Αυτή η εργασία περιγράφει τη διαδικασία για τη δημιουργία ενός Δείκτη Κοινωνικοοικονομικής Κατάστασης (SES) στην απουσία δεδομένων σχετικά με το εισόδημα ή την κατανάλωση, με τη χρήση της Ανάλυσης Κύριων Συνιστωσών (PCA) σε μεταβλητές που σχετίζονται με την ιδιοκτησία ανθεκτικών περιουσιακών στοιχείων, την πρόσβαση σε υποδομές και υπηρεσίες και χαρακτηριστικά της κατοικίας. Ένας από τους λόγους που χρησιμοποιήθηκε το PCA είναι διότι η συλλογή ακριβών δεδομένων για το εισόδημα είναι απαιτητική εργασία απαιτώντας εκτενείς πόρους για τις έρευνες νοικοκυριών. PCA (Principal Component Analysis) είναι μια στατιστική τεχνική για πολλές μεταβλητές η οποία χρησιμοποιείται για να μειώσουμε τον αριθμό των διαστάσεων(μεταβλητών) διατηρώντας όσο τον δυνατών περισσότερα χαρακτηριστικά για τα δεδομένα μας(διακύμανση). Ως αποτέλεσμα μας δίνει την δυνατότητα απεικόνισης και την εύρεση συσχετίσεων μεταξύ των εκάστοτε μεταβλητών. Τέλος είναι πολύ χρήσιμο όταν έχουμε να αναλύσουμε σύνθετες μεταβλητές δηλαδή ένα μεγάλο αριθμό υποψήφιων μεταβλητών που θα μπορούσαν να συλλεχθούν. Όπου κάποιες από αυτές ίσως να παρέχουν παρόμοιες πληροφορίες. Συνεπώς, μια φυσική προσέγγιση είναι να χρησιμοποιηθούν μέθοδοι όπως η ανάλυση κύριων συνιστωσών (PCA) για να προσπαθήσουν να οργανώσουν τα δεδομένα. Χρησιμοποιήθηκαν δεδομένα από τη Δημογραφική Έρευνα για την Υγεία (Demographic Health Survey - DHS). Στην παραπάνω ανάλυση χρησιμοποιήθηκαν οι ακόλουθες μεταβλητές ιδιοκτησία ανθεκτικών περιουσιακών στοιχείων, την πρόσβαση σε υποδομές και υπηρεσίες (π.χ. εγκαταστάσεις υγιεινής και πηγή νερού) και τα χαρακτηριστικά της κατοικίας (π.χ. αριθμός δωματίων για ύπνο και υλικό κατασκευής) και 27355 παρατηρήσεις(framedata dimensions ,[ observations \* variables]). Επιπλέον πραγματοποιήθηκαν περιγραφικές αναλύσεις για όλες τις μεταβλητές. Αξίζει να αναφερθεί ότι τα δεδομένα σε κατηγορική μορφή δεν είναι κατάλληλα για PCA, καθώς η αρίθμηση τους δεν έχει κανένα νόημα. Γι’αυτό τον λόγο οι ποιοτικές κατηγορικές μεταβλητές πρέπει να ανακωδικοποιηθούν σε δυαδικές μεταβλητές. Ένα άλλο ζήτημα είναι οι ελλείπουσες τιμές οι οποίες δημιουργούν πρόβλημα τόσο στην εφαρμογή του PCA όσο και στα αποτελέσματα. Εάν αντικαταστήσουμε το μέσο σε κάθε τιμή που λείπει αυτό μειώνει την διακύμανση που είναι πρόβλημα, επομένως πρέπει να είμαστε πολύ προσεκτική πως θα αντιμετωπίσουμε της ελλείπουσες τιμές. Από τα αποτελέσματα του πίνακα PCA στην συγκεκριμένη έρευνα έδωσαν βάση στο factor score. περιλαμβάνει τις βαθμολογίες των συντελεστών για κάθε μεταβλητή. Οι βαθμολογίες αυτές δηλώνουν πώς κάθε μεταβλητή σχετίζεται με την κύρια συνιστώσα. Συνήθως, μια μεταβλητή με θετική βαθμολογία συσχετίζεται με υψηλότερο κοινωνικοοικονομικό καθεστώς (SES), ενώ μια μεταβλητή με αρνητική βαθμολογία συσχετίζεται με χαμηλότερο SES. Ερμηνεύοντας τις βαρύτητες από το παράδειγμά μας, ένα νοικοκυριό στην αστική Βραζιλία με περισσότερα αγαθά, νερό από καλώδιο που φτάνει στην κατοικία, εγκατάσταση υγιεινής που οδηγεί σε αγωγό, ολοκληρωμένα επενδύσεις δαπέδου και υψηλό αριθμό δωματίων για ύπνο θα επιτύχει μια υψηλότερη βαθμολογία Κοινωνικοοικονομικής θέσης (SES). Το εύρημα είναι παρόμοιο για την αγροτική Βραζιλία, εκτός εάν περιλαμβάνει οποιαδήποτε εγκατάσταση υγιεινής και έναν καλά. Στην αστική Αιθιοπία, ένα νοικοκυριό με περισσότερα αγαθά και νερό από τον αγωγό προς την συγκυρία θα επιτύχει μια υψηλότερη βαθμολογία SES. Στην αγροτική Αιθιοπία, η κατοχή οποιουδήποτε αγαθού ή η πρόσβαση σε υποδομές όπως το νερό ή η υγιεινή θα οδηγήσει σε υψηλότερη βαθμολογία SES. Στην συνέχεια χρησιμοποιείται μια άλλη στατιστική μέθοδο, ανάλυση συστάδων που επιτρέπει τον χωρισμό παρατηρήσεων σε ομάδες βάση κάποιον χαρακτηριστικών. Η συγκεκριμένη έρευνα/δημοσίευση μας εισάγει μια νέα μεθοδολογία για το πως να χρησιμοποιήσουμε ένα Δείκτη Κοινωνικοοικονομικής Κατάστασης από PCA. Επισημαίνοντας το βασικό πλεονέκτημα το οποίο είναι η αποφυγή προβλημάτων μέτρησης που συνδέονται με τις μεθόδους που βασίζονται στο εισόδημα και την κατανάλωση, όπως η εποχικότητα. Επιπλέον σε σύγκριση με άλλες στατιστικές η PCA είναι απλή στον υπολογισμό της. Παρόλα τα προβλήματα, πρόσφατα έργα για τους δείκτες SES που βασίζονται στο PCA υποδεικνύουν ότι αυτοί μπορούν να επικυρωθούν και είναι ανθεκτικοί. Ωστόσο, είναι καθήκον του χρήστη να έχει υπόψη του ότι το PCA θα πρέπει να θεωρείται κυρίως ως μια σύντομη εμπειρική μέθοδος.