BECOME A MASTER OF WINE
WEATHER IS ALL YOUR NEED

Karry

1 描述性统计结果

到底是什么在影响波尔多葡萄酒的质量和价格呢， 也就是说为什么不同年份的葡萄酒价格会有这么大的差距呢？

> 在这里我们先对本次作业中“价格”这一概念做一定的说明：由于我们的数据中“价格”是相对于1961年的葡萄酒而言的，被规范化为100的数值，因此此处的葡萄酒“价格”本质是一个“价格指数”，也就意味着它不能表示葡萄酒在市场上的绝对价格到底是多少，只能表示相对于1961年葡萄酒而言相对价格的高低。

毋庸置疑，一定有一些价格上的差异是由酿酒的不同年份造成的。我们参照作业中的参考资料给出了两个自然的解释：

- 葡萄酒作为一种**投资品**，是“时间的玫瑰”。
- 葡萄酒作为一种**消费品**，天气不同造成味道不一，优质的葡萄酒千金难求。

这两点将分别在 2.1 和 2.2 中展开进行具体说明。

1.1 Returns to Holding Wine

老酒的持有时间更长，这需要对放弃消费葡萄酒的投资进行一定的回报。

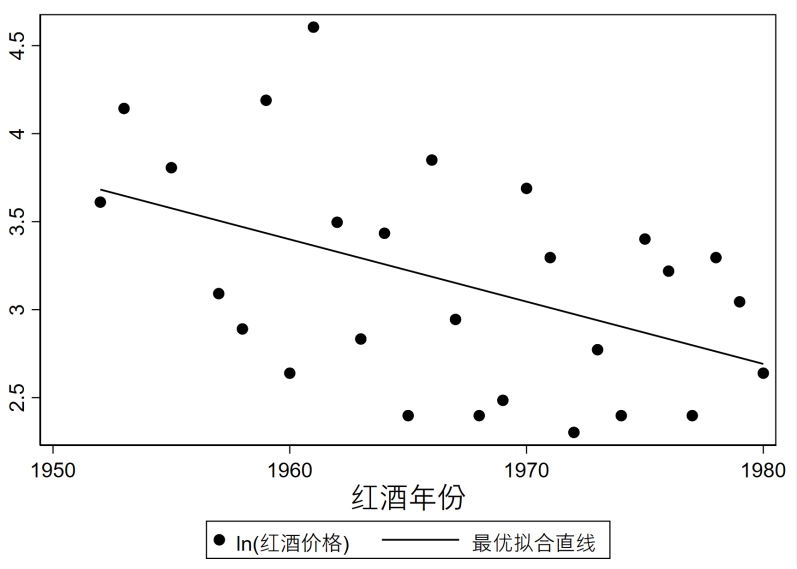

图 1. 葡萄酒价格与年份的散点图

为了检验这个假设，我们构建了图 1。图 1 是一个葡萄酒价格与年份的散点图。观察图中的数据点或最佳拟合线，很明显这两个变量之间存在负相关，最佳拟合线的斜率是 -0.035。这意味着：酒的生产时间越早，其价格就越高(或者如图 2. 中显示的酒龄越长价格越高)，这在一定程度上印证了我们的假设。但与此同时我们也可以从图 1 中看出：只用葡萄酒年份来解释不同年份的价格变化是明显不足够的。

