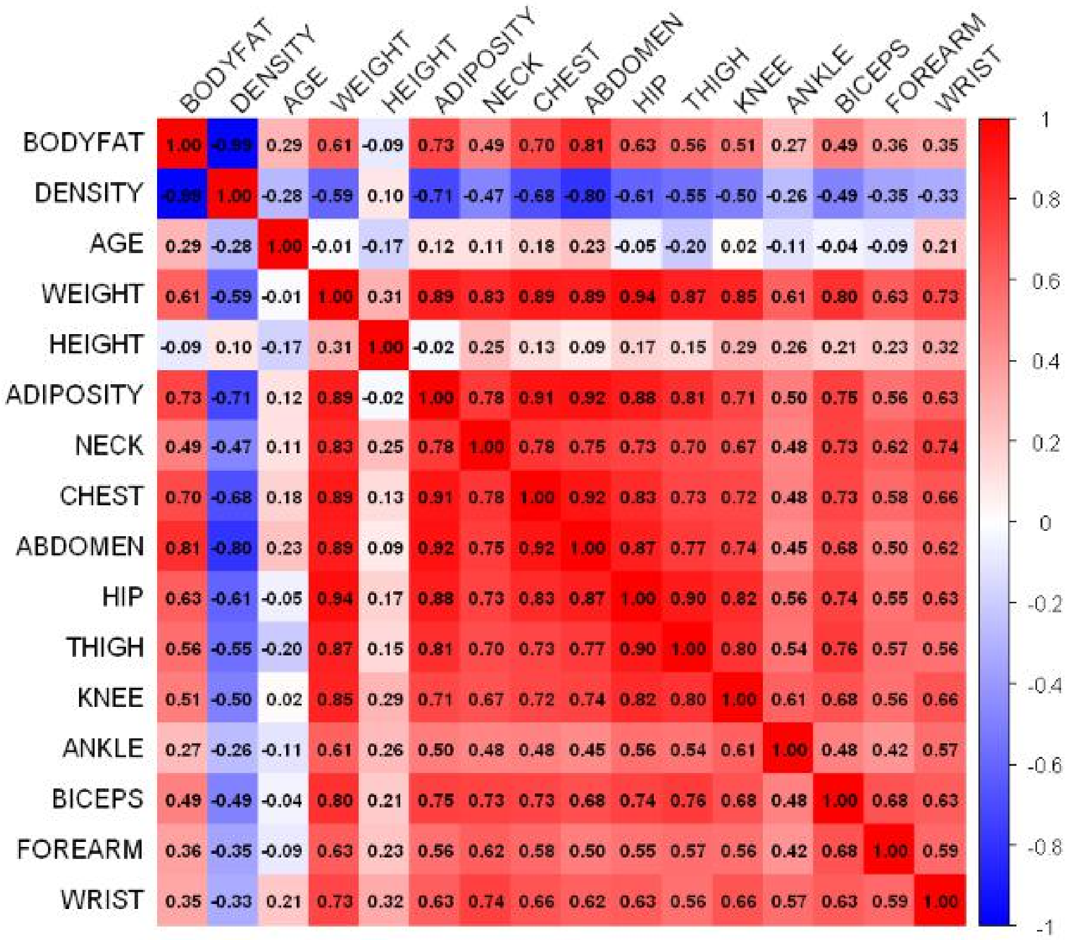
Body Fat Measurement Model

Group 4

Data Cleaning

- First, We check the corr. and lof scores.



