

# Διοίκηση Ποιότητας (Στατιστικός Έλεγχος Ποιότητας)

Σ. Μαλεφάκη  
Τμήμα Μηχανολόγων & Αεροναυπηγών Μηχανικών

2022–2023

## Εβδομαδιαίο Πρόγραμμα

### Θεωρία

▶ Τρίτη 13:00 - 15:00 (ΑΠ5)

### Ώρες γραφείου

▶ Τρίτη 11:00 - 13:00

▶ Τετάρτη 13:00 - 15:00

# Τρόποι Επικοινωνίας

☎: 2610 969486

🏠: Πολυώροφο κτήριο Μηχανολόγων και Αεροναυπηγών Μηχανικών,  
1ος όροφος, ΜΗ/Π102

✉: [smalefaki@upatras.gr](mailto:smalefaki@upatras.gr)

προσωπική σελίδα:

<http://www.des.upatras.gr/amm/smalefaki/index.htm>

σελίδα μαθήματος στο eclass:

<https://eclass.upatras.gr/courses/MECH1396/>

## ■ Βασικές Έννοιες & Αρχές Ποιότητας

Βασικοί ορισμοί και η σημασία, οι διαστάσεις & τα οφέλη της ποιότητας

## ■ Βασικά Εργαλεία της Διοίκησης Ποιότητας

ιστόγραμμα, φύλλα ελέγχου, διαγράμματα Παρετο, διαγράμματα διασποράς, διαγράμματα αιτίου/αποτελέσματος, διαγράμματα ροής, διαγράμματα ελέγχου και άλλα

## ■ Στατιστικός Έλεγχος Ποιότητας

Στατιστικός Έλεγχος Διεργασιών και Διαγράμματα Ελέγχου, διαγράμματα ελέγχου ιδιοτήτων, διαγράμματα ελέγχου μεταβλητών, Δειγματοληψία αποδοχής, χαρακτηριστική καμπύλη. Απλό πλάνο δειγματοληψίας, πολλαπλό πλάνο δειγματοληψίας Βασικοί πειραματικοί σχεδιασμοί Εφαρμογές με χρήση κατάλληλου λογισμικού.

## Προτεινόμενα Βιβλία

- ▶ Στατιστικός Έλεγχος Ποιότητας  
Μπερσίμης Σωτήριος, Ρακιτζής Αθανάσιος, Σαχλάς Αθανάσιος (2021)
- ▶ Στατιστικός έλεγχος ποιότητας  
Ταγαράς Γιώργος Ν. (2001)
- ▶ Διαχείριση και Στατιστικός Έλεγχος Ποιότητας,  
Χρήστος Κίτσος (2003)

## Προτεινόμενες σημειώσεις

- [http://www.unipi.gr/faculty/dantz/Statistical\\_Quality\\_Control.pdf](http://www.unipi.gr/faculty/dantz/Statistical_Quality_Control.pdf)

# Εισαγωγή στον Στατιστικό Έλεγχο Ποιότητας

Η έννοια της ποιότητας είναι άρρηκτα συνδεδεμένη με την καθημερινότητα μας.

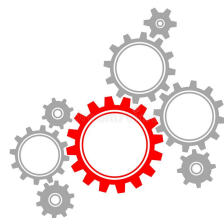
Κάθε άνθρωπος που αποφασίζει να αγοράσει ένα προϊόν (ή υπηρεσία) αντιμετωπίζει το πρόβλημα της επιλογής ανάμεσα σε ομοειδή προϊόντα (ή υπηρεσίες). Η επιλογή του προϊόντος από τον καταναλωτή καθορίζεται κυρίως από δύο παράγοντες, **την ποιότητα και την τιμή του**.

**Στόχος:** βέλτιστη δυνατή ποιότητα των παραγόμενων προϊόντων με το ελάχιστο δυνατό κόστος

Σημαντικό ρόλο στις διαδικασίες αυτές κατέχει ο Στατιστικός Έλεγχος Ποιότητας (Statistical Quality Control)

