

ΣΤΑΤΙΣΤΙΚΗ ΚΑΤΑΝΟΜΕΣ

Τυχαίες μεταβλητές

- Δεδομένου ενός πειράματος και των αποτελεσμάτων του (δειγματικός χώρος), μια τυχαία μεταβλητή αποτελεί μια αριθμητική περιγραφή των αποτελεσμάτων
 - Μια πραγματική συνάρτηση των αποτελεσμάτων του πειράματος
- Παραδείγματα
 - Ρίψη κέρματος: τυχαία μεταβλητή είναι τα κέρδη/ζημίες που σχετίζονται με το αποτέλεσμα (κορώνα/γράμματα)
 - Επένδυση: τυχαία μεταβλητή είναι τα κέρδη/ζημίες ενός επενδυτικού χαρτοφυλακίου

Διακριτές και συνεχείς τυχαίες μεταβλητές

- Διακριτές τυχαίες μεταβλητές λαμβάνουν τιμές από ένα πεπερασμένο σύνολο
 - Οι τιμές τους δεν είναι απαραίτητο να είναι ακέραιες: κέρδος €1,5 από τη ρίψη ενός κέρματος για γράμματα, ζημία €1,5 για κορώνα
- Συνεχείς τυχαίες μεταβλητές λαμβάνουν τιμές σε ένα διάστημα $[a, b]$
 - Οι πιθανές τιμές στο διάστημα αυτό δεν μπορούν να απαριθμηθούν

Περιγραφή διακριτών τυχαίων μεταβλητών

- Μια διακριτή τυχαία μεταβλητή X περιγράφεται από μια *συνάρτηση μάζας πιθανότητας* (probability mass function, PMF), $f(x)$, που δίνει την πιθανότητα η τυχαία μεταβλητή να λάβει μια συγκεκριμένη αριθμητική τιμή x , δηλαδή

$$f(x) = \Pr(X = x)$$

- Παράδειγμα

- Εάν στη ρίψη ενός ζαριού έρθει 6, τότε κερδίζετε €10, εάν έρθει 1, 3, 5, τότε χάνετε €5 (σε κάθε άλλο ενδεχόμενο, τα κέρδη είναι μηδενικά)
- Συνάρτηση μάζας πιθανότητας:

$$f(x) = \begin{cases} 1/6 & \text{εάν } x = 10 \\ 3/6 & \text{εάν } x = -5 \\ 2/6 & \text{εάν } x = 0 \end{cases}$$

Περιγραφή συνεχών τυχαίων μεταβλητών

- Η πιθανότητα μια συνεχής τυχαία μεταβλητή X να λάβει μια συγκεκριμένη τιμή είναι μηδέν
- Πρέπει να εξεταστεί η πιθανότητα η X να λάβει, κατά προσέγγιση, μια συγκεκριμένη τιμή x , εντός ενός διαστήματος $[x - dx, x + dx]$
- Συνάρτηση πυκνότητας πιθανότητας (probability density function, PDF) $f(x)$
 - Συνεχής συνάρτηση της πυκνότητας της πιθανότητας γύρω από μια αριθμητική τιμή x (πιθανότητα ανά μονάδα της τυχαίας μεταβλητής)
 - Υψηλή πυκνότητα υποδεικνύει ότι είναι αρκετά πιθανό η τυχαία μεταβλητή να λάβει τιμές κοντά στην τιμή x

Αθροιστική συνάρτηση κατανομής

- Από μια PMF ή PDF μπορεί να οριστεί η αθροιστική συνάρτηση κατανομής (cumulative distribution function, CDF)
 - Πιθανότητα μια τυχαία μεταβλητή X να λάβει τιμές που δεν ξεπερνούν μια αριθμητική τιμή x

$$F(x) = \Pr(X \leq x)$$

- Για διακριτές τυχαίες μεταβλητές:

$$F(x) = \sum_{t \leq x} f(t)$$

- Για συνεχείς τυχαίες μεταβλητές:

$$F(x) = \int_{-\infty}^x f(t) dt$$

- Μέσω της CDF μπορεί να υπολογιστεί οποιαδήποτε πιθανότητα
$$\Pr(X > x) = 1 - F(x) \quad \Pr(x_1 \leq X \leq x_2) = F(x_2) - F(x_1)$$

Κανονική κατανομή

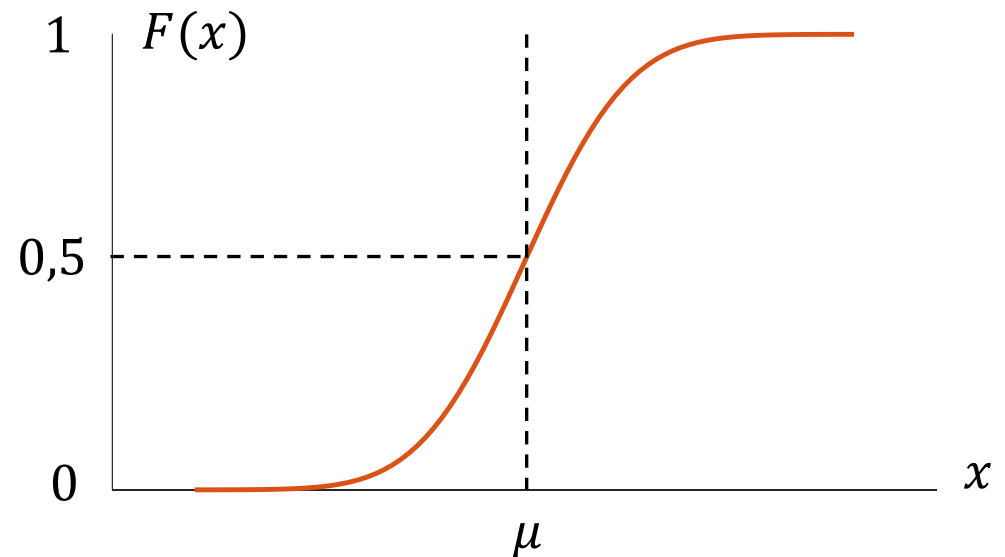
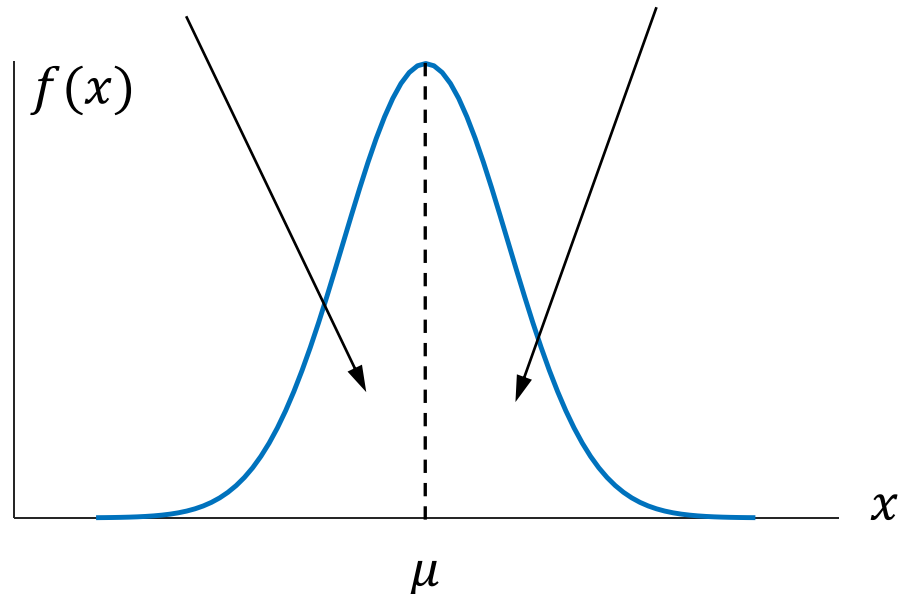
- Συμμετρική κατανομή (μέση τιμή = διάμεσος, μηδενική ασυμμετρία και κύρτωση)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

$$\Pr(-\infty < X \leq \mu) = 0,5$$

$$\Pr(\mu \leq X < \infty) = 0,5$$



Συναρτήσεις Excel

- Υπολογισμός πιθανότητας $\Pr(X \leq x)$ από την κανονική κατανομή $N(\mu, \sigma)$
=normdist($x; \mu; \sigma; \text{cumulative}$) ή =norm.dist($x; \mu; \sigma; \text{cumulative}$)
 - Cumulative = true (1) για υπολογισμούς από την αθροιστική κατανομή
 - Cumulative = false (0) για υπολογισμούς από την PDF
- Υπολογισμός τιμής x μιας τυχαίας μεταβλητής $X \sim N(\mu, \sigma)$, τέτοια ώστε $\Pr(X \leq x) = \alpha$
=norminv($\alpha; \mu; \sigma$) ή =norm.inv($\alpha; \mu; \sigma$)
- Ειδικές συναρτήσεις για $\mu = 0$ και $\sigma = 1$ (τυποποιημένη κανονική κατανομή)
=normsdist(z) ή =norm.s.dist(z)
=normsinv(α) ή =norm.s.inv(α)

Μετασχηματισμός z

- Κάθε κανονική κατανομή $N(\mu, \sigma)$ μπορεί να μετασχηματιστεί στην τυποποιημένη κανονική κατανομή $N(0, 1)$
- Μετασχηματισμός z

$$z = \frac{x - \mu}{\sigma}$$

- Συνάρτηση Excel: =standardize($x; \mu; \sigma$)
- Παράδειγμα: για μια τυχαία μεταβλητή $X \sim N(100, 50)$ υπολογίστε το μετασχηματισμό z για την τιμή $x = 180$

