

Προσθήκη Ανεξάρτητης Μεταβλητής στο Μοντέλο

Ας θυμηθούμε το Κενό Μοντέλο, $Y_i = b_0 + e_i$

$$y_i = \bar{y} + e_i$$

Ορισμός & Λειτουργία:

- Χρησιμοποιεί τον **μέσο όρο** ως τη μοναδική πρόβλεψη για κάθε παρατήρηση
- Ελαχιστοποιεί το **άθροισμα των τετραγώνων των αποκλίσεων** (Sum of Squares)

Το βασικό του μειονέκτημα:

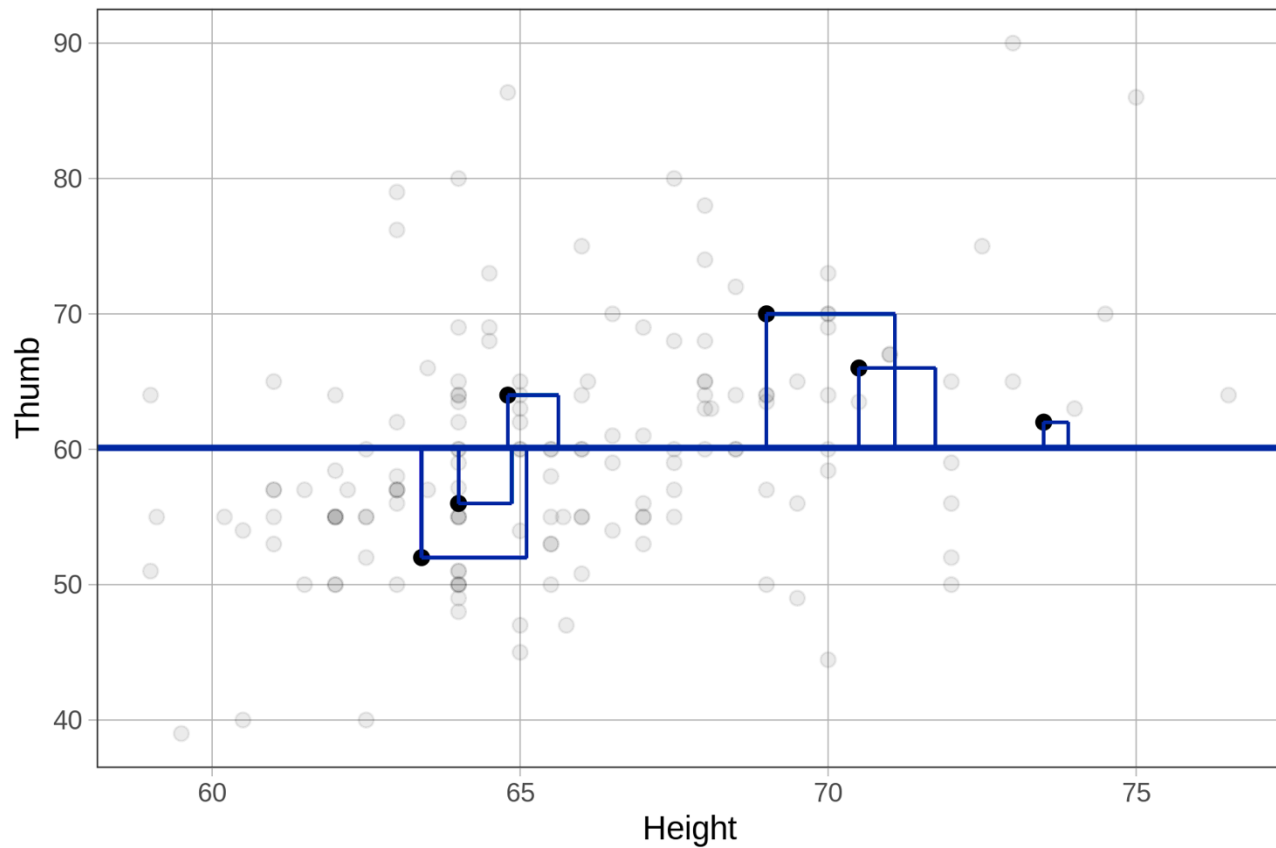
Δε χρησιμοποιεί άλλες μεταβλητές (ανεξάρτητες).

Δεν ερμηνεύει τη **μεταβλητότητα** (variability) των δεδομένων.

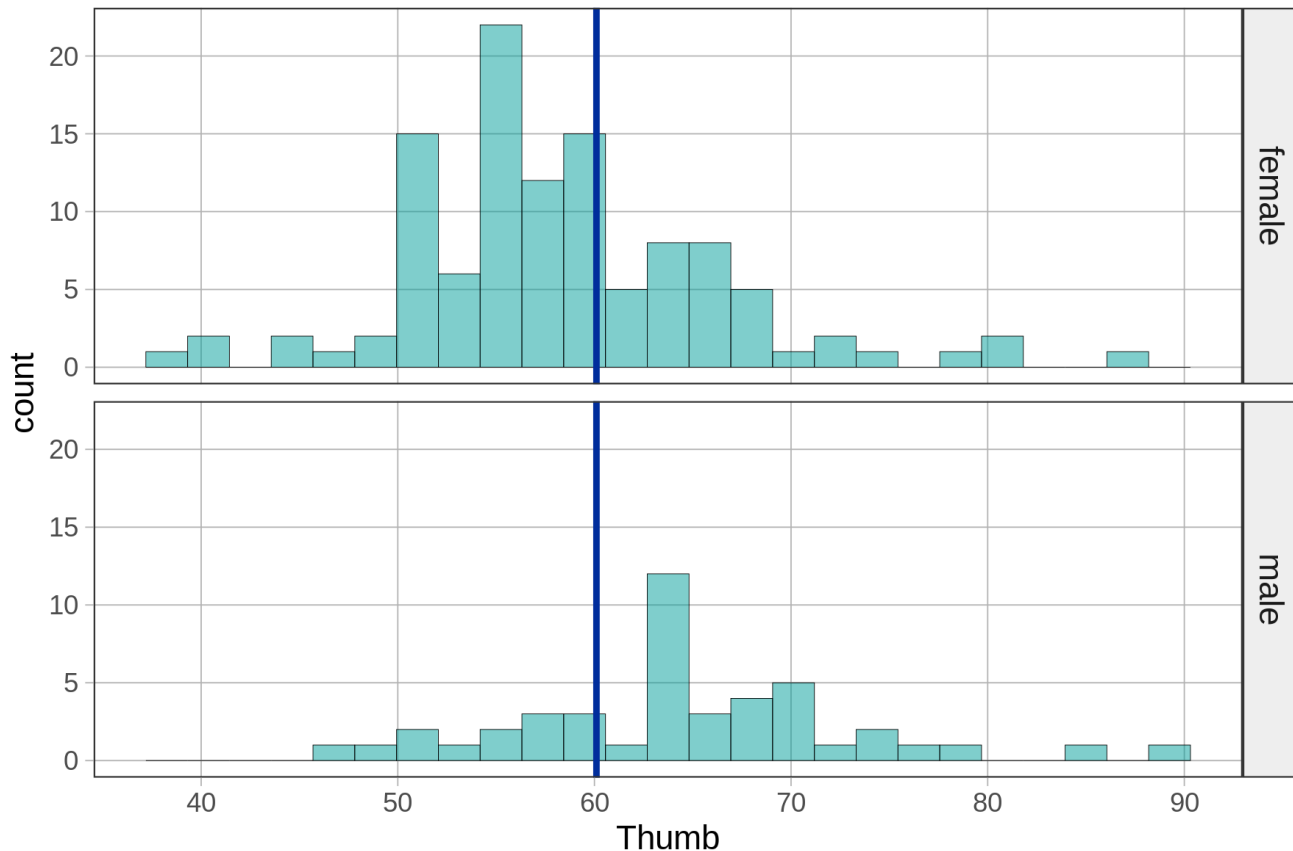
Ο Πραγματικός Σκοπός:

