\sqrt{\frac{\hat{p_1}(1-\hat{p_1})}{n_1} + \frac{\hat{p_2}(1-\hat{p_2})}{n_2}}\right)$$

Η. Συμπερασματολογία για δύο πληθυσμούς, μικρά δείγματα

Στατιστική συμπερασματολογία σχετικά με την παράμετρο $\mu_1 - \mu_2$ δύο πληθυσμών, όταν τα δείγματα που λαμβάνουμε είναι μικρότερα του 30.

Περίπτωση: $\sigma_1^2 = \sigma_2^2$ ή $\sigma_1^2 \neq \sigma_2^2$, με σ_1^2, σ_2^2 γνωστά.

Υπόθεση	Ελεγχοσυνάρτηση	Χωρίο Απόρριψης
$H_0 : \mu_1 - \mu_2 \geq \delta_0$ με $H_1 : \mu_1 - \mu_2 < \delta_0$	$Z = \dfrac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$K : Z \leq -z_\alpha$
$H_0 : \mu_1 - \mu_2 \leq \delta_0$ με $H_1 : \mu_1 - \mu_2 > \delta_0$	$Z = \dfrac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$K : Z \geq z_\alpha$
$H_0 : \mu_1 - \mu_2 = \delta_0$ με $H_1 : \mu_1 - \mu_2 \neq \delta_0$	$Z = \dfrac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$K : \lvert Z \rvert \geq z_{\alpha/2}$

$100(1-\alpha)\%$ Δ.Ε. για την $\mu_1 - \mu_2$

$$\left(\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Περίπτωση: $\sigma_1^2 = \sigma_2^2 = \sigma^2$ με σ^2 άγνωστο.

Υπόθεση	Ελεγχοσυνάρτηση	Χωρίο Απόρριψης
$H_0 : \mu_1 - \mu_2 \geq \delta_0$ με $H_1 : \mu_1 - \mu_2 < \delta_0$	$t = \dfrac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\left(\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$	$K : T \leq -t_{n_1+n_2-2,\alpha}$
$H_0 : \mu_1 - \mu_2 \leq \delta_0$ με $H_1 : \mu_1 - \mu_2 > \delta_0$	$t = \dfrac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\left(\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$	$K : T \geq t_{n_1+n_2-2,\alpha}$
$H_0 : \mu_1 - \mu_2 = \delta_0$ με $H_1 : \mu_1 - \mu_2 \neq \delta_0$	$t = \dfrac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\left(\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$	$K : \lvert T \rvert \geq t_{n_1+n_2-2,\alpha/2}$

$100(1-\alpha)\%$ Δ.Ε. για την $\mu_1 - \mu_2$

$$\left(\overline{X} - \overline{Y} - t_{n_1+n_2-2,\alpha/2}\sqrt{\left(\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},\ \overline{X} - \overline{Y} + t_{n_1+n_2-2,\alpha/2}\sqrt{\left(\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$$