

Εισαγωγή στην οικονομετρία (1404)

Διάλεξη 3

Άγγελος Αλεξόπουλος
Οικονομικό Πανεπιστήμιο Αθηνών -
Τμήμα Οικονομικής Επιστήμης
angelos@aueb.gr

Περιγραμμά Διάλεξης

➤ Βασικές Έννοιες Στατιστικής

- Τυχαίες μεταβλητές; Διακριτές τυχαίες μεταβλητές; Συνεχείς τυχαίες μεταβλητές; pdf; cdf; Πολυμεταβλητή κατανομή; Υπό συνθήκη κατανομή; Ανεξαρτησία
- Χαρακτηριστικά τυχαίων μεταβλητών (Μέση Τιμή - Διακύμανση – Ροπές)
- Συνδιακύμανση, Συντελεστής συσχέτισης
- Υπό συνθήκη μέσος και διακύμανση; Νόμος επαναλαμβανόμενων προσδοκιών

Επανάληψη – Μέρος 1^ο

Βασικές Έννοιες Στατιστικής

Τυχαία μεταβλητή

- Τυχαία μεταβλητή X (random variable, rv): η συνάρτηση που απεικονίζει τα δυνατά αποτελέσματα (ενδεχόμενα) ενός πειράματος τύχης
 - Διακριτή (discrete): αριθμήσιμο πλήθος τιμών
 - Συνεχής (continuous): σύνολο των πραγματικών αριθμών ή διαστήματα τιμών

Συνάρτηση πυκνότητας πιθανότητας

- Συνάρτηση πυκνότητας πιθανότητας ή κατανομή της τυχαίας μεταβλητής (probability density function, pdf)
- Έστω μια **διακριτή** τυχαία μεταβλητή X με πιθανές τιμές $x_1, x_2, x_3, \dots, x_n$ τότε

$$f(x) = P(X = x_i) \geq 0, \text{ με } i = 1, 2, \dots, n$$

$$\sum_i f(x) = 1$$

- Αν η X είναι μια **συνεχής** τυχαία συνεχής μεταβλητή, τότε

$$P(a < X < b) = \int_a^b f(x) dx$$

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

Συνάρτηση αθροιστικής κατανομής

- Συνάρτηση αθροιστικής κατανομής ή συνάρτηση κατανομής (cumulative distribution function, cdf)
- Η πιθανότητα μια τυχαία μεταβλητή X να παίρνει μια τιμή μικρότερη ή ίση της x

$$F(x) = P(-\infty \leq X \leq x) = \begin{cases} \sum_{u \leq x} f(u), & X \text{ διακριτή} \\ \int_{-\infty}^x f(u) du, & X \text{ συνεχής} \end{cases}$$

$$0 \leq F(x) \leq 1$$

Πολυμεταβλητές κατανομές

- Από κοινού κατανομή πιθανότητας (joint probability distribution) δύο διακριτών τυχαίων μεταβλητών, έστω X και Y
- η πιθανότητα οι τυχαίες μεταβλητές να λαμβάνουν ταυτόχρονα ορισμένες τιμές

$$f(x, y) = P(X = x, Y = y)$$

- Οριακή κατανομή πιθανότητας (marginal probability distribution)

$$f_x(x) = \sum_y f(x, y) \text{ οριακή pdf } X$$

$$f_y(y) = \sum_x f(x, y) \text{ οριακή pdf } Y$$

Υπό συνθήκη κατανομή

- Υπό συνθήκη κατανομή (conditional distribution)

Υπό συνθήκη pdf της X

$$f(x|y) = P(X = x|Y = y) = \frac{f(x, y)}{f_Y(y)}$$

Υπό συνθήκη pdf της Y

$$f(y|x) = P(Y = y|X = x) = \frac{f(x, y)}{f_X(x)}$$

Ανεξαρτησία

- Δύο τυχαίες μεταβλητές X και Y είναι ανεξάρτητες αν

$$f(x, y) = P(X = x, Y = y) = f_X(x)f_Y(y)$$

- Στην περίπτωση που οι X και Y έχουν την ίδια κατανομή και είναι ανεξάρτητες τότε

$$f(x, y) = P(X = x, Y = y) = f_X(x)f_Y(y) = f(x)f(y)$$

και οι X και Y είναι i.i.d. (independent and identically distributed), ανεξάρτητα και ομοιογενώς (πανομοιότυπα) κατανεμημένες

Μέσος ή αναμενόμενη τιμή (ροπή πρώτης τάξης)

