Ασκήσεις στις συνεχείς κατανομές

Πιθανότητες και Στατιστική

Τμήμα Πληροφορικής, Πανεπιστήμιο Πειραιώς

2023-2024

Άσκηση (Βλάβες σε ασανσερ)

Το κεντρικό ασανσέρ ενός εμπορικού κέντρου παθαίνει βλάβη κάθε τόσο, λόγω συχνής υπερφόρτωσης. Έχει παρατηρηθεί ότι ο χρόνος λειτουργίας μεταξύ δύο βλαβών ακολουθεί την εκθετική κατανομή με μέση τιμή 20 ημέρες.

i) Ποια είναι η πιθανότητα το ασανσερ να μην παρουσιάσει βλάβη για ένα μήνα;

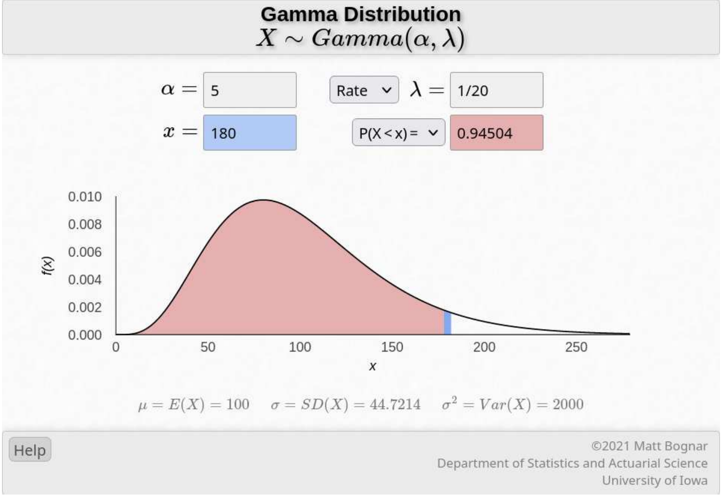
ii) Ποια είναι η πιθανότητα να παρουσιάσει λιγότερες από 5 βλάβες μέσα σε 180 ημέρες;

Λύση.
i) Ο χρόνος X λειτουργίας μεταξύ δύο βλαβών είναι ΤΜ που ακολουθεί την εκθετική κατανομή.
Γνωρίζουμε ότι $E(X) = 1/\lambda \Leftrightarrow 20 = 1/\lambda \Leftrightarrow \lambda = 1/20$.
Το ασανσερ δεν παρουσιάζει βλάβη για ένα μήνα όταν $X > 30$.

$$P(X > 30) = 1 - P(X \leq 30) = 1 - (1 - e^{-30\lambda}) = e^{-30/20} = 0.223.$$

ii) Έστω X_1, X_2, X_3, X_4, X_5 οι χρόνοι που μεσολαβούν μεταξύ της 1ης, 2ης, ..., 5ης βλάβης αντίστοιχα. Ο συνολικός χρόνος για να συμβούν και οι 5 βλάβες ισούται με $Y = X_1 + X_2 + X_3 + X_4 + X_5$.
Επειδή τα X_i, $i \in [5]$, ακολουθούν την εκθετική κατανομή $E(1/20)$ έπεται ότι η Y ακολουθεί την κατανομή Γάμμα με παραμέτρους $a = n = 5$ και $\theta = 1/\lambda = 1/20$, όπου η PDF της κατανομής Γάμμα δίδεται από τον τύπο: $f(x) = \begin{cases} \frac{\theta^a}{\Gamma(a)} x^{a-1} e^{-\theta x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$

Το ζητούμενο ενδεχόμενο προκύπτει όταν $Y > 180$. Χρησιμοποιώντας τις τιμές της CDF της κατανομής Γάμμα έχουμε ότι
$P(Y > 180) = 1 - P(Y \leq 180) = 1 - 0.94504 = 0.05496$



Άσκηση (Βάρη μαθητών)

Το 95% των μαθητών ενός σχολείου έχουν βάρος μικρότερο των 85 κιλών και το 10% μικρότερο των 55 κιλών. Αν το βάρος ακολουθεί κανονική κατανομή να βρεθεί το ποσοστό των μαθητών με βάρος μεγαλύτερο ή ίσο των 70 κιλών.