# Στατιστικός Έλεγχος Ποιότητας

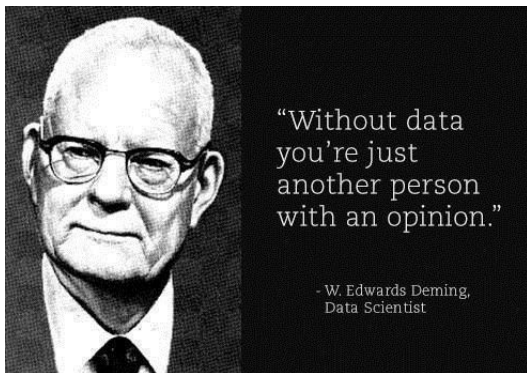
Ο Στατιστικός Έλεγχος Ποιότητας αποτελεί την παλαιότερη και γνωστότερη μέθοδο ελέγχου παραγωγικών διεργασιών για τη βελτίωση της ποιότητας των παραγόμενων προϊόντων. Ένας από τους βασικούς στόχους του είναι η έγκαιρη ανακάλυψη μη συμμορφωμένων με τις προδιαγραφές παραγόμενων προϊόντων η οποία σηματοδοτεί τη λήψη διορθωτικών ενεργειών για την απομάκρυνση των αιτιών που είναι υπεύθυνες για τις αποκλίσεις, συμβάλλοντας έτσι στη διατήρηση της ποιότητας των προϊόντων



# Η έννοια της ποιότητας

Edwards Deming (1900–1993)

Ποιότητα είναι, με χαμηλό κόστος, προβλέψιμη τυποποίηση και αξιοπιστία ενός προϊόντος, καθώς και η καταλληλότητα του για την αγορά.



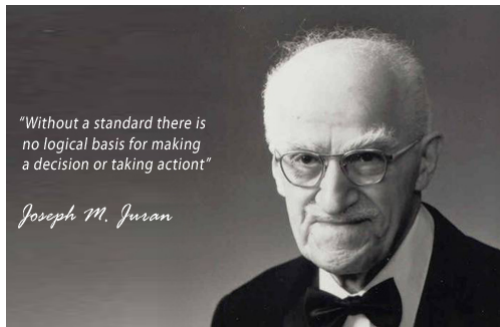


# Η έννοια της ποιότητας

Joseph Juran (1904–2008)

Ποιότητα ενός προϊόντος ή μιας υπηρεσίας είναι η καταλληλότητα του προς χρήση.

Ποιότητα είναι η προσαρμογή των χαρακτηριστικών ενός προϊόντος ή υπηρεσίας στις απαιτήσεις /προσδοκίες του πελάτη/χρήστη καταναλωτή.

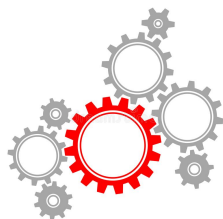


## Η έννοια της ποιότητας

Έτσι η ποιότητα μπορεί να 'μετρηθεί' μέσω της ικανοποίησης που προσφέρει το προϊόν στον καταναλωτή.

Ένας νεότερος ορισμός ορίζει την ποιότητα ως αντιστρόφως ανάλογη της μεταβλητότητας των χαρακτηριστικών της παραγωγικής διαδικασίας που προσδιορίζουν την ποιότητα του προϊόντος.

Ο Deming, υποστήριξε τη μείωση της μεταβλητότητας στις διαδικασίες ως τον πιο ενδεδειγμένο τρόπο για τη βελτίωση της ποιότητας.



## Οι διαστάσεις της ποιότητας

ΑΠΟΔΟΣΗ	Το προϊόν κάνει τη δουλειά για την οποία προορίζεται.
ΑΞΙΟΠΙΣΤΙΑ	Το προϊόν χρειάζεται συχνή επισκευή
ΔΙΑΡΚΕΙΑ	Η διάρκεια ζωής του προϊόντος
ΕΠΙΣΚΕΥΗ	Πόσο γρήγορη και οικονομική είναι η επισκευή του προϊόντος στην περίπτωση που εμφανιστεί βλάβη
ΑΙΣΘΗΤΙΚΗ	Πόσο ικανοποιητικό είναι από άποψη εμφάνισης (χρώμα, σχήμα, περιτύλιγμα, κτλ.) το προϊόν.
ΔΥΝΑΤΟΤΗΤΕΣ	Ποιες είναι οι επιπρόσθετες δυνατότητες του προϊόντος
ΦΗΜΗ ΕΠΙΧΕΙΡΗΣΗΣ	Ποια είναι η φήμη της εταιρείας
ΣΥΜΜΟΡΦΩΣΗ ΜΕ ΤΙΣ ΠΡΟΔΙΑΓΡΑΦΕΣ	Το προϊόν κατασκευάστηκε σύμφωνα με τις προδιαγραφές που έθεσε ο σχεδιαστής του

# Συνιστώσες του Στατιστικού Ελέγχου Ποιότητας

Ο Στατιστικός Έλεγχος Ποιότητας αποτελείται από ένα σύνολο μεθόδων στατιστικής ανάλυσης δεδομένων. Το σύνολο αυτό μπορεί να χωριστεί σε τρία βασικά υποσύνολα που το καθένα περιέχει στατιστικές μεθόδους προσανατολισμένες σε διαφορετικές φάσεις της παραγωγικής διαδικασίας.

