ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Σχολή Οικονομικών Επιστημών - Τμήμα Οικονομικής Επιστήμης
Στατιστική Ι - Χειμερινό Εξάμηνο 2018-2019
Διδάσκων: Α. Λουκά *email: loukaalex@aueb.gr*
Βοηθός: Δ. Σαρρή, *email: sarridan@aueb.gr*

Problem Set 6

Άσκηση 6.1

Μία εταιρία παράγει μαρμελάδα και τη διοχετεύει στο εμπόριο σε βάζα. Η κατανομή του καθαρού βάρους (σε γραμμάρια) αυτών των βάζων είναι κανονική με μέσο 500 και τυπική απόκλιση 10. Υποθέσατε ότι παίρνετε ένα τυχαίο δείγμα 4 βάζων, ποια είναι η πιθανότητα το μέσο καθαρό βάρος των 4 αυτών βάζων να είναι λιγότερο από 490 γραμμάρια; (Δίνεται ότι $P(Z \leq 2) = 0.9772$, για $Z \sim N(0,1)$.)

Λύση

Έστω ότι έχουμε την τυχαία μεταβλητή X για το καθαρό βαρος σε γραμμάρια των βάζων. Από τα δεδομένα έχουμε ότι: $X \sim N(500, 10^2)$.
Ουσιαστικά θέλω να υπολογίσω την πιθανότητα:

$$P(\bar{X} < 490) \tag{1}$$

όπου $\bar{X}$ το μέσο καθαρό βάρος $\mu_{\bar{X}}$, του τυχαίου δείγματος πλήθους $n = 4$.
Ο μέσος και η διακύμανση της $\bar{X}$ υπολογίζονται ως εξής:
$\mu_{\bar{X}} = E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n\mu}{n} = \mu = 500$
$\sigma^2_{\bar{X}} = Var(\bar{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} = \frac{100}{4} = 25$
Άρα $\bar{X} \sim N(500, 25)$.

$$P(\bar{X} < 490) = P(\bar{X} - \mu_{\bar{X}} < 490 - \mu_{\bar{X}}) = P(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{490 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}) =$$
$$P(Z < \frac{490 - 500}{5}) = P(Z < -2) = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$$

Άσκηση 6.2

Έστω ότι παίρνετε ένα τυχαίο δείγμα 100 καινούριων διαμερισμάτων που πωλήθηκαν το περασμένο έτος, όπου η τιμή πώλησης είναι μια τυχαία μεταβλητή X και ακολουθεί την κανονική κατανομή με μέσο $\mu = 50$ και τυπική απόκλιση $\sigma = 10$.
(α) Ποια είναι η πιθανότητα η μέση τιμή πώλησης στο δείγμα να είναι μεταξύ 50 και 51 ;
(β) Αν η διακύμανση είναι άγνωστη και δίνεται ότι η εκτίμησή της από τον εκτιμητή $\hat{S}^2 =$

$\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$, είναι $s=11.13$ να βρείτε την πιθανότητα η μέση τιμή πώλησης στο δείγμα να είναι μεταξύ 50 και 51.
(Δίνεται ότι $Pr(Z\leq 1)=0.8415$, για $Z\sim N[0,1]$, $P(0<T<0.88)=0.309$ για $T\sim t_{v-1}$)

Λύση

(α) Από τα δεδομένα έχουμε ότι: $X\sim N(50,10^2)$.
Ουσιαστικά θέλω να υπολογίσω την πιθανότητα:

$$P(50<\bar{X}<51) \tag{2}$$

όπου $\bar{X}$ η μέση τιμή πώλησης $\mu_{\bar{X}}$, του τυχαίου δείγματος πλήθους $n=100$.
Ο μέσος και η διακύμανση της $\bar{X}$ υπολογίζονται ως εξής:
$\mu_{\bar{X}}=E(\bar{X})=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\frac{n\mu}{n}=\mu=50$
$\sigma_{\bar{X}}^2=Var(\bar{X})=Var(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n^2}\sum_{i=1}^{n}Var(X_i)=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}=\frac{100}{100}=1$
Άρα $\bar{X}\sim N(50,1)$.

$$P(50<\bar{X}<51)=P(50-\mu_{\bar{X}}<\bar{X}-\mu_{\bar{X}}<51-\mu_{\bar{X}})=P(\frac{50-\mu_{\bar{X}}}{\sigma_{\bar{X}}}<\frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}}<\frac{51-\mu_{\bar{X}}}{\sigma_{\bar{X}}})=$$
$$P(\frac{50-50}{1}<Z<\frac{51-50}{1})=P(0<Z<1)=P(Z<1)-P(Z<0)=0.8415-0.5=0.3413$$

