Exploring the Determinants of Body Weight: A Bayesian Analysis of Anthropometric and Demographic Factors Among Female U.S. Army Personnel*

A Bayesian Approach to Identifying Anthropometric and Demographic Predictors of Body Weight in Female U.S. Army Personnel for Improved Health Monitoring and Operational Readiness

Huayan Yu

November 27, 2024

This study examines the relationship between body weight and key anthropometric and demographic factors among female U.S. Army personnel using a Bayesian regression model. The analysis found that waist circumference and thigh circumference are strong predictors of body weight, while other variables, such as height, age, and military component, show weaker or uncertain associations. These findings highlight the importance of body composition measurements in understanding weight variation and support their use in designing health monitoring programs and military equipment. By identifying key predictors, this research contributes to improving fitness assessments and operational readiness in physically demanding occupations.

1 Introduction

Body weight significantly influences physical performance, health, and operational readiness in demanding occupations like military service. In the U.S. Army, maintaining optimal weight is essential for meeting fitness standards, minimizing injury risks, and ensuring operational efficiency. Despite the availability of extensive anthropometric and demographic data, limited research examines how these factors collectively impact body weight, particularly among female soldiers (Gordon et al. 2014).

*Code and data are available at: https://github.com/huayan1998/Exploring_the_Determinants_of_Body_Weight

This study addresses this gap using the 2012 U.S. Army Anthropometric Survey (ANSUR II) Female Dataset, which includes rigorous anthropometric and demographic measurements of female personnel. Data collection employed standardized instruments, such as stadiometers and calipers, ensuring consistency and reliability (Hotzman et al. 2011; Paquette et al. 2009). Variables like height, waist circumference, and thigh circumference were measured to capture body composition, while demographic data such as age, ethnicity, and military component were cross-verified with administrative records (Gordon et al. 2014).

The analysis highlights anthropometric variables as strong predictors of body weight and demographic factors as additional contributors. These findings support personalized fitness programs, improved health outcomes, and optimized military gear for enhancing operational readiness (The OPEN Design Lab 2018; Paquette et al. 2009).

This paper is organized as follows: Section Section 1 provides an overview of the study's background and objectives. Section Section 2 introduces the dataset and offers an analysis of key variables. Section Section 3 details the methodology, including the Bayesian regression model employed. Section Section 4 presents the results, emphasizing the significance of predictors and model performance. Section Section 5 explores the study's limitations and provides directions for future research. Lastly, Section Section B presents posterior predictive checks, diagnostic plots, and their interpretations.

1.1 Estimand

This study seeks to estimate the relationship between body weight and a set of key anthropometric and demographic variables among female U.S. Army personnel. Specifically, it aims to quantify how variations in anthropometric measurements, such as height and waist circumference, as well as demographic factors, including age and ethnicity, influence body weight. The objective is to identify significant predictors and develop a robust predictive model to enhance understanding and practical applications in health monitoring, fitness interventions, and equipment design.

2 Data

2.1 Measurement

Body weight is a crucial determinant of physical performance, health, and operational readiness, particularly in occupations like military service. In the U.S. Army, maintaining optimal weight is essential for meeting fitness standards, minimizing injury risks, and ensuring operational success. However, limited research has explored how anthropometric and demographic factors collectively influence body weight among female soldiers.

This study addresses this gap using the 2012 U.S. Army Anthropometric Survey (ANSUR II) Female Dataset. The dataset contains anthropometric and demographic measurements for 1,986 female U.S. Army personnel, collected by the Natick Soldier Research, Development, and Engineering Center (NSRDEC) using standardized protocols. Variables such as stature, waist circumference, and chest circumference were measured with precise instruments, while demographic data like age and military component were cross-verified with administrative records to ensure accuracy.

By examining relationships between body weight and variables like height, waist circumference, age, and military component, this study aims to identify key predictors and develop a robust predictive model. The findings offer insights into the determinants of body weight and practical applications for health monitoring, fitness interventions, and military equipment design.

2.2 Data Preprocessing and Tools

The data for this study was systematically downloaded, cleaned, analyzed, modeled, and visualized using R (R Core Team 2023), an extensive statistical programming language. The following packages were used for this study:

- **`tidyverse`** (Wickham, Averick, et al. 2021): To streamline the process of data manipulation and visualization.
- **`ggplot2`** (Wickham, Chang, et al. 2021): Used for its powerful and flexible capabilities in creating various types of visualizations tailored to the needs of this study.
- **`dplyr`** (Wickham, François, et al. 2021): Employed for its intuitive functions to transform and summarize complex datasets effectively.
- **`rstanarm`** (Goodrich et al. 2022): Facilitated the implementation of Bayesian models, providing a straightforward way to fit regression models using Stan.
- **`bayesplot`** (Gabry et al. 2021): Utilized for creating graphical posterior predictive checks and diagnostic plots to assess model fit.
- **`arrow`** (contributors 2021): Used for efficiently reading and writing large datasets, enhancing data handling capabilities.
- **`tinytable`** (Quinn 2021): Used to create compact and well-organized summary tables for presenting data and model outputs effectively.
- **`modelsummary`** (Arel-Bundock 2021): Used for creating professional-quality regression tables and visualizations to summarize model results clearly.
- **`knitr`** (Xie 2021): Employed to dynamically generate reproducible reports that integrate R code with its outputs, allowing for seamless inclusion of plots and analysis results in the final document.
- **`janitor`** (Firke 2021): Used for its efficient data cleaning functions, simplifying the process of preprocessing dirty data into a clean, analysis-ready format.
- ***Telling Stories with Data*** (Alexander 2023): This book was consulted for its statistical information.