Τα τρία υποσύνολα είναι τα ακόλουθα:

- Σχεδιασμός και Ανάλυση Πειραμάτων (Design of Experiments)
- Στατιστικός Έλεγχος Διεργασιών (Statistical Process Control)
- Δειγματοληψία Αποδοχής (Acceptance Sampling)

# Συνιστώσες του Στατιστικού Ελέγχου Ποιότητας

- ▶ Ο **Σχεδιασμός και η Ανάλυση Πειραμάτων** περιέχει όλες εκείνες τις στατιστικές τεχνικές οι οποίες μας βοηθούν στην ανακάλυψη της επίδρασης που έχουν τα διάφορα επίπεδα των παραγόντων (μεταβλητών) που επηρεάζουν τις ποιοτικές παραμέτρους του τελικού προϊόντος και συνεπώς διαδραματίζει σημαντικό ρόλο στη βέλτιστη σχεδίαση της παραγωγικής διεργασίας.
- ▶ Ο **Στατιστικός Έλεγχος Διεργασιών** περιέχει στατιστικές τεχνικές που είναι απαραίτητες για τον έλεγχο της παραγωγικής διεργασίας κατά τη διάρκεια της παραγωγής των προϊόντων.
- ▶ Η **Δειγματοληψία Αποδοχής** περιέχει στατιστικές τεχνικές (δειγματοληπτικές) που είναι απαραίτητες για να αποφασίσουμε αν μια συγκεκριμένη παρτίδα προϊόντων θα γίνει δεκτή ή θα απορριφθεί.

# Στατιστικός έλεγχος Διεργασιών

## Βασικός Σκοπός:

Η μέτρηση και η βελτίωση της ποιότητας των διεργασιών.

Η ανίχνευση τυχόν αλλαγών στη μέση τιμή και στην τυπική απόκλιση των χαρακτηριστικών μιας διεργασίας.

Ουσιαστικά, μας ενδιαφέρει να δούμε αν η διεργασία έχει αλλάξει από τη στιγμή που πιστοποιήθηκε.

Στην πραγματικότητα κάνουμε ένα διαρκή έλεγχο υπόθετων ως προς το αν το δείγμα που έχουμε λάβει ανήκει στον ίδιο πληθυσμό με τα δείγματα που χρησιμοποιούμε για να πιστοποιηθεί η συγκεκριμένη διεργασία.

Εφαρμογή στατιστικών μεθόδων για τη μέτρηση και την ανάλυση της μεταβλητότητας στις τιμές χαρακτηριστικών της διεργασίας (Στατιστικός Έλεγχος Ποιότητας).

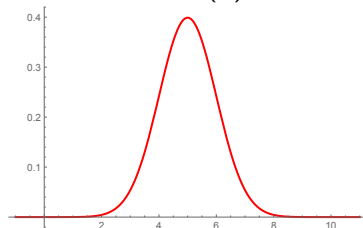
## Παράδειγμα

Ένα κοπτικό μηχάνημα σε μία διεργασία κόβει δισκία χαλαζία διαμέτρου  $8.00\text{mm}$ . Η κατανομή των τιμών της διαμέτρου είναι κανονική με τυπική απόκλιση  $0.01\text{mm}$ . Έστω ΑΟΠ και ΚΟΠ το ανώτερο και το κατώτερο όριο προδιαγραφών αντιστοίχως. Επιθυμούμε να κόψουμε δισκία με μέση τιμή διαμέτρου  $8.00\text{mm}$ , ποιες πρέπει να είναι οι προδιαγραφές του ελέγχου για τη διάμετρο του χαλαζία, ώστε να γίνεται αποδεκτό κατά μέσο όρο το 99.73% των δισκίων;

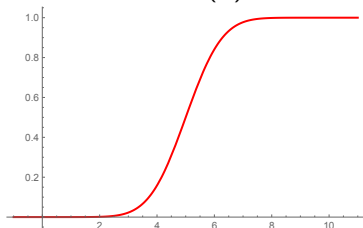
# Κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\}, \quad -\infty < x < \infty$$