- Λειτουργεί ως **μοντέλο αναφοράς**.
- Μας επιτρέπει να συγκρίνουμε αν τα πιο «σύνθετα» μοντέλα προσφέρουν πράγματι καλύτερη προσαρμογή στα δεδομένα.

Ας θυμηθούμε
το Κενό
Μοντέλο,
 $Y_i = b_0 + e_i$



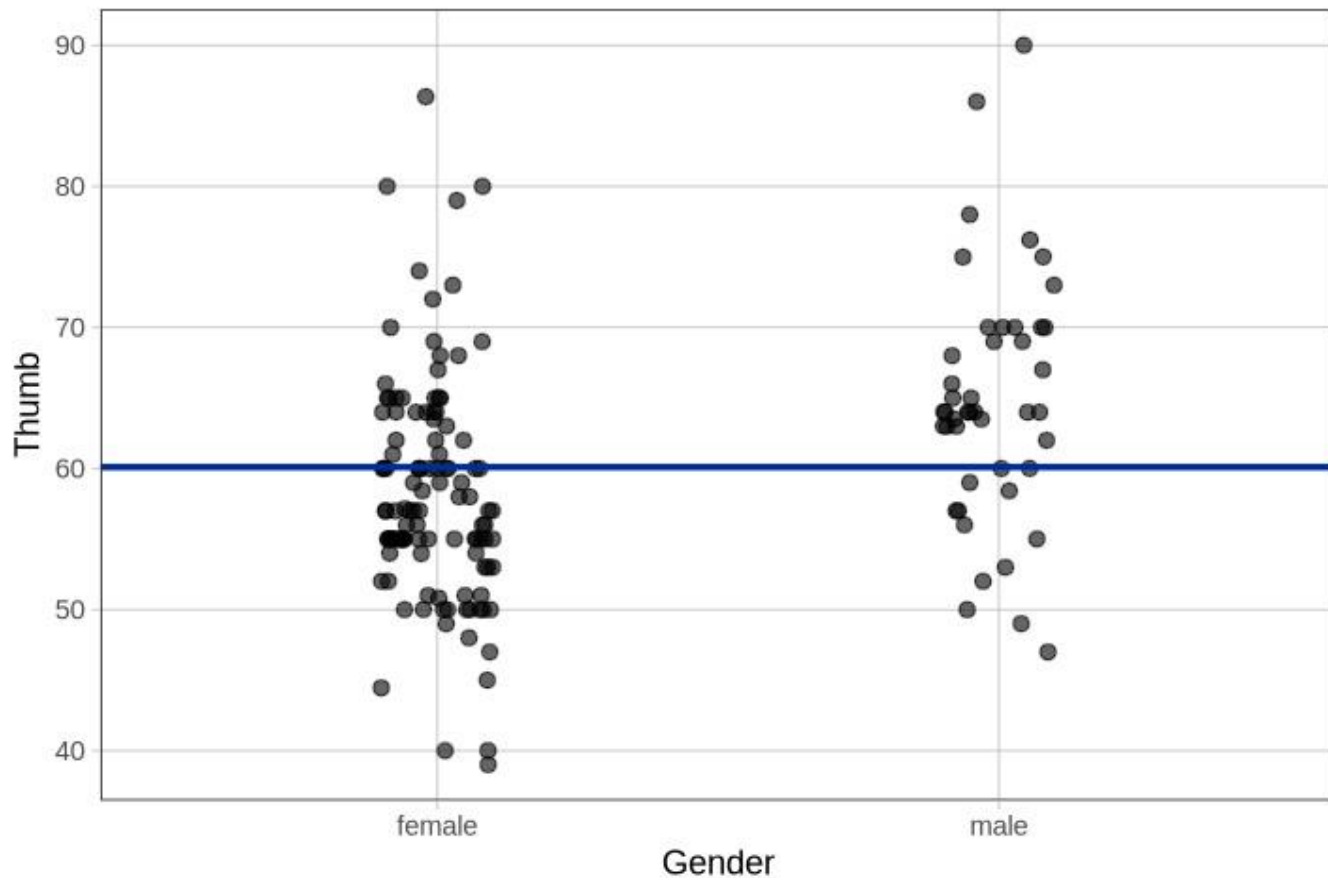
Εξήγηση της Μεταβλητότητας στο Μήκος Αντίχειρα



Σύμφωνα με το κενό μοντέλο, θα προβλέπαμε ότι το μήκος του αντίχειρα ενός μελλοντικού φοιτητή είναι 60.1 χιλιοστά (δηλαδή, ο συνολικός μέσος όρος του μήκους του αντίχειρα).

