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

Identifying Key Predictors to Enhance Health Monitoring and Operational Readiness

aa

November 24, 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 is a critical factor influencing physical performance, health, and operational readiness, especially in physically demanding occupations like military service. Within the U.S. Army, maintaining optimal weight is vital for soldiers to meet physical fitness standards, minimize injury risks, and ensure the success of military operations. Furthermore, excessive weight can increase healthcare costs and compromise operational efficiency, underscoring the importance of understanding the factors that contribute to variations in body weight. Despite the extensive data collected on anthropometric and demographic characteristics of military

*Code and data are available at:

personnel, there is limited research examining how these factors collectively influence body weight among U.S. Army soldiers, particularly female personnel.

This study aims to fill this gap by leveraging the 2012 U.S. Army Anthropometric Survey (ANSUR II) Female Dataset, a comprehensive collection of anthropometric and demographic measurements of female Army personnel. Specifically, the study focuses on identifying key anthropometric and demographic predictors of body weight and developing a predictive model to enhance health monitoring, fitness interventions, and equipment design. By examining relationships between body weight and variables such as height, waist circumference, and demographic factors like age and ethnicity, this research offers new insights into the determinants of body weight in a physically demanding workforce.

Our findings reveal that a combination of anthropometric measurements, including height, waist circumference, and upper and lower body circumferences, significantly predict body weight. Demographic factors, such as age and ethnicity, also contribute to variations in weight, highlighting the interplay between biological and environmental factors. The predictive model developed in this study demonstrates strong performance, offering practical applications for designing personalized fitness programs, improving health outcomes, and optimizing military gear.

This paper is organized as follows: Section 2 reviews the methodology, including the selection of variables and statistical approach. Section 3 presents the results, focusing on the significance of predictors and model performance. Section 4 discusses the implications of the findings for military readiness, health interventions, and equipment design. Finally, Section 5 concludes with the study's contributions, limitations, and suggestions for future research.

1.1 Estimand

This study estimates the relationship between body weight and key anthropometric and demographic variables among female U.S. Army personnel. It quantifies how variations in measurements like height and waist circumference, and factors such as age and ethnicity, influence body weight, aiming to identify significant predictors and develop a robust predictive model.

2 Data

2.1 Measurement

The dataset used in this study is the 2012 U.S. Army Anthropometric Survey (ANSUR II) Female Dataset, which provides detailed anthropometric and demographic measurements of 1,986 female U.S. Army personnel. The data collection was conducted by the Natick Soldier Research, Development and Engineering Center (NSRDEC) between 2010 and 2012, using standardized protocols to ensure consistency and reliability.

Anthropometric measurements were collected using precise tools such as stadiometers, measuring tapes, and calipers, ensuring accurate capture of variables like stature (height in cm), waist circumference, chest circumference, and biceps circumference (flexed). For example, stature was measured with participants standing upright, waist circumference at the narrowest point of the torso, and chest circumference at its widest point. These methods ensured that the physical measurements reflected real-world body compositions with minimal error.

Demographic variables, including age, ethnicity, and component (Active Duty, Reserves, or National Guard), were self-reported and cross-verified with administrative records to ensure accuracy. This meticulous process ensured that both physical and demographic data were reliable for analysis.

Data preprocessing for this study involved cleaning and organizing the dataset to focus on variables relevant to predicting body weight. This included removing implausible outliers, ensuring uniformity in measurement units, and verifying data completeness. The clean dataset was analyzed and modeled using R, leveraging the following packages:

- `tidyverse`: For efficient data manipulation and visualization.
- `ggplot2`: To create detailed and insightful visualizations.
- `dplyr`: For transforming and summarizing the dataset effectively.
- `broom`: For tidying model outputs and integrating them into analysis workflows.
- `caret`: To support the modeling and evaluation of predictive models.
- `knitr`: For creating dynamic, reproducible reports.