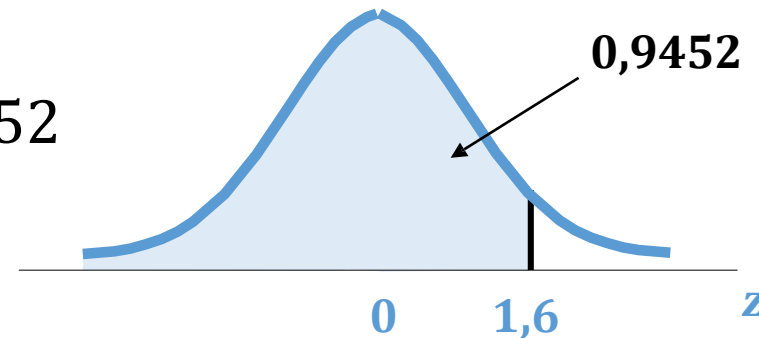
$$z = \frac{180 - 100}{50} = 1,6$$

- Η τιμή $x = 180$ είναι 1,6 τυπικές αποκλίσεις υψηλότερη από τη μέση τιμή

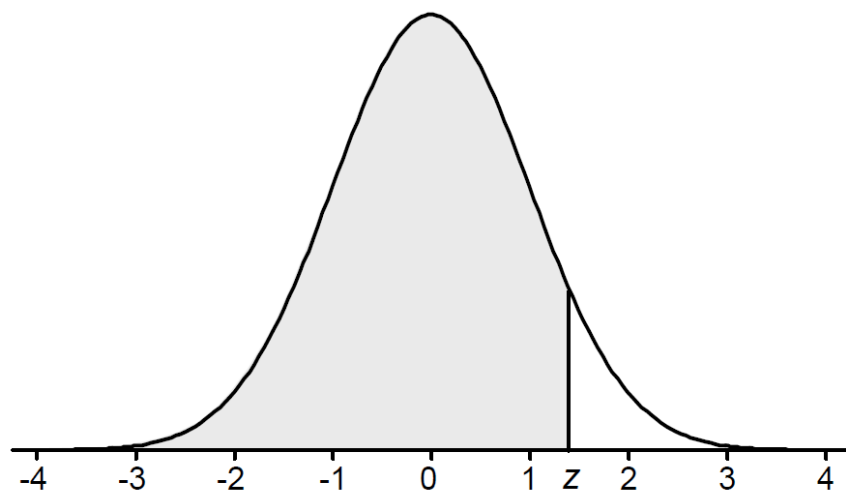
Πίνακες τυποποιημένης κανονικής κατανομή

- Για την τυποποιημένη κανονική κατανομή υπάρχουν διαθέσιμοι πίνακες πιθανοτήτων
 - Πριν τη χρήση ενός πίνακα πιθανοτήτων προσοχή στην πληροφορία που παρουσιάζει
- Ένας πίνακας παρουσιάζει αθροιστικές πιθανότητες της μορφής $\Pr(Z \leq z)$

Παράδειγμα: $\Pr(Z \leq 1,6) = 0,9452$

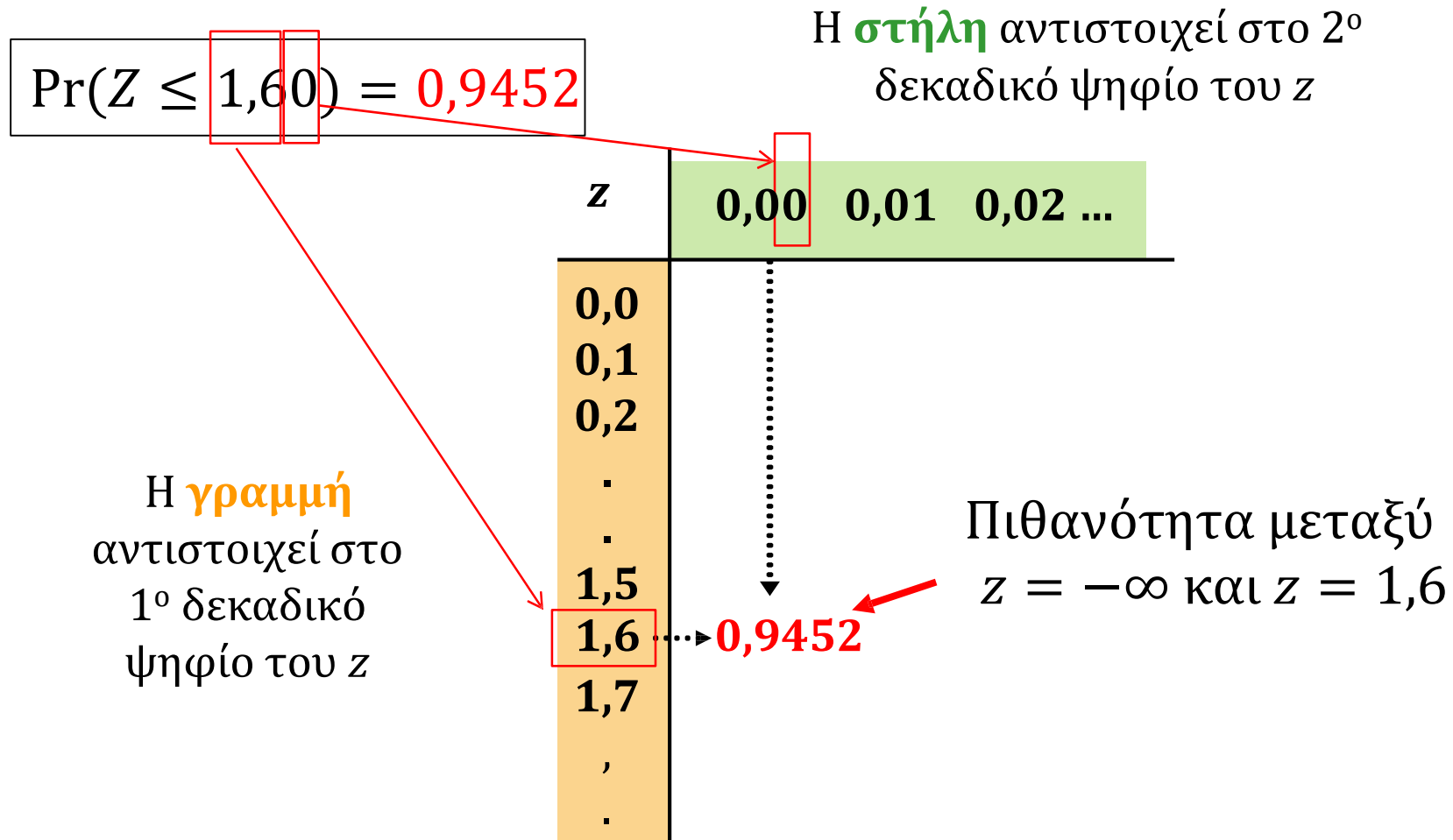


Πίνακες τυποποιημένης κανονικής κατανομή



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224

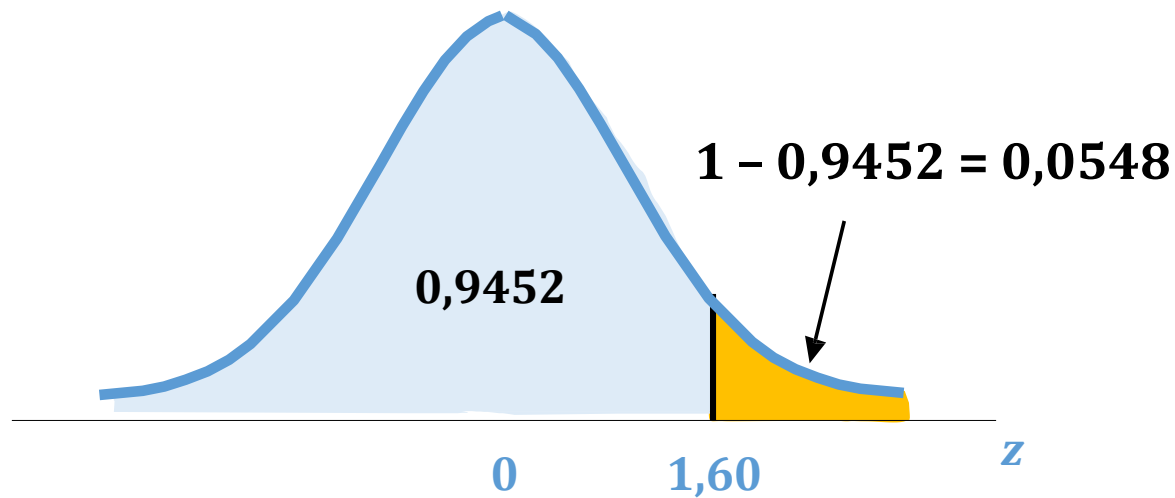
Πίνακες τυποποιημένης κανονικής κατανομή



Πιθανότητες από την κατανομή $N(0, 1)$

- Πιθανότητα από το άνω άκρο της κατανομής

$$\Pr(Z > 1,6) = 1 - \Pr(Z \leq 1,6) = 1 - 0,9452 = 0,0548$$

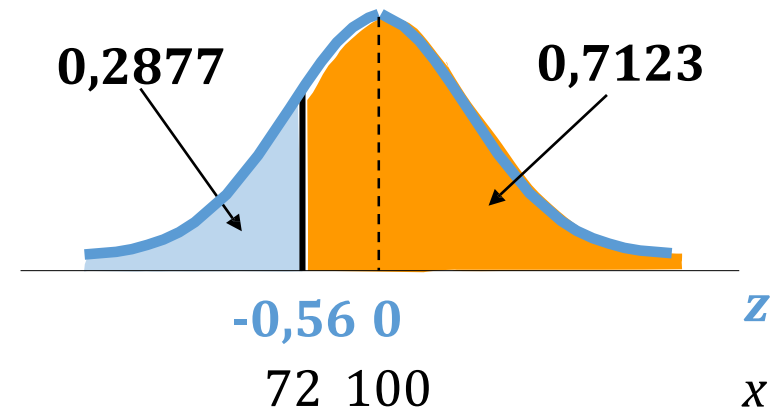


Πιθανότητες από την κατανομή $N(0, 1)$

- Για μία τυχαία μεταβλητή $X \sim N(100, 50)$ υπολογίστε τη πιθανότητα $\Pr(X > 72)$

$$Z = \frac{x - \mu}{\sigma} = \frac{72 - 100}{50} = -0,56$$

$$\begin{aligned}\Pr(X > 72) &= 1 - \Pr(X \leq 72) \\ &= 1 - \Pr(Z \leq -0,56) \\ &= 1 - 0,2877 \\ &= 0,7123\end{aligned}$$

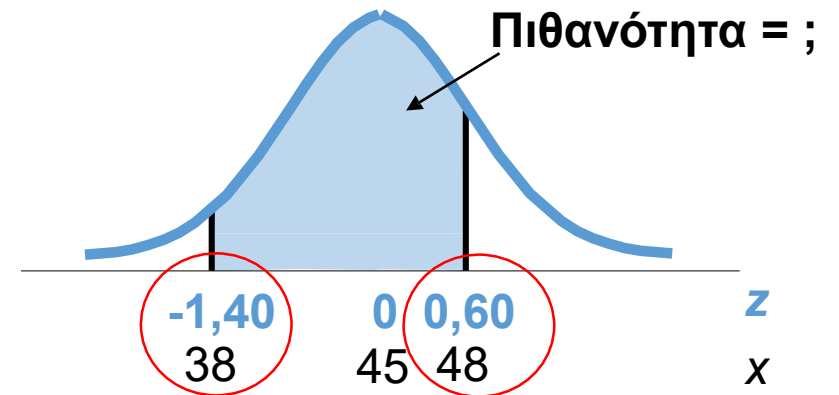


Υπολογισμός Excel: =1-norm.dist(72;100;50;1)

