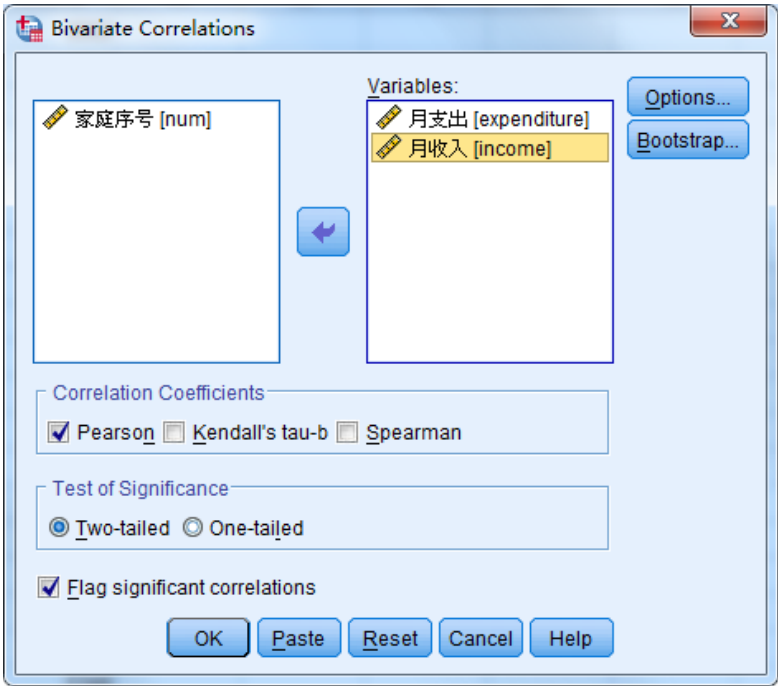


第九章 相关与回归分析

第一节 相关系数

一、使用 SPSS 软件计算相关系数

- 1.打开数据集：data9_01 40 户家庭每月日用杂货支出与月收入.sav
- 2.点击 Analyze——Correlate——Bivariate，选择变量 expenditure、income 进入 Variables，在 Correlation Coefficient 中选择 Pearson，点击 OK



结果如下：

Correlations			
		月支出	月收入
月支出	Pearson Correlation	1	.945**
	Sig. (2-tailed)		.000
	N	40	40
月收入	Pearson Correlation	.945**	1
	Sig. (2-tailed)	.000	
	N	40	40

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

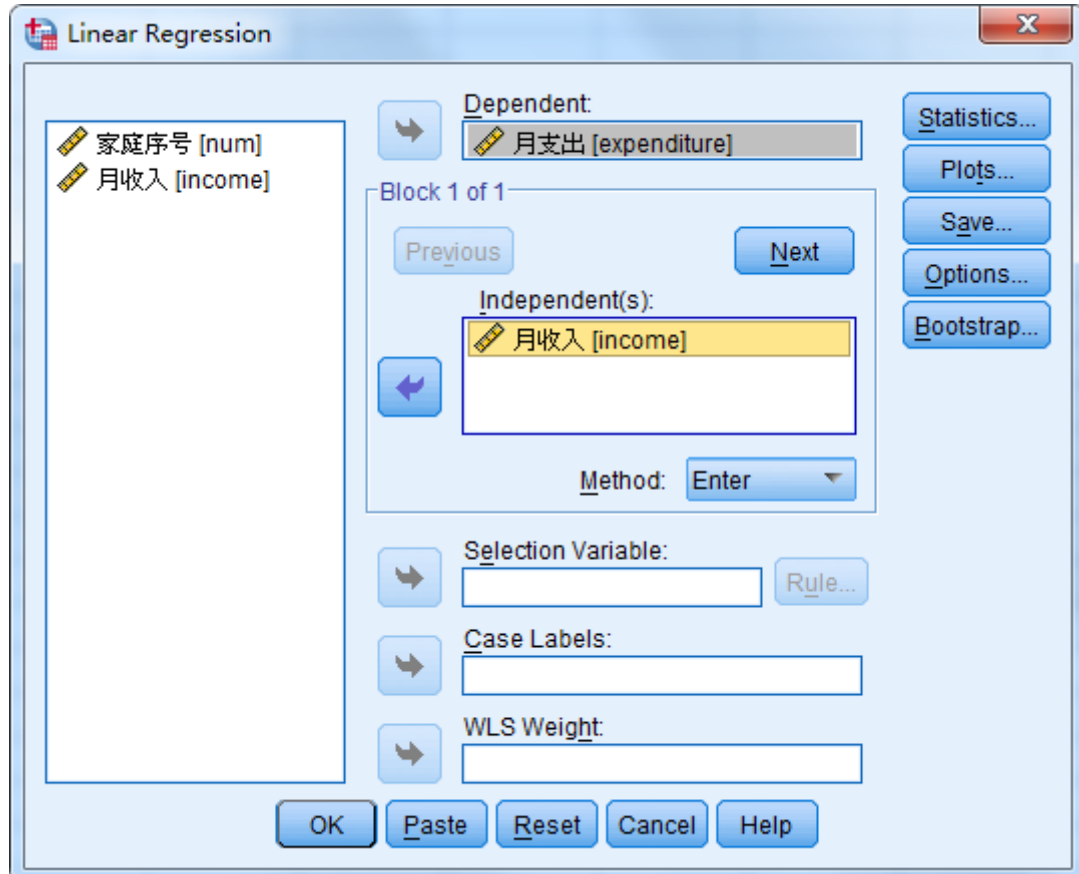
结果显示，40 户家庭日用杂货支出与月收入的相关系数达 0.945，且 P 值小于 0.001，可以认为家庭每月的日用杂货支出与月收入存在高度正相关关系。