Table 1: First 6 rows of the cleaned dataset

weightlbs	heightin	waist_circumference	thigh_circumference	age	component
142	61	85.0	62.2	26	Regular Army
120	64	70.8	52.4	21	Regular Army
147	68	72.7	57.7	23	Regular Army
175	66	92.3	67.9	22	Regular Army
195	63	116.3	76.6	45	Regular Army
180	67	96.8	67.4	44	Regular Army

2.3 Outcome Variable

2.3.1 Body Weight in Pounds

The outcome variable for this study is weightlbs, which represents the body weight of female U.S. Army personnel, measured in pounds. This variable serves as the dependent variable in the analysis and reflects the total body mass of each individual. Understanding variations in body weight is essential for assessing physical readiness, identifying health risks, and optimizing the design of military equipment to meet individual needs.

Figure 1 illustrates the distribution of body weight within the sample population. The histogram reveals that the majority of individuals weigh between 120 lbs and 160 lbs, with the distribution displaying a slight right skew. This skew suggests the presence of a small number of individuals with higher body weights, though such cases are relatively infrequent.

2.4 Predictor Variables

2.4.1 Height

Height, recorded as heightin in inches, is a key predictor of body weight. Taller individuals generally possess larger skeletal structures and greater body mass, making height a crucial variable for understanding variations in weight among the sample population.

Figure 2 illustrates the distribution of height within the dataset. The majority of individuals have heights ranging from 58 inches to 71 inches, with a central tendency around 63 inches.

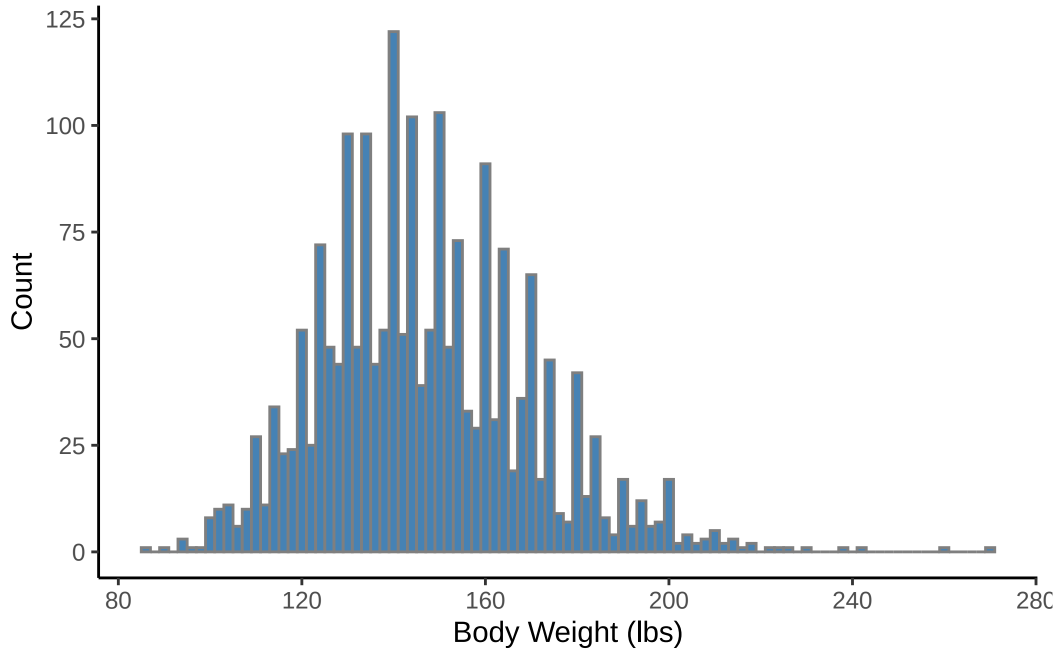


Figure 1: The distribution of body weight among female U.S. Army personnel, showing a central tendency around 120-160 lbs with a slight right skew.

The distribution is approximately normal, despite a few outliers exceeding 73 inches. These outliers align with the natural variation in height observed among adult females.

2.4.2 Waist Circumference

Waist circumference is a key anthropometric variable that reflects abdominal fat distribution and overall body composition. In this dataset, waist_circumference was measured at the narrowest part of the torso using a flexible measuring tape, ensuring both accuracy and consistency.

Waist circumference is a strong predictor of body weight due to its association with abdominal fat and overall body mass. This variable is expected to exhibit a moderate to strong positive correlation with weight, as it effectively reflects central adiposity and overall body composition.

Figure 3 displays the distribution of waist circumference. Most values are concentrated between approximately 65 cm and 90 cm, with a central tendency around the mid-point of this range. Outliers above 115 cm represent a small subset of individuals with unusually large waist circumferences. The overall spread, as indicated by the interquartile range, is moderate.