σππ  $f(x)$



ασκ  $F(x)$



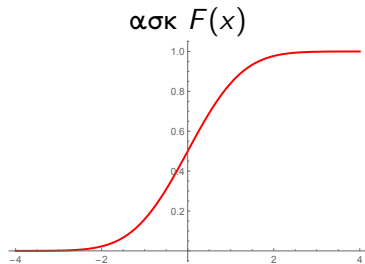
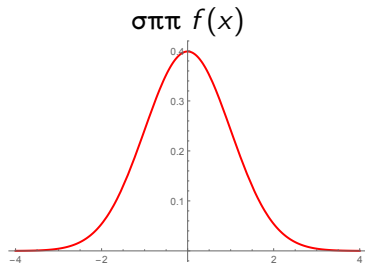
- $E(X) = \mu$
- $Var(X) = \sigma^2$
- $\alpha_3 = 0$  συμμετρική κατανομή
- $\alpha_4 = 3$  μεσόκυρτη κατανομή



# Τυπική Κανονική κατανομή $\mathcal{N}(0, 1)$

Αν  $\mu = 0$  και  $\sigma^2 = 1$  τότε έχουμε την τυπική κανονική κατανομή

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}x^2 \right\}, \quad -\infty < x < \infty$$



- $E(X) = 0$
- $Var(X) = 1$

# Κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$

$$F(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(u - \mu)^2\right\} du$$

Δεν μπορεί να βρεθεί αναλυτικά!

Υπάρχουν πίνακες που μας δίνουν την πιθανότητα  $P(X \leq x)$  για  $X \sim \mathcal{N}(0, 1)$  ( $P(X \leq x) = \Phi(x)$ )

z	0	0.01	0.02	0.03	0.04	0.05
0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894
1	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943

## Κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$

Πώς όμως βρίσκουμε πιθανότητες για μία οποιαδήποτε τ. μ.  
 $X \sim \mathcal{N}(\mu, \sigma^2)$ ;;;

### Θεώρημα

Αν  $X$  τ.μ. με  $X \sim \mathcal{N}(\mu, \sigma^2)$ , τότε η τυποποιημένη τ. μ.  $Z = \frac{X-\mu}{\sigma}$  ακολουθεί την τυπική κανονική κατανομή.

Άρα να θέλουμε να βρούμε τη πιθανότητα  $P(\alpha < X < \beta)$ ,  $X \sim \mathcal{N}(\mu, \sigma^2)$  τυποποιούμε την  $X$ , δηλαδή

$$\begin{aligned} P(\alpha < X < \beta) &= P\left(\frac{\alpha - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{\beta - \mu}{\sigma}\right) \\ &= P\left(\frac{\alpha - \mu}{\sigma} < Z < \frac{\beta - \mu}{\sigma}\right) \\ &= F\left(\frac{\beta - \mu}{\sigma}\right) - F\left(\frac{\alpha - \mu}{\sigma}\right) \end{aligned}$$

## Παράδειγμα

Ο υπεύθυνος παραγωγής γνωρίζει ότι το βάρος των συσκευασιών που παράγονται ακολουθεί την Κανονική κατανομή με μέση τιμή 400 γραμμάρια και τυπική απόκλιση 5 γραμμάρια. Επίσης, γνωρίζει ότι οι ανοχές που επιτρέπουν οι προδιαγραφές που θέτει το τμήμα σχεδιασμού είναι ίσες με 385 και 415 αντίστοιχα. Ο υπεύθυνος παραγωγής θέλει να υπολογίσει ποια είναι η πιθανότητα ένα προϊόν που επιλέγει να είναι εντός προδιαγραφών.

## Κανονική κατανομή $\mathcal{N}(\mu, \sigma^2)$

### Θεώρημα

Έστω  $X_1, X_2, \dots, X_n$  ανεξάρτητες τμ με  $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ . Τότε οποιοσδήποτε γραμμικός συνδυασμός  $c_1 X_1 + \dots + c_n X_n$  έχει επίσης κανονική κατανομή με μέση τιμή

$$E\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i \mu_i$$

και διασπορά

$$\text{Var}\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i^2 \sigma_i^2$$

## Άλλες συνεχείς κατανομές

- Ομοιόμορφη κατανομή
- Εκθετική κατανομή
- Κατανομή Γαμμά
- Weibull κατανομή
- $\chi^2$  κατανομή
- Κατανομή  $t$

## ...συνέχεια παραδείγματος...

Πολλές φορές δεν ενδιαφερόμαστε μόνο για το ποια είναι η πιθανότητα κάποιο δισκίο χαλαζία να μην πληρεί τις προδιαγραφές, αλλά να μας ενδιαφέρει η πιθανότητα