第八节 基于 SPSS 实现回归分析过程

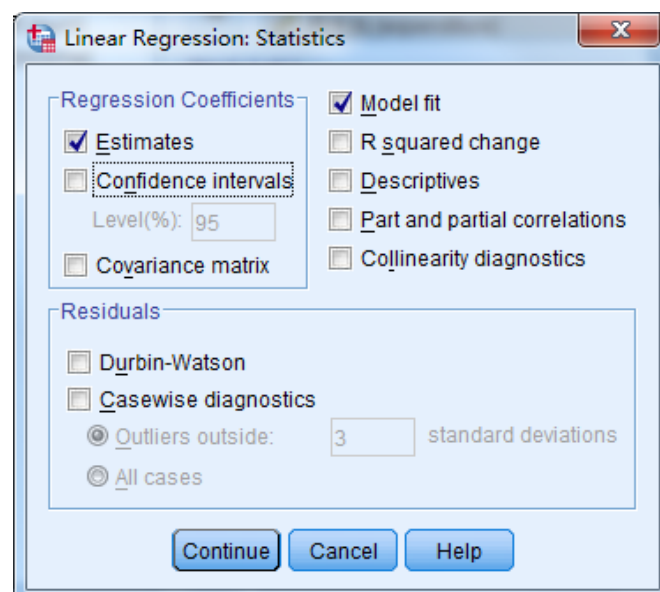
一、SPSS 拟合回归模型并进行显著性检验和识别异常值

1. 打开数据集：data9_01 40 户家庭每月日用杂货支出与月收入.sav

2. 点击 Analyze(分析)——Regression(回归)——Linear(线性)，选择变量 expenditure 进入 Dependent，选择 income 进入 independent



3. 点击 Statistics 后分别选中 Estimator 和 Model fit 用于计算回归模型并对其进行显著性检验



点击 Continue，点击 OK(确定)。

结果显示如下：

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	月收入 ^b	.	Enter

a. Dependent Variable: 月支出

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.945 ^a	.893	.890	82.328

a. Predictors: (Constant), 月收入

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2139306.141	1	2139306.141	315.626	.000 ^b
	Residual	257563.234	38	6777.980		
	Total	2396869.375	39			

a. Dependent Variable: 月支出

b. Predictors: (Constant), 月收入

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-74.367	53.122		-1.400	.170
	月收入	.134	.008	.945	17.766	.000

a. Dependent Variable: 月支出

说明回归模型为 $\hat{y}_i = -74.367 + 0.134x_i$, 判定系数 R Square 为 0.893

当进行样本异常值检验时，选择 Casewise diagnostics，并在 Outliers outside 的参数框中键入 2，点击 OK

结果显示如下：

Casewise Diagnostics^a

Case Number	Std. Residual	月支出	Predicted Value	Residual
11	-2.189	541	721.25	-180.252
12	2.282	1083	895.16	187.844