(β) Ο πληθυσμός από όπου προέρχεται το δείγμα, έχουμε από εκφώνηση ότι είναι κανονικός, όμως τώρα διαφοροποιείται το ερώτημα ως προς τη διακύμανση, που είναι θεωρητικά άγνωστη και δίνεται μια εκτίμηση για αυτήν. Ο μέσος δίνεται από ερώτημα (α)
Θα ισχύει γενικά ότι:

$$\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{v-1} \tag{3}$$

Εδώ δίνεται η εκτίμηση ότι $s=11.13$. Θα έχω ότι:

$$\frac{s}{\sqrt{n}}=\frac{11.36}{10}=1.136 \tag{4}$$

Άρα:

$$P(50<\bar{X}<51)=P(\frac{50-\mu_{\bar{X}}}{s/\sqrt{n}}<T<\frac{51-\mu_{\bar{X}}}{s/\sqrt{n}}=P(0<T<0.88)=0.309$$

(5)

Άσκηση 6.3

Έστω $\{X_1,X_2,...,X_{n=9}\}$ ένα τυχαίο δείγμα από την τυχαία μεταβλητή X που ακολουθεί την κανονική κατανομή με μέσο μ και διακύμανση σ^2. Για τον υπολογισμό του μέσου προτείνονται οι εκτιμητές $\bar{X}=\frac{X_1+X_2+...+X_9}{9}$ και $\mu^*=\frac{X_1+X_2}{2}$.
(α) Να δείξετε ότι και οι δύο εκτιμητές είναι αμερόληπτοι.
(β) Να προσδιορίσετε τον αποτελεσματικότερο εκτιμητή από τους δύο.

Λύση

(α) Από τα δεδομένα έχουμε ότι: $X \sim N(\mu, \sigma^2)$.
Ουσιαστικά για την αμεροληψία θα ελέγξω τις αναμενόμενες τιμές των δύο προτεινόμενων εκτιμητών:

$$E(\bar{X}) = \frac{E(X_1 + X_2 + ... + X_9)}{9} = \frac{E(X_1) + E(X_2) + ... + E(X_9)}{9} = \frac{9 \cdot \mu}{9} = \mu \qquad (6)$$

Άρα ο εκτιμητής $\bar{X}$ είναι αμερόληπτος.

$$E(\mu^*) = \frac{E(X_1 + X_2)}{2} = \frac{E(X_1) + E(X_2)}{2} = \frac{2 \cdot \mu}{2} = \mu \qquad (7)$$

Και ο εκτιμητής μ^* είναι αμερόληπτος.

(β) Για την αποτελεσματικότητα θα συγκρίνουμε τις δύο διακυμάνσεις, και ο εκτιμητής με την μικρότερη διακύμανση θα αποτελέσει τον πιο αποτελεσματικό. Άρα:

$$\sigma^2_{\bar{X}} = Var(\bar{X}) = Var(\frac{X_1 + X_2 + ... + X_9}{9}) = \frac{1}{9^2}(Var(X_1 + Var(X_2) + ... + Var(X_9)) = \frac{9\sigma^2}{9^2} = \frac{\sigma^2}{9} \qquad (8)$$

Ενώ για τον δεύτερο εκτιμητή:

$$\sigma^2_{\mu^*} = Var(\mu^*) = Var(\frac{X_1 + X_2}{2}) = \frac{1}{2^2}(Var(X_1) + Var(X_2)) = \frac{2\sigma^2}{2^2} = \frac{\sigma^2}{2} \qquad (9)$$

$Var(\mu^*) > Var(\bar{X})$ δηλαδή ο εκτιμητής $Var(\bar{X})$ είναι αποτελεσματικότερος.

Άσκηση 6.4

Έστω $\{X_1, X_2, ..., X_n\}$ τυχαίο δείγμα n ανεξάρτητων δοκιμών από την τυχαία μεταβλητή X που ακολουθεί την Bernoulli κατανομή με $E(X) = p, Var(X) = p(1-p)$, με (σταθερή) πιθανότητα επιτυχίας p , όπου $0 < p < 1$.
(α) Να βρεθεί ο εκτιμητής του p με τη μέθοδο των ροπών (MM).
(β) Να δειχθεί ότι ο εκτιμητής αυτός είναι αμερόληπτος.