- Το πρώτο δισκίο χαλαζία που θα είναι εκτός προδιαγραφών να είναι το δέκατο που θα εξετάσουμε.
- Από μια συσκευασία δέκα δισκίων χαλαζία τα δύο να είναι εκτός προδιαγραφών
- Πέντε διαδοχικά παραγόμενα δισκία κανένα να μην είναι εκτός προδιαγραφών

# Κατανομή Bernoulli

- πειράματα που έχουν δύο δυνατά αποτελέσματα (επιτυχία - αποτυχία)
- $X \in \{0, 1\}$
- $p$ : η πιθανότητα επιτυχίας
- $f(x) = p^x (1 - p)^{1-x}$ ,  $x \in \{0, 1\}$
- $E(X) = p$
- $Var(X) = p(1 - p)$



## Διωνυμική Κατανομή $B(n, p)$

Εκτελούμε  $n$  ανεξάρτητα πειράματα Bernoulli

- $X$  τμ η οποία δηλώνει πόσες επιτυχίες υπάρχουν στην ακολουθία των  $n$  ανεξάρτητων δοκιμών Bernoulli.
- $X \in \{0, 1, \dots, n\}$
- $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$
- $E(X) = np$
- $Var(X) = np(1-p)$

## Γεωμετρική Κατανομή $\mathcal{G}(p)$

Σε αυτή την περίπτωση εκτελούνται ανεξάρτητα πειράματα Bernoulli μέχρι να πραγματοποιηθεί επιτυχία για πρώτη φορά

- $X$ : τμ η οποία δηλώνει τον αριθμό των δοκιμών Bernoulli μέχρι να έχουμε για πρώτη φορά επιτυχία.
- $X \in 1, 2, \dots$
- $f(x) = p(1 - p)^{x-1}$
- $E(X) = p^{-1}$   
Περίοδος επαναφοράς
- $Var(X) = \frac{1-p}{p^2}$

## Ιδιότητα Έλλειψης Μνήμης

Εάν το γεγονός  $A$  δεν συμβεί μέχρι τη δοκιμή  $k$  τότε η πιθανότητα ότι δε θα συμβεί στις επόμενες  $\lambda$  δοκιμές είναι ίδια με την πιθανότητα ότι το γεγονός  $A$  δεν θα συμβεί στις πρώτες  $\lambda$  δοκιμές, δηλαδή

$$P(X > k + \lambda | X > k) = P(X > \lambda)$$

**Απόδειξη:**

Είναι η μοναδική διακριτή κατανομή με αυτή την ιδιότητα.

## Αρνητική Διωνυμική Κατανομή (Pascal) $\mathcal{NB}(r, p)$

Γενίκευση της Γεωμετρικής κατανομής

- $X$  τμ η οποία παριστάνει τον αριθμό δοκιμών που απαιτούνται μέχρι να συμπληρωθούν ακριβώς  $r$  επιτυχίες για πρώτη φορά.
- η  $X$  μπορεί να ερμηνευτεί σαν το άθροισμα  $r$  γεωμετρικών τμ.
- $X \in r, r + 1, \dots$
- $f(x) = \binom{r+x-1}{r-1} p^r (1-p)^x$
- $E(X) = \frac{rp}{1-p}$
- $Var(X) = \frac{rp}{(1-p)^2}$

## Υπεργεωμετρική κατανομή $HG(N, n, k)$

Όταν κάνουμε δειγματοληψία χωρίς επανατοποθέτηση από έναν πεπερασμένο πληθυσμό με έναν μέρος του να έχει κάποιο χαρακτηριστικό που μας ενδιαφέρει

- $X$ : τμ που μετράει τον αριθμό των επιτυχιών σε ένα δείγμα προκαθορισμένου μεγέθους
- $N$ : το μέγεθος του πληθυσμού
- $k$ : το πλήθος των ατόμων του πληθυσμού που έχουν το χαρακτηριστικό που μας ενδιαφέρει
- $n$ : το μέγεθος του δείγματος που παίρνω
- $X \in \{\max\{0, n + k - N\}, \dots, \min\{k, n\}\}$
- 

$$P(X = x) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$$

# Κατανομή Poisson $\mathcal{P}(\lambda)$

$X$  παριστάνει τον αριθμό πραγματοποίησης του  $A$  σε ένα ορισμένο χρονικό ή τοπικό διάστημα

- αριθμός ατυχημάτων σε ένα εργοστάσιο μέσα σε 6 μήνες
- αριθμός σεισμών σε μία χώρα μέσα σε ένα χρόνο
- αριθμός βλαβών σε ένα ηλεκτρικό σύστημα μέσα σε ένα μήνα λειτουργίας του.
- αριθμός κλίσεων σε ένα τηλεφωνικό κέντρο μέσα σε 20 λεπτά.