a. Dependent Variable: 月支出

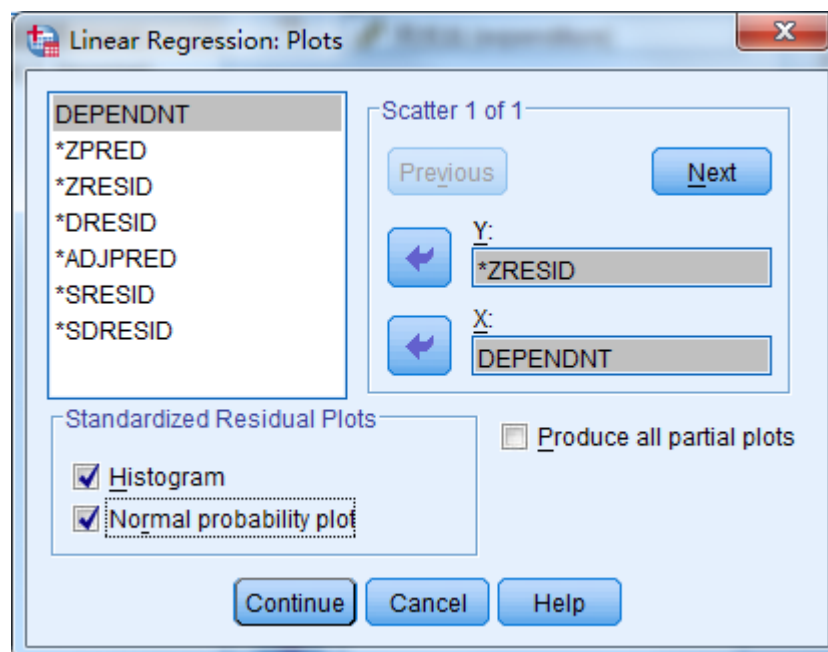
Residuals Statistics^a

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	513.08	1248.72	840.62	234.209	40
Residual	-180.252	187.844	.000	81.266	40
Std. Predicted Value	-1.399	1.742	.000	1.000	40
Std. Residual	-2.189	2.282	.000	.987	40

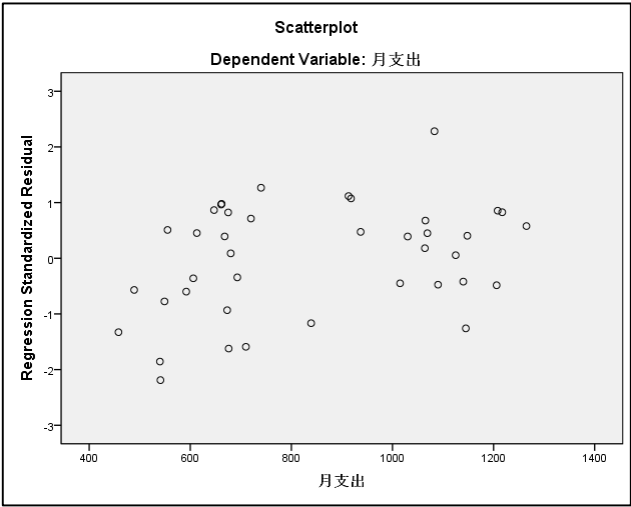
a. Dependent Variable: 月支出

说明第 11 和第 12 个样本点被检测为异常值。

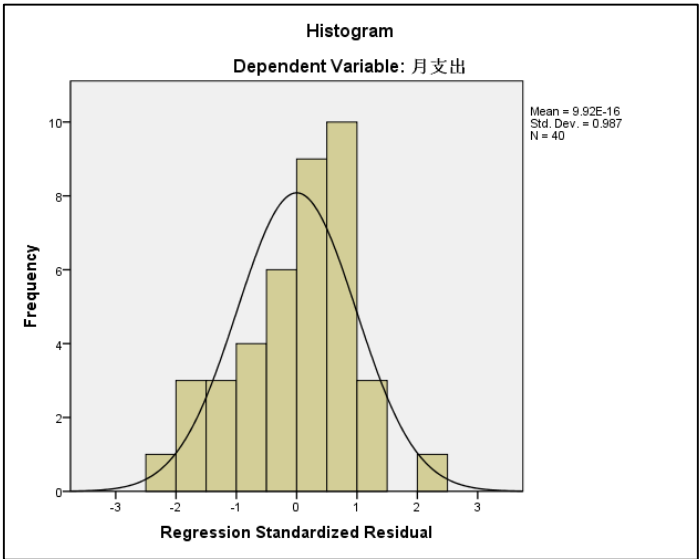
绘制残差图，点击 Plot，选择 *ZRESID 进入 y 轴，选择 DEPENDNT 进入 X 轴，并选择 Histogram 和 Normal probability plot，点击 Continue 和 OK



