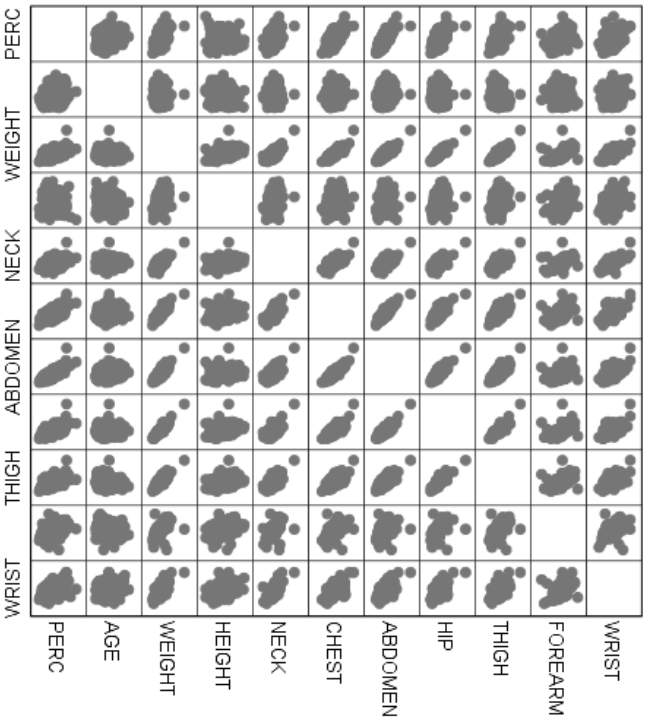
Jingyu Xiang
1537572

STATS 252

LABS5

1.
(a)The purpose of the study is to establish a generalized body fat percentage prediction equation by using simple measurement techniques. The response variable is Percent body fat. There are 10 explanatory variables. To generalize your conclusion to the population of interest, the relationship between the variables as described by the population regression equation must be approximately linear. the standard deviations of y values must be approximately the same, the y values must be normally distributed, the observations of the response variable are independent of one another, and no serious outliers in the regression model.
(b)Abdomen 2 circumference (cm) is better predictor of the body fat percentage for men. The accuracy of those measurements matters. To reduce measurements error, we can do a single measurement several times and take the average.

2.
(a)



(b)
Overall, a linear model is appear to be appropriate for describing the all relationships; For PERC and CHEST, PERC and ABDOMEN and PERC and HIP, the relationship seems to be strong. And the strongest relationship is between PERC and ABDOMEN. Their are outliers in some plots.

3.
(a)Correlations

		PERC	AGE	WEIGHT	HEIGHT	NECK	CHEST	ABDOMEN	HIP	THIGH	FOREARM	WRIST
PERC	Pearson Correlation	1	.291**	.612**	-.025	.491**	.703**	.804**	.625**	.560**	.361**	.347**
	Sig. (2-tailed)		.000	.000	.690	.000	.000	.000	.000	.000	.000	.000
	N	252	252	252	252	252	252	252	252	252	252	252
AGE	Pearson Correlation	.291**	1	-.013	-.245**	.114	.176**	.226**	-.050	-.200**	-.085	.214**
	Sig. (2-tailed)	.000		.840	.000	.072	.005	.000	.426	.001	.178	.001