Πιθανότητες από την κατανομή $N(0, 1)$

- Μία ομάδα εργαζομένων έχει μέσο ετήσιο εισόδημα €45000 με τυπική απόκλιση €5000
 - Ποια είναι η πιθανότητα ένας εργαζόμενος να έχει εισόδημα μεταξύ €38000 και €48000;

$$z_{38} = \frac{x - \mu}{\sigma} = \frac{38 - 45}{5} = -1,4$$
$$z_{48} = \frac{x - \mu}{\sigma} = \frac{48 - 45}{5} = 0,6$$

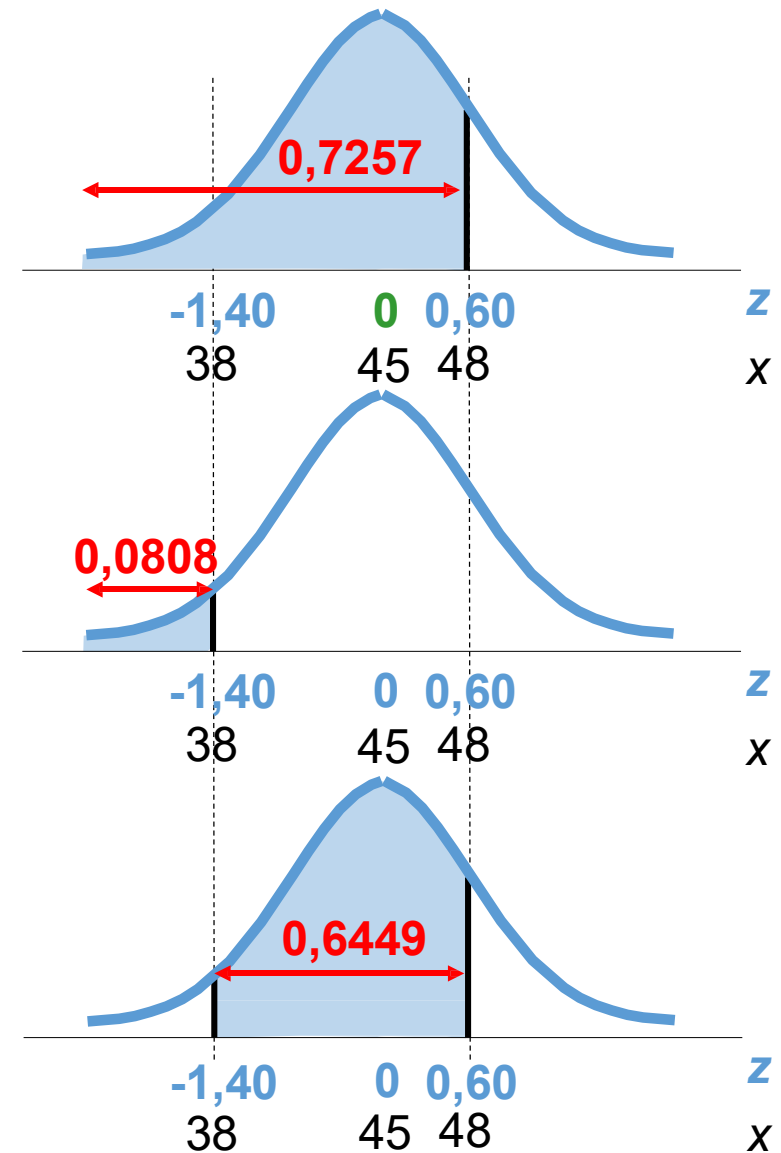


Πιθανότητες από την κατανομή $N(0, 1)$

$$\begin{aligned}\Pr(38 \leq x \leq 48) &= \Pr(-1,4 \leq z \leq 0,6) \\ &= \Pr(z \leq 0,6) - \Pr(z \leq -1,4) \\ &= 0,7257 - 0,0808 \\ &= 0,6449\end{aligned}$$

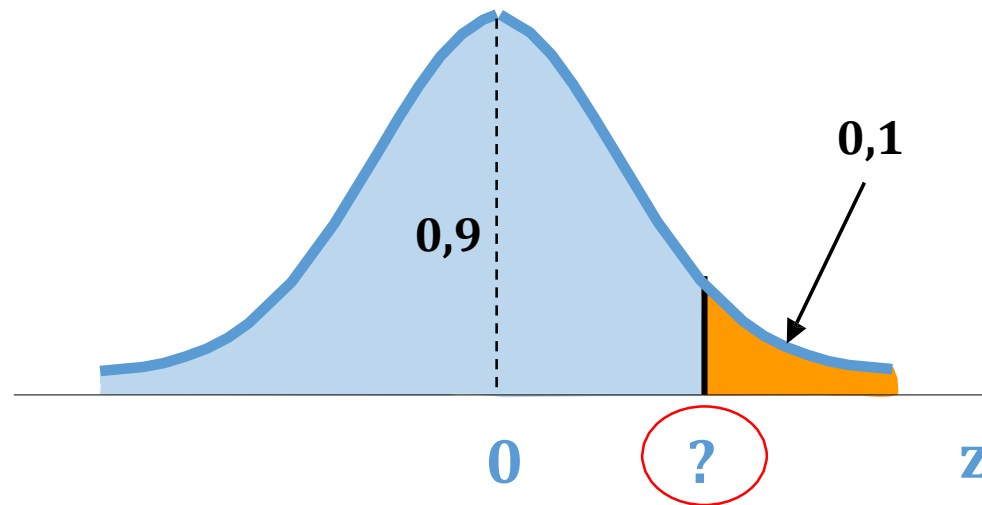
Υπολογισμός Excel :

`=norm.dist(48;45;5;1)-norm.dist(38;45;5;1)`



Παράδειγμα αντίστροφου υπολογισμού

- Για μία τυχαία μεταβλητή $X \sim N(100, 50)$ υπολογίστε μια οριακή τιμή x τέτοια ώστε $\Pr(X \leq x) = 0,9$
 1. Εύρεση από τον πίνακα, τιμής z που αντιστοιχεί σε πιθανότητα 0,9



Παράδειγμα αντίστροφου υπολογισμού

FIRST DIGIT OF z

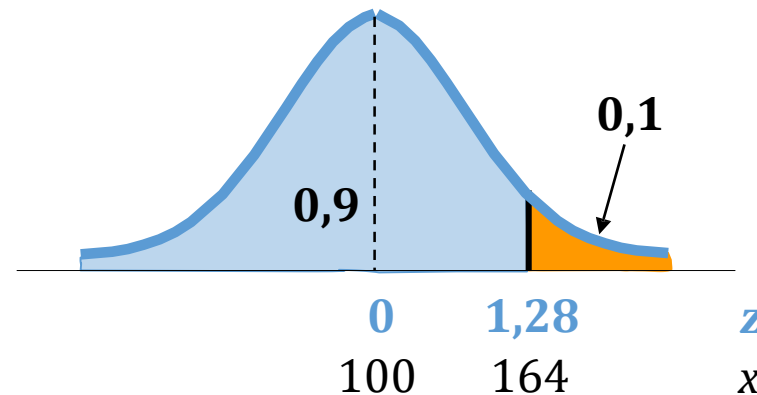
SECOND DIGIT OF z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

Παράδειγμα αντίστροφου υπολογισμού

- Για μία τυχαία μεταβλητή $X \sim N(100, 50)$ υπολογίστε μια οριακή τιμή x τέτοια ώστε $\Pr(X \leq x) = 0,9$
 - Εύρεση τιμής x που αντιστοιχεί σε $z = 1,28$

$$z = \frac{x - \mu}{\sigma}$$
$$x = \mu + z\sigma$$
$$x = 100 + (1,28)(50)$$
$$= 164$$



Υπολογισμός στο Excel:

=norm.inv(πιθανότητα; μ ; σ)
=norm.inv(0,9;100;50) -> 164,08

Διωνυμική κατανομή

- Διακριτή κατανομή που περιγράφει την πιθανότητα k «επιτυχιών» σε n διαδοχικές δοκιμές
- Κάθε δοκιμή έχει μόνο δύο αποτελέσματα: «αποτυχία» ή «επιτυχία»
- Το αποτέλεσμα μίας δοκιμής είναι ανεξάρτητο από τις προηγούμενες
- Η πιθανότητα «επιτυχίας» σε κάθε δοκιμή είναι p και η πιθανότητα «αποτυχίας» $q = 1 - p$

$$\Pr(X = k) = \frac{n!}{k! (n - k)!} p^k q^{n-k}$$

- Αναμενόμενος αριθμός «επιτυχιών»: $\mu = np$
- Διακύμανση: $\sigma^2 = npq$
- Συνάρτηση Excel: =binom.dist(k ; n ; p ; true/false) (true για αθροιστική κατανομή, false για συνάρτηση μάζας πιθανότητας)

Αθροιστικές πιθανότητες διωνυμικής κατανομής $Pr(X \leq k)$

[illegible]

Παράδειγμα

- Ένας ασφαλιστής συνεργάζεται με κάποια αντιπροσωπεία αυτοκινήτων
- Ο ασφαλιστής έχει **ασφαλίσει 15 αυτοκίνητα** της αντιπροσωπείας τα οποία έχουν το ίδιο έτος κατασκευής και προέρχονται από την ίδια κατασκευάστρια εταιρεία
- Από παλιότερα δεδομένα που έχει συλλέξει η εταιρεία, έχει εκτιμηθεί ότι κάθε αυτοκίνητο με αυτά τα χαρακτηριστικά **έχει πιθανότητα 5%** να παρουσιάσει βλάβη σε ένα έτος

α. Ποια η πιθανότητα σε ένα έτος:

- να παρουσιάσουν βλάβη **2** αυτοκίνητα;
- να μην παρουσιάσει βλάβη **κανένα** από τα 15 αυτοκίνητα;
- να παρουσιάσουν βλάβη **τουλάχιστον 2** αυτοκίνητα;