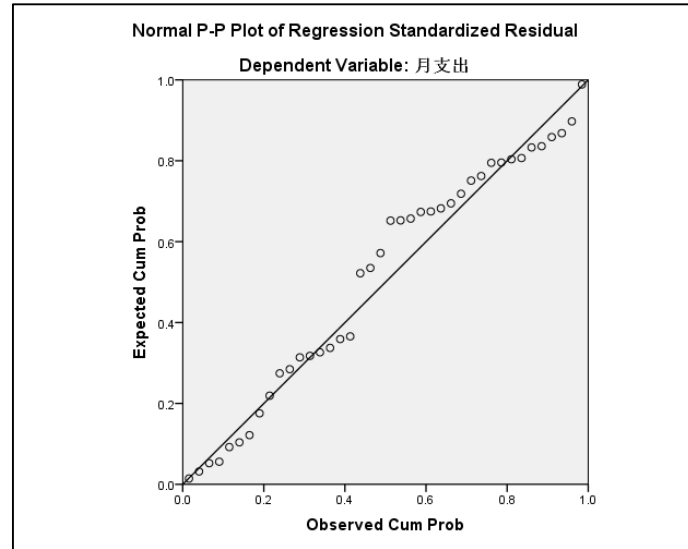
结果显示如下：
得到标准版残差图



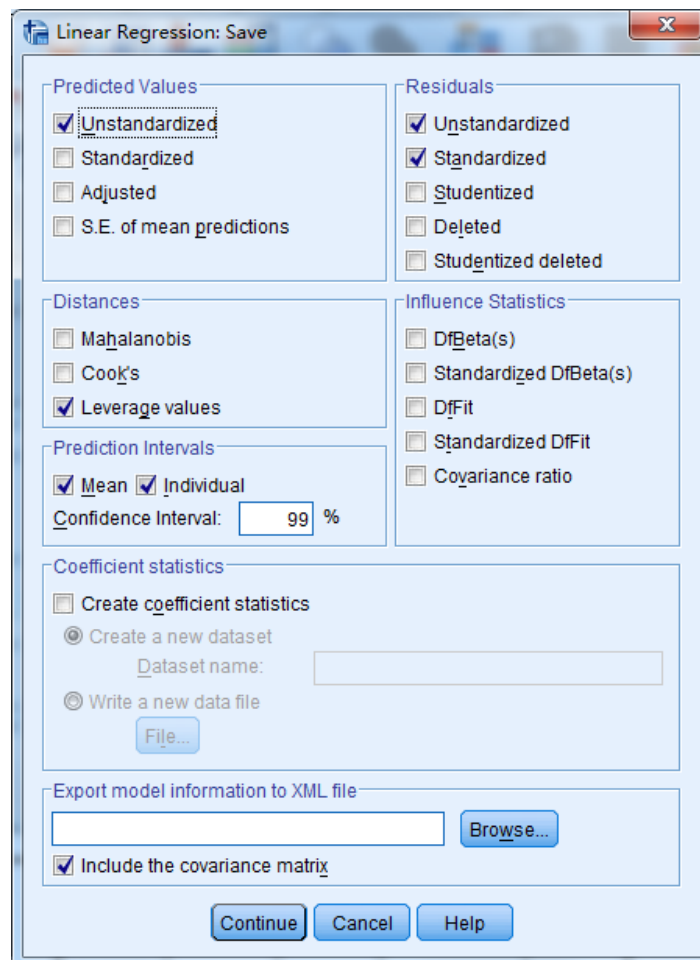
标准化残差的直方图：



和正态概率图：



4.根据回归方程进行预测，在Linear Regression中点击save



5.在Predicted Values中选择Unstandardized，用于输出为标准化的预测值

(PRE_1) ；在Prediction Intervals中选择Mean和Individual，并设置99%的置信

度，用于实现对总体均值和个别值的区间估计（均值下限LMCI_1、均值上限UMCI_1、单值下限LICI_1、单值上限UICI_1），在Residual中选择Unstandardized和Standardized，用于得到残差或标准化残差的计算结果（残差RES_1、标准化残差ZRE_1）；在Distances中选择Leverage values，用于得到杠杆率（LEV_1），Continue，OK。

