



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：

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

- 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：

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

- 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

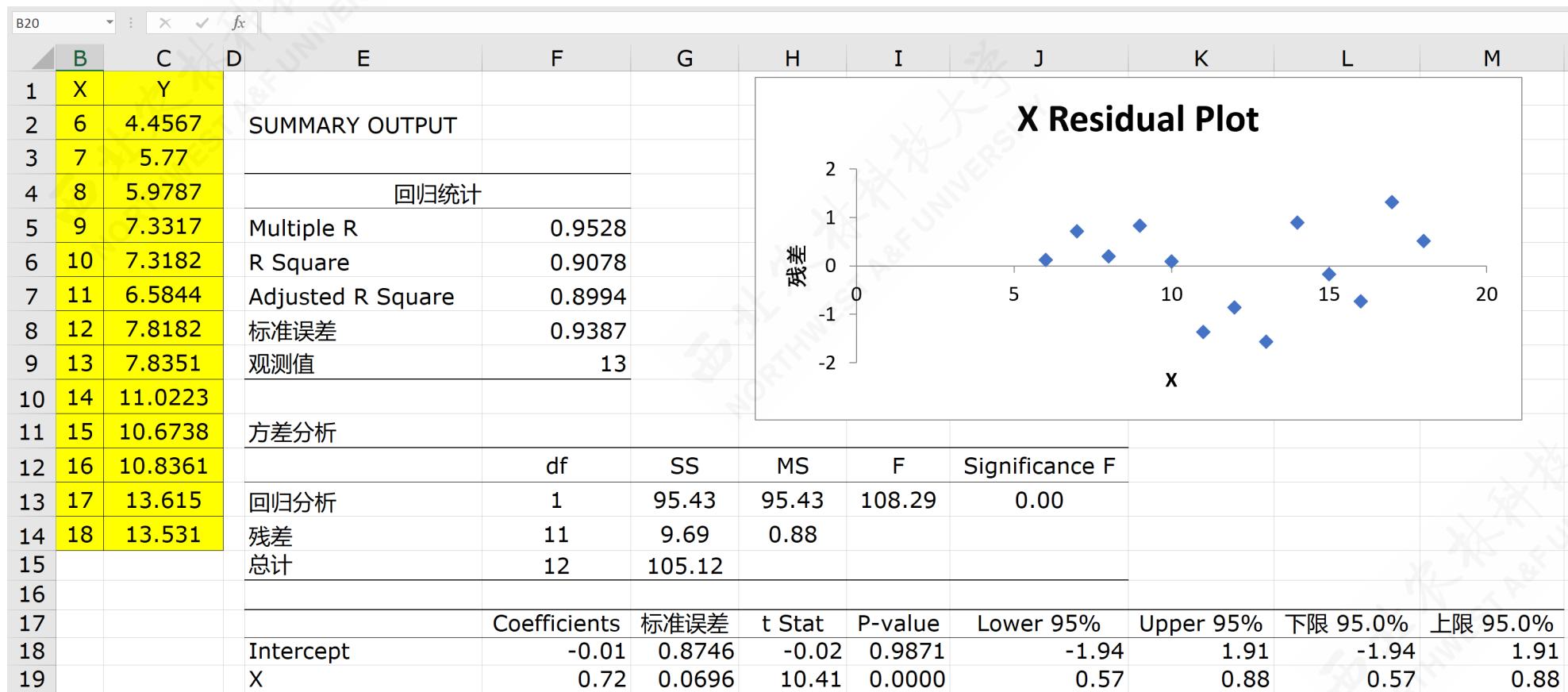
term	estimate	std. error	statistic	p. value
(Intercept)	-0.01	0.87	-0.02	0.99
X	0.72	0.07	10.41	0.00