	N	252	252	252	252	252	252	252	252	252	252	252
WEIGHT	Pearson Correlation	.612**	-.013	1	.487**	.831**	.894**	.893**	.941**	.869**	.630**	.730**
	Sig. (2-tailed)	.000	.840		.000	.000	.000	.000	.000	.000	.000	.000
	N	252	252	252	252	252	252	252	252	252	252	252
HEIGHT	Pearson Correlation	-.025	-.245**	.487**	1	.321**	.227**	.189**	.372**	.339**	.322**	.398**
	Sig. (2-tailed)	.690	.000	.000		.000	.000	.003	.000	.000	.000	.000
	N	252	252	252	252	252	252	252	252	252	252	252
NECK	Pearson Correlation	.491**	.114	.831**	.321**	1	.785**	.759**	.735**	.696**	.624**	.745**
	Sig. (2-tailed)	.000	.072	.000	.000		.000	.000	.000	.000	.000	.000
	N	252	252	252	252	252	252	252	252	252	252	252
CHEST	Pearson Correlation	.703**	.176**	.894**	.227**	.785**	1	.913**	.829**	.730**	.580**	.660**
	Sig. (2-tailed)	.000	.005	.000	.000	.000		.000	.000	.000	.000	.000
	N	252	252	252	252	252	252	252	252	252	252	252
ABDOMEN	Pearson Correlation	.804**	.226**	.893**	.189**	.759**	.913**	1	.881**	.770**	.494**	.620**
	Sig. (2-tailed)	.000	.000	.000	.003	.000	.000		.000	.000	.000	.000
	N	252	252	252	252	252	252	252	252	252	252	252
HIP	Pearson Correlation	.625**	-.050	.941**	.372**	.735**	.829**	.881**	1	.896**	.545**	.630**
	Sig. (2-tailed)	.000	.426	.000	.000	.000	.000	.000		.000	.000	.000
	N	252	252	252	252	252	252	252	252	252	252	252
THIGH	Pearson Correlation	.560**	-.200**	.869**	.339**	.696**	.730**	.770**	.896**	1	.567**	.559**
	Sig. (2-tailed)	.000	.001	.000	.000	.000	.000	.000	.000		.000	.000
	N	252	252	252	252	252	252	252	252	252	252	252
FOREARM	Pearson Correlation	.361**	-.085	.630**	.322**	.624**	.580**	.494**	.545**	.567**	1	.586**
	Sig. (2-tailed)	.000	.178	.000	.000	.000	.000	.000	.000	.000		.000
	N	252	252	252	252	252	252	252	252	252	252	252
WRIST	Pearson Correlation	.347**	.214**	.730**	.398**	.745**	.660**	.620**	.630**	.559**	.586**	1

Sig. (2-tailed)	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000
N	252	252	252	252	252	252	252	252	252	252	252

** Correlation is significant at the 0.01 level (2-tailed).

All 10 explanatory variables except HEIGHT have at least weak relationship with the PERC since all of their correlation coifficients are greater than 0.3. The relationship between PERC and AGE, WEIGHT, NECK, CHEAST, ABDOMEN, HIP, THIGH, FOREARM, WRIST are linear since their p-value <0.05. Positive relationship is shown in all pairs except PERC and HEIGHT(negative r).

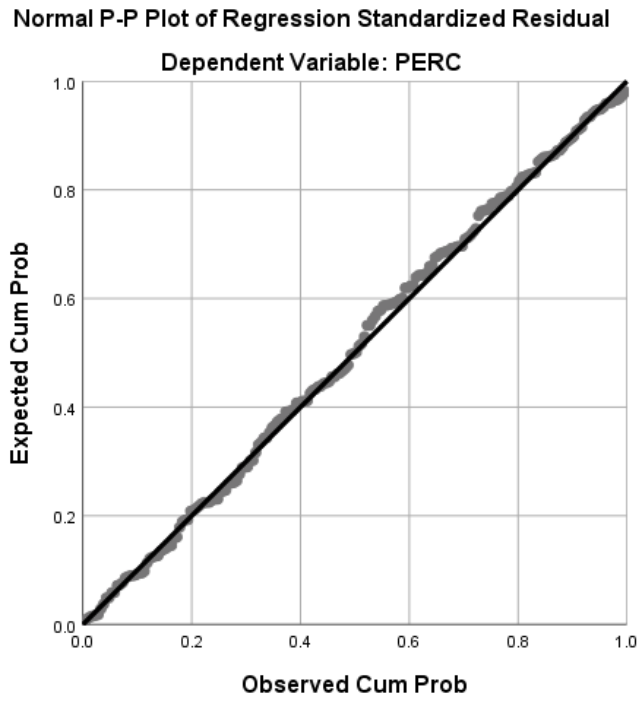
(b)
Use ABDOMEN, because it has greatest r.
Multicollinearity can be a problem in this case because some explanatory variables can be predicted by others according to the matrix.