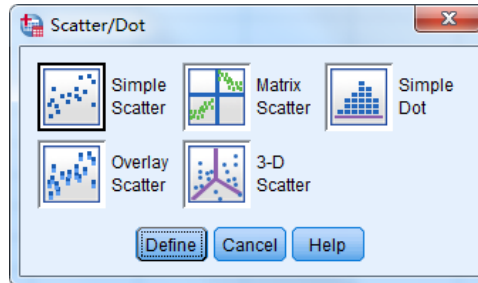
结果如下：

xpenditure	income	PRE_1	RES_1	ZRE_1	LEV_1	LMCI_1	UMCI_1	LICI_1	UICI_1
1148	8882	1114.71046	33.28954	.40435	.03512	1059.97582	1169.44510	884.85990	1344.56102
489	4558	535.83533	-46.83533	-.56888	.04342	477.44076	594.22990	305.08585	766.58482
1208	9053	1137.60307	70.39693	.85507	.04123	1080.15378	1195.05236	907.09099	1368.11515
1065	8094	1009.21703	55.78297	.67757	.01329	965.53627	1052.89778	781.74525	1236.68880
1015	8414	1052.05700	-37.05700	-.45011	.02090	1004.23171	1099.88229	823.75313	1280.36087
1125	8925	1120.46708	4.53292	.05506	.03661	1065.05803	1175.87613	890.45499	1350.47917
1206	9862	1245.90787	-39.90787	-.48474	.07678	1174.68844	1317.12730	1011.58415	1480.23155
613	4856	575.73006	37.26994	.45270	.03280	522.05988	629.40024	346.13065	805.32946
661	4899	581.48668	79.51332	.96581	.03139	528.47522	634.49813	352.04036	810.93300
606	5304	635.70602	-29.70602	-.36082	.01963	588.54577	682.86626	407.54053	863.87150

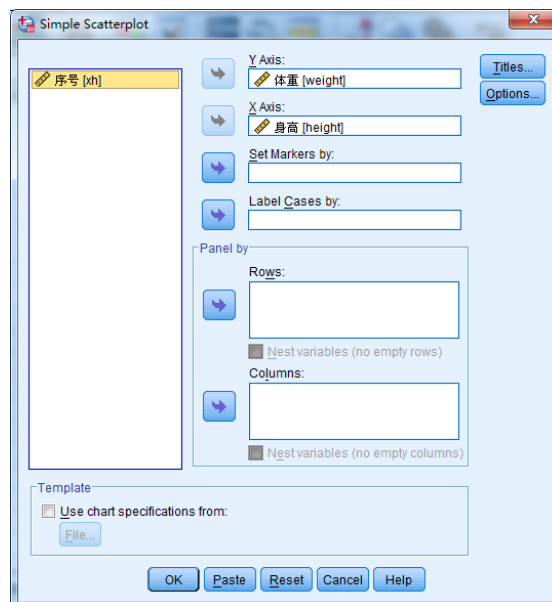
习题——计算题

一、

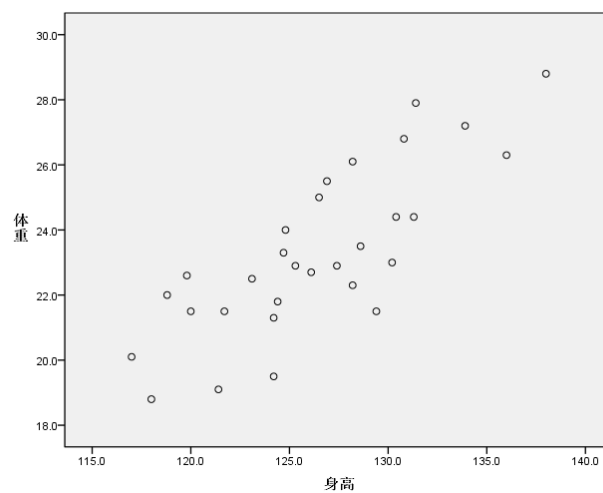
(1)、1.绘制散点图：点击Graphs——Scatter，选择Simple Scatter，点击Define