Λύση. Έστω X το βάρος ενός μαθητή του σχολείου. Η X ακολουθεί την κανονική κατανομή $N(\mu, \sigma^2)$. Θα υπολογίσουμε τα μ, σ^2 από τα δεδομένα της εκφώνησης.
Επειδή το 95% των μαθητών έχουν βάρος μικρότερο των 85 κιλών, έπεται ότι $P(X < 85) = 0.95$.
Το εργαλείο για τον υπολογισμό είναι η ιδιότητα ότι αν $X \sim N(\mu, \sigma^2)$ τότε $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$.

Βάσει αυτού:

$$P(X < 85) = P(X \leq 85) = P(\frac{X-\mu}{\sigma} \leq \frac{85-\mu}{\sigma}) = P(Z \leq \frac{85-\mu}{\sigma}) = 0.95$$

δηλαδή $\Phi(\frac{85-\mu}{\sigma}) = 0.95 \Leftrightarrow \Phi(\frac{85-\mu}{\sigma}) = \Phi(1.65)$.

Άρα, $\frac{85-\mu}{\sigma} = 1.65$ (1).

Επίσης, επειδή το 10% των μαθητών έχει βάρος μικρότερο των 55 κιλών, έπεται ότι $P(X < 55) = 0.1$ οπότε

$$P(X < 55) = P(X \leq 55) = P(\frac{X-\mu}{\sigma} < \frac{55-\mu}{\sigma}) = P(Z \leq \frac{55-\mu}{\sigma}) = 0.1$$

δηλαδή $\Phi(\frac{55-\mu}{\sigma}) = 0.1 \Leftrightarrow 1 - \Phi(\frac{\mu-55}{\sigma}) = 0.1 \Leftrightarrow \Phi(\frac{\mu-55}{\sigma}) = 0.90 \Leftrightarrow \Phi(\frac{\mu-55}{\sigma}) = \Phi(1.28)$.

Άρα, $\frac{\mu-55}{\sigma} = 1.28$ (2).

Λύνοντας το σύστημα (1) και (2) προκύπτει ότι

$$\mu = 65.29 \text{ και } \sigma = 11.94$$

Το ζητούμενο ποσοστό ισούται με την πιθανότητα

$$\begin{aligned}
P(X > 70) &= 1 - P(X \leq 70) \\
&= 1 - P(\frac{X - 65.29}{11.94} \leq \frac{70 - 65.29}{11.94}) \\
&= 1 - P(Z \leq 0.394) \\
&= 1 - \Phi(0.394) = 1 - \Phi(0.39) = 1 - 0.65 = 0.35
\end{aligned}$$

Άρα, περίπου το 35% των μαθητών έχει βάρος πάνω απο 70 κιλά.

Άσκηση (Υπόθεση πατρότητας)

Σε μια δίκη αναγνώρισης πατρότητας ο δικηγόρος του άνδρα καταθέτει ότι μπορεί να αποδείξει ότι ο πελάτης του ήταν εκτός χώρας σε ένα διάστημα μεταξύ της 290ης και 240ης ημέρας πριν την ημερομηνία γέννησης του παιδιού της πρώην συντρόφου του. Δεδομένου ότι ο χρόνος κύησης (σε μέρες) ακολουθεί την κανονική κατανομή με μέση τιμή $\mu = 270$ και $\sigma^2 = 100$, να εκτιμηθεί πόσο ισχυρό είναι το επιχείρημα που κατέθεσε ο δικηγόρος υπέρ του άνδρα.

Λύση. Έστω X ο χρόνος κύησης του εμβρύου. Αν ο άνδρας είναι ο πατέρας του παιδιού πρέπει $X > 290$ ή $X < 240$. Το ενδεχόμενο να συμβεί αυτό ισούται με

$$\begin{aligned}
&P(X > 290 \text{ ή } X < 240) = P(X > 290) + P(X < 240) \\
&= P\left(\frac{X-270}{10} > \frac{290-270}{10}\right) + P\left(\frac{X-270}{10} < \frac{240-270}{10}\right) \\
&= P(Z > 2) + P(Z < -3) = 1 - P(Z \leq 2) + P(Z \leq -3) \\
&= 1 - \Phi(2) + \Phi(-3) = 1 - \Phi(2) + 1 - \Phi(3) = 0.241.
\end{aligned}$$