Λύση

(α) Γενικά για να βρούμε τους εκτιμητές των παραμέτρων με τη μέθοδο των ροπών (MM), ορίζουμε τη ροπή k τάξης ως προς το μηδέν ως:
$\mu_k = E(X^k)$, $k = 1, 2, 3, ..$ όπου είναι μία συνάρτηση των παραμέτρων. Σε αυτήν την περίπτωση της παραμέτρου p. Η μέθοδος των ροπών εξισώνει τις ροπές του πληθυσμού με τις ροπές του δείγματος
$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$, $k = 1, 2, 3, ...$ Στη συγκεκριμένη περίπτωση θα έχω:

$\mu = E(X) = p$

$M_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$.
Συνεπώς εξισώνουμε και παίρνουμε: $p_{\hat{MM}} = \bar{X}$

(β)

Για την αμεροληψία θα απαιτήσω: $E(p_{\hat{MM}} = p)$. Άρα:

$$E(\hat{p}_{MM}) = E(\bar{X}) = E(\frac{\sum_{i=1}^{n} X_i}{n}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n \cdot p}{n} = p \qquad (10)$$

Άσκηση 6.5

Έστω $\{X_1, X_2, ..., X_n\}$ ένα τυχαίο δείγμα από την τυχαία μεταβλητή X που ακολουθεί την κανονική κατανομή με μέσο μ και διακύμανση σ^2. Να βρεθούν οι εκτιμητές των παραμέτρων (μ, σ^2) τη μέθοδο των ροπών (MM).

Λύση

Σε αυτήν την εφαρμογή πρέπει να εκτιμηθούν δύο παράμετροι, άρα και χρειαζόμαστε δύο εξισώσεις.
Για την πρώτη ροπή θα έχω:

$$M_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \tag{11}$$

Επίσης, $E(X) = \mu = \frac{1}{n}\sum_{i=1}^{n} X_i$. Δηλαδή:

$$\hat{\mu}_{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \tag{12}$$

Για τη διακύμανση γνωρίζουμε ότι:

$$Var(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2 \tag{13}$$

Χρησιμοποιώντας την εξίσωση (9) θα πάρω:

$$Var(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 \tag{14}$$

Προκύπτει συνεπώς ότι:

$$\hat{\sigma^2}_{MM} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{15}$$

Άσκηση 6.6

[Ε.Γ. Τσιώνας «Στατιστική με Εφαρμογές στα Οικονομικά» Αθήνα 2009]. Έστω $\{X_1, X_2, ..., X_n\}$ ένα τυχαίο δείγμα μεγέθους n από την τυχαία μεταβλητή X που ακολουθεί την κανονική κατανομή με μέσο μ και διακύμανση σ^2. Θεωρούμε τρεις εκτιμητές του μέσου $\mu_1 = n^{-1}\sum_{i=1}^{n} X_i$, $\mu_2 = (n+1)^{-1}\sum_{i=1}^{n} X_i$ και $\mu_3 = \frac{1}{2}X_1 + \frac{1}{2n}\sum_{i=1}^{n} X_i$. Να συγκρίνετε τους εκτιμητές σε όρους αμεροληψίας και όρους MSE.

Λύση

Για αμεροληψία:

$$E(\mu_1) = E(n^{-1}\sum_{i=1}^{n} X_i) = \frac{\sum_{i=1}^{n} E(X_i)}{n} = \frac{n \cdot \mu}{n} = \mu \tag{16}$$

άρα ο μ_1 είναι αμερόπληπτος.

$$E(\mu_2) = E((n+1)^{-1}\sum_{i=1}^{n} X_i) = \frac{\sum_{i=1}^{n} E(X_i)}{n+1} = \frac{n \cdot \mu}{n+1} \neq \mu \tag{17}$$

άρα ο μ_2 δεν είναι αμερόπληπτος.

$$E(\mu_3) = E(\frac{1}{2}X_1 + \frac{1}{2n}\sum_{i=1}^{n} X_i) = \frac{1}{2}E(X_1) + \frac{\sum_{i=1}^{n} E(X_i)}{2n} = \frac{1}{2}\mu + \frac{n \cdot \mu}{2n} = \mu \tag{18}$$

άρα ο μ_3 είναι αμερόπληπτος.

Για το MSE θα υπολογίσω την ποσότητα $MSE(\hat{\mu}) = E[(\hat{\mu} - \mu)^2] = Var(\hat{\mu}) + Bias(\hat{\mu})^2$ για κάθε ένα από τα $\hat{\mu}$.