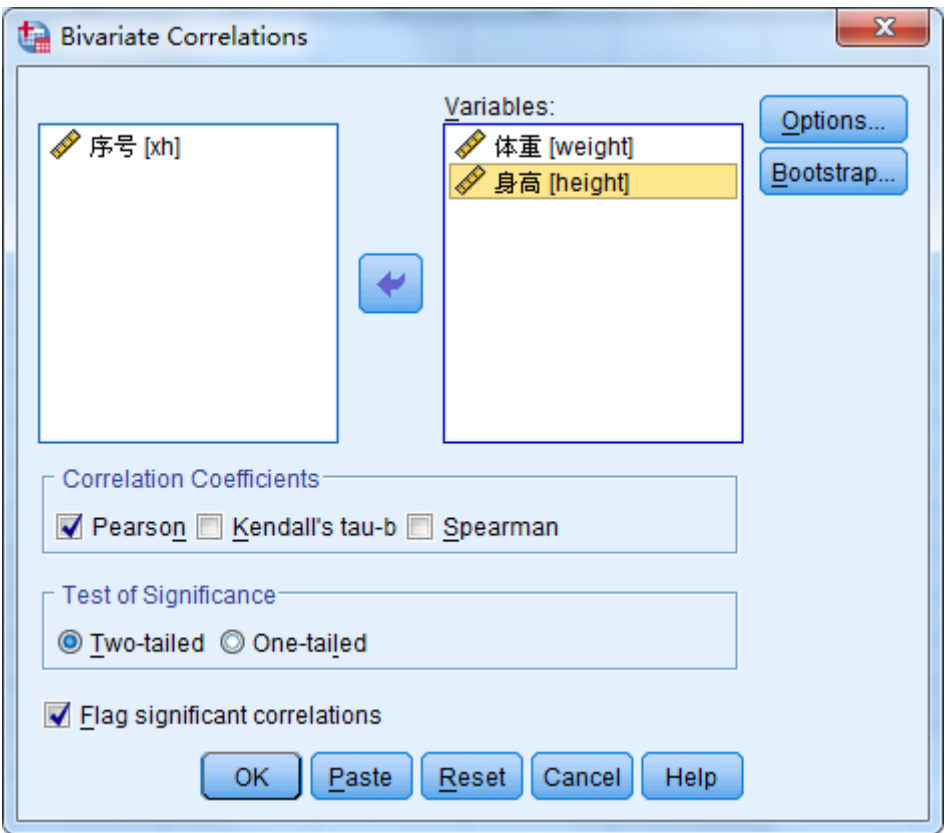
2.将weight选入Y Axis， height选入 X Axis. OK



结果如下：



(2)、相关系数：Analyze——Correlate，将weight和height选入Variables，在Correlation Coefficient中选择Pearson，OK



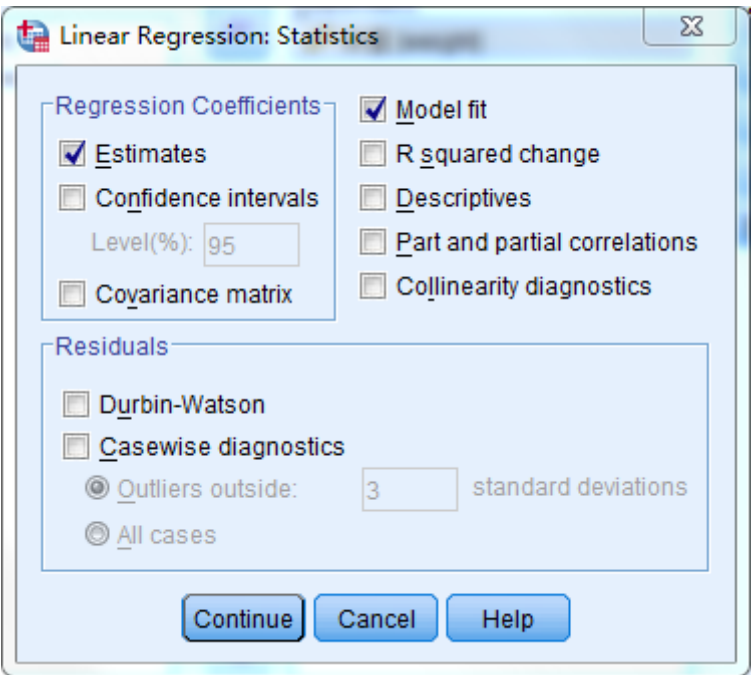
结果如下：

Correlations			
		体重	身高
体重	Pearson Correlation	1	.803**
	Sig. (2-tailed)		.000
	N	30	30
身高	Pearson Correlation	.803**	1
	Sig. (2-tailed)	.000	
	N	30	30
**. Correlation is significant at the 0.01 level (2-tailed).			