Table 1: Table of 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

2.2 Outcome Variable

2.2.0.1 Body Weight in Pounds

The outcome variable for this study is weightlbs, which measures the body weight of female U.S. Army personnel in pounds. This variable represents the total body mass of an individual and serves as the dependent variable in the analysis. Understanding variations in body weight is crucial for examining physical readiness, health risks, and the design of military equipment tailored to individual needs.

@fig-weight-distribution illustrates the distribution of body weight among the sample population. The histogram shows that most individuals weigh between 120 lbs and 160 lbs, with the distribution displaying a slight right skew. This skew indicates the presence of a few individuals with higher body weights, though these are relatively uncommon.

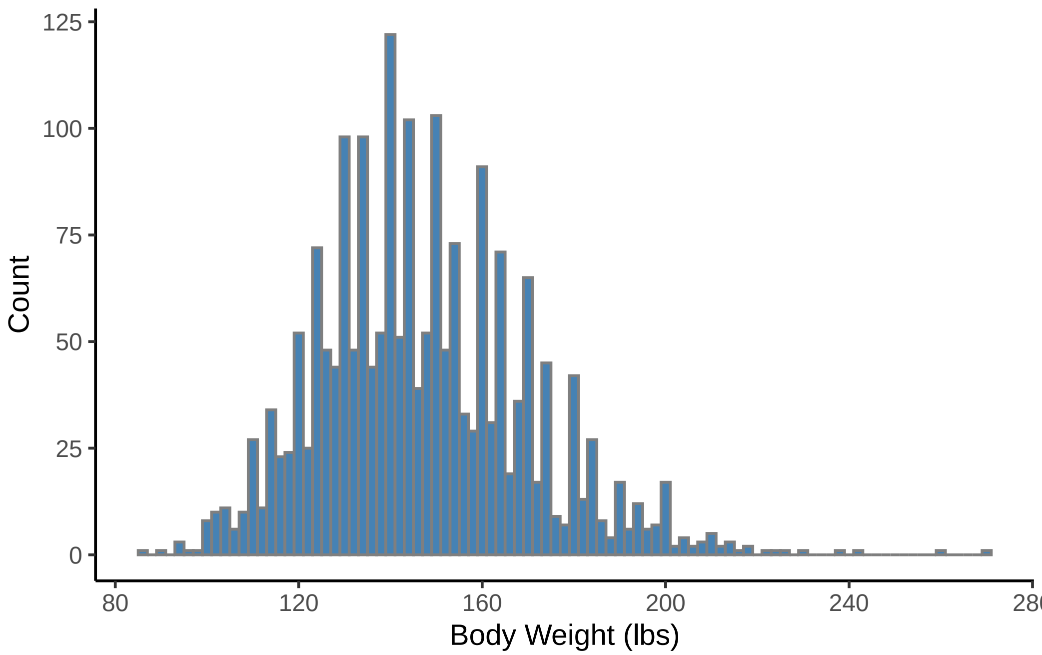


Figure 1: The distribution of body weight among female U.S. Army personnel, showing a central tendency around 120-160 lbs with a slight right skew.

2.3 Predictor Variables

2.3.0.1 Height

Height, measured as heightin in inches, is one of the primary predictors of body weight. Taller individuals typically have larger skeletal structures and more body mass, making height a critical variable in understanding variations in weight.

@fig-height-distribution shows the distribution of height among the sample population. Most individuals have heights clustered between 58 inches and 71 inches, with a central tendency around 63 inches. The distribution appears approximately normal if despite the outliers of person above 73 inches, which aligns with expected variations in height for adult females.