- Έστω X η τυχαία μεταβλητή τότε

$$\mu = E(X) = \begin{cases} \sum_x x f(x), & \text{αν η } X \text{ είναι διακριτή} \\ \int_{-\infty}^{\infty} x f(x) dx, & \text{αν η } X \text{ είναι συνεχής} \end{cases}$$

Ιδιότητες αναμενόμενης τιμής

- Αν b είναι μία σταθερά τότε $E(b) = b$
- Αν a και b είναι σταθερές και X μια τυχαία μεταβλητή τότε
$$E(aX + b) = aE(X) + b$$
- Αν X και Y δύο ανεξάρτητες (independent) τυχαίες μεταβλητές τότε
$$E(XY) = E(X)E(Y)$$

ωστόσο $E\left(\frac{X}{Y}\right) \neq \frac{E(X)}{E(Y)}$

Διακύμανση (ροπή δεύτερης τάξης)

- Έστω X μια τυχαία μεταβλητή και $E(X) = \mu$ τότε η διακύμανση δίνεται ως

$$\text{var}(X) = \sigma_X^2 = E(X - \mu)^2 = EX^2 - [E(X)]^2$$

$$\begin{aligned}\text{var}(X) &= \sigma_X^2 = E(X - \mu)^2 \\ &= E(X)^2 - 2\mu E(X) + \mu^2 \\ &= E(X)^2 - 2\mu^2 + \mu^2 = E(X)^2 - \mu^2\end{aligned}$$

- και $\sigma_X = \sqrt{\sigma_X^2}$ η τυπική απόκλιση της X

Ιδιότητες διακύμανσης

- Η διακύμανση μιας σταθεράς είναι μηδέν, $var(b) = 0$
- Αν a και b είναι σταθερές και X μια τυχαία μεταβλητή τότε
$$var(aX + b) = a^2 var(X)$$
- $var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$
- Αν X και Y δύο ανεξάρτητες (independent) τυχαίες μεταβλητές τότε
$$var(X + Y) = var(X) + var(Y)$$
$$var(X - Y) = var(X) + var(Y)$$
- Αν X και Y δύο ανεξάρτητες τυχαίες μεταβλητές και a και b είναι σταθερές τότε

$$var(aX + bY) = a^2 var(X) + b^2 var(Y)$$

Χαρακτηριστικά κατανομών - Ροπές

Ασυμμετρία και κύρτωση

➤ Ασυμμετρία, S ή sk (skewness)

$$S = \frac{E(X - \mu)^3}{\sigma_X^3}$$

- $S = 0$, συμμετρική κατανομή
- $S > 0$, θετική ασυμμετρία, μακριά δεξιά κατάρτηξη
- $S < 0$, αρνητική ασυμμετρία, μακριά αριστερή κατάρτηξη

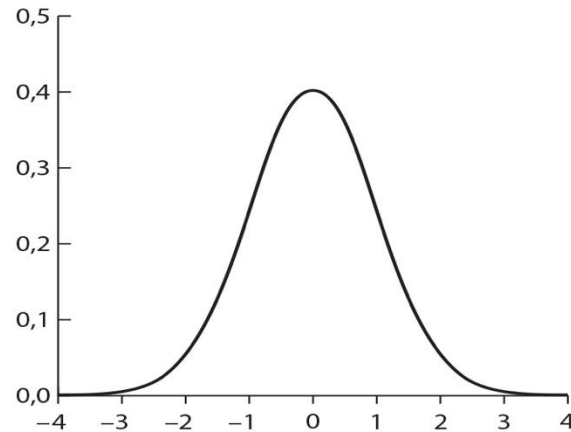
➤ Κύρτωση, K ή k (kurtosis)

$$K = \frac{E(X - \mu)^4}{\sigma_X^4}$$

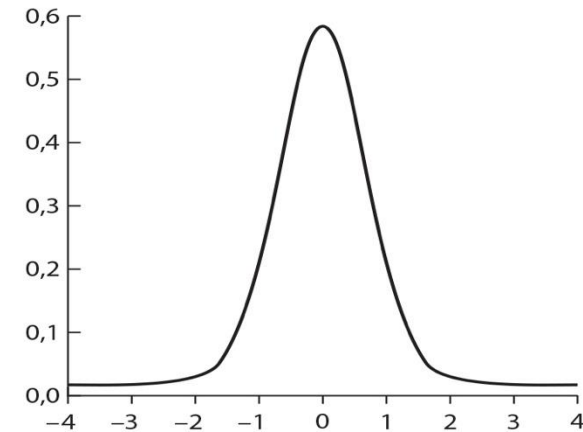
- $K = 3$, μεσόκυρτη (κανονική κατανομή)
- $K > 3$, λεπτόκυρτη, μακριά και λεπτά άκρα
- $K < 3$, πλατύκυρτη, κοντά και πλατιά άκρα