	BODYFAT	DENSITY	AGE	WEIGHT	HEIGHT	ADIPOSITY	NECK	CHEST	ABDOMEN	HIP	THIGH	KNEE	ANKLE	BICEPS	FOREARM	WRIST
BODYFAT	1.00	-0.99	0.29	0.61	-0.09	0.73	0.49	0.70	0.81	0.63	0.56	0.51	0.27	0.49	0.36	0.35
DENSITY	-0.99	1.00	-0.28	-0.59	0.10	-0.71	-0.47	-0.68	-0.80	-0.61	-0.55	-0.50	-0.26	-0.49	-0.35	-0.33
AGE	0.29	-0.28	1.00	-0.01	-0.17	0.12	0.11	0.18	0.23	-0.05	-0.20	0.02	-0.11	-0.04	-0.09	0.21
WEIGHT	0.61	-0.59	-0.01	1.00	0.31	0.89	0.83	0.89	0.89	0.94	0.87	0.85	0.61	0.80	0.63	0.73
HEIGHT	-0.09	0.10	-0.17	0.31	1.00	-0.02	0.25	0.13	0.09	0.17	0.15	0.29	0.26	0.21	0.23	0.32
ADIPOSITY	0.73	-0.71	0.12	0.89	-0.02	1.00	0.78	0.91	0.92	0.88	0.81	0.71	0.50	0.75	0.56	0.63
NECK	0.49	-0.47	0.11	0.83	0.25	0.78	1.00	0.78	0.75	0.73	0.70	0.67	0.48	0.73	0.62	0.74
CHEST	0.70	-0.68	0.18	0.89	0.13	0.91	0.78	1.00	0.92	0.83	0.73	0.72	0.48	0.73	0.58	0.66
ABDOMEN	0.81	-0.80	0.23	0.89	0.09	0.92	0.75	0.92	1.00	0.87	0.77	0.74	0.45	0.68	0.50	0.62
HIP	0.63	-0.61	-0.05	0.94	0.17	0.88	0.73	0.83	0.87	1.00	0.90	0.82	0.56	0.74	0.55	0.63
THIGH	0.56	-0.55	-0.20	0.87	0.15	0.81	0.70	0.73	0.77	0.90	1.00	0.80	0.54	0.76	0.57	0.56
KNEE	0.51	-0.50	0.02	0.85	0.29	0.71	0.67	0.72	0.74	0.82	0.80	1.00	0.61	0.68	0.56	0.66
ANKLE	0.27	-0.26	-0.11	0.61	0.26	0.50	0.48	0.48	0.45	0.56	0.54	0.61	1.00	0.48	0.42	0.57
BICEPS	0.49	-0.49	-0.04	0.80	0.21	0.75	0.73	0.73	0.68	0.74	0.76	0.68	0.48	1.00	0.68	0.63
FOREARM	0.36	-0.35	-0.09	0.63	0.23	0.56	0.62	0.58	0.50	0.55	0.57	0.56	0.42	0.68	1.00	0.59
WRIST	0.35	-0.33	0.21	0.73	0.32	0.63	0.74	0.66	0.62	0.63	0.56	0.66	0.57	0.63	0.59	1.00

Individual (IDNO)	LOF Scores(>2)
39	5.616425
42	3.042846

- We impute some individuals.

Individual (IDNO)	Abnormal column	Original Obs.	Imputed Obs.	Reasons	Imputation Method
31	ANKLE	33.9(cm)	23.8(cm)	ANKLE's histogram distribution anomaly, Leverage and Cook's distance check	KNN
86	ANKLE	33.7(cm)	22.5(cm)	ANKLE's histogram distribution anomaly, Leverage and Cook's distance check	KNN
221	WEIGHT	153.25(lbs)	173.20(inches)	Mismatch between BMI, HEIGHT, and WEIGHT	BMI and HEIGHT calculate
42	HEIGHT	29.5(inches)	69.43(inches)	Mismatch between BMI, HEIGHT, and WEIGHT	BMI and WEIGHT calculate

Individual (IDNO)	Abnormal column	Original Obs.	Imputed Obs.	Reasons	Imputation Method

172	BODYFAT
	DENSITY
		1.9(%)
		1.0983(gm/cm3)
			6.99(%)
			1.083(gm/cm3)
				BODYFAT and DENSITY calculated using ‘Siri’s equation’ mismatch, Abnormal BODYFAT	Regression by other columns

182	BODYFAT
	DENSITY
		0(%)
		1.1089(gm/cm3)
			4.53(%)
			1.089(gm/cm3)
				BODYFAT and DENSITY calculated using ‘Siri’s equation’ mismatch, Abnormal BODYFAT	Regression by other columns

48	BODYFAT	6.4(%)	14.14(%)	BODYFAT and DENSITY calculated using ‘Siri’s equation’ mismatch	DENSITY calculate
76	BODYFAT	18.3(%)	14.09(%)	BODYFAT and DENSITY calculated using ‘Siri’s equation’ mismatch	DENSITY calculate
96	DENSITY	1.0991(gm/cm3)	1.059(gm/cm3)	BODYFAT and DENSITY calculated using ‘Siri’s equation’ mismatch	BODYFAT calculate

- We deleted **five individuals** (39,54,163,175,216.) due to Leverage and Cook’s distance check.