Παράδειγμα

- $\Pr(X = 2) = \frac{15!}{2!(15-2)!} 0,05^2 0,95^{15-2} = 0,135$
 - Από πίνακα της αθροιστικής διωνυμικής κατανομής ($n = 15, p = 0,05$):
 $\left. \begin{array}{l} \Pr(X \leq 1) = 0,829 \\ \Pr(X \leq 2) = 0,964 \end{array} \right\} \Rightarrow \Pr(X = 2) = \Pr(X \leq 2) - \Pr(X \leq 1) = 0,135$
 - Υπολογισμός Excel: =BINOM.DIST(2;15;0,05;FALSE)
- $\Pr(X = 0) = \frac{15!}{0!(15-0)!} 0,05^0 0,95^{15} = 0,463$
 - Υπολογισμός Excel: =BINOM.DIST(0;15;0,05;FALSE)
- $\Pr(X \geq 2) = 1 - [\Pr(X = 0) + \Pr(X = 1)] = 0,171$
 - Από πίνακα της αθροιστικής διωνυμικής κατανομής ($n = 15, k = 1, p = 0,05$):
 $\Pr(X \leq 1) = 0,829 \Rightarrow 1 - \Pr(X \leq 1) = 0,171 = \Pr(X \geq 2)$
 - Υπολογισμός Excel: =1-BINOM.DIST(1;15;0,05;TRUE)

Παράδειγμα

- Αν κάθε αυτοκίνητο που παρουσιάζει βλάβη μέσα στο έτος λαμβάνει αποζημίωση 500 ευρώ, ποιο είναι το συνολικό ποσό που αναμένεται να πληρώσει η ασφαλιστική εταιρεία;
- Ποια είναι η τυπική απόκλιση για τη συνολική αποζημίωση που θα πληρώσει η εταιρεία;

$$Y = 500X \quad \mu_Y = 500\mu_X = 500(15)(0,05) = 375$$

$$\sigma_Y^2 = 500^2 \sigma_X^2 = 500(15)(0,05)(0,95) = 178125$$

$$\sigma_Y = 422,05$$

Κατανομή Poisson

- Τυχαιά ενδεχόμενα εμφανίζονται κατά διαστήματα χρόνου ή χώρου ως εξής
 - Σε κάθε διάστημα το ενδεχόμενο συμβαίνει ή δεν συμβαίνει
 - Η εμφάνιση του ενδεχομένου σε ένα διάστημα είναι ανεξάρτητη από τα προηγούμενα
 - Η πιθανότητα εμφάνισης του ενδεχομένου είναι σταθερή στα διαστήματα
- Η κατανομή Poisson περιγράφει το αριθμό των συμβάντων, δεδομένου το μέσου αριθμού λ των συμβάντων σε ένα προκαθορισμό διάστημα

$$\Pr(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

- Μέση τιμή και διακύμανση: $\mu = \sigma^2 = \lambda$
- Συνάρτηση Excel: =poisson.dist(k ; λ ; true/false) (true για αθροιστική κατανομή, false για συνάρτηση μάζας πιθανότητας)

Αθροιστικές πιθανότητες κατανομής Poisson

$Pr(X \leq k)$

k	λ													
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1	1,25	1,5	1,75	2
0	0,905	0,819	0,741	0,670	0,607	0,549	0,497	0,449	0,407	0,368	0,287	0,223	0,174	0,135
1	0,995	0,982	0,963	0,938	0,910	0,878	0,844	0,809	0,772	0,736	0,645	0,558	0,478	0,406
2	1,000	0,999	0,996	0,992	0,986	0,977	0,966	0,953	0,937	0,920	0,868	0,809	0,744	0,677
3	1,000	1,000	1,000	0,999	0,998	0,997	0,994	0,991	0,987	0,981	0,962	0,934	0,899	0,857
4	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,999	0,998	0,996	0,991	0,981	0,967	0,947
5	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,998	0,996	0,991	0,983

Παράδειγμα

- Μια βιοτεχνία λειτουργεί 8 ώρες/ημέρα και δέχεται 24 παραγγελίες / ημέρα
- Αν ο αριθμός των παραγγελιών ακολουθεί την κατανομή Poisson
 1. Ποια είναι η πιθανότητα να υπάρξουν 4 παραγγελίες την επόμενη ώρα;
 2. Ποια είναι η πιθανότητα να υπάρξουν το πολύ 2 παραγγελίες την επόμενη ώρα;
 3. Ποιος είναι ο αναμενόμενος αριθμός παραγγελιών σε διάστημα 2 ωρών;
 4. Ποια είναι η πιθανότητα να υπάρξουν τουλάχιστον 3 παραγγελίες σε 2 ώρες;

Παράδειγμα

Ο μέσος αριθμός παραγγελιών ανά ώρα είναι $\lambda = \frac{24}{8} = 3$

1. $\Pr(X = 4) = \frac{e^{-3} 3^4}{4!} = 0,168$

Από πίνακα αθροιστικής κατανομής Poisson ($\lambda = 3$):

$$\left. \begin{array}{l} \Pr(X \leq 3) = 0,647 \\ \Pr(X \leq 4) = 0,815 \end{array} \right\} \Rightarrow \Pr(X = 4) = \Pr(X \leq 4) - \Pr(X \leq 3) = 0,168$$

Υπολογισμός Excel: =POISSON.DIST(4;3;FALSE)

2. $\Pr(X \leq 2) = \Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2) = 0,4232$

Από πίνακα αθροιστικής κατανομής Poisson ($\lambda = 3, k = 2$):

$$\Pr(X \leq 2) = 0,423$$

Υπολογισμός Excel: =POISSON.DIST(2;3;TRUE)

Παράδειγμα

3. $2\lambda = 6$ παραγγελίες ανά 2 ώρες

4. $\Pr(X \geq 3) = 1 - [\Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2)] = 0,9380$

$$\Pr(X = 0) = \frac{e^{-6}6^0}{0!} = 0,0025$$

$$\Pr(X = 1) = \frac{e^{-6}6^1}{1!} = 0,0149$$

$$\Pr(X = 2) = \frac{e^{-6}6^2}{2!} = 0,0446$$

Από πίνακα αθροιστικής κατανομής Poisson ($\lambda = 6, k = 2$):

$$\Pr(X \leq 2) = 0,062 \Rightarrow 1 - \Pr(X \leq 2) = 0,938 = \Pr(X \geq 3)$$

Υπολογισμός Excel: =1-POISSON.DIST(2;6;TRUE)

Διαστήματα εμπιστοσύνης

- Ένα Δ.Ε. είναι μια εκτίμηση με τη μορφή ενός διαστήματος $[L, U]$ που δείχνει το εύρος εντός του οποίου αναμένεται να βρίσκεται μία παράμετρος θ ενός πληθυσμού

$$\Pr(L \leq \theta \leq U) = 1 - \alpha$$

- Κάθε Δ.Ε. χαρακτηρίζεται από ένα επίπεδο σημαντικότητας α (significance level) που ορίζει το επίπεδο εμπιστοσύνης $1 - \alpha$ (confidence level)
 - Το επίπεδο εμπιστοσύνης αναπαριστά την πιθανότητα το διάστημα εμπιστοσύνης να περιλαμβάνει την εξεταζόμενη παράμετρο του πληθυσμού
 - Συνηθισμένα επίπεδα εμπιστοσύνης: 90%, 95%, 99% ($\alpha = 10\%, 5\%, 1\%$)
 - Το εύρος του διαστήματος γίνεται μεγαλύτερο όσο το επίπεδο εμπιστοσύνης αυξάνει

Διαστήματα εμπιστοσύνης

- Γενικός υπολογισμός
(εκτίμηση σημείου) $\pm$ (κρίσιμη τιμή)(τυπικό σφάλμα)
 - Εκτίμηση σημείου: εκτίμηση δείγματος
 - Κρίσιμη τιμή: τιμή-όριο από στατιστικούς πίνακες
 - Τυπικό σφάλμα: τυπική απόκλιση της εκτίμησης σημείου
 - Το γινόμενο (κρίσιμη τιμή)(τυπικό σφάλμα) αναφέρεται ως περιθώριο σφάλματος (margin of error)
- Συνηθισμένα Δ.Ε.
 - Μέση τιμή
 - Περίπτωση 1: γνωστή διακύμανση
 - Περίπτωση 2: άγνωστη διακύμανση
 - Αναλογίες

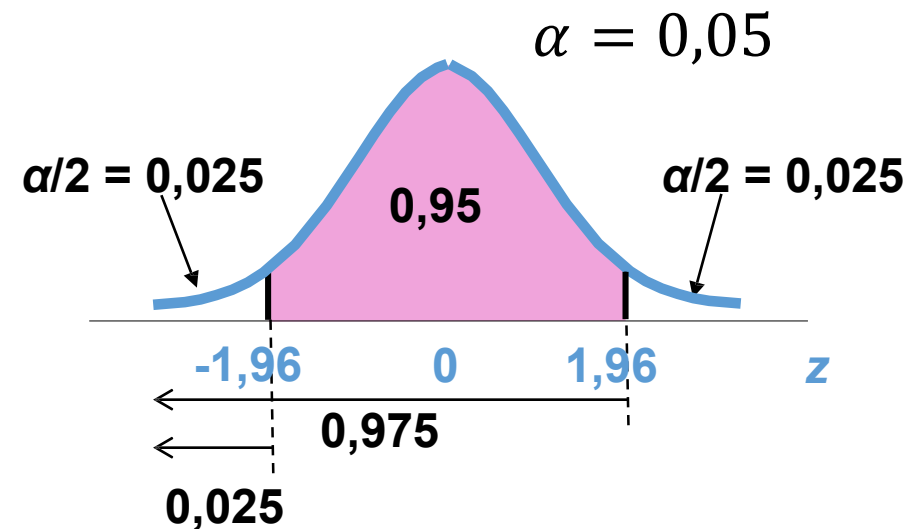
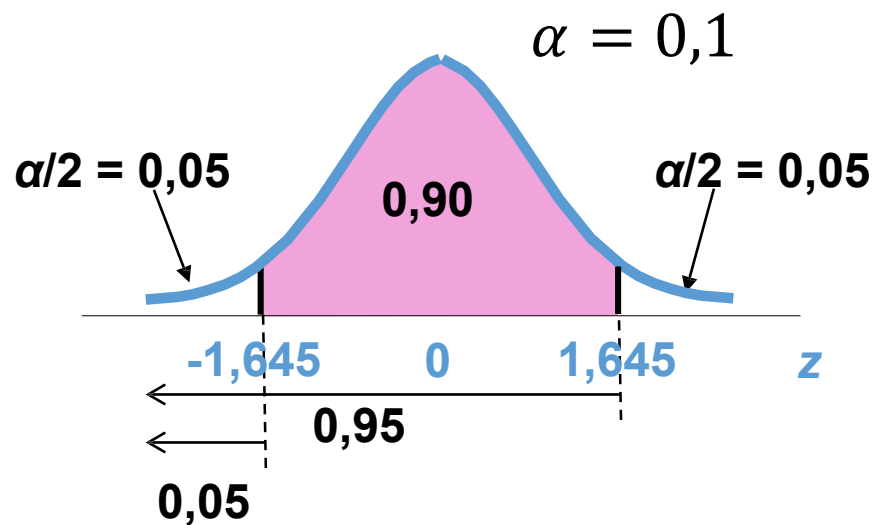
Δ.Ε. για τη μέση τιμή (σ γνωστό)

$$\bar{x} \pm z_{1-\alpha/2} \sigma_{\bar{x}}$$

- $\bar{x}$ = μέση τιμή δείγματος
- $\sigma_{\bar{x}} = \sigma / \sqrt{n}$ τυπικό σφάλμα μέσης τιμής
- $z_{1-\alpha/2}$ = κρίσιμη τιμή από την κανονική κατανομή, τέτοια ώστε $\Pr(Z > z_{1-\alpha/2}) = \alpha/2$