ΣΧΗΜΑ 2.3

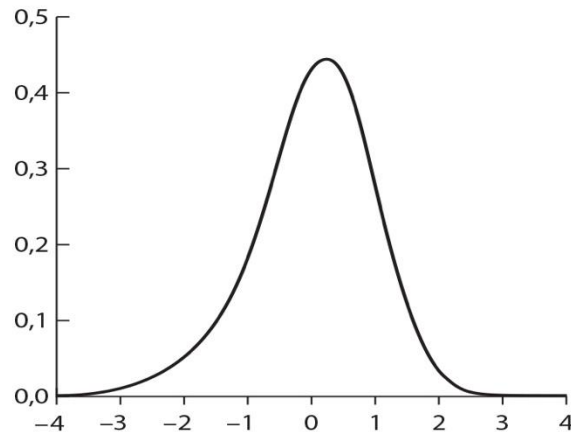
Τέσσερις κατανομές με διαφορετική ασυμμετρία και κύρτωση



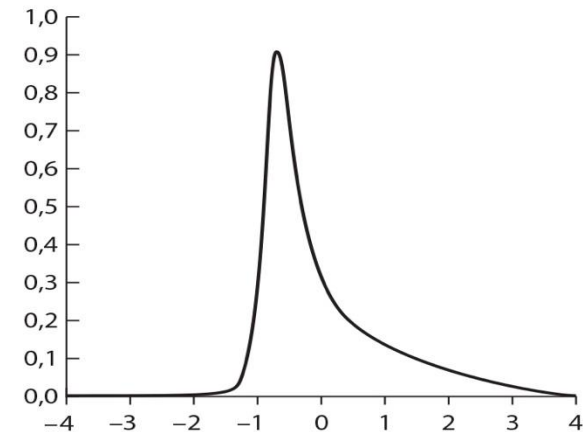
(α) Ασυμμετρία = 0, κύρτωση = 3



(β) Ασυμμετρία = 0, κύρτωση = 20



(γ) Ασυμμετρία = -0,1 και κύρτωση = 5



(δ) Ασυμμετρία = 0,6 και κύρτωση = 5

Όλες αυτές οι κατανομές έχουν μέσο 0 και διακύμανση 1. Οι κατανομές με ασυμμετρία 0 (δηλαδή η α και η β) είναι συμμετρικές, ενώ οι κατανομές με μη μηδενική ασυμμετρία (η γ και η δ) δεν είναι συμμετρικές. Οι κατανομές με κύρτωση που υπερβαίνει το 3 (οι β και δ) έχουν λεπτά άκρα.

Μέτρα εξάρτησης ή συσχέτισης

Συνδιακύμανση (covariance)

- Αν X και Y δύο τυχαίες μεταβλητές με μ_X και μ_Y , η συνδιακύμανση μεταξύ των δύο μεταβλητών δίνεται ως εξής

$$\text{cov}(X, Y) = E\{(X - \mu_X)(Y - \mu_Y)\} = E(XY) - \mu_X\mu_Y$$

- Αν X και Y δύο ανεξάρτητες τυχαίες μεταβλητές τότε η συνδιακύμανση τους είναι ίση με το μηδέν

$$\text{cov}(X, Y) = E(XY) - \mu_X\mu_Y = 0 \text{ καθώς } E(XY) = E(X)E(Y) = \mu_X\mu_Y$$

Μέτρα εξάρτησης ή συσχέτισης

Συσχέτιση (correlation)