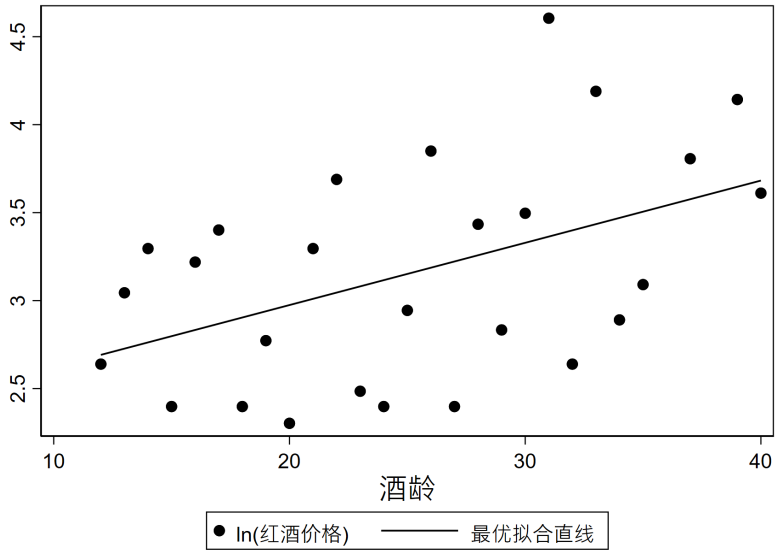

图 2. 葡萄酒价格与酒龄的散点图

1.2 Weather ！！！

好的天气成就精品葡萄，精品葡萄成就优质葡萄酒。

一般来说，任何水果的质量都取决于生产该水果的生长季节的天气，葡萄亦不例外。世界上任何地方的酿酒师都不约而同地认定波尔多葡萄酒的高质量年份对应的特征有以下三条：

- 8月和 9月是干燥的，即收获时的降雨量很小。
- 生长季节是温暖的，即夏天的温度较高。
- 前一个冬天是潮湿的，即冬天降雨量较高。

以上假设在图 3 中得到了较为充分的验证。该图显示了每个年份的夏季温度从低到高，从下到上；以及收获的雨水从低到高，从左到右；其中虚线分别表明两个维度的平均值。我们发现：售价高于平均价格的年份显示为实心黑色菱形——这些点大多数于左上象限，而售价低于平均价格的年份显示为红色空心正方形——这些点大多数处于右下象限，这充分表明表

明正是炎热、干燥的夏季造就了品质优秀的葡萄，这些葡萄被酿造成优质的葡萄酒，千金难求。

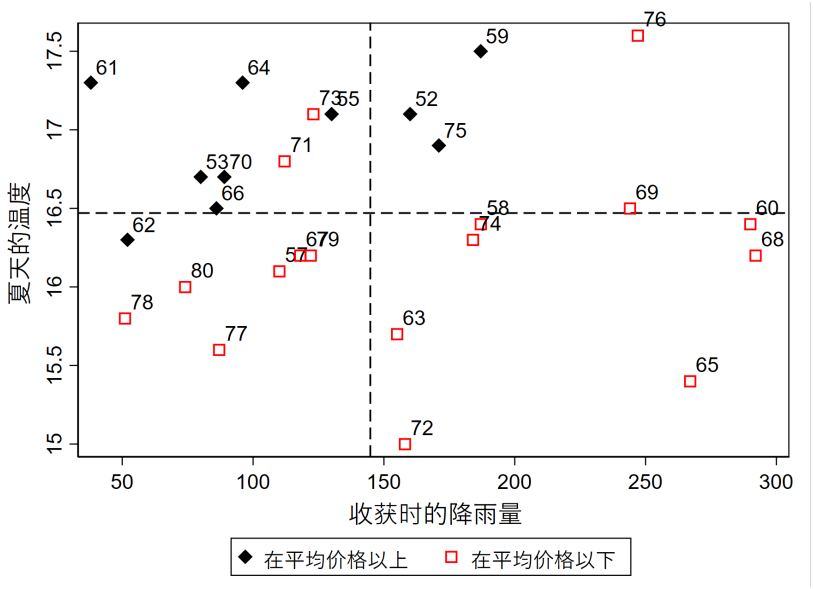

图 3. 葡萄收获时降雨量夏天温度与价格散点图

1.3 Conclusion

在上述描述性统计的基础上，我们的两大假设得到了初步的验证——**酒龄和天气**是影响葡萄酒价格的重要因素。但这些因素到底如何定量影响着葡萄酒价格？我们能否根据过去的数据构建出一个模型来精确地预测葡萄酒价格？这些都是本次作业要解答的核心问题，我们在第2节给出了定量模型的求解过程，在第3节给出了预测结果。

2 模型建立

基于上述分析我们建立了三种线性模型，并分别进行回归求解。本节所用到的所有变量简写均和作业中保持一致。为了便于阅读，我们便在此列出变量名及其实际含义:

变量名	含义
price	葡萄酒的相对价格
sum	夏天的温度
har	收获时的降雨量
sep	九月的气温
win	冬天的降雨量
age	酒龄

表 1. 相关变量说明

2.1 三种模型及求解结果

Model 1

$$ln(price) = \beta_0 + \beta_1 age + \mu \qquad (1)$$

```
. reg lnPrice age
```