ή εναλλακτικά

- αριθμός τυπογραφικών λαθών ανά σελίδα
- αριθμός αυτοκινητιστικών ατυχημάτων σε ένα διάστημα 100 χιλιομέτρων
- αριθμός ραγισμάτων ανά τετραγωνικό μέτρο σε μία μεταλλική κατασκευή

# Κατανομή Poisson $\lambda$

- $X \in \{0, 1, \dots\}$
- $f(x) = e^{-\lambda} \frac{\lambda^x}{x!}$
- $E(X) = \lambda$  (μέσος αριθμός εμφάνισης του A στην στη μονάδα του χρόνου ή του χώρου, με άλλα λόγια η συχνότητα εμφάνισης του A)
- $Var(X) = \lambda$
- Η κατανομή Poisson μπορεί να χρησιμοποιηθεί σαν μία προσέγγιση της Διωνυμική  $B(n, p)$  με  $\lambda = np$ , όταν το  $n$  είναι πολύ μεγάλο και το  $p$  πολύ μικρό. Στην πράξη χρησιμοποιούμε αυτή την προσέγγιση όταν  $n > 30$  και  $np < 5$ .

## Πολυωνυμική κατανομή

Έστω ένα πείραμα τύχης με  $k$  δυνατά αποτελέσματα (αμοιβαία αποκλειόμενα)  $A_1, \dots, A_k$ . Έστω  $p_i$ ,  $i = 1, \dots, k$  η πιθανότητα πραγματοποίησης του  $A_i$  η οποία είναι σταθερή σε κάθε εκτέλεση του πειράματος και  $\sum_{i=1}^k p_i = 1$ .

$X_1, \dots, X_k$  τμ που παριστάνουν τον αριθμό πραγματοποίησης των γεγονότων  $A_1, \dots, A_k$  αντίστοιχα στις  $n$  επαναλήψεις του πειράματος τύχης.

Οι  $X_1, \dots, X_k$  δεν είναι ανεξάρτητες αφού  $X_1 + \dots + X_k = n$ .

$$P(X_1 = x_1, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}, \quad \sum_{i=1}^k x_i = n$$

$k = 2$  Διωνυμική κατανομή

Η κάθε μία μεταβλητή  $X_i$  μπορεί να θεωρηθεί σαν Διωνυμική κατανομή με πιθανότητα επιτυχίας  $p_i$ .



## Συνοψίζοντας...

- **Διωνυμική:**
  - Άθροισμα  $n$  ανεξάρτητων κατανομών Bernoulli
  - Ειδική περίπτωση της Πολυωνυμικής για  $k = 2$ .
- **Γεωμετρική:** Ιδιότητα έλλειψης μνήμης.
- **Αρνητική Γεωμετρική:** Άθροισμα  $r$  ανεξάρτητων γεωμετρικών κατανομών.
- **Υπεργεωμετρική:** Όταν το  $N$  είναι πολύ μεγάλο και το  $n$  πολύ μικρό (στην πράξη το  $n$  είναι μικρό όταν  $n < 0.05N$  προσεγγίζεται από τη Διωνυμική κατανομή).
- **Poisson:** Προσέγγιση της Διωνυμικής όταν το  $n$  είναι μεγάλο και το  $p$  μικρό.

# Κατανομή του $\bar{X}$

## Θεώρημα

Έστω  $X_1, X_2, \dots, X_n$  ένα τυχαίο δείγμα από κανονικά κατανεμημένο πληθυσμό με πεπερασμένη μέση τιμή  $\mu$  και διασπορά  $\sigma^2$  τότε

- $E(\bar{X}) = \mu$  και  $Var(\bar{X}) = \sigma^2/n$
- $E(S^2) = \sigma^2$
- $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$

Τι γίνεται όμως όταν δεν έχουμε κανονικά κατανεμημένους πληθυσμούς; ☰



# Κατανομή του $\bar{X}$

## Θεώρημα

Έστω  $X_1, X_2, \dots, X_n$  ένα τυχαίο δείγμα από κανονικά κατανεμημένο πληθυσμό με πεπερασμένη μέση τιμή  $\mu$  και διασπορά  $\sigma^2$  τότε

- $E(\bar{X}) = \mu$  και  $Var(\bar{X}) = \sigma^2/n$
- $E(S^2) = \sigma^2$
- $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$

## Θεώρημα

Έστω  $X_1, X_2, \dots, X_n$  τ. μ. από κανονική κατανομή με μέση τιμή  $\mu$  και διασπορά  $\sigma^2$ , τότε

- $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$

Τι γίνεται όμως όταν δεν έχουμε κανονικά κατανεμημένους πληθυσμούς; ☰ ↻ 🔍

## Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ.)