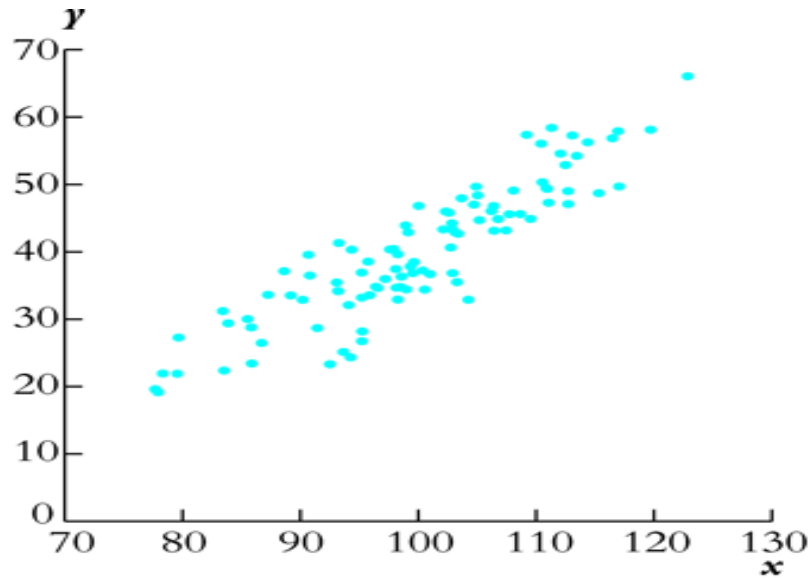
- Ο συντελεστής συσχέτισης ρ μετράει τη γραμμική συσχέτιση μεταξύ δύο μεταβλητών

