



统计学原理(Statistic)

胡华平

西北农林科技大学

经济管理学院数量经济教研室

huhuaping01@hotmail.com

2021-05-16

西北农林科技大学

第五章 相关和回归分析

5.1 变量间关系的度量

5.2 回归分析的基本思想

5.3 OLS方法与参数估计

5.4 假设检验

5.5 拟合优度与残差分析

5.6 回归预测分析

5.7 回归报告解读

5.7 回归报告解读

方程表达式

表格表达式

统计软件



回归分析的形式

课程要求：会熟练、正确阅读统计软件给出的各类分析报告，理解其中的关键信息和内涵。这些分析报告包括：传统的多元回归分析报告；以及各种计量检验的辅助分析报告（如异方差white检验报告）等。

根据统计软件的不同（`stata`；`Eview`；`R`；`Excel` ……），各种分析报告呈现形式略有差异，但基本要素和信息都大抵一致。

给定如下一元回归模型：

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$



回归分析的形式（多行方程表达法）

形式1：多行方程表达法（整理好的精炼报告）：根据统计软件的原始报告，选取最关键的信息，经过整理并以多行样本回归方程（SRF）的形式呈现。例如：

$$\begin{aligned} \hat{Y} &= -0.01 && + 0.72X \\ (t) & (-0.0165) && (10.4065) \\ (se) & (0.8746) && (0.0696) \\ (\text{fitness}) & R^2 = 0.9078; \bar{R}^2 = 0.8994 \\ & F^* = 108.29; p = 0.0000 \end{aligned}$$

- 第1行表示样本回归函数（回归系数）
- 第2行(t)表示回归系数对应的样本t统计量（ $t_{\hat{\beta}_i}^*, i \in 1, 2, \dots, k$ ）
- 第3行(se)表示回归系数对应的样本标准误差（ $S_{\hat{\beta}_i}, i \in 1, 2, \dots, k$ ）



回归分析的形式（多行方程表达法）

形式1：多行方程表达法（整理好的精炼报告）：根据统计软件的原始报告，选取最关键的信息，经过整理并以多行样本回归方程（SRF）的形式呈现。例如：

$$\begin{aligned} \hat{Y} &= -0.01 && + 0.72X \\ (t) & (-0.0165) && (10.4065) \\ (se) & (0.8746) && (0.0696) \\ (\text{fitness}) & R^2 = 0.9078; \bar{R}^2 = 0.8994 \\ & F^* = 108.29; p = 0.0000 \end{aligned}$$

- 第4行(fitness)表示回归模型拟合情况和统计检验的简要信息，其中 R^2 表示判定系数， \bar{R}^2 表示调整判定系数，F表示模型整体显著性检验中的样本F统计量值（ F^* ），p表示样本F统计量值对应的概率值。



回归分析的形式（表格列示法）

形式2：表格列示法（整理好的精炼报告）：根据统计软件的原始报告，往往是选取最关键的信息，经过整理以表格形式呈现，表格列示法的形式呈现为：

表格列示法

term	estimate	std. error	statistic	p. value
(Intercept)	-0.01	0.87	-0.02	0.99
X	0.72	0.07	10.41	0.00

- 第1列：term表示回归模型中包含的变量，也即 $X_{2i}, X_{3i}, \dots, X_{ki}$ ，其中截距项默认为(Intercept)。
- 第2列：estimate表示回归系数的估计值，也即 $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ 。
- 第3列：std.error表示回归系数对应的样本标准误差，也即 $S_{\hat{\beta}_i}, i \in 1, 2, \dots, k$ 。
- 第4列：statistic表示回归系数对应的样本t统计量，也即 $t_{\hat{\beta}_i}^*, i \in 1, 2, \dots, k$ 。
- 第5列：p.value表示回归系数样本t统计量对应的概率值，也即 $Pr(t = t_{\hat{\beta}_i}^*) = p$



(示例) Excel软件原始报告：全貌

形式3：原始报告：分析软件如EViews、R、STATA、Excel等直接自动生成的多元回归分析报告。Excel软件原始分析报告形式如下：

	B	C	D	E	F	G	H	I	J	K	L	M
1	X	Y										
2	6	4.4567	SUMMARY OUTPUT									
3	7	5.77										
4	8	5.9787		回归统计								
5	9	7.3317	Multiple R		0.9528							
6	10	7.3182	R Square		0.9078							
7	11	6.5844	Adjusted R Square		0.8994							
8	12	7.8182	标准误差		0.9387							
9	13	7.8351	观测值		13							
10	14	11.0223										
11	15	10.6738	方差分析									
12	16	10.8361			df	SS	MS	F	Significance F			
13	17	13.615	回归分析		1	95.43	95.43	108.29	0.00			
14	18	13.531	残差		11	9.69	0.88					
15			总计		12	105.12						
16												
17					Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
18			Intercept		-0.01	0.8746	-0.02	0.9871	-1.94	1.91	-1.94	1.91
19			X		0.72	0.0696	10.41	0.0000	0.57	0.88	0.57	0.88