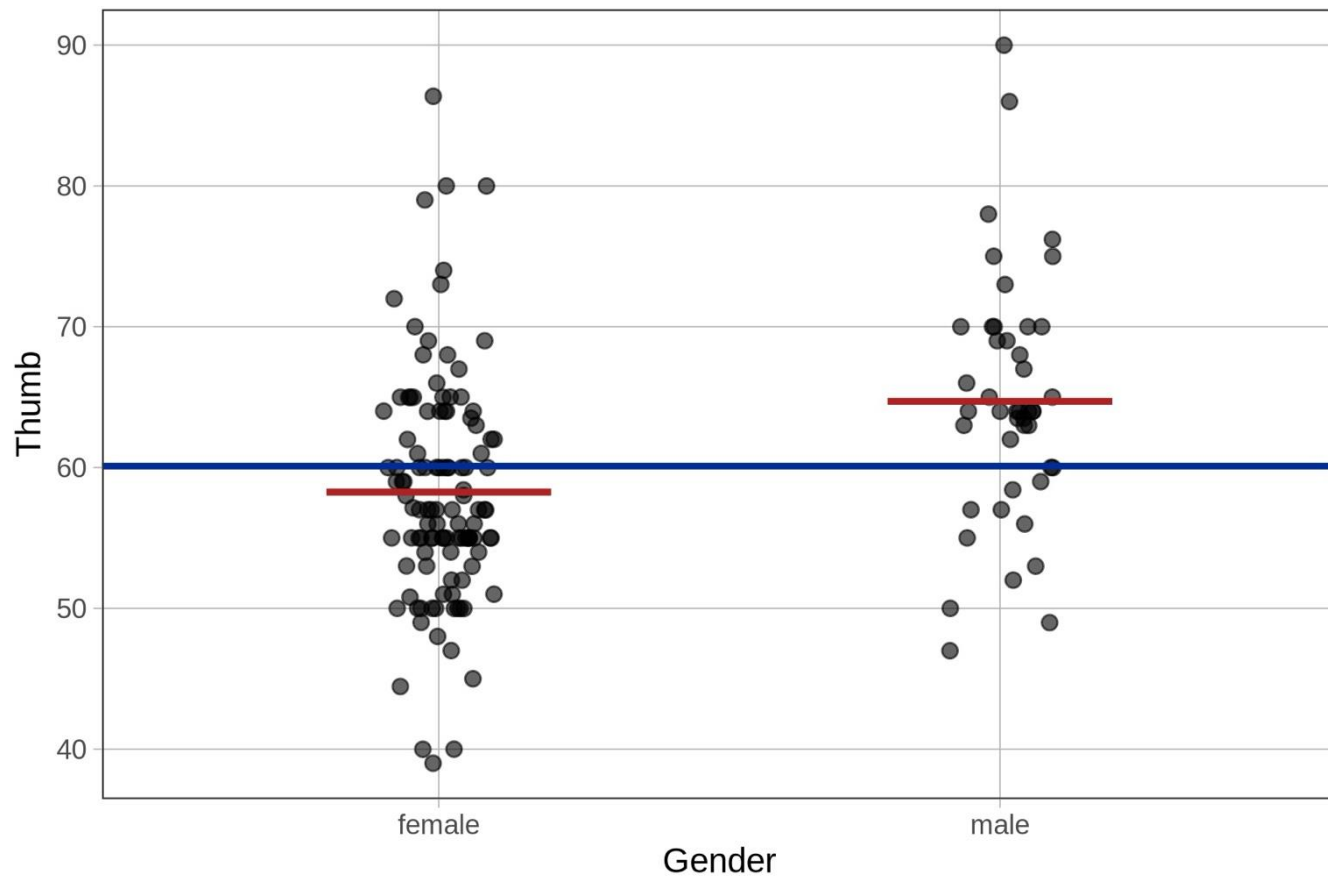
Πώς θα μπορούσε η γνώση του φύλου ενός φοιτητή να αλλάξει την πρόβλεψή μας για το μήκος του αντίχειρά του/της;

Εξήγηση της Μεταβλητότητας στο Μήκος Αντίχειρα



Μήκος αντίχειρα = Φύλο + Σφάλμα

$\text{lm}(\text{Thumb} \sim \text{Gender}, \text{data} = \text{Fingers})$



Με βάση το
διάγραμμα, τι θα
προέβλεπε το
μοντέλο του Φύλου
(Gender) για το
μήκος αντίχειρα

(α) μιας φοιτήτριας;

(β) ενός φοιτητή;

```
favstats(Thumb ~ Gender, data=Fingers)
```

	Gender	min	Q1	median	Q3	max	mean	sd	n	missing
1	female	39	54	57	63.125	86.36	58.25585	8.034694	112	0
2	male	47	60	64	70.000	90.00	64.70267	8.764933	45	0

	Gender	Thumb	Gender_predict
1	male	66.00	64.70267
2	female	64.00	58.25585
3	female	56.00	58.25585
4	male	58.42	64.70267
5	female	74.00	58.25585
6	female	60.00	58.25585

Call:

```
lm(formula = Thumb ~ Gender, data = Fingers)
```

Coefficients:

(Intercept) Gendermale

58.256

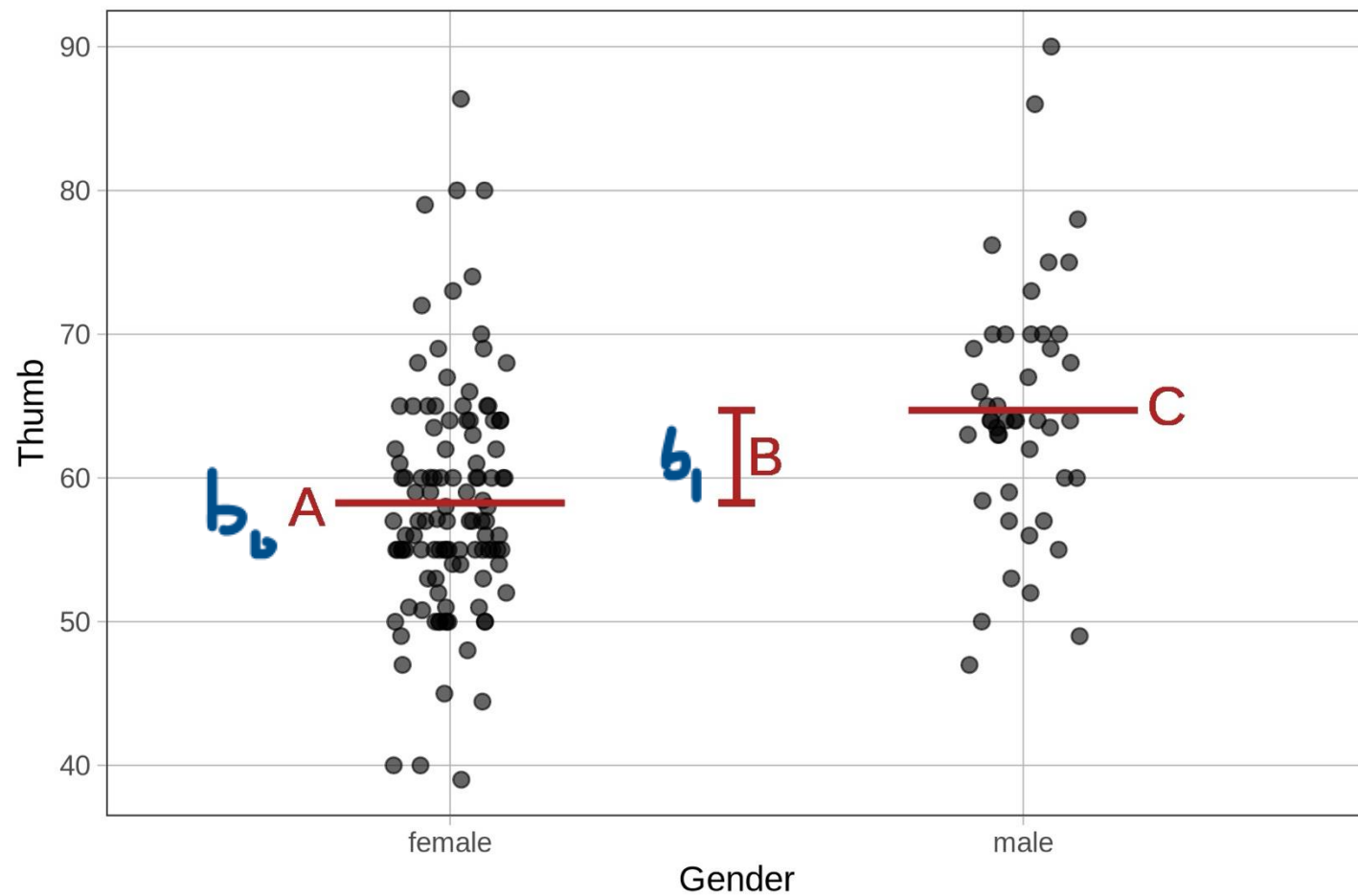
6.447

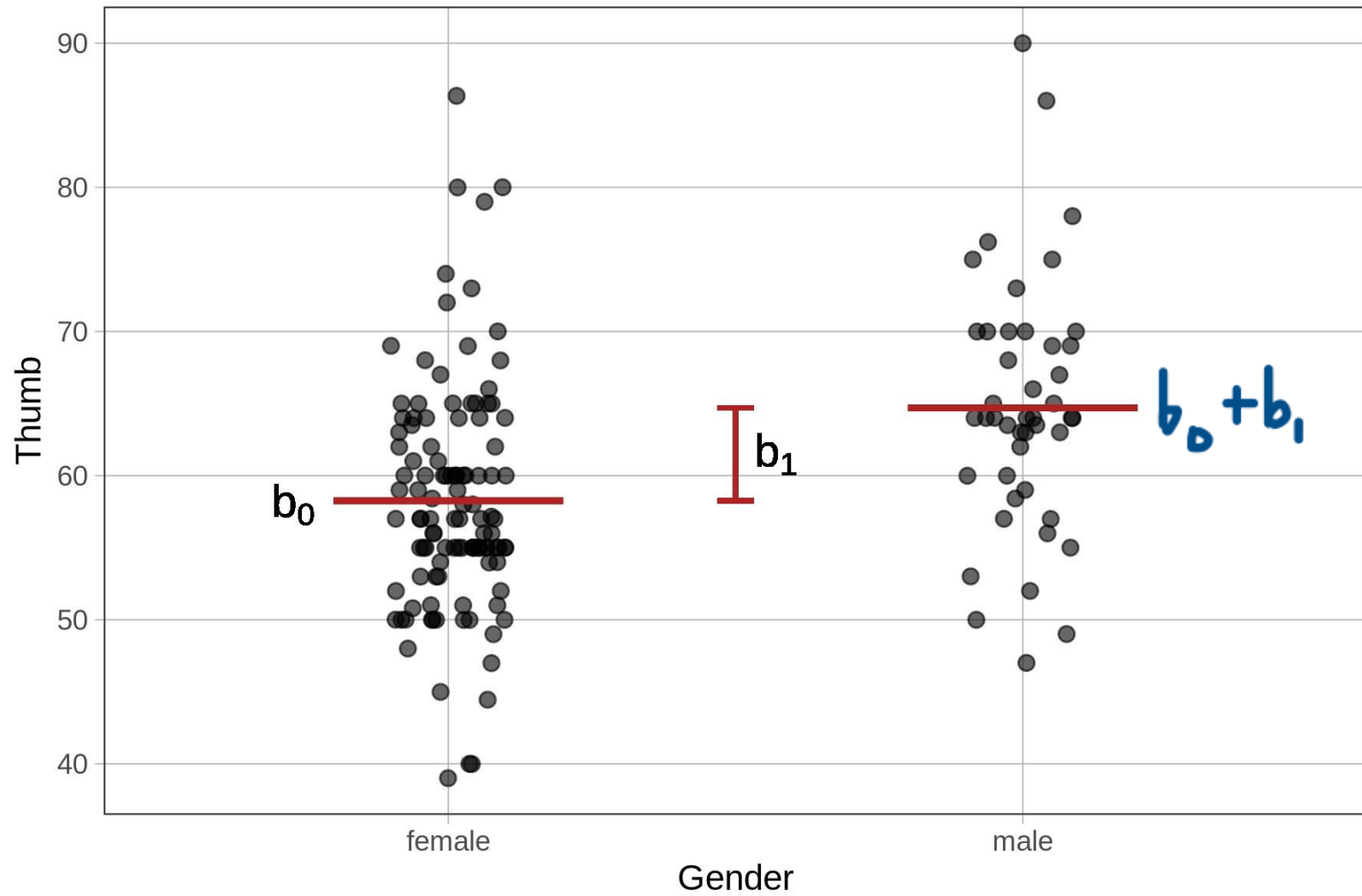
$$Y_i = b_0 + b_1 X_i + e_i$$

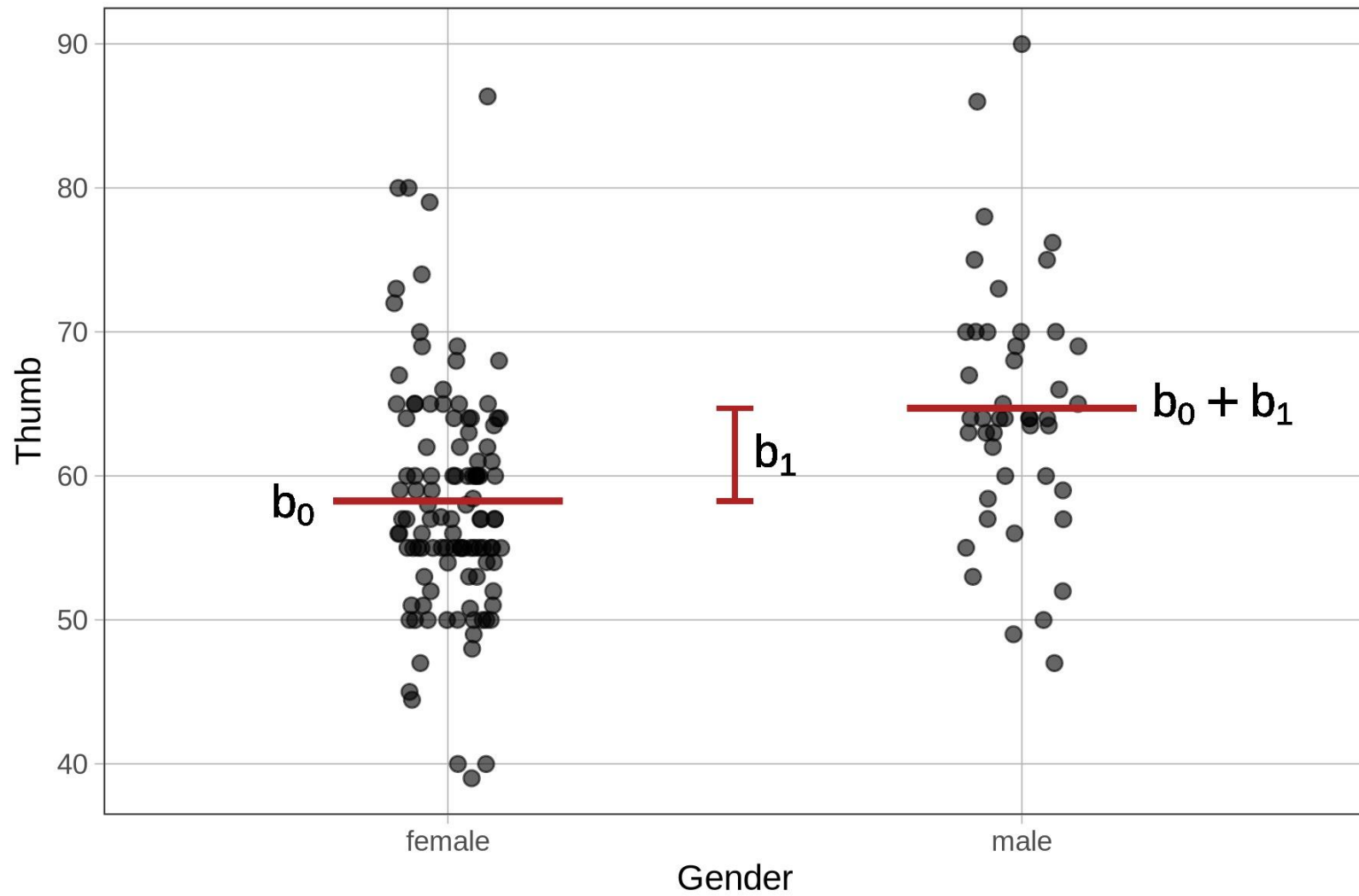
$$\text{Thumb}_i = b_0 + b_1 \text{Gender}_i + e_i$$

$$y_i = 58.256 + 6.447 \cdot 1 + e_i$$