Επομένως, η πιθανότητα το παιδί να έχει συλληφθεί στο διάστημα που απουσίαζε ο άντρας ισούται με

$P(240 \leq X \leq 290) = 1 - P(X > 290 \text{ ή } X < 240) = 1 - 0.241 = 0.759 = 76\%$

Άσκηση (Άθροισμα τιμών ομοιόμορφης κατανομής)

Μια γεννήτρια τυχαίων αριθμών παράγει ομοιόμορφα 100 τυχαίους αριθμούς στο διάστημα $(0, 1)$. Να βρεθεί η πιθανότητα το άθροισμα αυτών των αριθμών να είναι μεγαλύτερο ή ίσο του 52;

Λύση. Έστω X_1, X_2, ..., X_{100} οι τυχαίοι αριθμοί που παράγονται από την γεννήτρια. Για κάθε $i \in [100]$, ισχύει ότι η X_i ακολουθεί την ομοιόμορφη κατανομή $U(0, 1)$.
Υπενθυμίζεται ότι αν $X \sim U(a, b)$ τότε

$$E(X) = \frac{a+b}{2} \text{ και } Var(X) = \frac{(b-a)^2}{12}.$$

Επομένως, η X_i, $i \in [100]$, έχει

$$\text{μέση τιμή } \mu = \frac{0+1}{2} = 1/2 \text{ και διακύμανση } \sigma^2 = \frac{(1-0)^2}{12} = 1/12.$$

Έστω $S = X_1 + X_2 + \cdots + X_{100}$ το ζητούμενο άθροισμα. Ψάχνουμε την πιθανότητα $P(S > 52)$.

Υπενθυμίζεται ότι σύμφωνα με το Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ.) αν οι X_1, X_2, ..., X_n είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την ίδια κατανομή με μέση τιμή μ και διακύμανση σ^2 τότε η τ.μ. $X_1 + X_2 + \cdots + X_n$ ακολουθεί προσεγγιστικά την κατανομή $N(n \cdot \mu, n \cdot \sigma^2)$.

Από το Κ.Ο.Θ. προκύπτει ότι η κατανομή της S προσεγγίζεται από την κανονική κατανομή

$$N(\mu, \sigma^2) = N(100 \cdot \frac{1}{2}, 100 \cdot \frac{1}{12}) = N(50, \frac{100}{12})$$

Επομένως,

$$\begin{aligned} P(S \geq 52) &= 1 - P(S < 52) = 1 - P(S \leq 52) \\ &= 1 - P(\frac{S - 50}{\sqrt{100/12}} \leq \frac{52 - 50}{\sqrt{100/12}}) \\ &= 1 - P(Z \leq \frac{2\sqrt{3}}{5}) = 1 - \Phi(\frac{2\sqrt{3}}{5}) = 1 - \Phi(0.69) \\ &= 1 - 0.755 = 0.245. \end{aligned}$$

Άσκηση (Άθροισμα ενδείξεων ζαριού)

Ρίχνουμε ένα ζάρι 10 φορές. Να βρεθεί η πιθανότητα το άθροισμα των ενδείξεων να είναι μεταξύ του 30 και 40 (συμπεριλαμβανομένων των άκρων).

Λύση. Έστω X_i η ένδειξη της i-οστής ρίψης του ζαριού, $i \in [10]$ και έστω

$$S = X_1 + X_2 + \cdots + X_{10}.$$

Για να χρησιμοποιήσουμε το Κ.Ο.Θ. αρχικά θα βρούμε την μέση τιμή και την διακύμανση της X_i:
Για την μέση τιμή έχουμε:

$$E(X_i) = \frac{1+2+3+4+5+6}{6} = \frac{6 \cdot 7}{2 \cdot 6} = \frac{7}{2}.$$

Για την διακύμανση θα χρησιμοποιήσουμε τον τύπο:

$$\mathrm{Var}(X_i) = E(X_i^2) - E^2(X_i)$$

Ισχύει ότι

$$E(X_i^2) = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{91}{6} \text{ και } E^2(X_i) = \left(\frac{7}{2}\right)^2 = \frac{49}{4}$$

οπότε $\mathrm{Var}(X_i) = \dfrac{182 - 147}{12} = \dfrac{35}{12}$.