Κ.Ο.Θ.

Έστω  $X_1, X_2, \dots, X_n$  τυχαίο δείγμα από μία κατανομή με μέση τιμή  $\mu$  και διασπορά  $\sigma^2$ . Αν το  $n$  είναι αρκετά μεγάλο (για  $n \rightarrow \infty$ ) τότε

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1) \text{ προσεγγιστικά}$$

ή ισοδύναμα

$$\frac{\sum X_i - n\mu}{\sigma \sqrt{n}} \sim \mathcal{N}(0, 1) \text{ προσεγγιστικά}$$

Με βάση το Κ.Ο.Θ έχουμε ότι η δειγματική μέση τιμή ακολουθεί προσεγγιστικά την κανονική κατανομή ανεξάρτητα από την κατανομή του πληθυσμού που παίρνουμε το δείγμα.

## Έλεγχοι Υποθέσεων

Ένα κοπτικό μηχάνημα σε μία διεργασία κόβει δισκία χαλαζία διαμέτρου  $8.00\text{mm}$ . Η κατανομή των τιμών της διαμέτρου είναι κανονική με τυπική απόκλιση  $0.01\text{mm}$ . Έστω ΑΟΠ και ΚΟΠ το ανώτερο και το κατώτερο όριο προδιαγραφών αντιστοίχως. Επιθυμούμε να κόψουμε δισκία με μέση τιμή διαμέτρου  $8.00\text{mm}$ , ποιες πρέπει να είναι οι προδιαγραφές του ελέγχου για τη διάμετρο του χαλαζία, ώστε να γίνεται αποδεκτό κατά μέσο όρο το  $99.73\%$  των δισκίων; Ελέγχουμε ένα τυχαίο δείγμα 100 δισκίων χαλαζία και παίρνουμε μέση τιμή  $7.73\text{mm}$ . Με βάση αυτά τα δεδομένα μπορούμε να ισχυριστούμε ότι η μέση διάμετρος έχει αλλάξει σε επίπεδο σημαντικότητας  $0.0027$ ;

# Έλεγχοι Υποθέσεων

Η διαδικασία του στατιστικού ελέγχου

- 1 Ορίζουμε τη μηδενική υπόθεση ( $H_0 : \theta = \theta_0$ )
- 2 Ορίζουμε την εναλλακτική υπόθεση ( $H_1 : \theta \neq \theta_0$  ή  $H_1 : \theta < \theta_0$  ή  $H_1 : \theta > \theta_0$ )
- 3 Υπολογίζουμε την τιμή της στατιστικής συνάρτησης στο δείγμα μας
- 4 Ορίζουμε την περιοχή απόρριψης της  $H_0$
- 5 Εξάγουμε τα συμπεράσματα μας.

# Σφάλματα

- ❶ **Σφάλμα τύπου I** (να απορρίψουμε τη μηδενική υπόθεση ενώ είναι σωστή)

$$\alpha = P(\text{απόρριψη } H_0 | H_0 \text{ σωστή})$$

$\alpha$  : επίπεδο σημαντικότητας ή μέγεθος του ελέγχου

- ❷ **Σφάλμα τύπου II** (να αποδεχτούμε τη μηδενική υπόθεση ενώ είναι λάθος)

$$\beta = P(\text{αποδοχή } H_0 | H_0 \text{ λάθος})$$

Η πιθανότητα σφάλματος τύπου II εξαρτάται από την ακριβή τιμή της παραμέτρου  $\theta$  που ισχύει στην πραγματικότητα, άρα  $\beta(\theta)$  με πεδίο ορισμού την περιοχή τιμών της  $\theta$  που καθορίζει η  $H_1$ . Η συνάρτηση  $\beta(\theta)$  λέγεται χαρακτηριστική καμπύλη του ελέγχου.