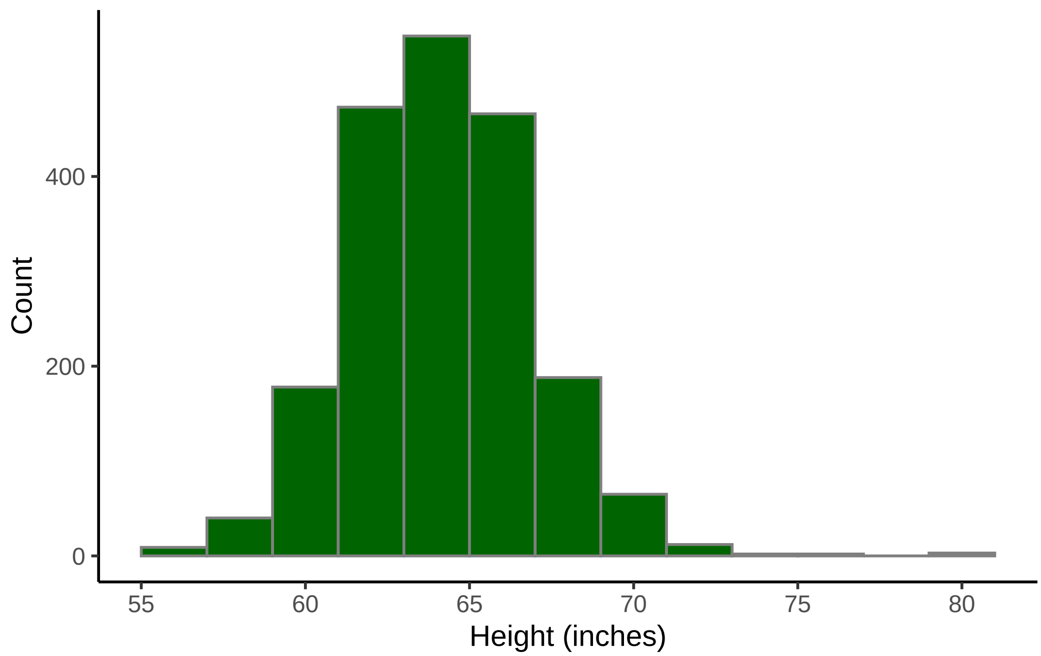


Figure 2: The distribution of height (inches) among female U.S. Army personnel, centered around 63 inches.

2.3.0.2 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 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 show a moderate to strong positive correlation with weight, particularly as it reflects adiposity and central body composition.

Figure 3 The violin plot shows that most values are concentrated between approximately 65 cm and 90 cm, with a central tendency around the mid-point of this range. Outliers are present above 115 cm, indicating a small number of individuals with unusually large waist circumferences. The overall spread (interquartile range) is moderate.

2.3.0.3 Thigh Circumference

Thigh circumference measures the size of the upper leg, reflecting 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, ensuring consistency.

Figure 3 shows the distribution is slightly narrower than that of waist circumference, with most values falling between 50 cm and 65 cm, The median appears to be around 61 cm. A few outliers exist above 75 cm, representing individuals with larger thigh circumferences.

Thigh circumference provides insight into lower body muscle and fat distribution, making it an important predictor of weight. This variable is expected to have a positive correlation with body weight, as larger thigh circumferences typically indicate greater body mass.