X Residual Plot

The plot displays residuals on the y-axis (ranging from -2 to 2) against the independent variable X on the x-axis (ranging from 0 to 20). The data points show a clear upward trend, suggesting a positive linear relationship between X and the residuals.



(示例) Excel软件原始报告：参数估计

	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.0%
Intercept	-0.01	0.8746	-0.02	0.9871	-1.94	1.91	-1.94	1.91
X	0.72	0.0696	10.41	0.0000	0.57	0.88	0.57	0.88

Excel回归分析的参数估计结果



(示例) Excel软件原始报告 : 拟合优度

回归统计	
Multiple R	0.9528
R Square	0.9078
Adjusted R Square	0.8994
标准误差	0.9387
观测值	13

Excel回归分析的拟合优度



(示例) Excel软件原始报告：方差分解

方差分析	df	SS	MS	F	Significance F
回归分析	1	95.43	95.43	108.29	0.00
残差	11	9.69	0.88		
总计	12	105.12			

Excel回归分析的ANOVA表结果



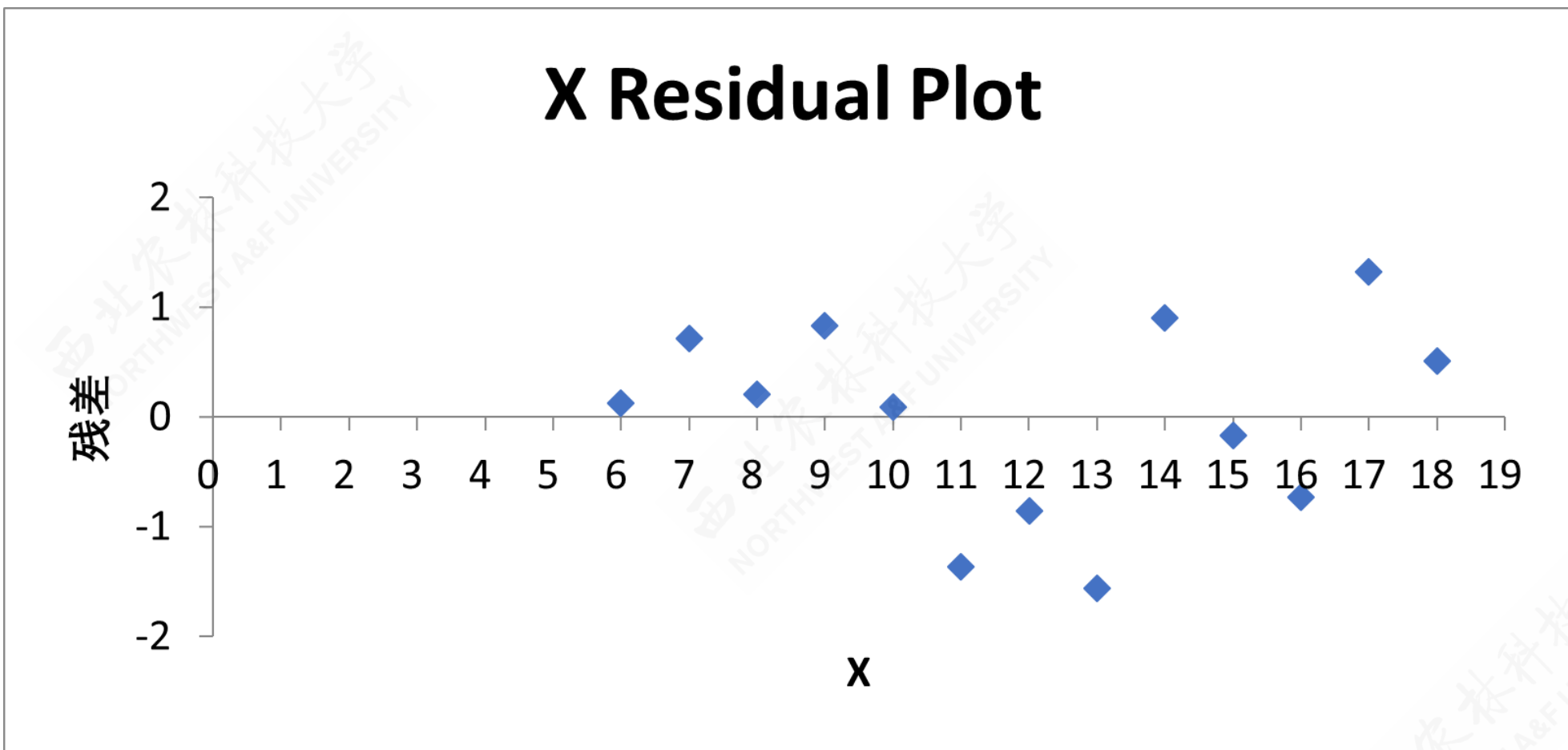
(示例) Excel软件原始报告 : 残差表

RESIDUAL OUTPUT			
观测值	预测 Y	残差	标准残差
1	4.330	0.127	0.141
2	5.054	0.716	0.796
3	5.778	0.200	0.223
4	6.502	0.829	0.923
5	7.227	0.092	0.102
6	7.951	-1.366	-1.520
7	8.675	-0.857	-0.953
8	9.399	-1.564	-1.740
9	10.123	0.899	1.001
10	10.847	-0.173	-0.193
11	11.571	-0.735	-0.818
12	12.295	1.320	1.469
13	13.019	0.512	0.569

Excel回归分析的预测和残差结果



(示例) Excel软件原始报告 : 残差图



Excel回归分析的残差图



(示例) Excel软件操作步骤：配置数据分析模块

西北农林科技大学
NORTHWEST A&F UNIVERSITY

西北农林科技大学
NORTHWEST A&F UNIVERSITY

西北农林科技大学
NORTHWEST A&F UNIVERSITY



(示例) Excel软件操作步骤：回归分析的操作步骤

The screenshot displays the Microsoft Excel interface with the '数据' (Data) ribbon selected. A '回归' (Regression) dialog box is open, and the '数据分析' (Data Analysis) task pane is also visible. The spreadsheet contains the following data:

	A	B	C