由结果可知，两者相关系数为0.803，相对应的P值小于0.001，认为身高和体重存在高度正相关关系。

(3)、1点击Analyze(分析)——Regression(回归)——Linear(线性)，选择变量weight进入Dependent，选择height进入independent，点击Statistics，在Regression

Coefficients中选中Estimates，并选中Model fit，Continue，OK。



结果如下：

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.803 ^a	.645	.632	1.5505

a. Predictors: (Constant), 身高

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	122.208	1	122.208	50.836	.000 ^b
	Residual	67.311	28	2.404		
	Total	189.519	29			

a. Dependent Variable: 体重

b. Predictors: (Constant), 身高

Double-click to activate

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-26.615	7.007		-3.798	.001
	身高	.395	.055	.803	7.130	.000

a. Dependent Variable: 体重

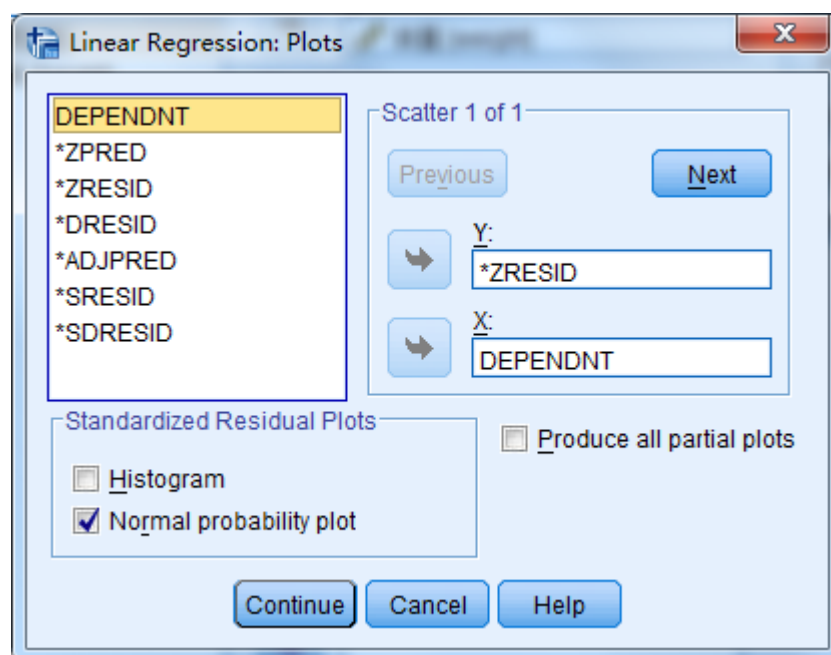
由结果可知，回归方程为 $\text{weight} = 0.395 \times \text{height} - 26.615$ ，斜率的含义身高每增加或减少一个单位，平均体重增加或减少0.395个单位。

(4)、同样由上图答案可知，F统计量的值为50.836，相应的P值小于0.001，所以认为在0.05的显著性水平下，身高与体重存在上述的线性关系。拟合有度评价指标R方等于0.645，可以说明样本回归方程可以解释总离差平方和中64.5%的变异性。