Source	SS	df	MS		
				Number of obs	= 27
				F(1, 25)	= 6.90
Model	2.21353304	1	2.21353304	Prob > F	= 0.0145
Residual	8.01702429	25	.320680972	R-squared	= 0.2164
				Adj R-squared	= 0.1850
Total	10.2305573	26	.393482974	Root MSE	= .56629

lnPrice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0353828	.0134675	2.63	0.014	.0076461	.0631196
_cons	2.267031	.3562595	6.36	0.000	1.533301	3.000761

图 4. Model 1 的求解结果

Model 2

$$ln(price) = \beta_0 + \beta_1 age + \beta_2 sum + \beta_3 har + \beta_4 win + \mu \qquad (2)$$

```
reg lnPrice age sum har win
```

Source	SS	df	MS	Number of obs	=	27
				F(4, 22)	=	27.01
Model	8.49977773	4	2.12494443	Prob > F	=	0.0000
Residual	1.7307796	22	.0786718	R-squared	=	0.8308
				Adj R-squared	=	0.8001
Total	10.2305573	26	.393482974	Root MSE	=	.28048

lnPrice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0243519	.0069947	3.48	0.002	.0098458	.038858
sum	.6187109	.0943199	6.56	0.000	.4231034	.8143185
har	-.0037482	.0007915	-4.74	0.000	-.0053896	-.0021069
win	.0011972	.000474	2.53	0.019	.0002143	.0021802
_cons	-7.831138	1.662956	-4.71	0.000	-11.2799	-4.382377

图 5. Model 2 的求解结果

Model 3

$$ln(price) = \beta_0 + \beta_1 age + \beta_2 sum + \beta_3 har + \beta_4 win + \beta_5 sep + \mu$$

```
reg lnPrice age sum har win sep
```

Source	SS	df	MS	Number of obs	=	27
				F(5, 21)	=	20.67
Model	8.50253084	5	1.70050617	Prob > F	=	0.0000
Residual	1.72802649	21	.082286976	R-squared	=	0.8311
				Adj R-squared	=	0.7909
Total	10.2305573	26	.393482974	Root MSE	=	.28686

lnPrice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.024604	.0072852	3.38	0.003	.0094536	.0397544
sum	.6071434	.1153446	5.26	0.000	.3672711	.8470157
har	-.0036647	.0009294	-3.94	0.001	-.0055975	-.0017319
win	.0011768	.0004974	2.37	0.028	.0001424	.0022112
sep	.0101573	.0555306	0.18	0.857	-.1053249	.1256395
_cons	-7.819812	1.701863	-4.59	0.000	-11.35903	-4.280595

图 6. Model 2 的求解结果

> 注：三种模型均采用 OLS 进行求解。具体的求解代码已在上面几个图中展现，不再赘述细致过程。

2.2 模型结果分析

为了更好地展示回归结果，我们又制作了表 2 报告了葡萄酒价格对酒龄和天气变量的回归结果。结果表明：在一个包括四个变量的模型中，葡萄酒的酒龄、夏天的温度、收获时的降雨量以及冬天的降雨量，可以解释波尔多葡萄酒年份平均价格的大约80%的变化，仅仅分析酒龄的

影响，得出的模型只能解释20%多一点。这表明天气是决定葡萄酒年份质量和成熟期价格的一个极其重要的因素。

表 2: ln(price) 对酒龄和天气变量的回归结果

独立变量	(1)		(2)		(3)	
酒龄	0.0353*	(0.0134)	0.0243**	(0.0070)	0.025*	(0.00729)
夏天的温度	-		0.0619**	(0.094)	0.607*	(0.115)
季节(4月-9月)						
收货时的降雨量	-		-0.004**	(0.00079)	-0.004 **	(0.00093)
冬天的降雨量	-		0.0011*	(0.00047)	0.001*	(0.000497)
季节(10月-3月)						
九月的气温	-		-		0.010	(0.056)
R^2	0.2164		0.831		0.831	
均方根误差	0.56629		0.280		0.287	