1	obs	X	Y
2	1	6	4.4567
3	2	7	5.77
4	3	8	5.9787
5	4	9	7.3317
6	5	10	7.3182
7	6	11	6.5844
8	7	12	7.8182
9	8	13	7.8351
10	9	14	11.0223
11	10	15	10.6738
12	11	16	10.8361
13	12	17	13.615
14	13	18	13.531
15			

Numbered callouts in the image indicate the following steps:

- 1: Clicking the '数据' (Data) ribbon.
- 2: Clicking the '数据分析' (Data Analysis) button.
- 3: Selecting '回归' (Regression) in the '数据分析' task pane.
- 4: Clicking the '确定' (OK) button in the '数据分析' task pane.
- 5: Clicking the selection icon for the Y input range.
- 6: Clicking the selection icon for the X input range.
- 7: Checking the '标志' (Labels) checkbox.
- 8: Selecting '新工作簿' (New Workbook) under '输出选项' (Output Options).
- 9: Checking the '残差图' (Residuals Plot) checkbox.
- 10: Clicking the '确定' (OK) button in the '回归' dialog box.

Excel“回归分析”的操作步骤



回归分析的形式 (EViews软件原始报告)

形式3: 原始报告: 分析软件如EViews、R、STATA等直接自动生成的多元回归分析报告。EViews软件原始分析报告形式如下: 抬头区域

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.014453	0.874624	-0.016525	0.9871
X	0.724097	0.069581	10.40648	0.0000

R-squared	0.907791	Mean dependent var	8.674708
Adjusted R-squared	0.899409	S.D. dependent var	2.959706
S.E. of regression	0.938704	Akaike info criterion	2.852004
Sum squared resid	9.692810	Schwarz criterion	2.938920
Log likelihood	-16.53803	Hannan-Quinn criter.	2.834139
F-statistic	108.2948	Durbin-Watson stat	1.737984
Prob(F-statistic)	0.000000		

- **Dependent Variable: Y:** 因变量
- **Method: Least Squares:** 分析方法
- **Date: 03/09/19 Time: 10:55:** 分析的时间
- **Sample: 1 13:** 样本范围
- **Included observations: 13:** 样本数n





回归分析的形式 (EViews软件原始报告)