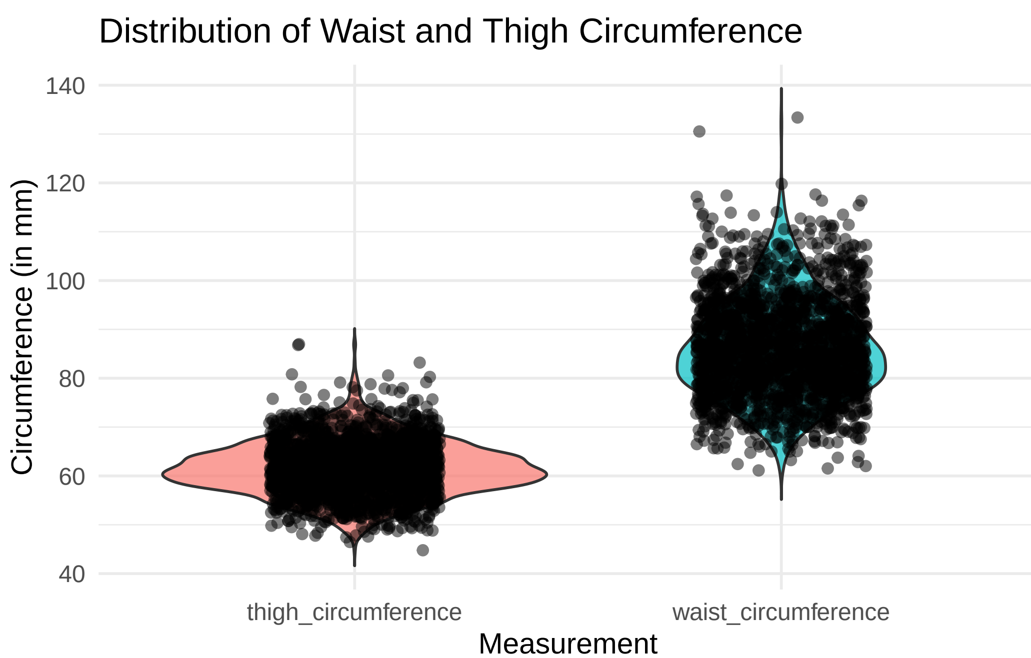


Figure 3: Figure: Violin plots showing the distribution of waist and thigh circumferences, with individual data points overlaid to provide a detailed visualization of the data spread.

2.3.0.4 Age

Age is a critical demographic variable in this study, reflecting biological and lifestyle factors that influence body weight. As individuals age, changes in metabolism, muscle mass, and fat

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

Table 3: Summary of Age Distribution

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

The table showing distribution of age among female U.S. Army personnel, with most individuals between 22 and 34 years old.

distribution can impact body composition. In the dataset, age is measured in years and was self-reported by participants, with additional verification through administrative records to ensure accuracy.

Table 2 illustrates the distribution of age within the sample population. Most individuals are between 22 and 34 years old, with a noticeable peak around the late 20s.

2.3.0.5 Component

Component refers to the military branch or service type (e.g., Active Duty, Reserves, or National Guard) of the female U.S. Army personnel. This demographic variable is categorical and provides insight into lifestyle and activity differences that may influence body weight. For instance, Active Duty personnel may have different physical fitness levels compared to those in the Reserves or National Guard due to varying levels of daily activity and training regimens.

@tbl-component-distribution shows the distribution of military components among female U.S. Army personnel, categorized as Regular Army, Army Reserve, and National Guard. The Regular Army comprises the majority (51.36%, 1,020 individuals), reflecting high physical activity levels due to daily duties and training. The National Guard represents 42.65% (847 individuals), with more variable schedules balancing civilian and military roles. The Army Reserve, the smallest group (5.99%, 119 individuals), likely experiences intermittent physical activity due to less frequent training requirements.

3 Model

3.1 Bayesian Regression Model

To analyze the relationship between body weight and selected predictors, a Bayesian regression model was constructed using the cleaned ANSUR II Female data. The Bayesian regression model uses a Gaussian likelihood function. The model is based on a total of 1000 observations, randomly sampled from the cleaned dataset. We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022), taking advantage of its efficient Bayesian

Table 4: The table showing the distribution of military components (Regular Army, Reserves, National Guard) among female U.S. Army personnel.

Table 5: Summary of components Distribution

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

Distribution of military components are categorized as Regular Army, Reserves, and National Guard.

estimation techniques for continuous outcome variables. The model uses weight (kg) as the dependent variable and the following predictors:

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

In the Bayesian regression model implemented using the rstanarm package, default priors are applied to the model parameters to ensure robust and reliable inference. These priors are designed to be weakly informative, balancing the need for regularization with 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 prior helps stabilize the intercept estimate without imposing overly strong assumptions about its central value.
- 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 priors are intentionally weakly informative to provide flexibility while maintaining some level of regularization, ensuring that the model remains robust to overfitting. The rstanarm package handles the specification and implementation of these priors, making Bayesian inference accessible for a wide range of applications. By incorporating these priors, the model can better capture the relationships between predictors and body weight, producing credible parameter estimates and uncertainty quantification.