Από το Κεντρικό Οριακό Θεώρημα έπεται ότι η κατανομή της S προσεγγίζεται από την κανονική κατανομή

$$N(10 \cdot \frac{7}{2}, 10 \cdot \frac{35}{12}) = N(35, \frac{350}{12}),$$

οπότε

$$\begin{aligned} P(30 \le S \le 40) &= P\left(\frac{30-35}{\sqrt{350/12}} \le \frac{S-35}{\sqrt{350/12}} \le \frac{40-35}{\sqrt{350/12}}\right) \\ &= P(-5/(5\sqrt{7/6}) \le Z \le 5/(5\sqrt{7/6})) \\ &= P(-\sqrt{6/7} \le Z \le \sqrt{6/7}) \\ &= \Phi(\sqrt{6/7}) - \Phi(-\sqrt{6/7}) = 2\Phi(\sqrt{6/7}) - 1 \\ &= 2\Phi(0.93) - 1 = 2 \cdot 0.824 - 1 = 0.648. \end{aligned}$$

Άσκηση (Δεύτερο πιάτο στη Λέσχη)

Ο αριθμός των φοιτητών που ζητάνε δεύτερο πιάτο φαγητού στην φοιτητική λέσχη ακολουθεί την κατανομή Poisson με μέση τιμή 100. Ο μάγειρας έχει ετοιμάσει επιπλέον φαγητό που αντιστοιχεί σε 120 πιάτα. Να βρεθεί η πιθανότητα να μην φτάσουν οι επιπλέον μερίδες.

Λύση. Έστω X ο αριθμός των φοιτητών που ζητάνε δεύτερο πιάτο φαγητού.

Υπενθυμίζεται ότι όταν η X ακολουθεί την κατανομή Poisson με παράμετρο λ τότε

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

και

$$E(X) = \mathrm{Var}(X) = \lambda.$$

Επειδή $E(X) = 100$ έπεται ότι $\lambda = 100$, άρα η ζητούμενη πιθανότητα ισούται με

$$P(X > 120) = \sum_{k=120}^{\infty} e^{-100}\frac{100^k}{k!}.$$

Μπορούμε να προσεγγίσουμε την κατανομή Poisson χρησιμοποιώντας την επόμενη ιδιότητα:
Αν οι X_1, X_2, ..., X_n είναι ανεξάρτητες τ.μ. που ακολουθούν τις κατανομές Poisson με παραμέτρους λ_1, λ_2, ..., λ_n τότε η τ.μ. $X_1 + X_2 + \cdots + X_n$ ακολουθεί την κατανομή Poisson με παράμετρο $\lambda_1 + \lambda_2 + \cdots + \lambda_n$.
Αν θεωρήσουμε ότι η X είναι άθροισμα των ανεξάρτητων τ.μ. X_1, X_2, ..., X_{100} που ακολουθούν την κατανομή Poisson με παράμετρο $\lambda = 1$ τότε βάσει του Κ.Ο.Θ. η $X = X_1 + X_2 + \cdots + X_{100}$ προσεγγιστικά ακολουθεί την κανονική κατανομή $N(100 \cdot 1, 100 \cdot 1) = N(100, 100)$.
Άρα,

$$
\begin{aligned}
P(X > 120) &= 1 - P(X \leq 120) \\
&= 1 - P(\frac{X - 100}{\sqrt{100}} \leq \frac{120 - 100}{\sqrt{100}}) = 1 - P(Z \leq 2) \\
&= 1 - \Phi(2) = 1 - 0.977 = 0.023 = 2.3\%.
\end{aligned}
$$

Ασκήσεις

Άσκηση (Μέτρηση απόστασης άστρου)

Ένας αστρονόμος θέλει να υπολογίσει την απόσταση d (σε έτη φωτός) ενός μακρινού άστρου από την Γη. Σε κάθε μέτρηση της απόστασης υπεισέρχονται σφάλματα τα οποία ακολουθούν την κανονική κατανομή με διακύμανση $\sigma^2 = 4$ έτη φωτός. Γι' αυτό το λόγο θα κάνει πολλές μετρήσεις της απόστασης και θα θεωρήσει ως απόσταση την μέση τιμή των μετρήσεων.
Να βρεθεί ο αριθμός n των μετρήσεων που πρέπει να κάνει έτσι ώστε η εκτιμώμενη απόσταση να απέχει από την πραγματική απόσταση το πολύ ± 0.5 έτη φωτός με πιθανότητα τουλάχιστον 95%.