4.
PERC = B0+B1(AGE)+B2(WEIGHT)+B3(HEIGHT)+B4(NECK)+B5(CHEST)+B6(ABDOMEN)+B7(HIP)+B8(THIGH)+B9(FOREARM)+B10(WRIST)

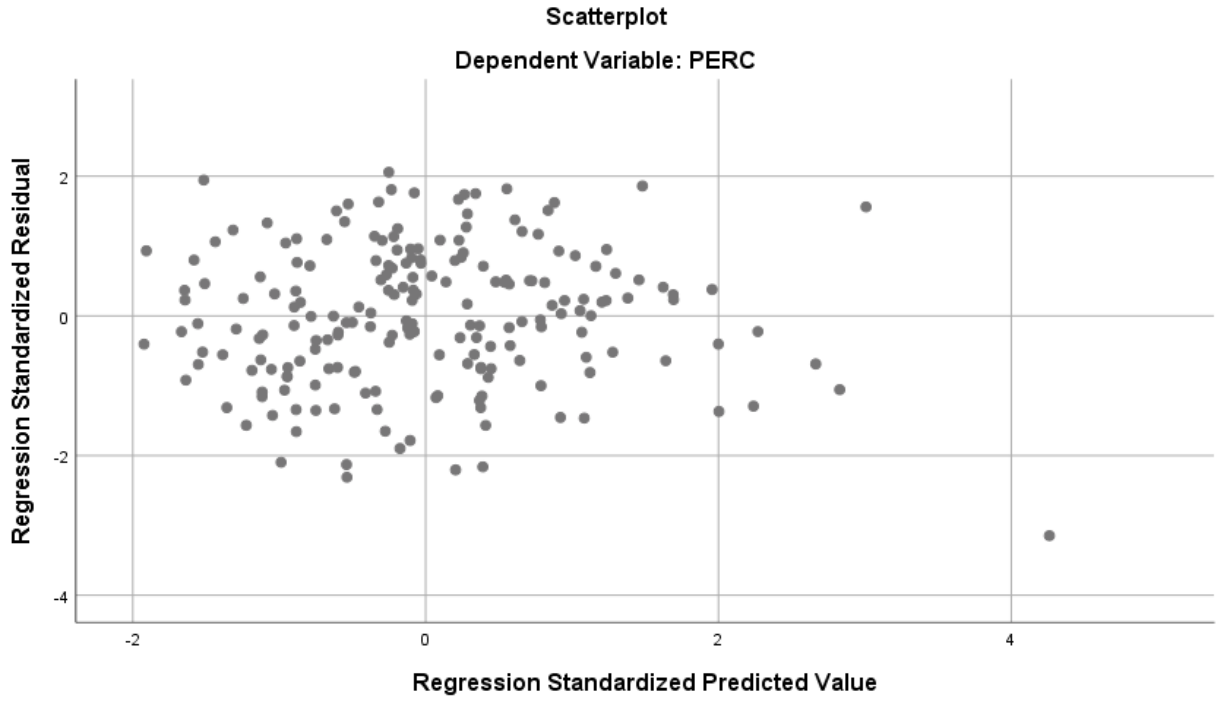
Assumptions:1. The relationship between the variables should be approximately linear.
2.equal standard deviations of the y values(PERC).
3.populations of explanatory variables should be normal.
4.no serious outliers
5. Independent observations of responsible variables.

5.
(a)
4 are included, the order they are added is ABDOMEN, WEIGHT, NECK, FOREARM.
(b)
PERC = 0.968(ABDOMEN)-0.130(WEIGHT)-0.726(NECK)+0.529(FOREARM)-34.757
(c)
percentage of the variation in PERC: 0.715
(d)
H0: the fitted model as a whole
Ha: the fitted model not as a whole
F=122.460
P-val = 0.000 < 0.05 reject H0.
So the fitted model not as a whole.
(e)
ABDOMEN and WEIGHT contributes most.

6.
(a)



There is no evidence that the assumption of normality is violated
(b)



The residuals are decentralized, outliers exist. Assumption of equal variance is not violated.
(c)
The 20th observation is an influential case.
studentized residuals=-3.7340
Cook's distance=1.13760
and leverages=0.28472

7.
(a)
PERC=0.924(ABDOMEN)-0.084(WEIGHT)-1.522(WRIST)-23.525
R^2=0.734
(b)
PERC = 0.924*86.4-0.084*184.75-1.522*18.2-23.525=13.0892
(c)
95% confidence interval for the mean percentage of body fat of the individuals:
(11.94839, 14.27628)
95% prediction interval for an individual with the body circumference measurements and body weight:
(4.57179, 21.65289)
(d)
H0: evidence of an association between the body fat percentage and the lower body measurements (THIGH and HIP) after accounting for age, weight, height, neck, chest, abdomen, forearm and wrist?
Ha: No evidence of an association between the body fat percentage and the lower body measurements (THIGH and HIP) after accounting for age, weight, height, neck, chest, abdomen, forearm and wrist?

8.
H0: There is no mean difference between actual and predicted percent fat.
Ha: There is mean difference between actual and predicted percent fat.
P-val = 0.817.
P-val >0.05 Don't not reject H0.
There is not sufficient evidence that there is mean difference between actual and predicted percent fat.

95%CI =(-1.37857, 1.09277)
We are 95% confident that the mean difference between actual and predicted percent fat is between -1.37857 and 1.09277.