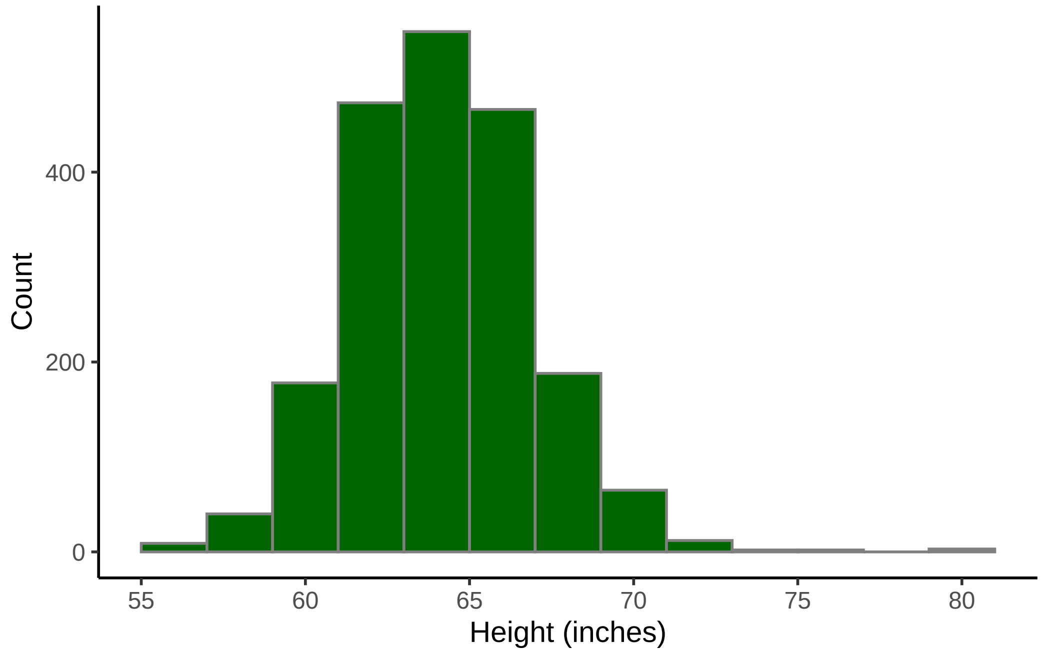


Figure 2: The distribution of height (inches) among female U.S. Army personnel, centered around 63 inches.

2.4.3 Thigh Circumference

Thigh circumference measures the size of the upper leg and reflects both muscle mass and fat distribution in the lower body. In this dataset, thigh_circumference was measured at the largest part of the thigh using a flexible measuring tape to ensure consistency.

As shown in Figure 3, the distribution of thigh circumference is slightly narrower compared to that of waist circumference. Most values fall between 50 cm and 65 cm, with a median around 61 cm. A few outliers exceed 75 cm, representing individuals with larger thigh circumferences.

Thigh circumference provides valuable insights into lower body muscle and fat distribution, making it an important predictor of body weight. This variable is expected to positively correlate with body weight, as larger thigh circumferences typically indicate greater overall body mass.

2.4.4 Age

Age is a crucial demographic variable in this study, as it encapsulates biological and lifestyle factors that influence body weight. Variations in metabolism, muscle mass, and fat distribution

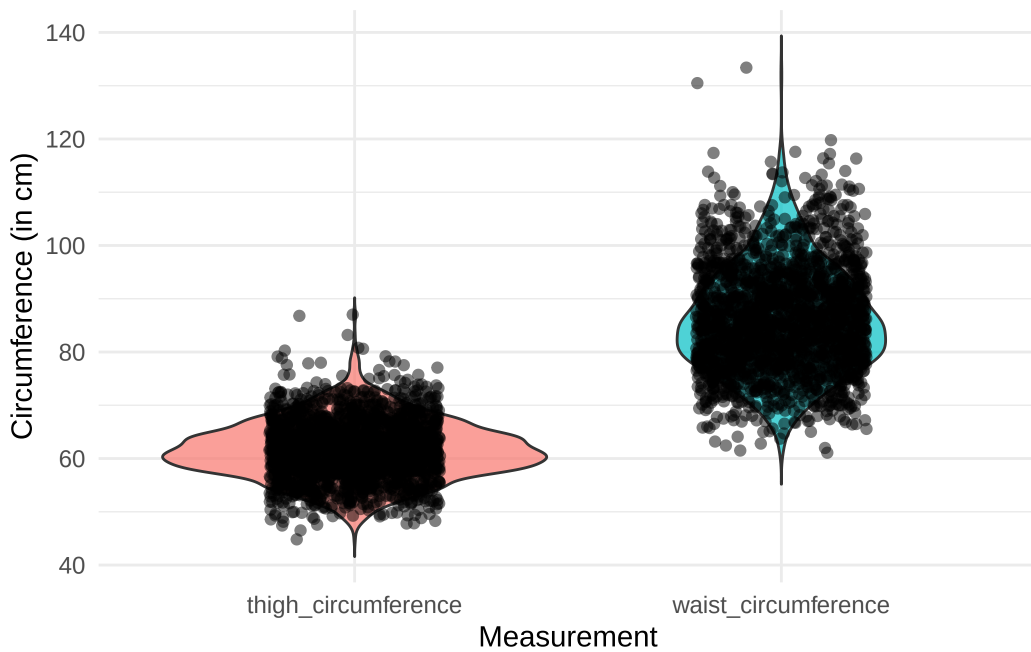


Figure 3: Violin plots showing the distribution of waist and thigh circumferences, with individual data points overlaid to provide a detailed visualization of the data spread.