Λύση. Έστω $X_1, X_2, \ldots, X_n$ οι μετρήσεις των αποστάσεων του άστρου από την Γη.
Υποθέτουμε ότι $X_i \sim N(d, 4)$ για κάθε $i \in [n]$.
Θεωρούμε την τ.μ.

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Ψάχνουμε τον **ελάχιστο** n ώστε $P(\left|\overline{X} - d\right| \leq 0.5) \geq 0.95$.

Από το Κεντρικό Οριακό Θεώρημα προκύπτει ότι $\overline{X}_n \sim N(d, \frac{4}{n})$.
Άρα,

$$P(-0.5 \leq \overline{X}_n - d \leq 0.5) = 0.95 \Leftrightarrow P(\frac{-0.5}{\sqrt{4/n}} \leq \frac{\overline{X}_n - d}{\sqrt{4/n}} \leq \frac{0.5}{\sqrt{4/n}}) = 0.95$$

$$\Leftrightarrow P(-\frac{\sqrt{n}}{4} \leq Z \leq \frac{\sqrt{n}}{4}) = 0.95 \Leftrightarrow \Phi(\frac{\sqrt{n}}{4}) - \Phi(-\frac{\sqrt{n}}{4}) = 0.95$$

$$\Leftrightarrow 2\Phi(\frac{\sqrt{n}}{4}) - 1 = 0.95 \Leftrightarrow 2\Phi(\frac{\sqrt{n}}{4}) = 1.95 \Leftrightarrow \Phi(\frac{\sqrt{n}}{4}) = 0.975$$

$$\Leftrightarrow \Phi(\frac{\sqrt{n}}{4}) = \Phi(1.96) \Leftrightarrow \frac{\sqrt{n}}{4} = 1.96 \Leftrightarrow \sqrt{n} = 7.84 \Leftrightarrow n = 61.46$$

δηλαδή για να πετύχουμε την ζητούμενη ακρίβεια πρέπει να γίνουν τουλάχιστον 62 μετρήσεις της απόστασης.

Άσκηση (Instagram)

Υποθέστε ότι ο αριθμός των φωτογραφιών που δημοσιεύει στο instagram μια influencer μέσα σε μια μέρα είναι μια τυχαία μεταβλητή με μέση τιμή $\mu = 50$.
α) Τι μπορούμε να πούμε για την πιθανότητα ότι σε μια μέρα αναρτώνται πάνω από 75 φωτογραφίες;
β) Αν η διακύμανση σ^2 είναι γνωστό ότι ισούται με 25, τι μπορούμε να πούμε για την πιθανότητα ότι σε μια μέρα αναρτώνται μεταξύ 40 και 60 φωτογραφίες (χωρίς τα άκρα);

Λύση. Έστω X ο αριθμός των φωτογραφιών που δημοσιεύονται σε μια μέρα.
(Ανισότητα Markov) Υπενθυμίζεται ότι αν μια (διακριτή ή συνεχής) ΤΜ X λαμβάνει μόνο μη αρνητικές τιμές και έχει πεπερασμένη μέση τιμή $\mu = E(X)$, τότε

$$P(X \geq c) \leq \frac{\mu}{c}, \qquad \text{για κάθε πραγματική σταθερά } c > 0.$$

α) Από την ανισότητα Markov προκύπτει ότι

$$P(X > 75) \leq \frac{E(X)}{75} = \frac{50}{75} = \frac{2}{3}$$

(Ανισότητα Chebyshev) Υπενθυμίζεται ότι αν μια (διακριτή ή συνεχής) ΤΜ X έχει πεπερασμένη μέση τιμή μ και διακύμανση σ^2, τότε

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}, \qquad \text{για κάθε πραγματική σταθερά } c > 0.$$