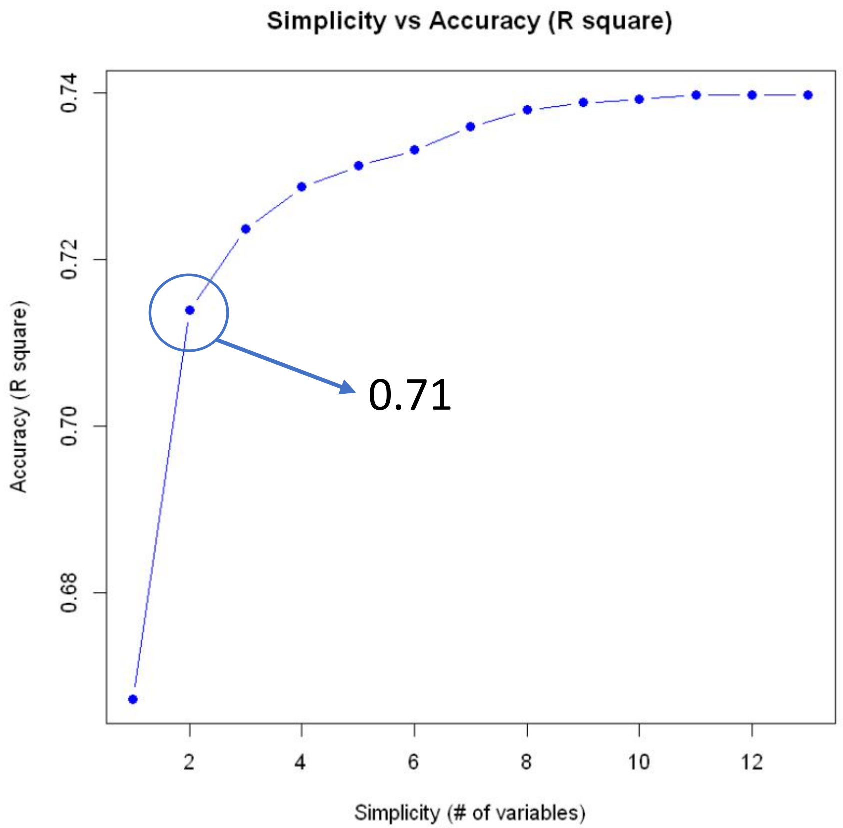
Individual (IDNO)	Leverage	Cook’s distance
39	0.466	0.409
54	0.172	0.020
163	0.167	0.018
175	0.273	0.025
216	0.134	0.030

Final Cleaned Data: **n=247** (from n=252) with p = 13 predictors

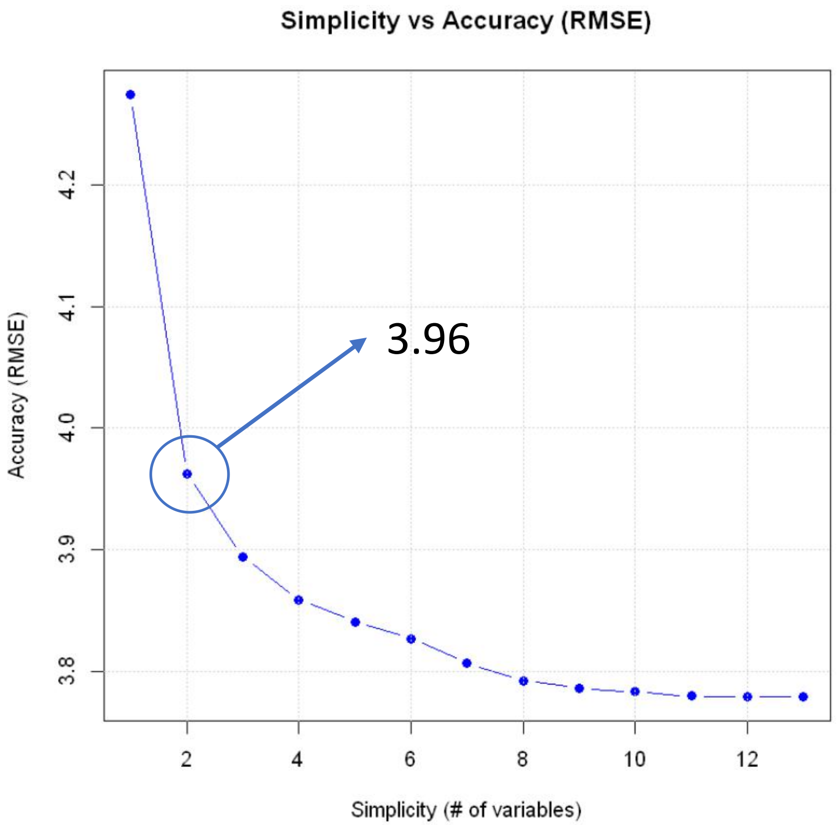
- Remove IDNO
- Remove DENSITY
- Remove BMI – highly collinear
- Predictors: Age (years),Weight (lbs),Height (inches), Adioposity (bmi), Neck circumference (cm), Chest circumference (cm), Abdomen 2 circumference (cm), Hip circumference (cm), Thigh circumference (cm), Knee circumference (cm), Ankle circumference (cm), Biceps (extended) circumference (cm), Forearm circumference (cm), Wrist circumference (cm)