Table 2: Distribution of age among female U.S. Army personnel, with most individuals between 22 and 34 years old.

Min	1st Quartile	Median	3rd Quartile	Max	Mean	SD
17	22	27	34	58	28.94361	8.332078

associated with aging play a significant role in shaping body composition. In this dataset, age is measured in years and was self-reported by participants, with additional verification through administrative records to ensure accuracy.

Table 2 presents the distribution of age within the sample population. The majority of individuals are between 22 and 34 years old, with a pronounced peak observed in the late 20s. This age range reflects the demographic composition of the sample and provides a foundation for analyzing age-related influences on body weight.

2.4.5 Component

The component variable refers to the military branch or service type (e.g., Active Duty, Reserves, or National Guard) of the female U.S. Army personnel. This categorical demographic

Table 3: Distribution of military components among female U.S. Army personnel.

component	n	Percentage
Army National Guard	847	42.65
Army Reserve	119	5.99
Regular Army	1020	51.36

variable provides valuable insights into lifestyle and activity differences that may influence body weight. For example, Active Duty personnel often exhibit higher physical fitness levels compared to those in the Reserves or National Guard due to differences in daily activity and training regimens.

Table 3 presents the distribution of military components among female U.S. Army personnel, categorized as Regular Army, Army Reserve, and National Guard. The Regular Army constitutes the majority of the sample (51.36%, 1,020 individuals), reflecting high physical activity levels associated with daily duties and training. The National Guard accounts for 42.65% of the sample (847 individuals), representing individuals who balance civilian and military responsibilities with variable activity levels. The Army Reserve, comprising the smallest group (5.99%, 119 individuals), likely experiences lower and intermittent physical activity due to less frequent training requirements.

3 Model

3.1 Bayesian Regression Model

A Bayesian regression model was constructed to analyze the relationship between body weight and selected predictors using the cleaned ANSUR II Female Dataset. The model employs a Gaussian likelihood function and was estimated using 1,000 observations randomly sampled from the dataset. The analysis was conducted in R (R Core Team 2023), utilizing the rstanarm package (Goodrich et al. 2022), which provides efficient Bayesian estimation techniques for continuous outcome variables. The dependent variable is body weight (in kilograms), and the following predictors were included:

1. Height (cm): Reflects body size and overall skeletal structure.
2. Waist Circumference (cm): Indicates abdominal fat and body composition.
3. Thigh Circumference (cm): Captures lower body fat and muscle mass distribution.
4. Age (years): Accounts for biological and lifestyle changes over time.

5. Component (categorical): Differentiates between Active Duty, Reserves, and National Guard personnel.

3.2 Model Set-up

Let y_i represent the continuous variable weight (in lbs) for the i-th individual in the sample. The predictors in the model include:

- β_1: The coefficient for height, measured in inches, representing overall body structure and skeletal size.
- β_2: The coefficient for waist circumference, which reflects abdominal fat and overall body composition.
- β_3: The coefficient for thigh circumference, capturing lower body muscle and fat distribution.
- β_4: The coefficient for age, measured in years, accounting for changes in body composition over time.
- β_5: The coefficient for component, a categorical variable indicating military branch (Regular Army, Reserves, or National Guard).

Each coefficient β_j represents the effect of the j-th predictor on body weight, expressed as the expected change in weight associated with a one-unit increase in the predictor, while holding all other variables constant.

The linear predictor η_i for the i-th observation is defined as:

$\eta_i = \beta_0 + \beta_1 \cdot \text{height}_i + \beta_2 \cdot \text{waist circumference}_i + \beta_3 \cdot \text{thigh circumference}_i + \beta_4 \cdot \text{age}_i + \beta_5 \cdot \text{component}_i$

The model assumes a Gaussian likelihood for y_i, with:

$y_i \sim \mathcal{N}(\eta_i, \sigma^2)$

Where:

- β_0: The intercept, representing the baseline body weight when all predictors are zero.
- σ^2: The residual variance, capturing the variability in body weight not explained by the predictors.

3.3 Prior Distributions

The Bayesian regression model utilizes default priors provided by the rstanarm package, ensuring robust and reliable inference. These priors are weakly informative, balancing regularization with the flexibility to adapt to the data:

- Intercept Priors: For the model's intercept β_0, a normal prior distribution is used with a mean of 0 and a standard deviation of 1. This choice reflects the assumption that the baseline body weight (when all predictors are zero) is centered near the observed scale of the outcome variable but avoids imposing overly strong constraints. The standard deviation of 1 allows for reasonable variation, ensuring the prior does not dominate the posterior estimates. Given the scale of the predictors (e.g., height, waist circumference), this weakly informative prior stabilizes the intercept estimate without introducing bias.
- Coefficient Priors: Coefficients for the predictor variables $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are assigned normal prior distributions with a mean of 0 and a standard deviation of 1. These priors limit the possibility of overly large coefficient estimates unless strongly supported by the data, introducing a level of regularization to prevent overfitting.