- ❸ **Ισχύς** ενός ελέγχου ονομάζεται η πιθανότητα απόρριψης της  $H_0$  όταν η  $H_0$  είναι πραγματικά λάθος.

$$\gamma = P(\text{απόρριψη } H_0 | H_0 \text{ λάθος}) = 1 - \beta$$

$$\gamma(\theta) = 1 - \beta(\theta)$$

# Σφάλματα

- 1 Επιλέγουμε και καθορίζουμε το  $\alpha$  και βάση αυτού ορίζεται η περιοχή απόρριψης της μηδενικής υπόθεσης. Το  $\alpha$  είναι συνήθως πολύ μικρό
- 2 Καθορίζοντας κατάλληλο μέγεθος δείγματος, εκτός της προκαθορισμένης τιμής του  $\alpha$ , μπορούμε να προκαθορίσουμε και το  $\beta(\theta)$
- 3 Παρατηρούμενο επίπεδο σημαντικότητας ή  $p$  τιμή ( $p$ -value) του ελέγχου ονομάζεται η πιθανότητα να παρατηρηθεί μία τιμή της στατιστικής συνάρτησής μεγαλύτερη (πιο ακραία) από αυτή που έδωσε το συγκεκριμένο δείγμα

$$P(Z > |z| | H_0 \text{ σωστή})$$

όπου  $z$  η τιμή του στατιστικού για το συγκεκριμένο δείγμα.

Η παραπάνω πιθανότητα αναφέρεται σε μονόπλευρους ελέγχους ενώ για δίπλευρους ελέγχους διπλασιάζεται



## Γενική ιδέα

Έστω ότι θέλουμε να ελέγξουμε την υπόθεση

$$H_0 : \theta = \theta_0 \quad \text{κατά} \quad H_1 : \theta \neq \theta_0$$

Ο πληθυσμός με βάση κάποιον κανόνα χωρίζεται σε δύο περιοχές, στην περιοχή απόρριψης της  $H_0$  και στην περιοχή μη απόρριψης της. Κατόπιν, γίνεται ο έλεγχος βάσει του κριτηρίου που ορίστηκε και αποφασίζεται αν η  $H_0$  απορρίπτεται ή όχι.

Έστω  $\hat{\theta}$  εκτιμητής του  $\theta$

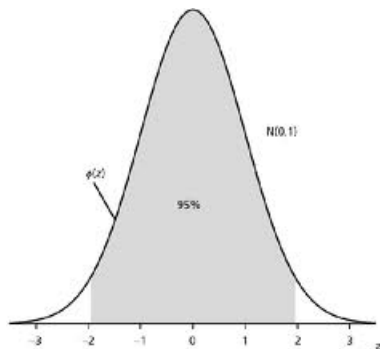
Ξέρουμε ότι  $\hat{\theta} \sim \mathcal{N}(\theta, \tau^2)$

Υπό την μηδενική υπόθεση έχουμε

$$\hat{\theta} \sim \mathcal{N}(\theta_0, \tau^2) \Rightarrow Z = \frac{\hat{\theta} - \theta_0}{\tau} \sim \mathcal{N}(0, 1)$$

Άρα αν η  $H_0$  είναι σωστή, δηλαδή αν το δείγμα μας προέρχεται από έναν πληθυσμό με μέση τιμή  $\theta_0$  τότε η μεταβλητή  $Z$  θα παίρνει τιμές με μεγάλη πιθανότητα  $(1 - \alpha)$  γύρω από το 0.

# Γενική ιδέα



Ποια είναι η περιοχή απόρριψης όταν η εναλλακτική υπόθεση είναι

$$H_1 : \theta > \theta_0?$$

Ποια είναι η περιοχή απόρριψης όταν η εναλλακτική υπόθεση είναι

$$H_1 : \theta < \theta_0?$$

# Έλεγχος υπόθεσης για το μέσο $\mu$ ενός πληθυσμού

με γνωστή διασπορά

$$H_0 : \mu = \mu_0 \quad \text{κατά} \quad H_1 : \mu \neq \mu_0 \quad \text{ή} \quad \mu < \mu_0 \quad \text{ή} \quad \mu > \mu_0$$

$$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n) \quad \text{ή ισοδύναμα} \quad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

- 1 αν  $H_1 : \mu \neq \mu_0$  τότε απορρίπτω τη μηδενική υπόθεση αν  $z < -z_{\alpha/2}$  ή  $z > z_{\alpha/2}$
- 2 αν  $H_1 : \mu > \mu_0$  τότε απορρίπτω τη μηδενική υπόθεση αν  $z > z_{\alpha}$
- 3 αν  $H_1 : \mu < \mu_0$  τότε απορρίπτω τη μηδενική υπόθεση αν  $z < -z_{\alpha}$