[1] 所有的回归都是针对本次作业数据中价格（自然对数）对天气和酒龄进行回归。
[2] ⋆表示在 5%的显著性水平下显著;⋆⋆表示在 1%的显著性水平下显著。

当然我们也分别对三个模型的结果做出以下细致解释：

Model 1

- 其他条件不变时，每增加一年的酒龄葡萄酒价格增加 3.53%。

Model 3

- 其他条件不变时，每增加一年的酒龄，葡萄酒价格增加 2.43%。
- 其他条件不变时，每增加一单位夏天的气温，葡萄酒价格增加 61.9%。
- 其他条件不变时，每增加一单位收获时的降雨量，葡萄酒价格减少 4%。
- 其他条件不变时，每增加一单位冬季的降雨量，葡萄酒价格增加 1.1%。

Model 3

- 其他条件不变时，每增加一年的酒龄，葡萄酒价格增加 2.5%。
- 其他条件不变时，每增加一单位夏天的气温，葡萄酒价格增加 60.7%。
- 其他条件不变时，每增加一单位收获时的降雨量，葡萄酒价格减少 3.665%。
- 其他条件不变时，每增加一单位冬季的降雨量，葡萄酒价格增加 1.177%。
- 其他条件不变时，每增加一单位冬季的降雨量，葡萄酒价格增加 1.157%。

值得一提的是: 将 Model 1 和 Model 2，3 作比较我们发现，酒龄的偏效应明显被高估了，这是为什么呢？

3 预测结果

第 1 节中我们对主要因素进行了描述性统计完成了定性分析，第 2 节我们分别建立模型并求解模型。我们选取 Model 3(毕竟它有最高的 **R 方**)预测 80 年代葡萄酒的相对价格进而回答了本次作业中的两个问题即：

- (a) 1986年是一个好年份吗？
- (b) 1982年和1983年在事后被认为是极好的年份。你的模型是否支持，它们与1961年相比如何？

当然我们还做了一些有意思的拓展，并提出了未来可供深入探讨的新问题。

3.1 80 年代葡萄酒相对价格的预测结果

葡萄酒年份	预测价格
1981	23.996
1982	30.645
1983	40.692
1984	16.723
1985	32.481
1986	11.44
1987	18.673
1988	35.376
1989	52.17367
1990	56.63228
1991	23.53701

表 3. 80年代葡萄酒相对价格的预测结果

对作业中两个问题的回答：

- (a) 从我们预测的结果来看 1986 年是一个极其糟糕的年份。我们可以从表 2 中看到 1986 年的葡萄酒相对价格仅为 11.44， 几乎只是 1961 年的十分之一！同时我们也可以从图 7. 中得到这一结论——1986年的降雨量和温度均在平均线以下(注意：图7. 只是在 图2. 中增加了1981-1991年的年份数据，但根据历史上的正常降雨量和温度数据，将轴线保持在同一位置上)。
- (b) **1982年 和 1983年的确不是很坏的年份，但也绝对不像某些专家说的是极好的年份，我们的模型反映：它们的价格不及 1961 年的二分之一！** 这一结论同样可以在图7. 中得到证明，可以看到 1982 年的干燥程度不及平均水平，1983年干燥程度也不及1961年。这的确是一个很反直觉的的结果，因为我们总是开口闭口就是82年拉菲，但现在看来这些酒有炒作之嫌，无精品之实。