αλν εἶναι
αλν τρῶς





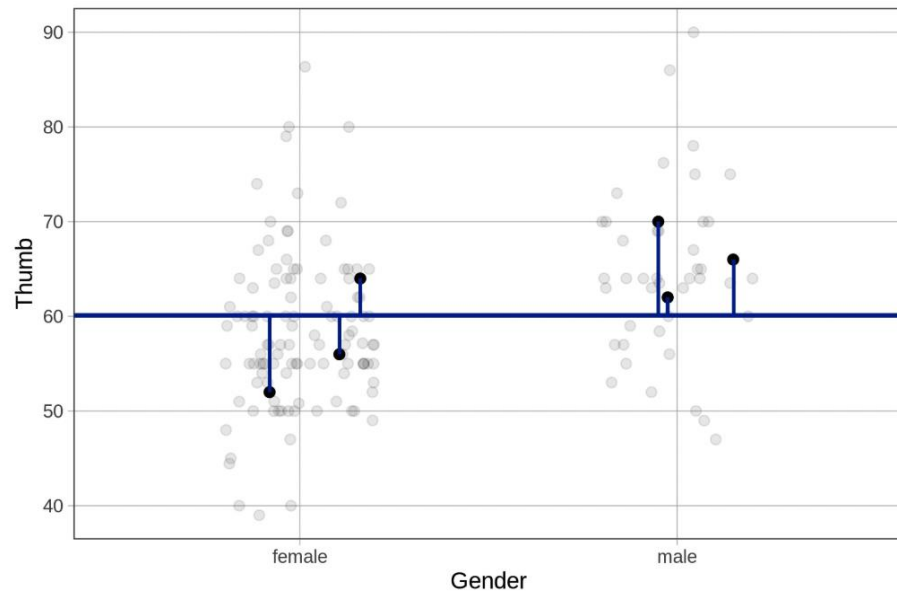


$$y_i = b_0 + b_1 X_i + b_2 Z_i + b_3 W_i$$

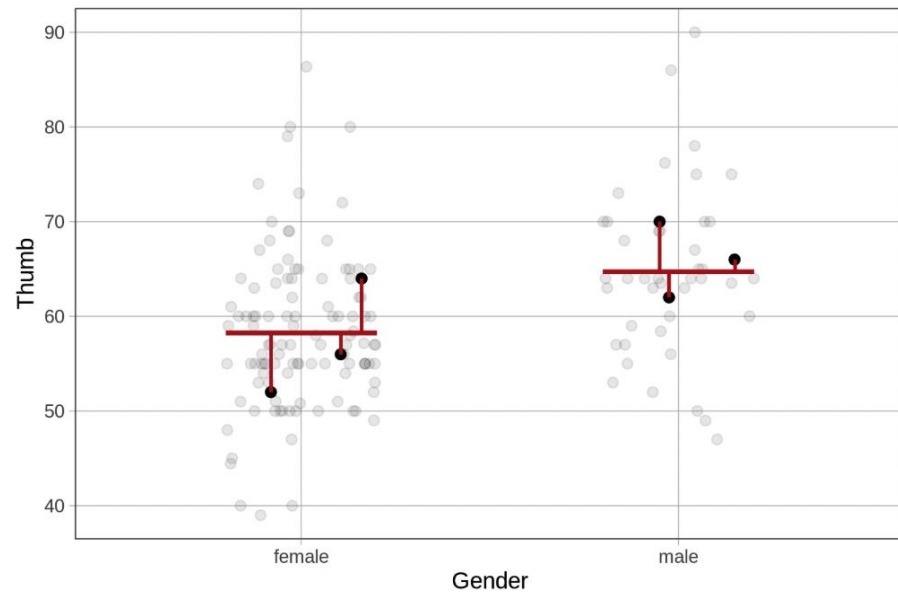
Ποιές από τις παρακάτω εξισώσεις αντιστοιχούν στο προσαρμοσμένο μοντέλο του Φύλου;
(Υπάρχουν περισσότερες από μία σωστές απαντήσεις.)