(5)、略

(6)、略

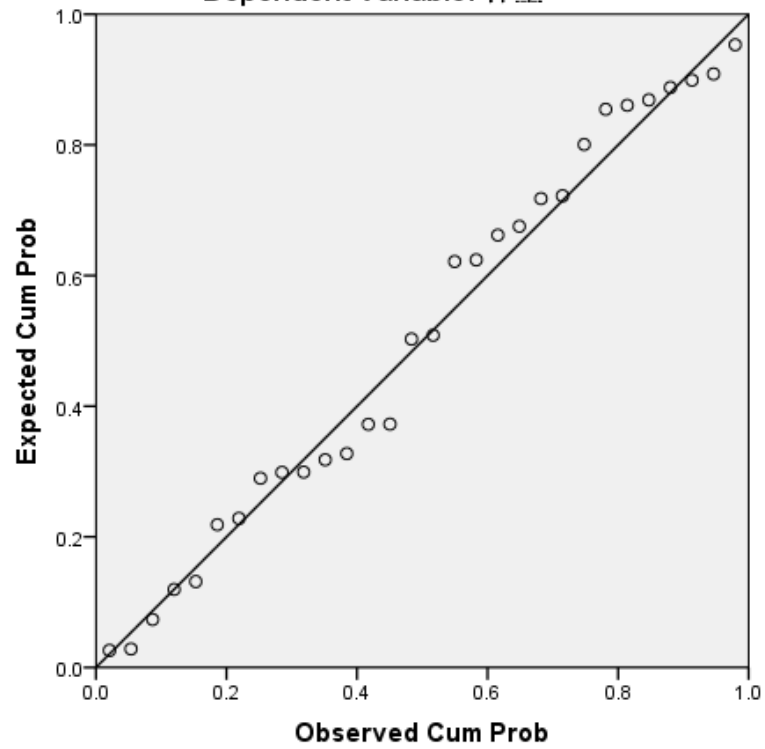
(7)、标准化残差图：点击Analyze(分析)——Regression(回归)——Linear(线性)，选择变量weight进入Dependent, 选择height进入independent, 点击Plots, 将ZERSID选入Y，DEPENDNT选入X，在Standardized Residual Plots中选择Normal Probability plot, Continue, OK



结果如下

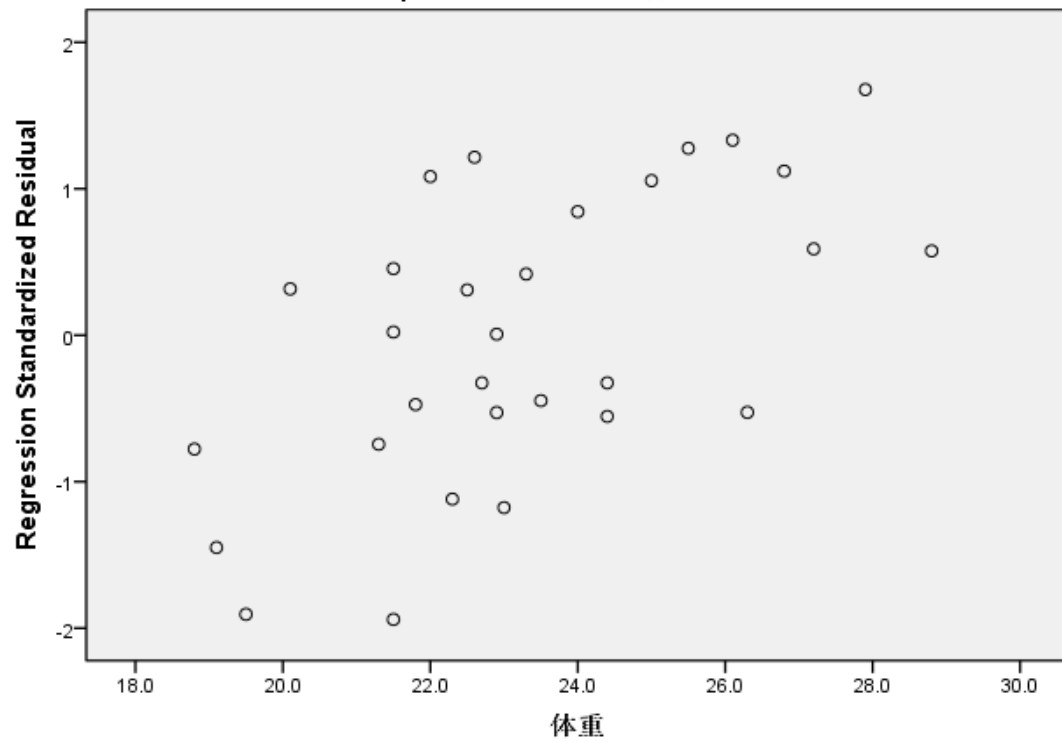
Normal P-P Plot of Regression Standardized Residual

Dependent Variable: 体重



Scatterplot

Dependent Variable: 体重



根据残差图可以看出，几乎所有的点都落在-2和+2之间，从正态概率图可以看出，没有发现明显违背正态性假定的证据。

二、 略

三、 略