Final Model

BODYFAT = - 41.547 - 0.124*WEIGHT (lbs) + 0.894*ABDOMEN (cm)


Simplicity vs Accuracy (R square)

0.71


Simplicity vs Accuracy (RMSE)

3.96

Model Diagnostics

Consider the assumptions of linear regression:

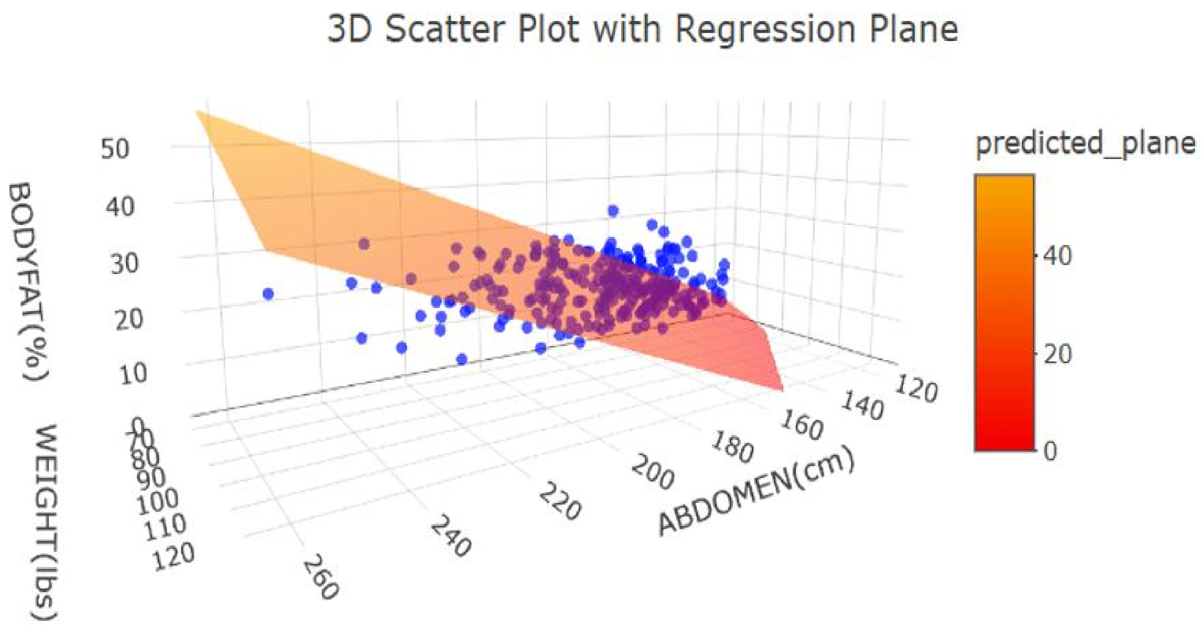
Assumption	Test	criteria
Normality of residuals	Shapiro-Wilk test / Q-Q plot	$p\text{-value} = 0.240 > 0.05$
Independence of residuals	Durbin-Watson test	$p\text{-value} = 0.134 > 0.05$
Homoscedasticity	Breusch-Pagan test / residual plot	$p\text{-value} = 0.496 > 0.05$
Multicollinearity	Variance Inflation Factor (VIF)	$\text{VIF} = 4.27 < 10$

Statistical Analysis

Predictors	P-value
WEIGHT	1.27e-09
ABDOMEN	<2e-16
Overall Model(under F-test)	<2.2e-16