- ☐ A: $\text{Thumb}_i = 6.4 + 58.3\text{Gendermale}_i + e_i$
- ☐ B: $\text{Thumb}_i = 6.4 + 6.4\text{Gendermale}_i + e_i$
- ☐ C: $\text{Thumb}_i = 58.3 + 6.4\text{Gendermale}_i + e_i$
- ☐ D: $\text{Gender}_i = 58.3 + 6.4\text{Thumb}_i + e_i$
- ☐ E: $Y_i = 58.3 + 6.4X_i + e_i$

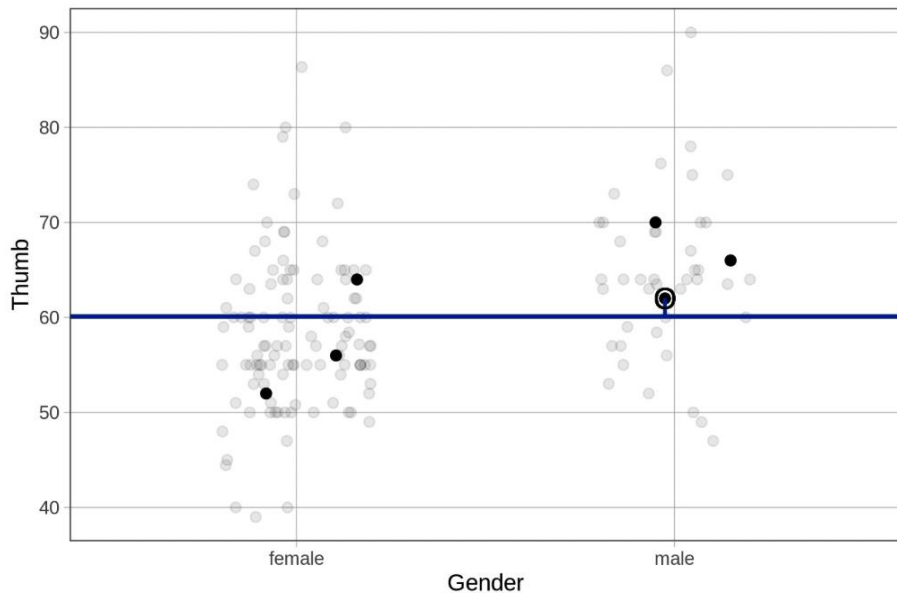
Σφάλματα από το κενό μοντέλο



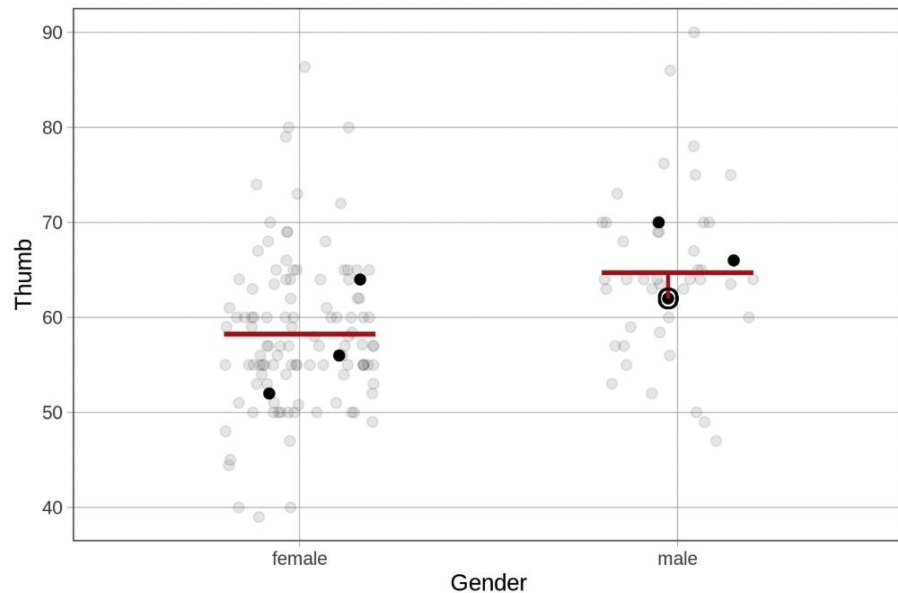
Σφάλματα από το μοντέλο του Φύλου

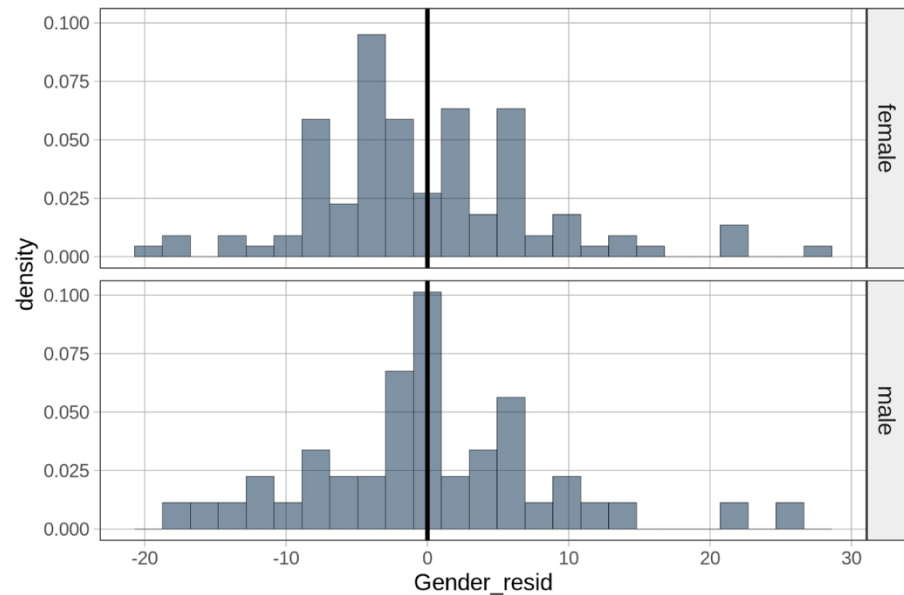
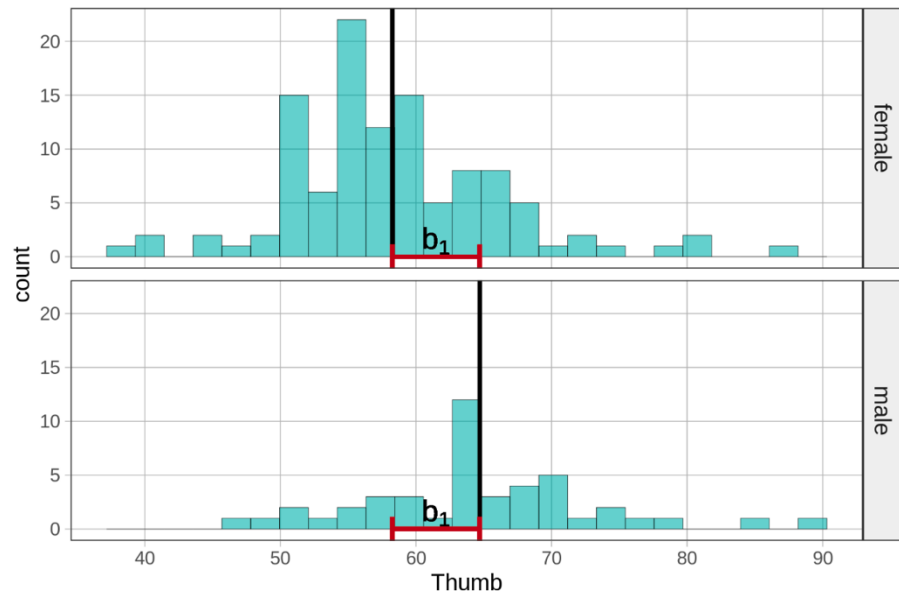