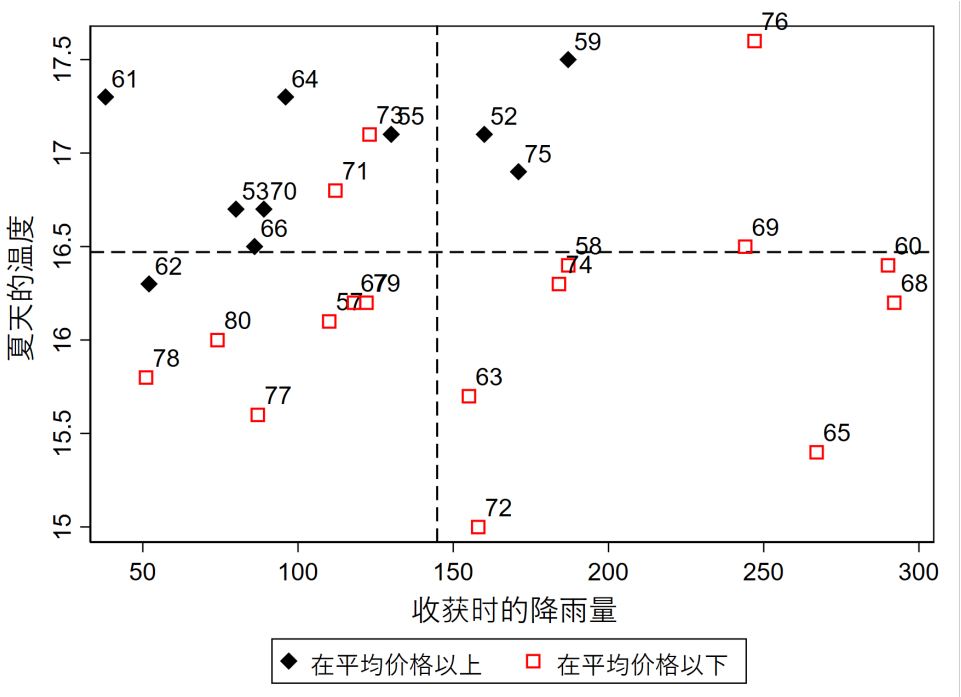

图 7. 葡萄收获时降雨量夏天温度与价格散点图

3.2 一些有趣的发现和问题

根据老师发送的 A425_wine.xlsx 文件，我们可以得到 80 年代各个年份的葡萄酒实际价格，这个地方是葡萄酒的绝对价格，在数值上和我们预测的结果没有可比性，但是我们如果仅仅站在纵向比较（即只是比较不同年份酒价的高低）的角度上来说的话可以发现一个很有趣的现象如图 8. 所示：我们预测的价格和实际价格的走势基本相同，这在一定程度上也说明了我们模型的有效性。

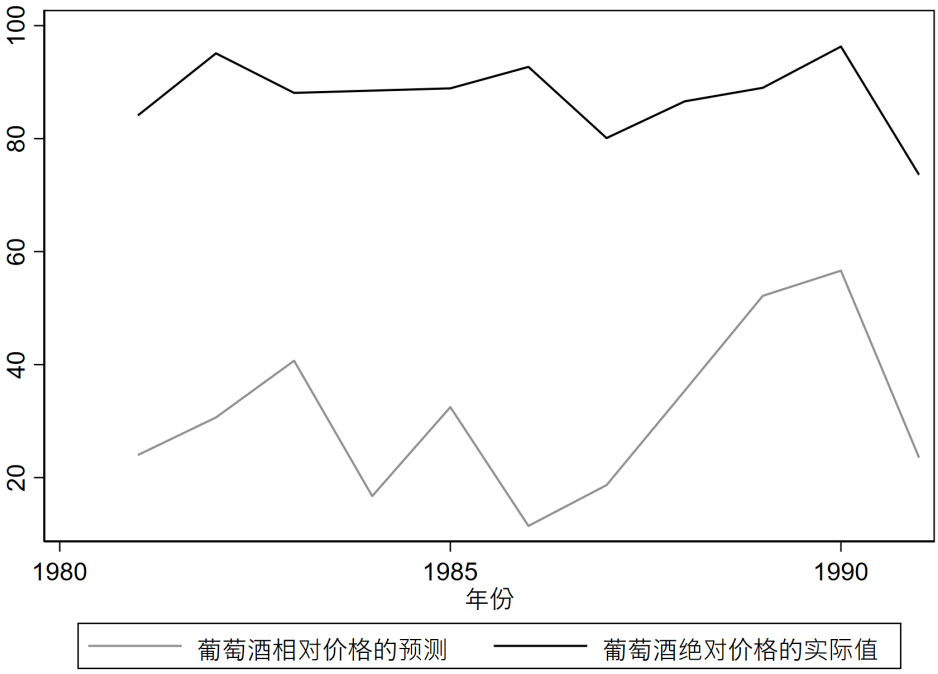

图 8. 预测和实际值的趋势走向

我们在本次作业的分析中提出了几个问题，另外我们也发现一些单独的问题值得讨论，我们在这汇总起来，以为后续的讨论提供方向：

- 在Model 1 中酒龄的偏效应明显被高估了，这是为什么呢？
- 怎么才精准预测某一年份葡萄酒的绝对价格呢？
- 既然这个模型这么有效，那它是否对葡萄酒的定价效率有影响？更准确的说：葡萄酒的定价是否更准确，价格波动率是否会下降？