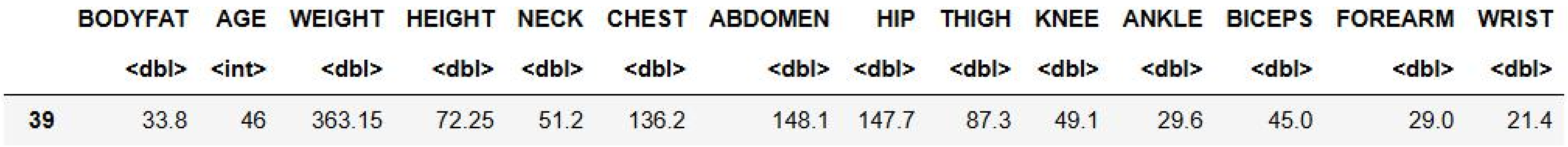
Residual Standard Error	3.987

R^2	0.714

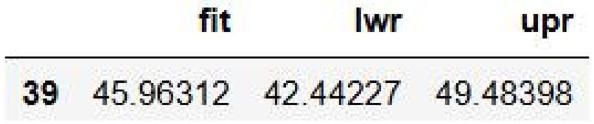


Robustness and Sensitivity

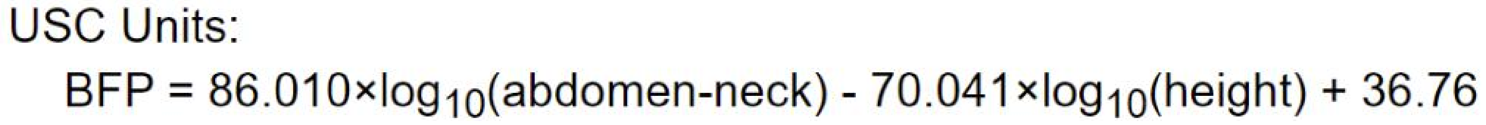
- **Robustness**



	BODYFAT	AGE	WEIGHT	HEIGHT	NECK	CHEST	ABDOMEN	HIP	THIGH	KNEE	ANKLE	BICEPS	FOREARM	WRIST
	<dbl>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
39	33.8	46	363.15	72.25	51.2	136.2	148.1	147.7	87.3	49.1	29.6	45.0	29.0	21.4

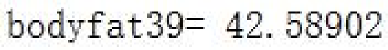


	fit	lwr	upr
39	45.96312	42.44227	49.48398



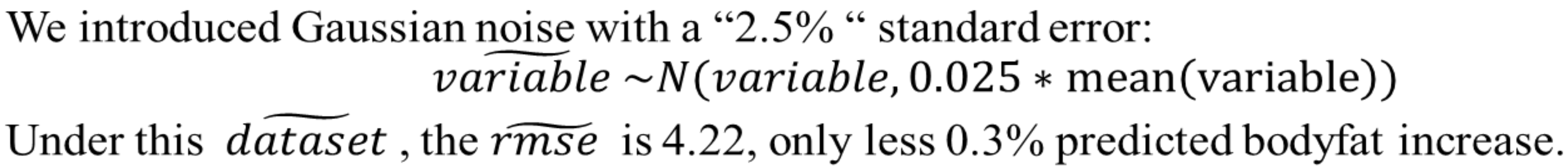
USC Units:

BFP = 86.010×$\log_{10}$(abdomen-neck) - 70.041×$\log_{10}$(height) + 36.76



```
bodyfat39= 42.58902
```

- **Sensitivity**



We introduced Gaussian noise with a "2.5% " standard error:

$$\widetilde{variable} \sim N(variable, 0.025 * \text{mean(variable)})$$

Under this $\widetilde{dataset}$, the $\widetilde{rmse}$ is 4.22, only less 0.3% predicted bodyfat increase.

Strengths and Weaknesses

Final Model:

Body Fat ~ -41.547 - 0.124* WEIGHT(lbs) + 0.894*ABDOMEN(cm)

- **Strengths**
  - A linear model with only two variables
  - Explains ABDOMEN of variation in body fat
  - Satisfies all the fundamental assumptions of linear regression

- **Weaknesses**
  - Less intuitively when interprete the coef. of WEIGHT

Conclusion

Fight with mess data.