$$\rho = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

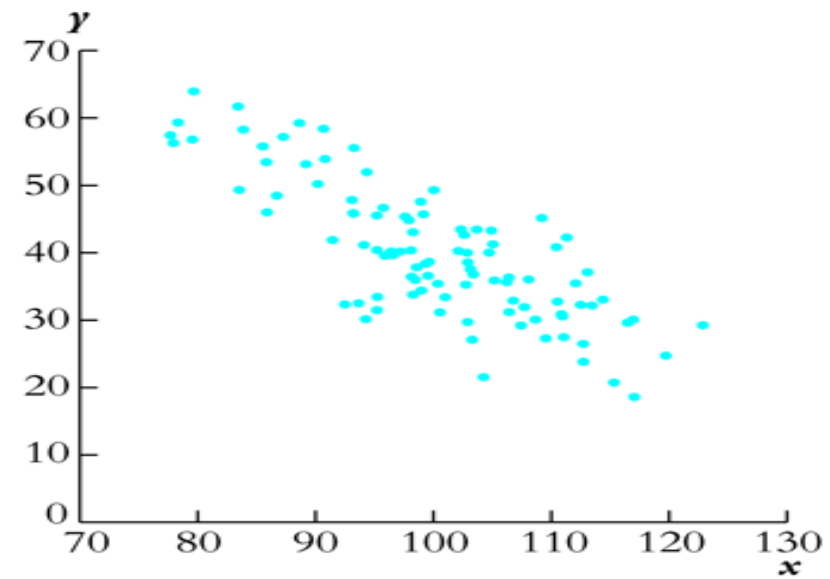
$$-1 \leq \text{corr}(X, Y) \leq 1$$

- $\text{cov}(X, Y) = \rho \sigma_X \sigma_Y$

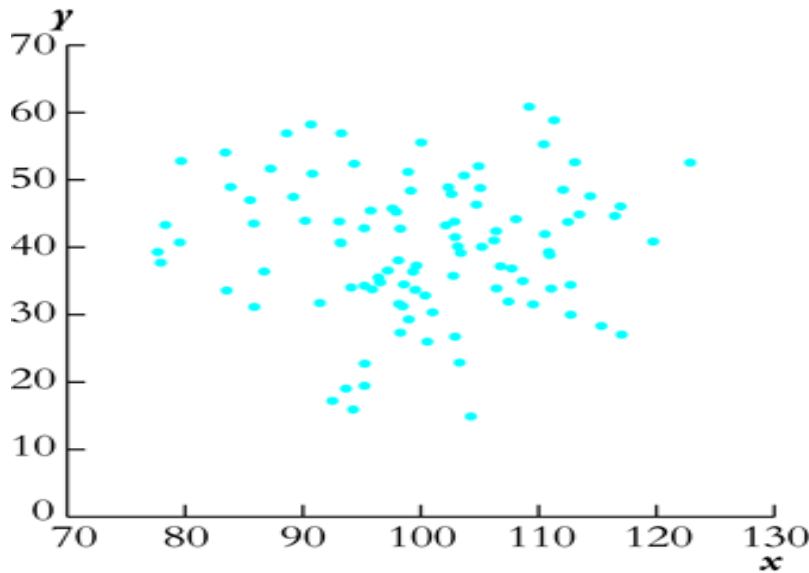
Μέτρα εξάρτησης ή συσχέτισης



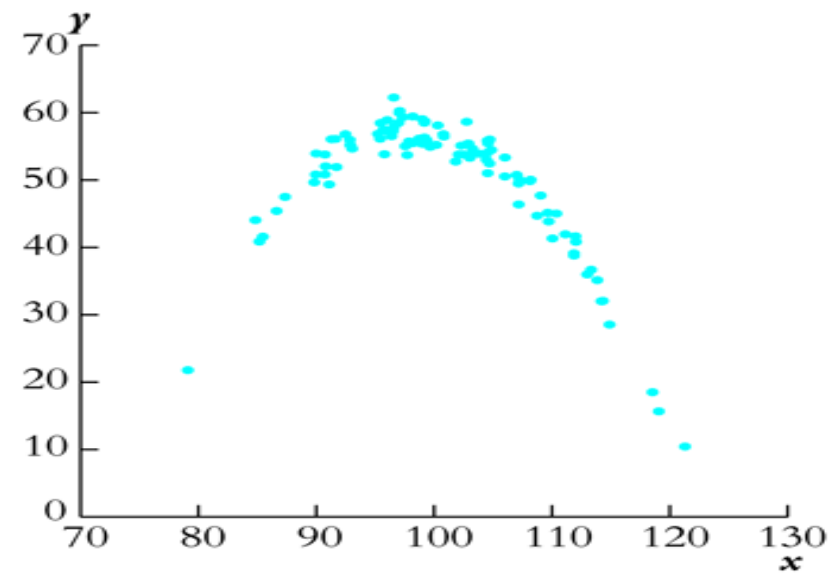
(a) Correlation = +0.9



(b) Correlation = -0.8



(c) Correlation = 0.0



(d) Correlation = 0.0 (quadratic)

Διακύμανση και συσχετιζόμενες μεταβλητές

- Αν X και Y δύο εξαρτημένες (dependent) τυχαίες μεταβλητές τότε

$$\begin{aligned} \text{var}(X + Y) &= \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y) \\ &= \text{var}(X) + \text{var}(Y) + 2\rho\sigma_X\sigma_Y \end{aligned}$$

και

$$\begin{aligned} \text{var}(X - Y) &= \text{var}(X) + \text{var}(Y) - 2\text{cov}(X, Y) \\ &= \text{var}(X) + \text{var}(Y) - 2\rho\sigma_X\sigma_Y \end{aligned}$$

Υπό συνθήκη μέσος & διακύμανση

- Αν $f(x, y)$ η από κοινού pdf δύο τυχαίων μεταβλητών X και Y , ο υπό συνθήκη μέσος (ή υπό συνθήκη προσδοκία) της X , με δεδομένο ότι $Y = y$, δίνεται ως εξής

$$E(X|Y = y) = \begin{cases} \sum_x x f(x|Y = y) & \text{αν } X \text{ διακριτή} \\ \int_{-\infty}^{\infty} x f(x|Y = y) & \text{αν } X \text{ συνεχής} \end{cases}$$

- Υπό συνθήκη διακύμανση

$$var(X|Y = y) = E \left\{ \left([X - E(X|Y = y)]^2 \right) | Y = y \right\}$$

- Αν X και Y είναι ανεξάρτητες τότε $E(X|Y) = E(X)$, $cov(X, Y) = 0$ ($corr(X, Y) = 0$) και $var(X|Y) = var(X)$

Ο νόμος των επαναλαμβανόμενων προσδοκιών

- Ο νόμος των επαναλαμβανόμενων προσδοκιών (law of iterated expectations)
- Η αναμενόμενη τιμή της Y είναι η προσδοκία της δεσμευμένης προσδοκίας της Y όταν δίνεται η X

$$E(Y) = E[E(Y|X)]$$

Περίληψη

➤ Βασικές έννοιες στατιστικής:

- Τυχαίες μεταβλητές; Διακριτές τυχαίες μεταβλητές; Συνεχείς τυχαίες μεταβλητές; pdf; cdf; Πολυμεταβλητή κατανομή; Υπό συνθήκη κατανομή; Ανεξαρτησία
- χαρακτηριστικά τυχαίων μεταβλητών (Μέση Τιμή - Διακύμανση – Ροπές)
- Συνδιακύμανση, Συντελεστής συσχέτισης
- Υπό συνθήκη μέσος και διακύμανση; Νόμος επαναλαμβανόμενων προσδοκιών

Βιβλιογραφία

- ✓ Stock & Watson, κεφ. 1-2
- ✓ Τζαβαλή, κεφ. 1, παράρτημα Β
- ✓ Gujarati & Porter, Εισαγωγή, παράρτημα Α

Ευχαριστώ!

Ερωτήσεις?