Υπολογισμοί Excel

1. Κρίσιμη τιμή ($z_{1-\alpha/2}$):
=NORM.INV(1 - $\alpha/2$; 0; 1)
2. Περιθώριο σφάλματος ($z_{1-\alpha/2} \sigma_{\bar{x}}$):
=CONFIDENCE.NORM(α ; σ ; n)



Επίπεδα εμπιστοσύνης & σημαντικότητας

Επίπεδο εμπιστοσύνης $(1 - \alpha)\%$	Επίπεδο σημαντικότητας $\alpha\%$	Κρίσιμη τιμή $z_{1-\alpha/2}$
80%	20%	1,280
90%	10%	1,645
95%	5%	1,960
98%	2%	2,330
99%	1%	2,575

Παράδειγμα

- Από ένα δείγμα 12000 συνδρομητών μίας εταιρείας κινητής τηλεφωνίας εκτιμήθηκε ότι ο μέσος μηνιαίος χρόνος ομιλίας είναι 586,4 λεπτά

Ποιο είναι το 90% διάστημα εμπιστοσύνης για τον μέσο μηνιαίο χρόνο ομιλίας, εάν θεωρηθεί ότι η τυπική απόκλιση του μηνιαίου χρόνου ομιλίας είναι γνωστή και ίση με 588,9 λεπτά;

1) Τυπικό σφάλμα: $\sigma_{\bar{x}} = \sigma / \sqrt{n} = 588,9 / \sqrt{12000} = 5,376$

2) $z_{1-\alpha/2} = z_{0,95} = 1,645$

Υπολογισμός Excel: =norm.inv(0,95;0;1)

3) Διάστημα εμπιστοσύνης:

$$\begin{aligned} & [\bar{x} - z_{0,95}\sigma_{\bar{x}}, \bar{x} + z_{0,95}\sigma_{\bar{x}}] = \\ & [586,4 - 1,645(588,9), 586,4 + 1,645(588,9)] = \\ & [577,58, 595,27] \end{aligned}$$

Δ.Ε. για τη μέση τιμή (σ άγνωστο)

- Η τυπική απόκλιση σ του πληθυσμού σπανίως είναι γνωστή
- Αντί αυτής χρησιμοποιούμε τη δειγματική εκτίμηση s
- Με την αλλαγή αυτή ο μετασχηματισμός z για τη μέση τιμή

$$\frac{\bar{x} - \mu}{s/\sqrt{n}}$$

δεν ακολουθεί την κανονική κατανομή, αλλά την κατανομή t του Student (William Sealy Gosset, 1908) με $\nu = n - 1$ βαθμούς ελευθερίας (degrees of freedom)

- Εκτίμηση Δ.Ε.: $\bar{x} \pm t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}$
- Υπολογισμοί Excel:
 - Κρίσιμη τιμή ($t_{n-1, 1-\alpha/2}$): =T.INV.2T(α ; $n - 1$)
 - Περιθώριο σφάλματος ($t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}$): =CONFIDENCE.T(α ; s ; n)

**Ποσοστιαία σημεία
κατανομής t**
Τιμές $t_{\nu,1-\alpha}$ για τις οποίες
 $\Pr(T \leq t_{\nu,1-\alpha}) = 1 - \alpha$
Για $1 - \alpha < 0,5$ ισχύει η σχέση
 $t_{\nu,1-\alpha} = -t_{\nu,\alpha}$

ν	$1 - \alpha$										
	0,55	0,6	0,7	0,8	0,9	0,95	0,975	0,98	0,99	0,995	0,999
1	0,158	0,325	0,727	1,376	3,078	6,314	12,706	15,895	31,821	63,657	318,309
2	0,142	0,289	0,617	1,061	1,886	2,920	4,303	4,849	6,965	9,925	22,327
3	0,137	0,277	0,584	0,978	1,638	2,353	3,182	3,482	4,541	5,841	10,215
4	0,134	0,271	0,569	0,941	1,533	2,132	2,776	2,999	3,747	4,604	7,173
5	0,132	0,267	0,559	0,920	1,476	2,015	2,571	2,757	3,365	4,032	5,893
6	0,131	0,265	0,553	0,906	1,440	1,943	2,447	2,612	3,143	3,707	5,208
7	0,130	0,263	0,549	0,896	1,415	1,895	2,365	2,517	2,998	3,499	4,785
8	0,130	0,262	0,546	0,889	1,397	1,860	2,306	2,449	2,896	3,355	4,501
9	0,129	0,261	0,543	0,883	1,383	1,833	2,262	2,398	2,821	3,250	4,297
10	0,129	0,260	0,542	0,879	1,372	1,812	2,228	2,359	2,764	3,169	4,144
11	0,129	0,260	0,540	0,876	1,363	1,796	2,201	2,328	2,718	3,106	4,025
12	0,128	0,259	0,539	0,873	1,356	1,782	2,179	2,303	2,681	3,055	3,930
13	0,128	0,259	0,538	0,870	1,350	1,771	2,160	2,282	2,650	3,012	3,852
14	0,128	0,258	0,537	0,868	1,345	1,761	2,145	2,264	2,624	2,977	3,787
15	0,128	0,258	0,536	0,866	1,341	1,753	2,131	2,249	2,602	2,947	3,733
16	0,128	0,258	0,535	0,865	1,337	1,746	2,120	2,235	2,583	2,921	3,686
17	0,128	0,257	0,534	0,863	1,333	1,740	2,110	2,224	2,567	2,898	3,646
18	0,127	0,257	0,534	0,862	1,330	1,734	2,101	2,214	2,552	2,878	3,610
19	0,127	0,257	0,533	0,861	1,328	1,729	2,093	2,205	2,539	2,861	3,579
20	0,127	0,257	0,533	0,860	1,325	1,725	2,086	2,197	2,528	2,845	3,552

Παράδειγμα

- Δεδομένα για την εβδομαδιαία υπερωριακή εργασία 18 διοικητικών στελεχών

6	21	17	20	7	0	8	16	29
3	8	12	11	9	21	25	15	16

- Ποιο είναι το 90% διάστημα εμπιστοσύνης για τη μέση εβδομαδιαία υπερωριακή εργασία των στελεχών;

Παράδειγμα

1. Εκτίμηση μέσης τιμής και τυπικής απόκλισης βάσει του δείγματος:

$$\bar{x} = 13,56 \text{ και } s = 7,8$$

2. Ορισμός βαθμών ελευθερίας

Δείγμα μεγέθους $n = 18$, οπότε $\nu = 17$ βαθμοί ελευθερίας

3. Εύρεση κρίσιμης τιμή από την κατανομή t ($\alpha = 0,1$)

$$t_{n-1, 1-\alpha/2} = t_{17, 0,95} = 1,74$$

Υπολογισμός Excel: =T.INV.2T(0,1;17)

4. Υπολογισμός διαστήματος εμπιστοσύνης

$$\left[\bar{x} - t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}} \right] = [10,36, 16,76]$$

Δ.Ε. για ποσοστά (αναλογίες)

- Εκτίμηση διαστήματος για το πραγματικό ποσοστό p των μονάδων ενός πληθυσμού που έχει μια συγκεκριμένη ιδιότητα
- Εκτίμηση ποσοστού από δείγμα μεγέθους n στο οποίο υπάρχουν X περιπτώσεις με την ιδιότητα που ενδιαφέρει

$$\hat{p} = X/n$$

- Διάστημα εμπιστοσύνης

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

Παράδειγμα

- Ένα πανεπιστήμιο έκανε μια έρευνα μεταξύ των αποφοίτων του και βρήκε ότι μεταξύ 135 αποφοίτων 10 ήταν άνεργοι ένα χρόνο μετά την αποφοίτηση
 - Ποιο είναι το 95% διάστημα εμπιστοσύνης για το ποσοστό των ανέργων αποφοίτων ένα χρόνο μετά την αποφοίτηση;

- Υπολογισμός αναλογίας δείγματος και τυπικού σφάλματος

$$\hat{p} = \frac{10}{135} = 0,074 \text{ και } s_p = \sqrt{\frac{0,074(1 - 0,074)}{135}} = 0,023$$

- Κρίσιμη τιμή: $z_{1-\alpha/2} = z_{0,975} = 1,96$
- Δ.Ε.: $[\hat{p} - z_{1-\alpha/2}s_p, \hat{p} + z_{1-\alpha/2}s_p] = [0,029, 0,119]$

Έλεγχοι υποθέσεων (hypothesis testing)

- Κάθε στατιστικό μέγεθος (μέση τιμή, διακύμανση, συσχέτιση, αναλογίες, ...) συνοδεύεται από ελέγχους υποθέσεων για την εξαγωγή στατιστικών συμπερασμάτων
- Γενική διαδικασία ενός στατιστικού ελέγχου
 1. Ορισμός της υπόθεσης και του επιπέδου εμπιστοσύνης
 2. Υπολογισμός στατιστικού μεγέθους και του τυπικού σφάλματος από ένα δείγμα
 3. Ορισμός στατιστικής συνάρτησης ελέγχου και της κατάλληλης κατανομής
 4. Εφαρμογή κανόνα απόφασης για την εξαγωγή στατιστικών συμπερασμάτων