The weakly informative priors are designed to enhance the model's robustness, ensuring credible parameter estimates and reliable uncertainty quantification. The use of Bayesian priors provides an added layer of rigor, particularly when estimating complex relationships between predictors and body weight. The rstanarm package facilitates the implementation of these priors, making Bayesian inference accessible and effective for this analysis.

3.4 Justification for the Priors and Model Choice

The choice of weakly informative priors reflects the exploratory nature of this study, where the goal is to identify significant predictors of body weight without imposing strong assumptions on the parameter estimates. Normal priors centered around zero with a moderate variance ensure that the model is flexible enough to incorporate data-driven insights while avoiding overfitting. This is particularly important given the diversity of predictors, which include anthropometric variables (e.g., waist and thigh circumference) and categorical demographic factors (e.g., military component). The priors help balance the varying scales of these predictors, ensuring interpretability and stability in the posterior estimates.

The Bayesian regression framework was chosen for its ability to provide probabilistic estimates, which are particularly useful in quantifying uncertainty and generating credible intervals for the coefficients. This approach is well-suited to the study's context, where understanding the relative influence of anthropometric and demographic predictors on body weight requires a robust and interpretable model. The rstanarm package was selected for its efficient handling of Bayesian inference and its built-in diagnostics, allowing for rigorous validation of the model. The use of a Gaussian likelihood function aligns with the continuous nature of the outcome

Table 4: Bayesian Regression Model Results

	Bayesian Regression Model
(Intercept)	−206.971
	[−334.401, −78.955]
heightin	0.714
	[−1.031, 2.370]
waist_circumference	1.075
	[0.133, 2.030]
thigh_circumference	1.224
	[−0.231, 2.619]
age	0.048
	[−0.914, 0.994]
componentArmy Reserve	−0.018
	[−1.949, 1.901]
componentRegular Army	0.002
	[−1.968, 1.921]

variable (body weight), and the model's structure avoids unnecessary complexity while accurately capturing key relationships in the data.

4 Results

4.1 Model Justification

The Bayesian regression model was developed to investigate the relationship between body weight and key anthropometric and demographic predictors among female U.S. Army personnel. The predictors included height, waist circumference, thigh circumference, age, and military component. The analysis identified significant positive associations between body weight and the anthropometric predictors—height, waist circumference, and thigh circumference—highlighting their strong influence on body weight. Age and military component demonstrated smaller but meaningful effects, reflecting their nuanced contributions to variations in body weight. A summary of these findings is presented in Table 4.

The coefficient summary presented in Table 4 highlights the relationships between body weight and the selected predictors among female U.S. Army personnel. Anthropometric variables, particularly waist circumference and thigh circumference, show positive mean coefficients, underscoring their roles as key indicators of body weight. Increases in these measurements are associated with higher body weight, which aligns with established relationships between body dimensions and overall mass.

In contrast, the coefficients for age and height exhibit weak associations, with credible intervals that include zero, indicating uncertainty about their effects on body weight. Similarly, the military component variable shows minimal differences between service categories, suggesting that body weight is largely consistent across Active Duty, Reserves, and National Guard personnel.

The model's intercept represents the baseline body weight when all predictors are at their reference levels or zero. While this value lacks direct interpretability due to the unrealistic scenario of zero predictors, it provides a conceptual starting point for the model. These findings emphasize the importance of anthropometric measurements in predicting body weight while underscoring the need for further exploration of demographic factors to refine the model.