β) Η πιθανότητα να μην συμβεί αυτό, δηλαδή $|X - 50| \geq 10$ μπορεί φράσεται από πάνω με την ανισότητα Chebyshev από όπου προκύπτει ότι

$$P(|X - 50| \geq 10) \leq \frac{\sigma^2}{10^2} = \frac{1}{4}$$

Άρα, η πιθανότητα συμβεί, δηλαδή $|X - 50| < 10$ φράσσεται από κάτω ως εξής:

$$P(|X - 50| < 10) = 1 - P(|X - 50| \geq 10) = 1 - \frac{1}{4} \geq \frac{3}{4}.$$

Άσκηση (Πάρτυ γενεθλίων στην Θεσσαλονίκη)

Σε ένα πάρτυ γενεθλίων έχουμε καλέσει 50 φίλους και έχουμε παραγγείλει 110 σάντουιτς (σουβλάκια με πίτα). Ο αριθμός των σάντουιτς που χρειάζεται κάθε άτομο για να χορτάσει ακολουθεί την κατανομή Poisson με μέση τιμή $\mu = 2$.

α) Να βρεθεί η πιθανότητα να μην φτάσουν τα σάντουιτς που παραγγείλαμε για να χορτάσουν όλα τα άτομα στο πάρτυ.

β) Πόσα σάντουιτς πρέπει να παραγγείλουμε ώστε με πιθανότητα 99% να χορτάσουν όλοι καλεσμένοι;

γ) Πόσους φίλους πρέπει να εξαιρέσουμε από την πρόσκληση ώστε με πιθανότητα 99% να χορτάσουν οι υπόλοιποι;

Λύση. Έστω X_i, $i \in [50]$, ο αριθμός των σάντουιτς που χρειάζεται ο i-οστός καλεσμένος για να χορτάσει.
Γνωρίζουμε ότι $X_i \sim Poisson(\lambda)$ όπου $\lambda = E(X) = \mathrm{Var}(X) = 2$.
Έστω S ο συνολικός αριθμός των σάντουιτς για να γιορτάσουν τα 50 άτομα, δηλαδή

$$S = X_1 + X_2 + \cdots + X_{50}$$

Βάσει του Κ.Ο.Θ. η S ακολουθεί προσεγγιστικά την κανονική κατανομή $N(50 \cdot 2, 50 \cdot 2) = N(100, 100)$.
α) Η ζητούμενη πιθανότητα είναι

$$P(S > 110) = 1 - P(S \leq 110) = 1 - P(\frac{S - 100}{\sqrt{100}} \leq \frac{110 - 100}{\sqrt{100}})$$
$$= 1 - \Phi(1) = 1 - 0.84 = 16\%.$$

β) Αν n είναι ο ζητούμενος αριθμός των σάντουιτς, ψάχνουμε τον ελάχιστο n ώστε

$$P(S \leq n) = 0.99 \Leftrightarrow P(\frac{S-100}{\sqrt{100}} \leq \frac{n-100}{\sqrt{100}}) = 0.99$$

$$\Leftrightarrow \Phi(\frac{n-100}{10}) = \Phi(2.33) \Leftrightarrow \frac{n-100}{10} = 2.33 \Leftrightarrow n = 124.$$

Άρα, πρέπει να παραγγείλει 14 σάντουιτς ακόμα.

γ) Έστω ότι καλεί k καλεσμένους. Ψάχνουμε τον μέγιστο k ώστε

$$P(S \leq 110) = 0.99 \text{ όπου } S = X_1 + X_2 + \cdots + X_k.$$

Βάσει του Κ.Ο.Θ. η S ακολουθεί προσεγγιστικά την κανονική κατανομή $N(k \cdot 2, k \cdot 2) = N(2k, 2k)$.
Επομένως,

$$P(S \leq 110) = 0.99 \Leftrightarrow P(\frac{S - 2k}{\sqrt{2k}} \leq \frac{110 - 2k}{\sqrt{2k}}) = 0.99$$

$$\Leftrightarrow \Phi(\frac{110 - 2k}{\sqrt{2k}}) = \Phi(2.33) \Leftrightarrow 110 - 2k = 2.33 \cdot \sqrt{2k} \Leftrightarrow n = 44.06.$$

Άρα, ο μέγιστος αριθμός καλεσμένων είναι $k = 44$, δηλαδή θα χρειαστεί να εξαιρέσει 6 άτομα από την πρόσκληση.