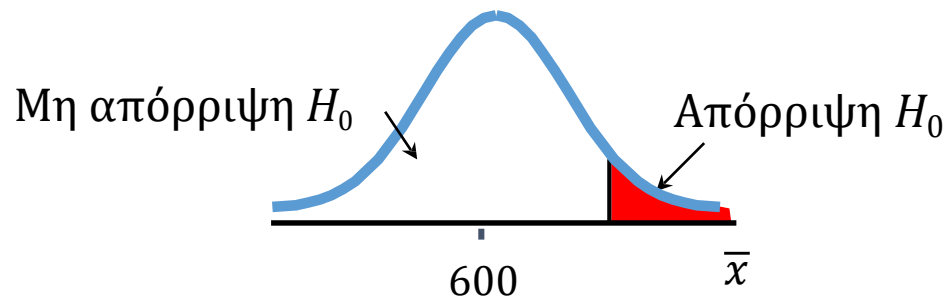
Μονόπλευροι στατιστικοί έλεγχοι

- Η εναλλακτική υπόθεση εκφράζεται σε μορφή $<$ ή $>$

$$H_0: \mu = 600$$

$$H_1: \mu > 600$$

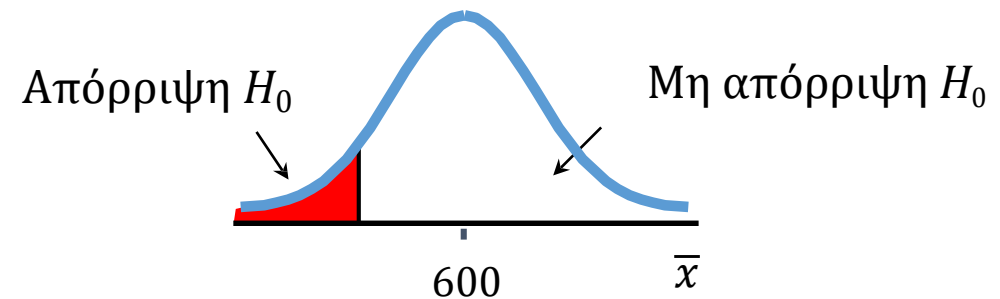
Υποθέτουμε ότι $\mu = 600$ εκτός εάν η μέση τιμή του δείγματος είναι σημαντικά μεγαλύτερη από 600



$$H_0: \mu = 600$$

$$H_1: \mu < 600$$

Υποθέτουμε ότι $\mu = 600$ εκτός εάν η μέση τιμή του δείγματος είναι σημαντικά μικρότερη από 600



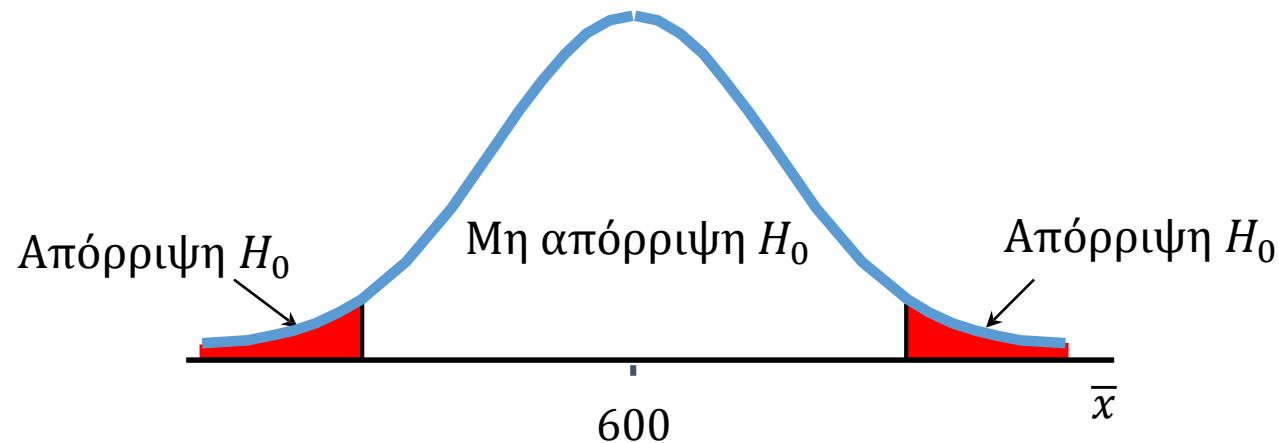
Αμφίπλευροι στατιστικοί έλεγχοι

- Η εναλλακτική υπόθεση εκφράζεται στη μορφή $\neq$

$$H_0: \mu = 600$$

$$H_1: \mu \neq 600$$

Υποθέτουμε ότι $\mu = 600$, εκτός εάν διαπιστωθεί ότι η μέση τιμή του δείγματος διαφέρει σημαντικά από την τιμή 600



Σφάλματα

- Σφάλμα τύπου I
 - Απόρριψη της μηδενικής υπόθεσης ενώ είναι αληθής
- Σφάλμα τύπου II
 - Μη απόρριψη της μηδενικής υπόθεσης ενώ είναι δεν είναι αληθής

Απόφαση	Πραγματική κατάσταση	
	H_0 αληθής	H_0 ψευδής
Απόρριψη H_0	Σφάλμα τύπου I $\Pr(\text{σφάλμα τύπου I}) = \alpha$	Σωστό συμπέρασμα
Μη απόρριψη H_0	Σωστό συμπέρασμα	Σφάλμα τύπου II $\Pr(\text{σφάλμα τύπου I}) = \beta$

Έλεγχος υποθέσεων μέσης τιμής ενός πληθυσμού (σ γνωστό)

- Δεξιόπλευρος έλεγχος ($H_0: \mu = \mu_0, H_1: \mu > \mu_0$)
 - Απόρριψη H_0 σε επίπεδο σημαντικότητας α εάν

$$z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{1-\alpha}$$

- Αριστερόπλευρος έλεγχος ($H_0: \mu = \mu_0, H_1: \mu < \mu_0$)
 - Απόρριψη H_0 σε επίπεδο σημαντικότητας α εάν

$$z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z_{1-\alpha}$$

- Αμφίπλευρος έλεγχος ($H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$)
 - Απόρριψη H_0 σε επίπεδο σημαντικότητας α εάν

$$z_{\bar{x}} < -z_{1-\alpha/2} \text{ ή } z_{\bar{x}} > z_{1-\alpha/2}$$

Παράδειγμα 1

- Ένας κατασκευαστής λαμπτήρων φωτισμού ισχυρίζεται ότι οι λαμπτήρες έχουν μέση διάρκεια ζωής τουλάχιστον 8000 ωρών
- Από ένα δείγμα 36 λαμπτήρων μετρήθηκε μια μέση διάρκεια 8120 ωρών
- Η τυπική απόκλιση του πληθυσμού είναι 500 ώρες
- Ερώτημα: με βάση τα δεδομένα του δείγματος, ευσταθεί ο ισχυρισμός της εταιρείας σε επίπεδο σημαντικότητας 5%;

Παράδειγμα 1

- **Βήμα 1:** Διατύπωση υπόθεσης

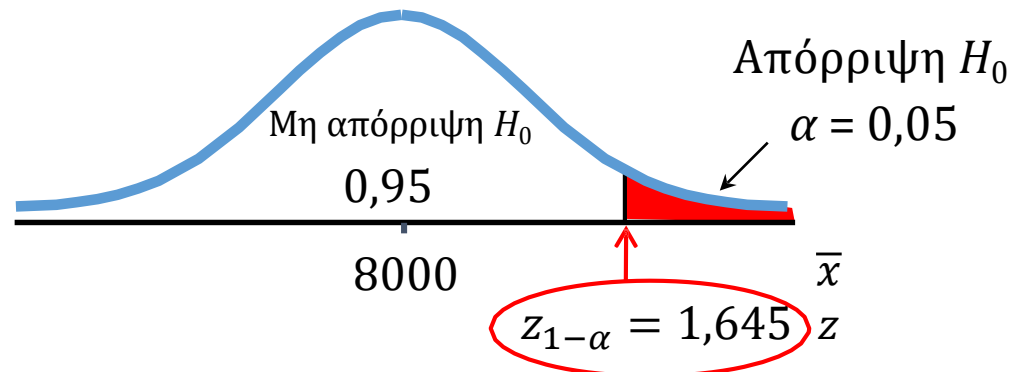
$$H_0: \mu = 8000 \text{ έναντι } H_1: \mu > 8000$$

- **Βήμα 2:** Ορισμός επιπέδου σημαντικότητας α

$$- \alpha = 0,05$$

- **Βήμα 3:** Κρίσιμη τιμή

– Δεξιόπλευρος έλεγχος



Παράδειγμα 1

- **Βήμα 4:** Κανόνας απόφασης

$$z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{8120 - 8000}{\frac{500}{\sqrt{36}}} = 1,44 < z_{1-\alpha} = 1,645$$

- Επομένως η H_0 δεν απορρίπτεται

Ο ισχυρισμός της εταιρείας ότι $\mu > 8000$ δεν υποστηρίζεται από τα δεδομένα του δείγματος

Παράδειγμα 2

- Μια εταιρεία κινητής τηλεφωνίας θεωρεί ότι ο μέσος μηνιαίος χρόνος ομιλίας των χρηστών είναι 600 λεπτά
- Για ένα δείγμα 49 συνδρομητών βρέθηκε ότι ο μέσος χρόνος ομιλίας είναι 690 λεπτά/μήνα
- Η τυπική απόκλιση του πληθυσμού είναι 300 λεπτά/μήνα
- Είναι σωστός ο ισχυρισμός της εταιρείας σε επίπεδο σημαντικότητας 5%;

Παράδειγμα 2

- **Βήμα 1:** Διατύπωση υπόθεσης

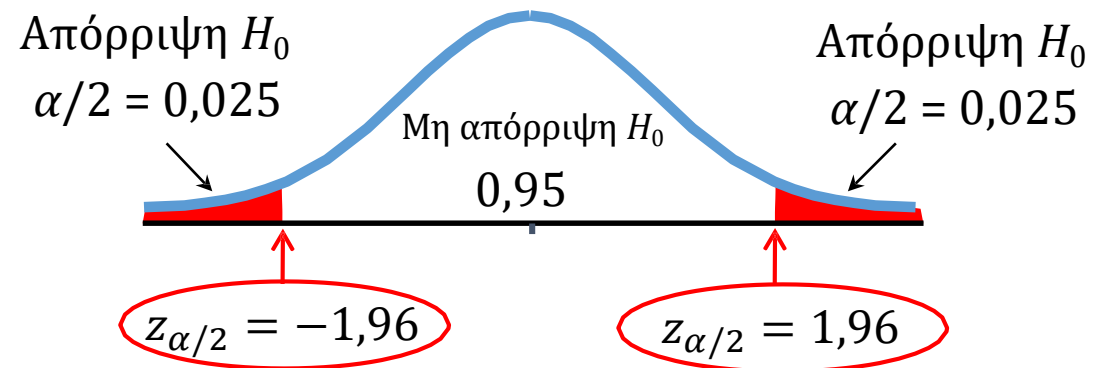
$$H_0: \mu = 600 \text{ έναντι } H_1: \mu \neq 600$$