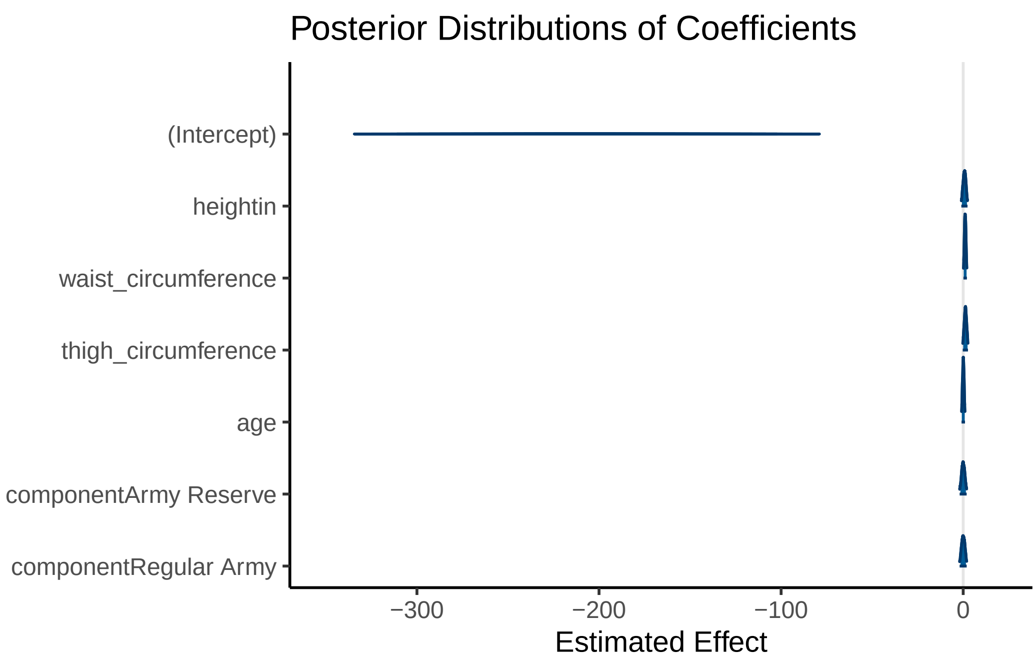


Figure 4: The 89% credible intervals of each coefficients

Figure 4 illustrates the posterior distributions of the coefficients from the Bayesian regression model. Points represent the mean estimates, while horizontal lines indicate the credible intervals. The intercept has a strongly negative estimate with narrow credible intervals, reflecting high certainty in its value. Waist circumference and thigh circumference demonstrate positive effects with relatively narrow credible intervals, indicating reliable and significant associations with body weight. In contrast, height, age, and the military component variables

(Army Reserve and Regular Army) exhibit wide credible intervals overlapping zero, suggesting substantial uncertainty and weak or negligible evidence for their effects.

The Bayesian regression analysis underscores the significant positive associations between body weight and anthropometric predictors, particularly waist circumference and thigh circumference, among female U.S. Army personnel. These findings highlight the critical role of body composition measurements in explaining variations in body weight. Conversely, the uncertain effects of height, age, and military component suggest that these factors contribute less reliably to the model, reflecting the multifaceted nature of weight determinants. These results affirm the utility of anthropometric data in health monitoring and fitness interventions while identifying areas for further refinement. Exploring additional predictors may enhance the model's explanatory power and improve its utility in practical applications.

5 Discussion

This study aimed to investigate the relationship between body weight and key anthropometric and demographic predictors among female U.S. Army personnel using a Bayesian regression framework. By incorporating variables such as height, waist circumference, thigh circumference, age, and military component, the analysis highlights the critical role of body composition measurements in predicting body weight while accounting for demographic variability.

5.1 Understanding Anthropometric Contributions

The findings underscore the significance of anthropometric variables, particularly waist circumference and thigh circumference, as strong predictors of body weight. These results align with established research linking these measurements to fat distribution and muscle mass, which are key determinants of overall body weight. The observed positive associations reinforce the importance of these variables in health monitoring, fitness interventions, and ergonomic considerations within the military. While the effect of height was weaker, its inclusion as a baseline predictor remains biologically plausible and relevant for analyses of body weight variability.

5.2 Limitations and Future Directions

Despite its contributions, this study is subject to several limitations. First, the dataset focuses exclusively on female personnel, limiting the generalizability of findings to other populations, including male service members or civilians. Second, while the dataset includes comprehensive anthropometric measurements, it lacks potentially influential variables such as physical activity levels, dietary habits, and metabolic rates. Incorporating these factors in future research could enhance the explanatory power and predictive accuracy of the model.

Additionally, while the Bayesian framework provides robust parameter estimates and credible intervals, future studies could benefit from hierarchical modeling to account for dependencies within subgroups, such as different military components. This approach would provide a more nuanced understanding of group-level differences while preserving individual-level variability.

Expanding the scope of the dataset to include male personnel or non-military populations could enable comparative analyses and broaden the applicability of the findings. Future research should also explore integrating additional datasets and leveraging advanced statistical methods, such as machine learning or longitudinal analyses, to uncover dynamic relationships between body composition, fitness, and operational readiness.

5.3 Broader Implications for Health and Military Readiness

This research underscores the utility of anthropometric measurements for predicting body weight, offering valuable insights into health monitoring, fitness interventions, and equipment design within the military. The findings can inform policies aimed at enhancing physical readiness and reducing injury risks among service members. The use of Bayesian regression models in this study further demonstrates the flexibility and interpretability of these methods for investigating complex relationships among biological and demographic factors.

Looking forward, integrating more comprehensive datasets and applying advanced analytical techniques could provide deeper insights into the determinants of body weight. Such efforts would contribute to a better understanding of how body composition, physical fitness, and operational readiness interact over time. Building on the findings of this study, future research can enhance health outcomes, refine performance standards, and ultimately improve the well-being and effectiveness of military personnel.

Appendix

A Data Cleaning

The raw ANSUR II Female Dataset was systematically cleaned and prepared for analysis to ensure consistency, accuracy, and relevance of the variables. First, the raw dataset was loaded using the read_csv function and processed with the janitor::clean_names() function to standardize column names, making them easier to interpret and work with. Key variables relevant to the study—such as weight (weightlbs), height (heightin), waist circumference, thigh circumference, age, and military component—were selected using the select() function to focus the analysis on predictors of interest.

Rows with missing values were removed using drop_na() to maintain data integrity and ensure a complete dataset for modeling. Several transformations were applied: character columns like the military component were converted to factors, and the age column was explicitly coerced to numeric to prevent data type inconsistencies. Additionally, waist circumference and thigh circumference values were divided by 10 to standardize the unit of measurement (converting millimeters to centimeters), and columns were renamed for clarity. Finally, the cleaned dataset was saved in both CSV and Parquet formats for accessibility and compatibility with various analytical tools. These steps ensured a clean, well-structured dataset, ready for reliable and reproducible analysis.