Table 6: 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]

4 Results

4.1 Model Justification

The Bayesian regression model was developed to understand the relationship between body weight and anthropometric and demographic predictors among female U.S. Army personnel. The predictors included height, waist circumference, thigh circumference, age, and military component. The analysis revealed that anthropometric measurements, particularly height, waist circumference, and thigh circumference, have significant positive associations with body weight, while age and component showed smaller but meaningful effects. These findings are summarized in Table 6.

The coefficient summary presented in Table 6 reflects the relationships between body weight and selected predictors among female U.S. Army personnel. For instance, waist circumference and thigh circumference show positive mean coefficients, suggesting their roles as key indicators of body weight, with increases in these measurements associated with higher body weight. These findings align with the expected relationship between anthropometric dimensions and overall body mass.

Conversely, the coefficients for age and height exhibit weak associations, with credible intervals that include zero, indicating uncertainty about their effects on body weight. Additionally, the military component variable shows minimal differences between groups, suggesting that body weight is largely consistent across different service categories.

The intercept serves as the baseline of the model, representing the expected body weight when all predictors are at their reference levels or zero. While this value is not directly interpretable due to the unrealistic scenario of predictors being zero, it provides a starting point for the model. The results emphasize the significance of body composition measurements, while highlighting the need for further refinement in understanding the role of demographic factors.

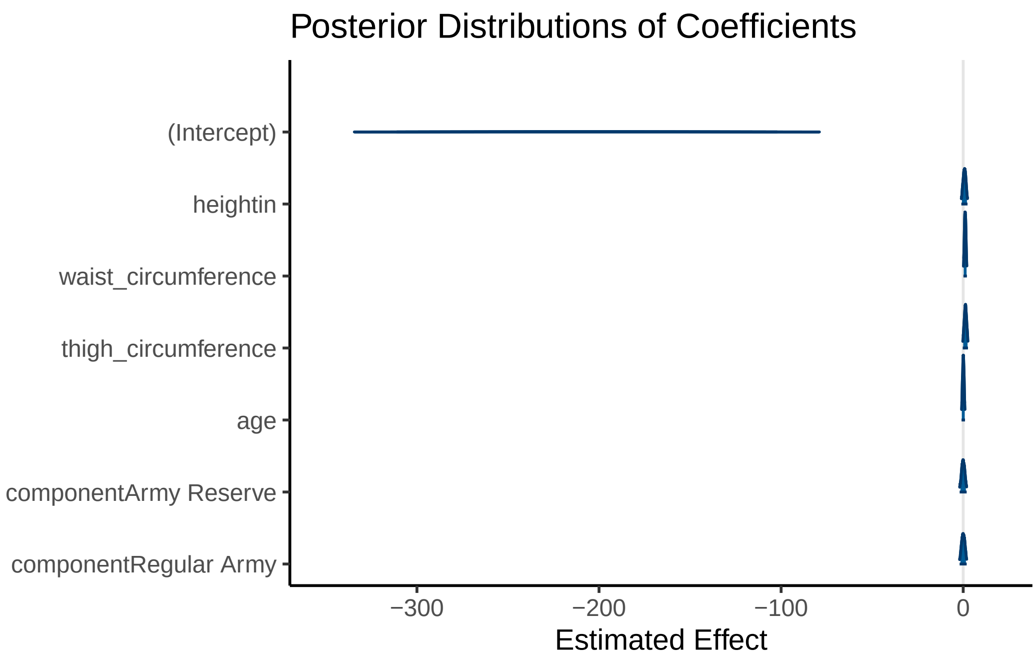


Figure 4: The 89% credible intervals of each coefficients

Figure 4 displays the posterior distributions of the coefficients from the Bayesian regression model, with points representing mean estimates and horizontal lines showing credible intervals. The intercept has a highly negative estimate with tight credible intervals, suggesting strong certainty. Waist circumference and thigh circumference have positive effects with narrower credible intervals, indicating more reliable associations with body weight. In contrast, height, age, and component variables (Army Reserve and Regular Army) have wide credible intervals overlapping zero, indicating substantial uncertainty and weak or negligible evidence for their effects.