Σύμφωνα με το κενό μοντέλο, αυτός ο φοιτητής έχει μήκος αντίχειρα μεγαλύτερο από το μέσο όρο

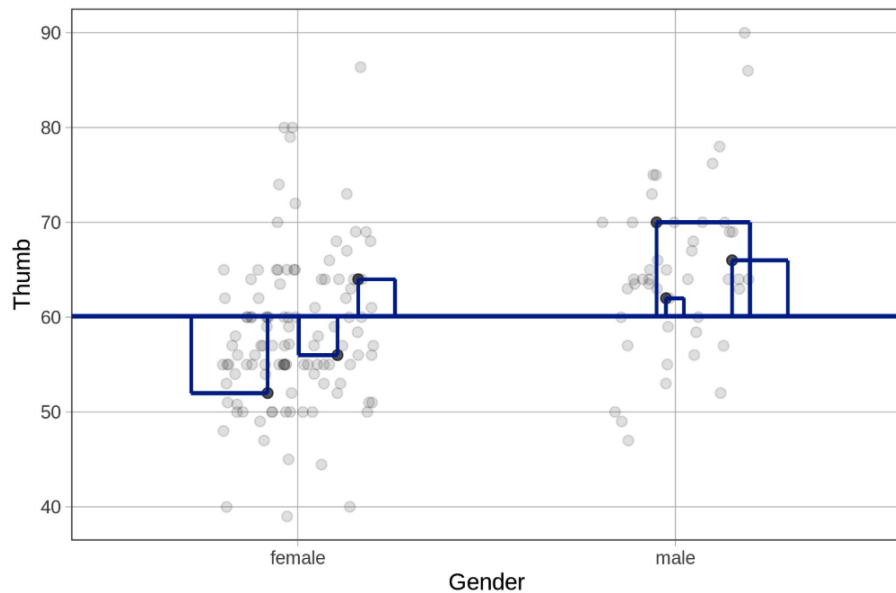


Σύμφωνα με το μοντέλο του φύλου, ο ίδιος φοιτητής έχει μήκος αντίχειρα ελαφρώς κάτω από το μέσο όρο με δεδομένο ότι είναι άντρας.

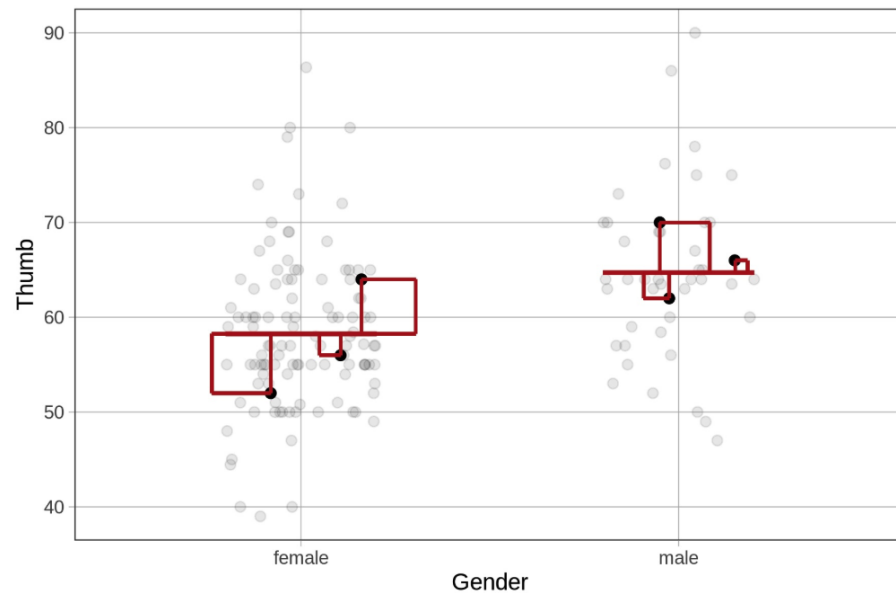




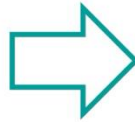
SS Total, Άθροισμα Τετραγώνων των Υπολοίπων από το κενό μοντέλο



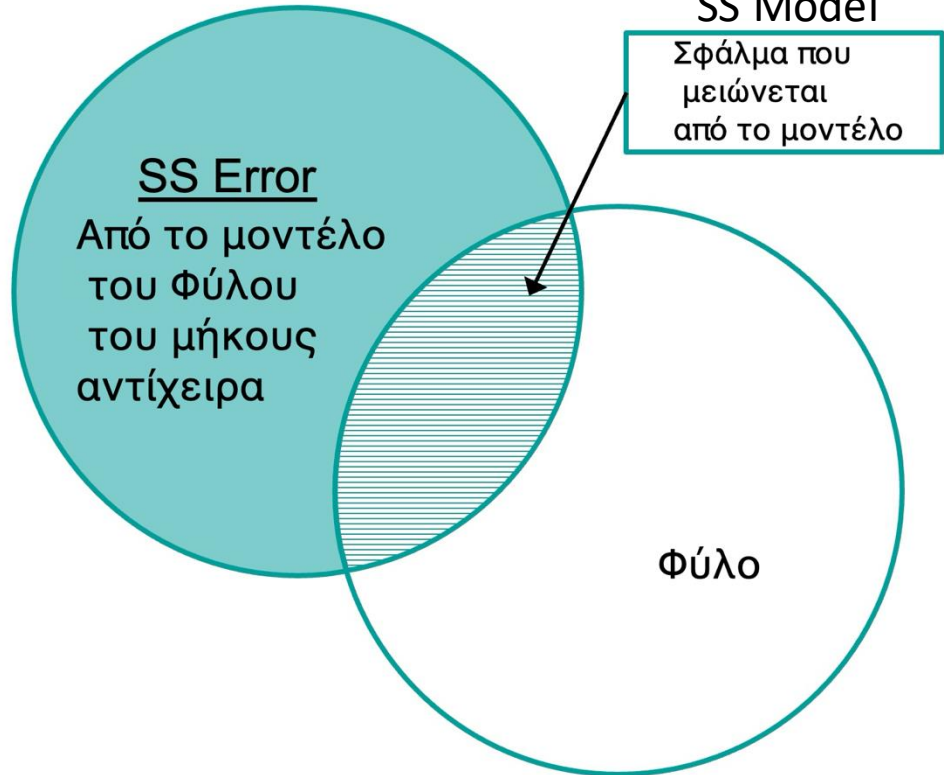
SS Error, Άθροισμα Τετραγώνων των Υπολοίπων από το μοντέλο του Φύλου