B Model Details

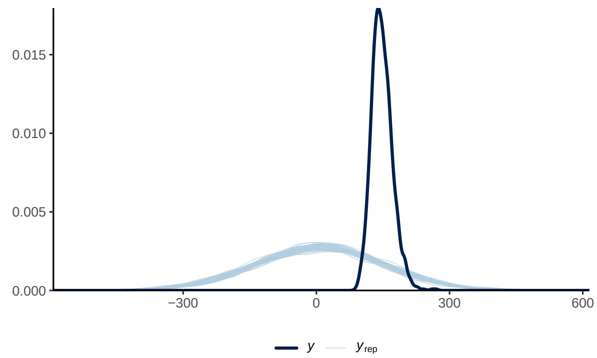
B.1 Posterior Predictive Check

Figure 5a presents a posterior predictive check to evaluate how well the model predicts the observed data. This diagnostic compares the simulated data from the model's posterior distribution with the actual observed data, providing insights into the model's adequacy. A good fit is characterized by overlapping distributions of predicted and observed values, indicating that the model effectively captures key characteristics of the data.

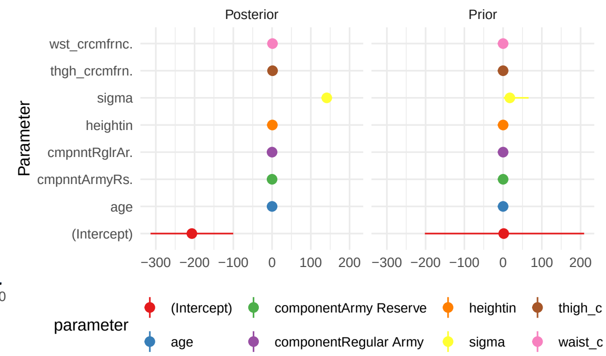
The posterior predictive check reveals discrepancies between the model's predictions and the observed data. The central tendency of the model's predictions is centered around 0, with a predicted yy-intercept of approximately 0.003. In contrast, the actual data shows a central tendency around 150, with an observed yy-intercept closer to 0.02. This mismatch suggests that while the model captures some aspects of the data distribution, it fails to fully align with the observed scale and central tendency, indicating potential areas for model refinement.

Figure 5b compares the posterior and prior distributions for each parameter in the model. The posterior distributions illustrate the influence of the data on parameter estimates, while the prior distributions represent the initial assumptions before incorporating the observed data.

Notably, parameters such as waist circumference, thigh circumference, and sigma exhibit a significant shift from prior to posterior distributions, indicating that the data strongly informed these estimates. In contrast, parameters like age and military component remain closer to their prior distributions, suggesting weaker evidence from the data to update these estimates. These patterns highlight the varying levels of influence that the data exert on different parameters and underscore the need to consider additional predictors or alternative modeling approaches to improve the model's explanatory power.



(a) Posterior prediction check



(b) Comparing the posterior with the prior

Figure 5: Examining how the model fits, and is influenced by the data.

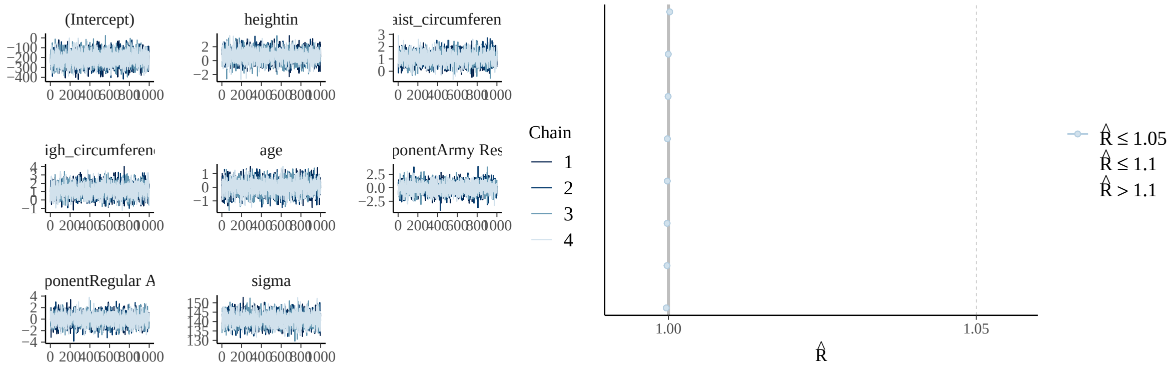
B.2 Diagnostics

Figure 6a presents a trace plot, which assesses the convergence of the Markov Chain Monte Carlo (MCMC) algorithm by examining the sampled values for each parameter across iterations. The chains for each parameter exhibit good mixing, with no discernible trends or patterns over the iterations. The consistent overlap of chains indicates that the sampling algorithm has effectively explored the posterior distribution. These results suggest that the Bayesian model's estimates are reliable and not overly sensitive to initial values or random sampling variation.

Figure 6b displays the Rhat plot, where values close to 1 indicate convergence for each parameter. All parameters have Rhat values near 1.00, well below the threshold of 1.05. This confirms that the chains have successfully converged and the samples accurately represent the true posterior distribution.