The Bayesian regression analysis highlights the significant positive associations between body weight and anthropometric predictors, particularly waist circumference and thigh circumference, among female U.S. Army personnel. These findings emphasize the importance of body composition measurements in understanding variations in weight. However, the weak and uncertain effects of height, age, and military component suggest that these factors contribute less reliably to the model, underscoring the complexity of weight determinants. The results reinforce the utility of anthropometric data for health monitoring and fitness interventions, while also identifying areas where further refinement and exploration of additional predictors could enhance the model's explatory power.

5 Discussion

This study sought to understand the relationship between body weight and anthropometric and demographic predictors among female U.S. Army personnel through a Bayesian regression framework. By selectively incorporating key variables, such as height, waist circumference, thigh circumference, age, and military component, this analysis identifies patterns that highlight the critical role of body composition measurements in predicting weight while accounting for demographic variability.

5.1 Understanding Anthropometric Contributions

The results emphasize the significance of anthropometric variables, particularly waist circumference and thigh circumference, in predicting body weight. These findings align with established literature, which associates these measurements with fat distribution and muscle mass, critical components of overall body weight. The observed positive associations reinforce the relevance of these variables for health monitoring, fitness interventions, and ergonomic considerations within the military. While height showed weaker effects, its contribution to body weight variability remains biologically plausible, suggesting its inclusion as a baseline predictor in similar analyses.

5.2 Limitations and Future Directions

Despite its contributions, this study is constrained by several limitations. First, the dataset focuses exclusively on female personnel, limiting the generalizability of findings to other populations. Additionally, while anthropometric measurements were comprehensive, other potentially influential variables, such as physical activity levels, dietary habits, or metabolic rates, were not included. Future research could integrate these variables to enhance the explanatory power of predictive models. Expanding the dataset to include male personnel or non-military populations could also provide comparative insights into the determinants of body weight across different contexts.

Moreover, while the Bayesian framework offers robust parameter estimates and credible intervals, incorporating hierarchical models could account for potential dependencies within subgroups (e.g., military components). This approach would allow for a deeper understanding of group-level differences while preserving individual-level variability.

5.3 Broader Implications for Health and Military Readiness

This research highlights the utility of anthropometric measurements for predicting body weight, providing valuable insights for health monitoring, fitness interventions, and equipment design within the military. The findings can inform policies aimed at optimizing physical readiness and reducing injury risks among service members. Furthermore, the approach used in this study demonstrates the potential of Bayesian regression models in health and military research, offering a flexible and interpretable framework for understanding complex relationships among biological and demographic factors.

Looking forward, integrating additional datasets and applying advanced statistical methods, such as machine learning or longitudinal analyses, could provide new insights into the dynamic relationships between body composition, fitness, and operational readiness. By building on the findings of this study, future research can enhance the understanding of weight determinants and contribute to improved health outcomes and performance standards within the military.