1. Correlation
2. Histogram
3. LOF
4. Relationship between data
5. KNN
6. Setting range
7. Regression
8. Outlier detection

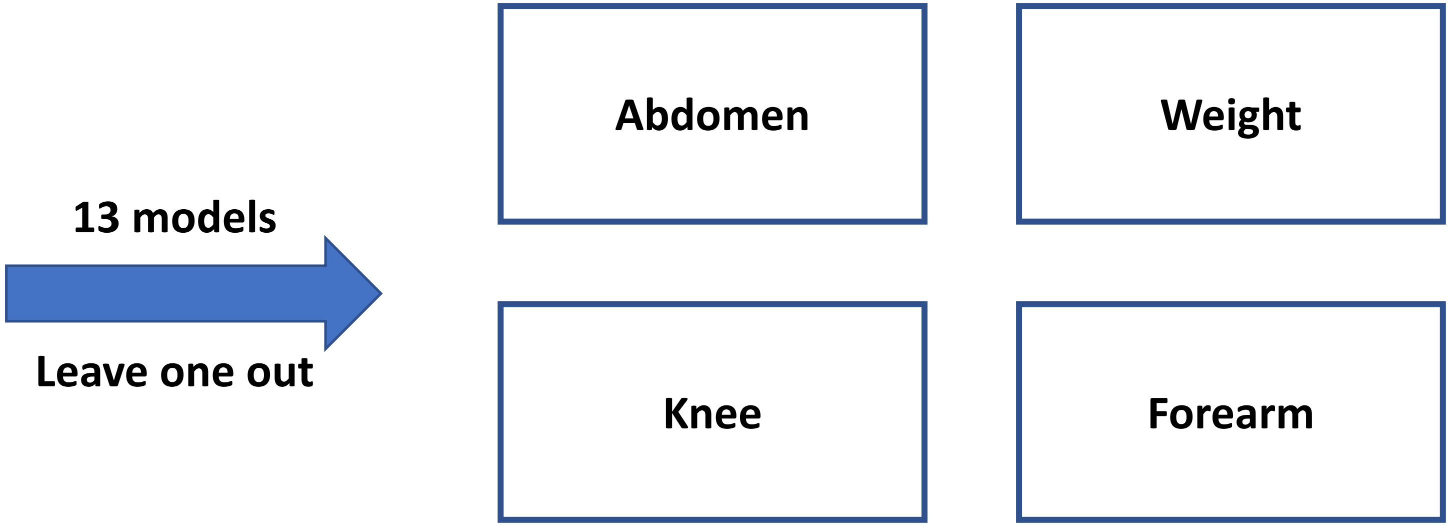
Checked one by one

Is this data correct?
How it can be wrong?
Why is value is so strange?
Which one is wrong among these 3?
How to impute it?
Why can't we delete it?
......

Conclusion

Compared RMSE in 8 different models

1. Forward model
2. Stepwise selection
3. Best subset selection
4. LASSO
5. Ridge
6. Elastic Net
7. Random forest
8. XGBoost