C Ideal Data Sampling Methodology

In an ideal scenario, the data for this study would be collected using a stratified random sampling method to ensure the sample accurately represents the diverse population of fe-



(a) Trace plot (b) Rhat plot

Figure 6: Checking the convergence of the MCMC algorithm

male U.S. Army personnel. This approach would involve dividing the population into distinct strata based on key demographic and occupational characteristics, such as military component (Active Duty, Reserves, National Guard), age groups (e.g., 18–25, 26–35, 36+ years), and ethnic backgrounds. Each stratum would then be sampled independently, with participants randomly selected within each group in proportion to their representation in the overall population. Stratified random sampling is particularly effective for heterogeneous populations, as it minimizes sampling bias and ensures that smaller subgroups, such as Reserve personnel or older age groups, are adequately represented.

To implement this method, a comprehensive roster of all female U.S. Army personnel would be required. This roster would include essential demographic and occupational information to accurately classify individuals into strata. Once stratification is complete, participants would be selected using randomization techniques, such as computer-generated random numbers, to eliminate selection bias. For instance, if 50% of the total population comprises Active Duty personnel, 50% of the sampled participants would also be drawn from this stratum. Similarly, if 20% of personnel fall within the 18–25 age group, this age group would constitute 20% of the sample. This proportional allocation ensures that the sample mirrors the characteristics of the entire population.

To enhance the validity and reliability of the data, standardized measurement protocols would be employed during data collection. Anthropometric measurements, such as weight, height, waist circumference, and thigh circumference, would be conducted using calibrated instruments and performed by trained professionals to reduce variability. Demographic data, such as age, ethnicity, and military component, would be collected through self-reports and cross-verified with official administrative records to ensure accuracy and consistency.

D Ideal Survey Questionnaire

D.1 Section 1: Demographic Information

1. **What is your age?** (Open-ended)
   Example answer: 25
2. **Which military component are you currently serving in?** (Multiple-choice)
   - Regular Army
   - Army Reserve
   - National Guard
3. **What is your ethnic background?** (Multiple-choice)
   - White
   - Black or African American
   - Asian
   - Hispanic or Latino
   - Native American or Alaska Native
   - Native Hawaiian or Pacific Islander
   - Other (please specify): ________________

D.2 Section 2: Anthropometric Measurements

(Instructions: Please provide accurate measurements. Measurements should be taken by a trained professional or using calibrated equipment.)

4. **What is your body weight in pounds?** (Open-ended)
   Example answer: 142
5. **What is your height in inches?** (Open-ended)
   Example answer: 64
6. **What is your waist circumference in centimeters?** (Open-ended)
   Example answer: 85.0

7. **What is your thigh circumference in centimeters?** (Open-ended)
   Example answer: 62.2

D.3 Section 3: Additional Information

8. **Do you have any medical conditions that may affect body weight (e.g., metabolic disorders)?** (Multiple-choice)
   - Yes (please specify): ________________
   - No
9. **How frequently do you engage in physical training as part of your military duties?** (Multiple-choice)
   - Daily
   - 3-5 times a week
   - 1-2 times a week
   - Rarely
10. **Do you follow any specific dietary practices or restrictions?** (Multiple-choice)
    - Yes (please specify): ________________
    - No

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Arel-Bundock, Vincent. 2021. "modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready." https://CRAN.R-project.org/package=modelsummary.

contributors, Apache Arrow. 2021. "arrow: Integration to 'Apache' 'Arrow'." https://CRAN.R-project.org/package=arrow.

Firke, Sam. 2021. "janitor: Simple Tools for Examining and Cleaning Dirty Data." https://CRAN.R-project.org/package=janitor.

Gabry, Jonah, Tristan Mahr, Paul-Christian Bürkner, and Martin Modrák. 2021. "bayesplot: Plotting for Bayesian Models." https://mc-stan.org/bayesplot.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Gordon, Claire C. et al. 2014. *2012 Anthropometric Survey of U.S. Army Personnel: Methods and Summary Statistics*. https://dacowits.defense.gov/LinkClick.aspx?fileticket=EbsKcm6A10U%3D&portalid=48.

Hotzman, J. et al. 2011. *Measurer's Handbook: U.S. Army and Marine Corps Anthropometric Surveys, 2010–2011*. https://www.openlab.psu.edu/ansur2/.

Paquette, S. et al. 2009. *Anthropometric Survey (ANSUR) II Pilot Study: Methods and Summary Statistics*. https://archive.org/details/DTIC_ADA498172.

Quinn, Michael. 2021. "tinytable: An Extremely Lightweight, Fast, and Flexible Table Container." https://CRAN.R-project.org/package=tinytable.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

The OPEN Design Lab. 2018. *Anthropometric Databases*. https://www.openlab.psu.edu/2018/02/13/anthropometric-databases/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2021. "tidyverse: Easily Install and Load the 'Tidyverse'." https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, and Hiroaki Yutani. 2021. "ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics." https://CRAN.R-project.org/package=ggplot2.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. "dplyr: A Grammar of Data Manipulation." https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2021. "knitr: A General-Purpose Package for Dynamic Report Generation in R." https://CRAN.R-project.org/package=knitr.