Κενό μοντέλο του μήκους αντίχειρα



Μοντέλο του Φύλου του μήκους αντίχειρα

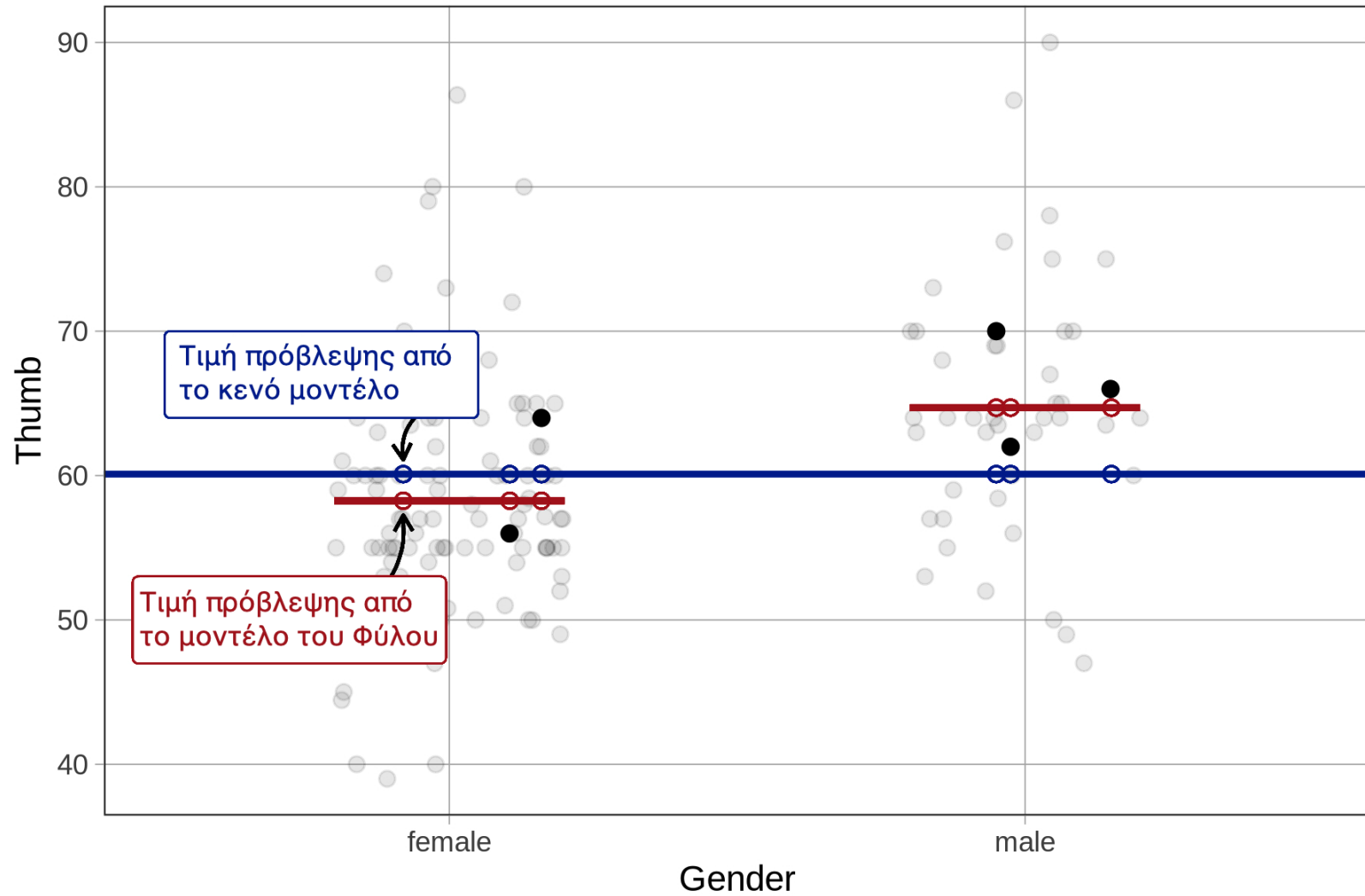


$$SS\ Total = SS\ Model + SS\ Error$$

Analysis of Variance Table (Type III SS)

Model: Thumb ~ Gender

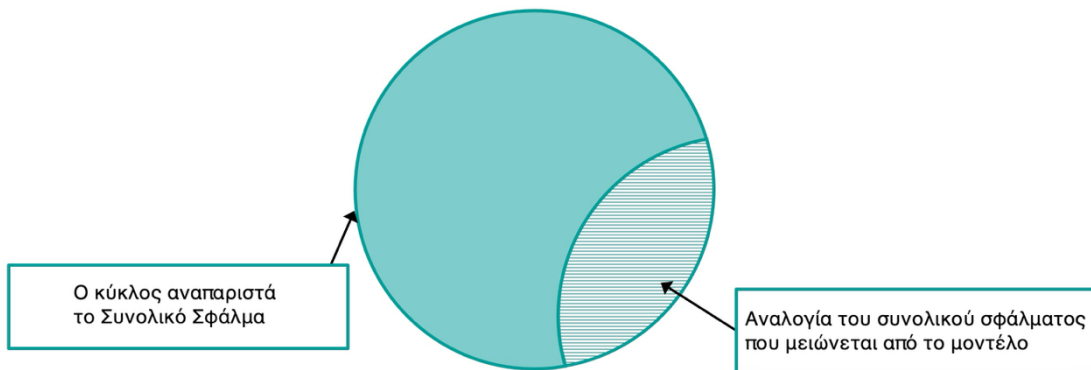
	SS	df	MS	F	PRE	p
Model (error reduced)	1334.203	1	1334.203	19.609	0.1123	.0000
Error (from model)	10546.008	155	68.039			
Total (empty model)	11880.211	156	76.155			



Χρήση της Αναλογικής Μείωσης του Σφάλματος (PRE) για τη Σύγκριση Δύο Μοντέλων

Υπολογίζεται ως το πηλίκο δύο αθροίσματων τετραγώνων. Είναι η αναλογία ή **το ποσοστό της συνολικής μεταβλητότητας της εξαρτημένης μεταβλητής που εξηγείται από την ανεξάρτητη μεταβλητή**

Μοντέλο του Φύλου του μήκους αντίχειρα



$$PRE = \frac{SS \text{ Model}}{SS \text{ Total}}$$