13 models

Leave one out

Abdomen

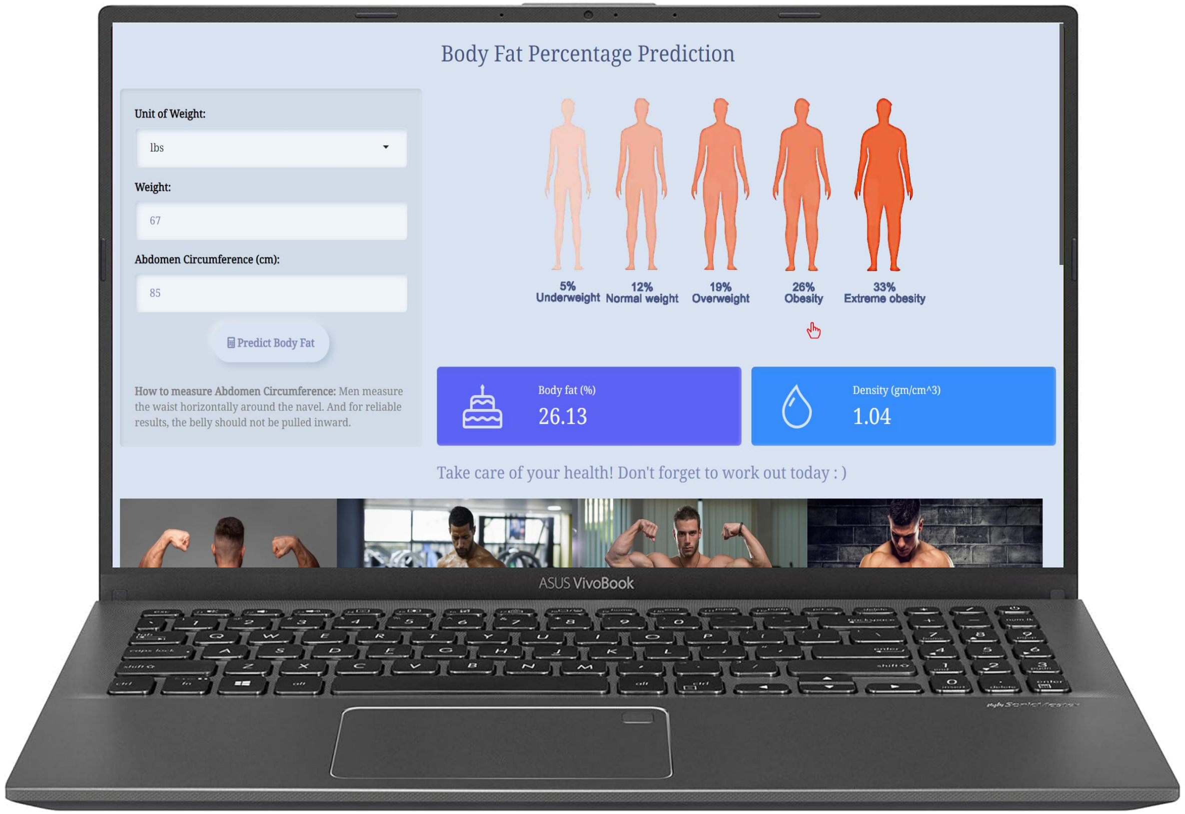
Weight

Knee

Forearm

Conclusion

Measure your body fat at any time (with our model) !



Body Fat Percentage Prediction

Unit of Weight: lbs

Weight: 67

Abdomen Circumference (cm): 85

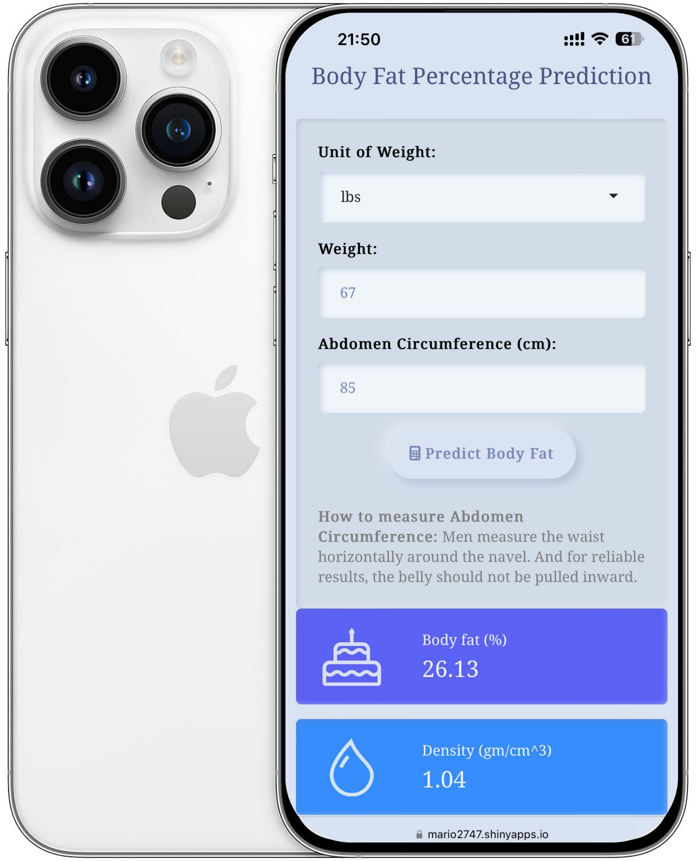
Predict Body Fat

How to measure Abdomen Circumference: Men measure the waist horizontally around the navel. And for reliable results, the belly should not be pulled inward.

Body fat (%) 26.13

Density (gm/cm^3) 1.04

Take care of your health! Don't forget to work out today :)



Body Fat Percentage Prediction

Unit of Weight: lbs

Weight: 67

Abdomen Circumference (cm): 85

Predict Body Fat

How to measure Abdomen Circumference: Men measure the waist horizontally around the navel. And for reliable results, the belly should not be pulled inward.

Body fat (%) 26.13

Density (gm/cm^3) 1.04

Body Fat Measurement Model

Group 4