- **Βήμα 2:** Ορισμός επιπέδου σημαντικότητας α

$$- \alpha = 0,05$$

- **Βήμα 3:** Κρίσιμη τιμή

– Αμφίπλευρος έλεγχος



Παράδειγμα 2

- **Βήμα 4:** Κανόνας απόφασης

$$z_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{690 - 600}{\frac{300}{\sqrt{49}}} = 2,1$$

- Επειδή $z_{\bar{x}} = 2,1 > z_{1-\alpha/2} = 1,96$ η H_0 απορρίπτεται

Ο ισχυρισμός της εταιρείας ότι $\mu = 600$ δεν υποστηρίζεται από τα δεδομένα του δείγματος

Έλεγχος υποθέσεων μέσης τιμής ενός πληθυσμού (σ αγνωστό)

- Χρησιμοποιείται η δειγματική τυπική απόκλιση ως εκτίμηση της τυπικής απόκλισης του πληθυσμού
- Η κρίσιμη τιμή και το επίπεδο σημαντικότητας (p -value) ορίζονται από την κατανομή t του Student
 - Υπόθεση: ο πληθυσμός είναι κανονικά κατανεμημένος

Έλεγχος υποθέσεων μέσης τιμής ενός πληθυσμού (σ αγνωστό)

- Δεξιόπλευρος έλεγχος ($H_0: \mu = \mu_0, H_1: \mu > \mu_0$)

- Απόρριψη H_0 σε επίπεδο σημαντικότητας α εάν

$$t_{\bar{x}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1, 1-\alpha}$$

- Υπολογισμός κρίσιμης τιμής $t_{n-1, 1-\alpha}$ στο Excel: =T.INV(1 - α ; $n - 1$)

- Αριστερόπλευρος έλεγχος ($H_0: \mu = \mu_0, H_1: \mu < \mu_0$)

- Απόρριψη H_0 σε επίπεδο σημαντικότητας α εάν

$$t_{\bar{x}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} < -t_{n-1, 1-\alpha}$$

- Υπολογισμός κρίσιμης τιμής $t_{n-1, 1-\alpha}$ στο Excel: =T.INV(1 - α ; $n - 1$)

- Αμφίπλευρος έλεγχος ($H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$)

- Απόρριψη H_0 σε επίπεδο σημαντικότητας α εάν

$$t_{\bar{x}} < -t_{n-1, 1-\alpha/2} \text{ ή } t_{\bar{x}} > t_{n-1, 1-\alpha/2}$$

- Υπολογισμός κρίσιμης τιμής $|t_{n-1, 1-\alpha/2}|$ στο Excel: =T.INV.2T(α ; $n - 1$)

Παράδειγμα

- Σύμφωνα με παλιότερα στατιστικά στοιχεία η μέση τιμή διανυκτέρευσης σε ένα ξενοδοχείο στο Λονδίνο είναι £188, αλλά λόγω του επικείμενου Brexit ένας πράκτορας πιστεύει ότι οι τιμές έχουν μειωθεί
- Από ένα δείγμα 25 ξενοδοχείων βρέθηκαν τα ακόλουθα
$$\bar{x} = 177,5 \text{ και } s = 25,4$$
- Είναι σωστός ο ισχυρισμός του πράκτορα σε επίπεδο σημαντικότητας $\alpha = 0,05$;

Παράδειγμα

- **Βήμα 1:** Διατύπωση υπόθεσης

$$H_0: \mu = 188 \text{ έναντι } H_1: \mu < 188$$

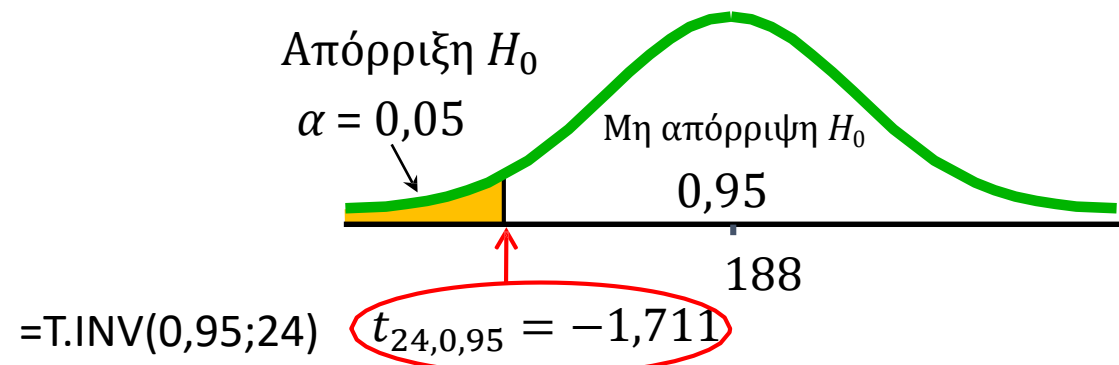
- **Βήμα 2:** Ορισμός επιπέδου σημαντικότητας α

- $\alpha = 0,05$

- **Βήμα 3:** Κρίσιμη τιμή

- Κατανομή t με $\nu = n - 1 = 24$ βαθμούς ελευθ.

- Αριστερόπλευρος έλεγχος



Παράδειγμα

- **Βήμα 4:** Κανόνας απόφασης

$$t_{\bar{x}} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{177,5 - 188}{\frac{25,4}{\sqrt{25}}} = -2,07 < t_{24,0,95} = -1,711$$

- Επομένως η H_0 απορρίπτεται

Ο ισχυρισμός του πράκτορα ότι $\mu < 188$ υποστηρίζεται από τα δεδομένα του δείγματος

Έλεγχος μέσων τιμών δύο ανεξάρτητων πληθυσμών (γνωστές τυπικές αποκλίσεις σ_1, σ_2)

- Δείγματα n_1 και n_2 παρατηρήσεων από δύο ανεξάρτητους πληθυσμούς με γνωστές τυπικές αποκλίσεις (σ_1, σ_2)
- Μέσες τιμές των δύο δειγμάτων $\bar{x}_1$ και $\bar{x}_2$
- Τυπικό σφάλμα της διαφοράς $\bar{d} = \bar{x}_1 - \bar{x}_2$

$$\sigma_{\bar{d}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Για μικρά δείγματα ($n_1, n_2 < 30$), θα πρέπει οι πληθυσμοί να είναι κανονικοί κατανεμημένοι
- Για μεγάλα δείγματα ($n_1, n_2 \geq 30$), ισχύει το κεντρικό οριακό θεώρημα

Έλεγχος μέσων τιμών δύο ανεξάρτητων πληθυσμών (γνωστές τυπικές αποκλίσεις σ_1, σ_2)

- Δεξιόπλευρος έλεγχος ($H_0: \mu_1 - \mu_2 = 0, H_1: \mu_1 - \mu_2 > 0$)
 - Απόρριψη H_0 σε επίπεδο σημαντικότητας α εάν

$$z_{\bar{x}} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{\bar{d}}/\sqrt{n}} > z_{1-\alpha}$$

- Αριστερόπλευρος έλεγχος ($H_0: \mu_1 - \mu_2 = 0, H_1: \mu_1 - \mu_2 < 0$)
 - Απόρριψη H_0 σε επίπεδο σημαντικότητας α εάν

$$z_{\bar{x}} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{\bar{d}}/\sqrt{n}} < -z_{1-\alpha}$$

- Αμφίπλευρος έλεγχος ($H_0: \mu_1 - \mu_2 = 0, H_1: \mu_1 - \mu_2 \neq 0$)
 - Απόρριψη H_0 σε επίπεδο σημαντικότητας α εάν

$$z_{\bar{x}} < -z_{1-\alpha/2} \text{ ή } z_{\bar{x}} > z_{1-\alpha/2}$$

Έλεγχος μέσων τιμών δύο ανεξάρτητων πληθυσμών (άγνωστες τυπικές αποκλίσεις σ_1, σ_2)

- Χρήση των δειγματικών τυπικών αποκλίσεων s_1, s_2
- Στατιστικός έλεγχος με την κατανομή t του Student
 - Για μικρά δείγμα θα πρέπει να γίνει η υπόθεση κανονικότητας των πληθυσμών
- Δύο περιπτώσεις: (α) $\sigma_1 = \sigma_2$, (β) $\sigma_1 \neq \sigma_2$

Έλεγχος μέσων τιμών δύο ανεξάρτητων πληθυσμών (Περίπτωση $\sigma_1 = \sigma_2$)

- Χρήση κατανομής t του Student με $n_1 + n_2 - 2$ β.ε.
- Στατιστική συνάρτηση ελέγχου

$$t_{\bar{d}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad \text{με} \quad s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}}$$

Έλεγχος μέσων τιμών δύο ανεξάρτητων πληθυσμών (Περίπτωση $\sigma_1 \neq \sigma_2$)