- 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：





(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

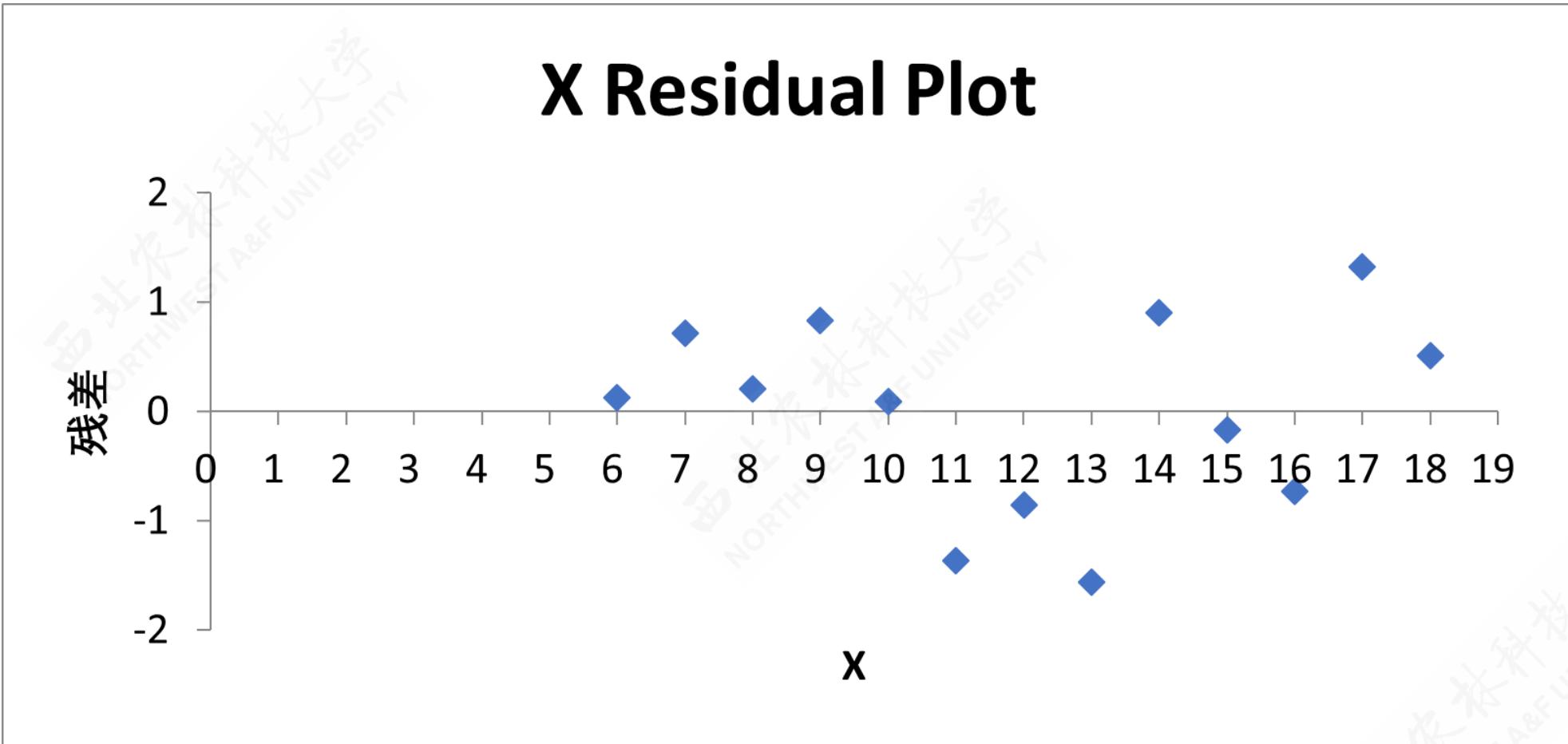
RESIDUAL OUTPUT

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



(示例) Excel软件操作步骤：回归分析的操作步骤

The screenshot shows the Microsoft Excel interface with several windows open, illustrating the steps for performing regression analysis.

1. Data Tab: The "Data" tab is selected in the ribbon. A red circle highlights the "数据分析" (Data Analysis) button in the "Analysis" group.

2. Data Analysis Dialog: The "数据分析" (Data Analysis) dialog box is open, showing various statistical tools. The "回归" (Regression) option is selected, highlighted with a red circle. A red circle also highlights the "确定" (OK) button.

3. Regression Dialog: The "Regression" dialog box is open, showing the input range (\$C\$1:\$C\$14) for Y values and (\$B\$1:\$B\$14) for X values, both highlighted with red circles. Other options like "标志" (Labels), "常数为零", and "置信度" are shown. The "输出选项" (Output Options) section has "新工作簿" (New Workbook) selected, highlighted with a red circle. Under "残差" (Residuals), "残差" and "标准残差" are checked, while "残差图" and "线性拟合图" are unchecked. The "正态分布" (Normal Distribution) section has "正态概率图" unchecked. Buttons for "确定" (OK), "取消" (Cancel), and "帮助" (Help) are visible.

4. Data Analysis Dialog: The "数据分析" (Data Analysis) dialog box is partially visible on the right, with its own "确定" (OK) button highlighted with a red circle.

5. Excel Worksheet: The main worksheet shows data from row 1 to 14. Column A is labeled "obs", column B is labeled "X", and column C is labeled "Y". The data includes values such as 4.4567, 5.77, 5.9787, etc. The formula bar shows \$D\$18.

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

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			

- 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

Equation: EQ_WAGE Workfile: CHPT2::wage\				
View	Proc	Object	Print	Name
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			

- 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 $S_{\hat{\beta}_i}, i \in 1, 2, \dots, k$ 。
- 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）

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		

- **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