$$MSE(\hat{\mu_1}) = Var(\hat{\mu_1}) + B(\hat{\mu_1})^2 = \frac{\sum_{i=1}^{n} Var(X_i)}{n^2} + 0 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \tag{19}$$

$$\begin{aligned} MSE(\hat{\mu_2}) &= Var(\hat{\mu_2}) + B(\hat{\mu_2})^2 = \frac{\sum_{i=1}^{n} Var(X_i)}{(n+1)^2} + (E(\hat{\mu_2}) - \mu)^2 \\ &= \frac{n\sigma^2}{(n+1)^2} + (\frac{n \cdot \mu}{n+1} - \mu)^2 = \frac{n\sigma^2}{(n+1)^2} + \frac{(n \cdot \mu - \mu(n+1))^2}{(n+1)^2} = \\ &= \frac{n\sigma^2 + \mu^2}{(n+1)^2} \end{aligned} \tag{20}$$

$$\begin{aligned} MSE(\hat{\mu_3}) &= Var(\hat{\mu_3}) + B(\hat{\mu_3})^2 = \frac{Var(X_1)}{4} + \frac{\sum_{i=1}^{n} Var(X_i)}{4n^2} + 0 \\ &= \frac{\sigma^2}{4} + \frac{n \cdot \sigma^2}{4n^2} = \frac{n\sigma^2 + \sigma^2}{4n} = \frac{\sigma^2(n+1)}{4n} \end{aligned} \tag{21}$$

Από τα τρία MSE για τους τρεις εκτιμητές, το μικρότερο σφάλμα το έχει ο πρώτος.

Άσκηση 6.7

Έστω $\{X_1, X_2, ..., X_n\}$ ένα τυχαίο δείγμα από την τυχαία μεταβλητή X με μέσο μ και τυπική απόκλιση σ, θεωρήστε τους ακόλουθους δύο εκτιμητές:

$\hat{\theta}_1 = \frac{5X_1 - 2X_2 + 4X_3}{7}$

$\hat{\theta}_2 = \frac{6X_1 + 2X_2 - 2X_3}{7}$.

(α) Εξετάστε τους εκτιμητές ως προς την αμεροληψία τους.

(β) Ποιός από τους δύο εκτιμητές είναι ο καλύτερος αν γνωρίζουμε ότι $\mu^2 > \sigma^2$;

Λύση

(α)

$$E(\hat{\theta}_1) = \frac{5E(X_1) - 2E(X_2) + 4E(X_3)}{7} = \frac{5\mu - 2\mu + 4\mu}{7} = \mu \tag{22}$$

άρα είναι αμερόληπτος.

$$E(\hat{\theta}_2) = \frac{6E(X_1) + 2E(X_2) - 2E(X_3)}{7} = \frac{6\mu + 2\mu - 2\mu}{7} = \frac{6\mu}{7} \neq \mu \tag{23}$$

άρα είναι μεροληπτικός ο εκτιμητής $\hat{\theta}_2$.

(β) Θα εξετάσουμε την διακύμανση των δύο εκτιμητών και τη μεροληψία του δεύτερου εκτιμητή.

$$Var(\hat{\theta}_1) = \frac{25Var(X_1) + 4Var(X_2) + 16Var(X_3)}{49} = \frac{45\sigma^2}{49} \tag{24}$$

$$Var(\hat{\theta}_2) = \frac{36Var(X_1) + 4Var(X_2) + 4Var(X_3)}{49} = \frac{44\sigma^2}{49} \tag{25}$$

Ο δεύτερος εκτιμητής παρότι μεροληπτικός έχει μικρότερη διακύμανση από τον αμερόληπτο εκτιμητή. Όσον αφορά τη μεροληψία του:

$$Bias(\hat{\theta_2}) = E(\hat{\theta_2}) - \mu = \frac{6\mu}{7} - \mu = -\frac{\mu}{7} \tag{26}$$

Αν θέλουμε αμερόληπτους εκτιμητές γενικά, τότε καλύτερος από τους δύο εκτιμητές είναι ο πρώτος.
Σε όρους MSE θα έχουμε:

$$MSE(\hat{\theta_1}) = Var(\hat{\theta_1}) + B(\hat{\theta_1})^2 = \frac{45\sigma^2}{49} = \frac{44\sigma^2}{49} + \frac{\sigma^2}{49} \tag{27}$$

$$MSE(\hat{\theta_2}) = Var(\hat{\theta_2}) + B(\hat{\theta_2})^2 = \frac{44\sigma^2}{49} + (\frac{-\mu}{7})^2 = \frac{44\sigma^2}{49} + \frac{\mu^2}{49} \tag{28}$$

Εφόσον μας δίνει ότι $\mu^2 > \sigma^2$:

$$\mu^2 > \sigma^2 \iff \frac{\mu^2}{49} > \frac{\sigma^2}{49} \iff \frac{44\sigma^2}{49} + \frac{\mu^2}{49} > \frac{44\sigma^2}{49} + \frac{\sigma^2}{49} \iff MSE(\hat{\theta_2}) > MSE(\hat{\theta_1}) \tag{29}$$

Συνεπώς, σε όρους MSE ο εκτιμητής $\hat{\theta_1}$ ελιναι καλύτερος, καθότι έχει μικρότερο σφάλμα.