- Χρήση κατανομής t του Student με

$$\beta. \varepsilon. = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

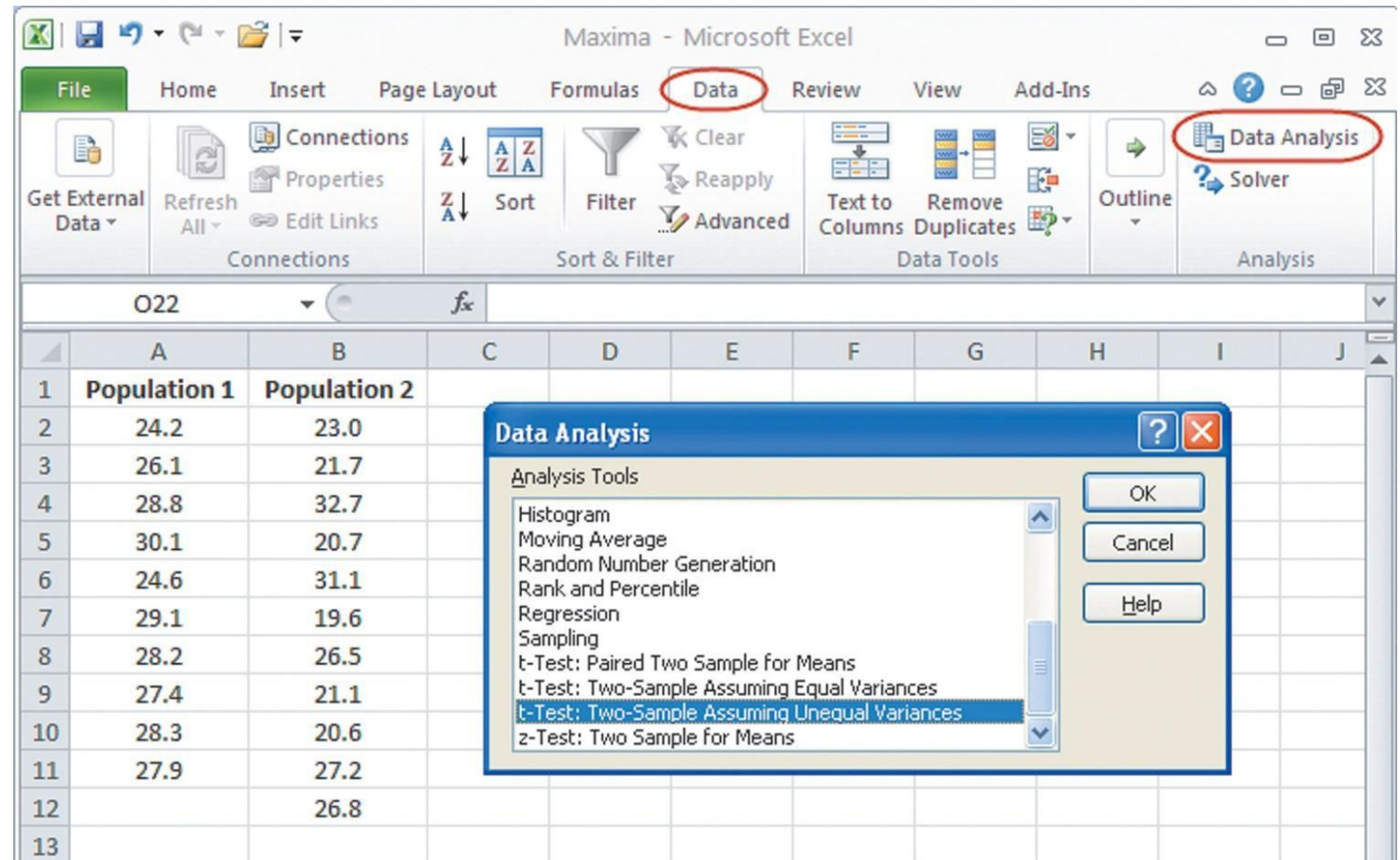
- Στατιστική συνάρτηση ελέγχου

$$t_{\bar{d}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Έλεγχος μέσων τιμών δύο ανεξάρτητων πληθυσμών

Υλοποίηση στο Excel

- Εισαγωγή δεδομένων των πληθυσμών σε διαφορετικές στήλες
- Από την καρτέλα **Data** επιλέγετε **Data Analysis**
- Επιλογή **t-test: Two-Sample Assuming Unequal Variances**



Έλεγχος μέσων τιμών δύο ανεξάρτητων πληθυσμών

Υλοποίηση στο Excel

	A	B	C	D	E	F	G	H	I
1	Population 1	Population 2							
2	24.2	23.0							
3	26.1	21.7							
4	28.8	32.7							
5	30.1	20.7							
6	24.6	31.1							
7	29.1	19.6							
8	28.2	26.5							
9	27.4	21.1							
10	28.3	20.6							
11	27.9	27.2							
12		26.8							
13									
14									
15									
16									

t-Test: Two-Sample Assuming Unequal Variances

Input

Variable 1 Range:

Variable 2 Range:

Hypothesized Mean Difference:

☒ Labels

Alpha:

Output options

☒ Output Range:

☐ New Worksheet Ply:

☐ New Workbook

OK Cancel Help

Έλεγχος μέσων τιμών δύο ανεξάρτητων πληθυσμών

Υλοποίηση στο Excel

	A	B	C	D	E	F
1	Population 1	Population 2		t-Test: Two-Sample Assuming Unequal Variances		
2	24.2	23.0				
3	26.1	21.7			Population 1	Population 2
4	28.8	32.7		Mean	27.47	24.63636364
5	30.1	20.7		Variance	3.729	20.24854545
6	24.6	31.1		Observations	10	11
7	29.1	19.6		Hypothesized Mean Difference	0	
8	28.2	26.5		df	14	
9	27.4	21.1		t Stat	1.904526432	
10	28.3	20.6		P(T<=t) one-tail	0.038797843	
11	27.9	27.2		t Critical one-tail	1.761310136	
12		26.8		P(T<=t) two-tail	0.077595685	
13				t Critical two-tail	2.144786688	
14						

Έλεγχος μέσων τιμών δύο εξαρτημένων πληθυσμών

- Για ένα δείγμα 603 εξαγορασμένων επιχειρήσεων εξετάστηκε η αποδοτικότητά τους πριν και μετά την εξαγορά
- Ερώτημα: πώς επηρέασε η εξαγορά την αποδοτικότητα των επιχειρήσεων;
- Δύο δείγματα (πριν & μετά την εξαγορά) που αφορούν τις ίδιες επιχειρήσεις

Firm ID	Πριν	Μετά	Διαφορά d
AT9010011	10,21	9,29	0,92
AT9070007	10,02	7,87	2,15
AT9090007	4,31	5,52	-1,21
AT9110048	-0,20	3,18	-3,38
AT9110242	34,82	35,70	-0,88
AT9110254	10,04	11,86	-1,82
AT9110441	-2,76	2,54	-5,30
AT9150003	21,58	16,78	4,81
.....			

Έλεγχος μέσων τιμών δύο εξαρτημένων πληθυσμών

- Μέση τιμή και τυπική απόκλιση διαφοράς

$$\bar{d} = \sum_{i=1}^n \frac{d_i}{n} = 0,75 \quad s_d = \sqrt{\frac{(d_i - \bar{d})^2}{n-1}} = 10,23$$

- Ορισμός υποθέσεων: $H_0: \mu_d = 0$ έναντι $H_1: \mu_d \neq 0$
- Έστω επίπεδο σημαντικότητας $\alpha = 0,05$
- Κανόνας απόφασης

$$t_{\bar{d}} = \frac{\bar{d}}{s_d/\sqrt{n}} = 1,809 < t_{n-1, 1-\alpha/2} = t_{602, 0,975} = 1,964$$

- Υπολογισμός Excel: =t.inv.2t(0,05;602)
- Δεν απορρίπτεται η H_0

Έλεγχος μέσων τιμών δύο εξαρτημένων πληθυσμών

Υλοποίηση στο Excel

- Εισαγωγή δεδομένων σε δύο στήλες
- Από την καρτέλα **Data** επιλέγετε **Data Analysis**
- Επιλογή **t-Test: Paired Two Sample for Means**

The screenshot shows the Microsoft Excel interface with the **Data** tab selected. The **Data Analysis** toolpak is installed, as evidenced by the **Data Analysis** button in the **Analyze** group. The **Data Analysis** dialog box is open, displaying a list of analysis tools. The tool **t-Test: Paired Two Sample for Means** is selected in the list. The background spreadsheet contains data for a t-test, with columns for Firm ID, ROA (t-1), and ROA (t+1).

	A	B	C	D	E	F	G	H	I
1	Firm ID	ROA (t-1)	ROA (t+1)						
2	AT9010011	10.21	9.29						
3	AT9070007	10.02	7.87						
4	AT9090007	4.31	5.52						
5	AT9110048	-0.20	3.18						
6	AT9110242	34.82	35.70						
7	AT9110254	10.04	11.86						
8	AT9110441	-2.76	2.54						
9	AT9150003	21.58	16.78						
10	AU0000172	12.82	16.25						
11	AU0040854	-1.90	2.33						
12	BE0202239	4.10	9.22						

Έλεγχος μέσων τιμών δύο εξαρτημένων πληθυσμών

Υλοποίηση στο Excel

	A	B	C	D
1	Firm ID	ROA (t-1)	ROA (t+1)	
2	AT9010011	10.21	9.29	
3	AT9070007	10.02	7.87	
4	AT9090007	4.31	5.52	
5	AT9110048	-0.20	3.18	
6	AT9110242	34.82	35.70	
7	AT9110254	10.04	11.86	
8	AT9110441	-2.76	2.54	
9	AT9150003	21.58	16.78	
10	AU0000172	12.82	16.25	
11	AU0040854	-1.90	2.33	
12	BE0202239	4.10	9.22	

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

Έλεγχος μέσων τιμών δύο εξαρτημένων πληθυσμών

Υλοποίηση στο Excel

	A	B	C
1	t-Test: Paired Two Sample for Means		
2			
3		ROA (t-1)	ROA (t+1)
4	Mean	3.223320066	2.46952073
5	Variance	110.1719852	133.8392454
6	Observations	603	603
7	Pearson Correlation	0.573402523	
8	Hypothesized Mean Difference	0	
9	df	602	
10	t Stat	1.808541782	
11	P(T<=t) one-tail	0.035510522	
12	t Critical one-tail	1.647388728	
13	P(T<=t) two-tail	0.071021045	
14	t Critical two-tail	1.963912434	

← Μονόπλευρος έλεγχος

← Αμφίπλευρος έλεγχος

Έλεγχοι ενός δείγματος για ποσοστά

- Δεδομένου του ποσοστού $\hat{p}$ από ένα δείγμα μεγέθους n πρέπει να εξεταστούν υποθέσεις για την πραγματική αναλογία p του πληθυσμού
 - Αριστερόπλευρος έλεγχος: $H_0: p = p_0$ έναντι $H_1: p < p_0$
 - Δεξιόπλευρος έλεγχος: $H_0: p = p_0$ έναντι $H_1: p > p_0$
 - Αμφίπλευρος έλεγχος: $H_0: p = p_0$ έναντι $H_1: p \neq p_0$
- Στατιστική συνάρτηση ελέγχου από την κανονική κατανομή

$$z_p = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

- Κανόνες απόφασης όπως στους προηγούμενους ελέγχους

Έλεγχοι αναλογιών (δύο ανεξάρτητα δείγματα)

- Δεδομένων των αναλογιών $\hat{p}_1$ και $\hat{p}_2$ από δύο δείγματα μεγέθους n_1 και n_2 πρέπει να εξεταστεί η σχέση των πραγματικών αναλογιών p_1, p_2 των πληθυσμών
 - Αριστερόπλευρος έλεγχος: $H_0: p_1 - p_2 = 0$ έναντι $H_1: p_1 - p_2 < 0$
 - Δεξιόπλευρος έλεγχος: $H_0: p_1 - p_2 = 0$ έναντι $H_1: p_1 - p_2 > 0$
 - Αμφίπλευρος έλεγχος: $H_0: p_1 - p_2 = 0$ έναντι $H_1: p_1 - p_2 \neq 0$
- Στατιστική συνάρτηση ελέγχου από την κανονική κατανομή
$$z_p = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\pi(1 - \pi)(1/n_1 + 1/n_2)}}, \quad \pi = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2}$$
- Κανόνες απόφασης όπως στους προηγούμενους ελέγχους