形式3: 原始报告: 分析软件如EViews、R、STATA等直接自动生成的多元回归分析报告。EViews软件原始分析报告形式如下: 三线表区域

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.014453	0.874624	-0.016525	0.9871
X	0.724097	0.069581	10.40648	0.0000

R-squared	0.907791	Mean dependent var	8.674708
Adjusted R-squared	0.899409	S.D. dependent var	2.959706
S.E. of regression	0.938704	Akaike info criterion	2.852004
Sum squared resid	9.692810	Schwarz criterion	2.938920
Log likelihood	-16.53803	Hannan-Quinn criter.	2.834139
F-statistic	108.2948	Durbin-Watson stat	1.737984
Prob(F-statistic)	0.000000		

- 第1列: Variable表示模型包含的变量, $X_{2i}, X_{3i}, \dots, X_{ki}$, 其中截距项默认为C。
- 第2列: Coefficient回归系数, 也即 $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$;
- 第3列: Std. Error回归系数的样本标准误差, 也即也即 $S_{\hat{\beta}_i}, i \in 1, 2, \dots, k$ 。
- 第4列: t-Statistic表示回归系数对应的样本t统计量, 也即 $t_{\hat{\beta}_i}^*, i \in 1, 2, \dots, k$;



回归分析的形式 (EViews软件原始报告)

形式3: 原始报告: 分析软件如EViews、R、STATA等直接自动生成的多元回归分析报告。EViews软件原始分析报告形式如下: 指标值区域 (左)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.014453	0.874624	-0.016525	0.9871
X	0.724097	0.069581	10.40648	0.0000

R-squared	0.907791	Mean dependent var	8.674708
Adjusted R-squared	0.899409	S.D. dependent var	2.959706
S.E. of regression	0.938704	Akaike info criterion	2.852004
Sum squared resid	9.692810	Schwarz criterion	2.938920
Log likelihood	-16.53803	Hannan-Quinn criter.	2.834139
F-statistic	108.2948	Durbin-Watson stat	1.737984
Prob(F-statistic)	0.000000		

- R-squared: 回归模型判定系数 R^2 。
- Adjusted R-squared: 回归模型调整判定系数 \bar{R}^2 。
- S.E. of regression: 回归模型的回归误差标准差 $\hat{\sigma}$ 。
- Sum squared resid: 回归模型的残差平方和RSS $RSS = \sum e_i^2$ 。
- Log likelihood: 回归模型的对数似然值。



回归分析的形式 (EViews软件原始报告)

形式3: 原始报告: 分析软件如EViews、R、STATA等直接自动生成的多元回归分析报告。EViews软件原始分析报告形式如下: 指标值区域(右)

Equation: EQ_WAGE Workfile: CHPT2::wage\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: Y
Method: Least Squares
Date: 03/09/19 Time: 10:55
Sample: 1 13
Included observations: 13

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.014453	0.874624	-0.016525	0.9871
X	0.724097	0.069581	10.40648	0.0000

R-squared	0.907791	Mean dependent var	8.674708
Adjusted R-squared	0.899409	S.D. dependent var	2.959706
S.E. of regression	0.938704	Akaike info criterion	2.852004
Sum squared resid	9.692810	Schwarz criterion	2.938920
Log likelihood	-16.53803	Hannan-Quinn criter.	2.834139
F-statistic	108.2948	Durbin-Watson stat	1.737984
Prob(F-statistic)	0.000000		

- Mean dependent var: Y的均值 \bar{Y} 。
- S.D. dependent var: Y的样本标准差 S_Y 。
- Akaike info criterion: 回归模型的AIC信息准则。
- Schwarz criterion: 回归模型的Schwarz准则。
- Hannan-Quinn criter.: 回归模型的Hannan-Quinn准则。



回归分析的形式 (R软件原始报告)

形式4: 原始报告: 分析软件如EViews、R、STATA等直接自动生成的多元回归分析报告。R软件原始分析报告形式如下:

```
Call:
lm(formula = mod_wage, data = data_wage)

Residuals:
    Min       1Q   Median       3Q      Max
-1.564 -0.735  0.127  0.716  1.320

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.0145    0.8746   -0.02    0.99
X              0.7241    0.0696   10.41 0.0000005 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.94 on 11 degrees of freedom
Multiple R-squared:  0.908,    Adjusted R-squared:  0.899
F-statistic: 108 on 1 and 11 DF,  p-value: 0.000000496
```

本节结束