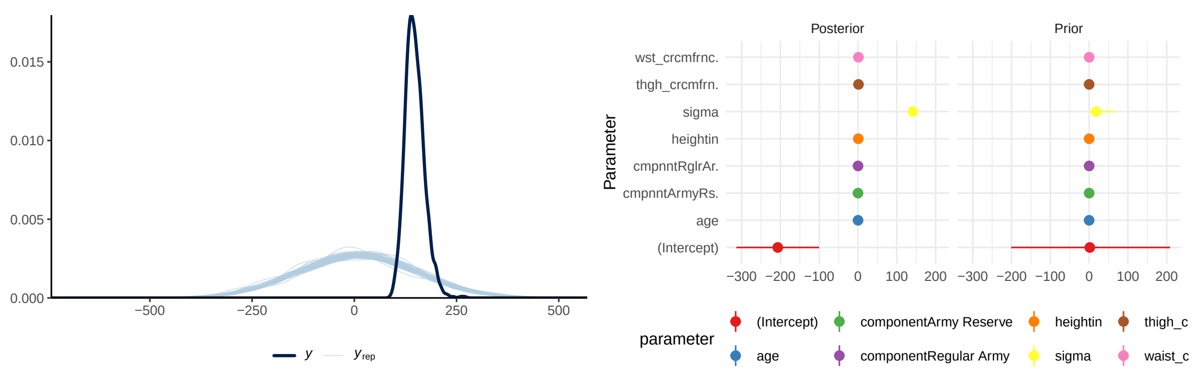
Appendix

A Model details

A.1 Posterior Predictive Check

In Figure 5a, we implement a posterior predictive check to assess how well the model predicts the observed data. This plot compares the simulated data from the model's posterior to the actual observed data, providing insight into the adequacy of the model's fit. A good fit would display overlapping distributions of the predicted and observed values, indicating that the model effectively captures the key characteristics of the data. The posterior predictive check reveals discrepancies between the model's predicted outcomes and the observed data. The central tendency of the model's predictions is centered around 0, with a predicted y intercept of approximately 0.003. In contrast, the actual data exhibits a central tendency around 150, with an observed y 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.

in Figure 5b, we compares the posterior and prior distributions for each parameter in the model. The posterior distributions reflect the data's influence on the estimates, while the prior distributions represent the initial assumptions before observing the data. Parameters such as waist circumference, thigh circumference, and sigma show a noticeable shift from prior to posterior, indicating that the data had a strong influence on these estimates. In contrast, parameters like age and military component remain closer to their prior distributions, reflecting weaker evidence from the data to update these estimates.



(a) Posterior prediction check

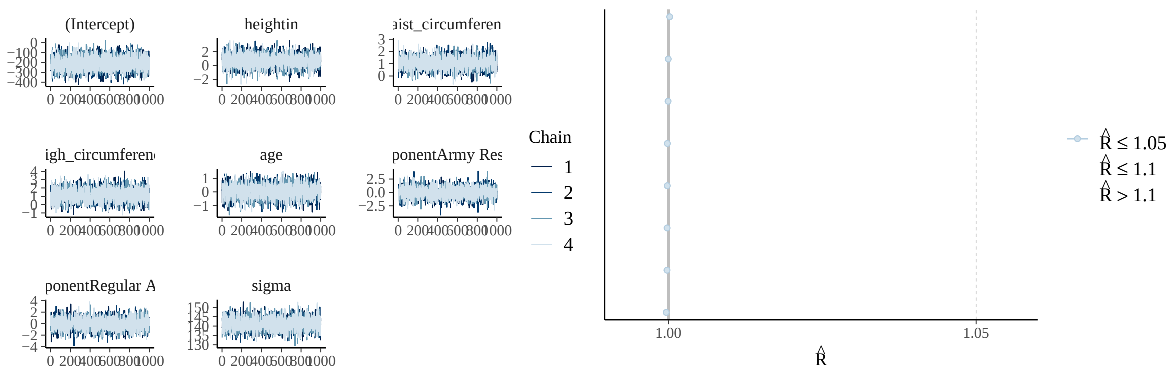
(b) Comparing the posterior with the prior

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

A.2 Diagnostics

Figure 6a shows a trace plot, which evaluates the convergence of the MCMC algorithm by examining the sampled values for each parameter across iterations. Each parameter's chain appears to mix well, with no clear trends or patterns over the iterations, indicating stable convergence. The chains overlap consistently, suggesting that the sampling algorithm has explored the posterior distribution effectively. This behavior implies that the Bayesian model's estimates are reliable and not overly sensitive to initial values or random sampling variation.

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



(a) Trace plot

(b) Rhat plot

Figure 6: Checking the convergence of the MCMC algorithm

References

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

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