

# 数理统计学习笔记

## 大作业（一）完成指导

慕弋云子

December 4, 2020

关于学习笔记的常规内容，可见本文 GitHub Repo 的地址，在这里你可以找到之前或之后的内容，不会迷路：[www.github.com/Muyiyunzi/ShuLiTongJi\\_BUAA](https://www.github.com/Muyiyunzi/ShuLiTongJi_BUAA)

本文旨在帮助大家事半功倍地完成大作业的相关内容，对于个中原理，我们可以不求甚解（孙老师的八字方针原话：**点到为止，独立完成**），虽然百度上有一些之前学长学姐的作业，但是毕竟是付费的，大家还得照猫画虎、邯郸学步，排个版也会花费不少的功夫。

于是，我在 GitHub Repo 中上传了.docx 的文件，大家可以借此省去排版的功夫，照着这篇文档，也可以“傻瓜式”地快速完成这份大作业。

孙老师的原话是，“如果查重率到了七八十，那我可就得处理你了”，所以我的理解是：**只要数据使用的不同**，那些原理什么的抄一抄（注好引用），用软件跑一下得到结果自己分析分析，**就一定不会被查**。另外孙老师还说，要**结构完整，敢下判断**，所以说除了数据一定要自己找一下之外，剩下的就是按照这篇《完成指导》所给的图去索骥，只要结构完整，就绝对不会为难大家。

## 1 SPSS 的安装

这个大作业的主要目的是**让大家学习使用统计软件**。而 SPSS 这个软件易于破解，操作清晰，用于完成大作业简直再合适不过。作为一个六系学生我当然也想过用 Python，但想了想研究调库也得有段时间，还是不如 SPSS 一键生成来得方便。所以在此墙裂推荐大家使用 SPSS。

我所使用的软件版本是 25，各个版本的操作应该都是大同小异，由于我使用的是 Windows 系统，所以其他系统的童鞋就得自己找一下破解，并跳过本节的剩余内容了。在这里我贴一个 Windows 系统的度盘链接方便大家直接下载：**提取码：beay**。按照下面的指示完成安装和破解。

下载后双击运行.exe 文件即可，是否安装 Python 的补充文件、路径等可以自行随缘选择。安装完毕后，不要立即激活，然后将 crack.rar 解压，将 crack 目录下的 lservrc 文件替换到安装目录下。例如我 SPSS 的安装路径是 D:\SPSS\，那么就将文件复制到这个路径下并选择替换。



图 1: IBM SPSS 的安装界面

替换完毕后，可以选择使用搜索栏（或 Win+S）搜索 SPSS，或是在安装路径下打开 Stats.exe 即可运行 SPSS。

## 2 数据的选择、处理与查找

### 2.1 数据查找

这里主要推荐两个地方吧。一个是 [UCI Machine Learning Repository](#)，这里面有很多用于机器学习的数据集。点进“View All Dataset”的链接中，左侧的 Default Task 框框中，有一个 Regression，大家可以自行选择数据。

Name	Data Types	Default Task	Attribute Types	# Instances	# Attributes	Year
UCI 3D Road Network (North Jutland, Denmark)	Sequential, Text	Regression, Clustering	Real	434874	4	2013
UCI Air Quality	Multivariate, Time-Series	Regression	Real	9358	15	2016
UCI Air quality	Multivariate, Time-Series	Regression	Real	9358	15	2016
UCI Airfoil Self-Noise	Multivariate	Regression	Real	1503	6	2014
UCI Alcohol QCM Sensor Dataset	Multivariate	Classification, Regression, Clustering	Real	125	8	2019
UCI Algerian Forest Fires Dataset	Multivariate	Classification, Regression	Real	244	12	2019

图 2: UCI ML Repo 的 Regression Task 界面

再一个就是 **data fountain**，在顶部菜单栏处点击数据集即可进行检索。这里面的数据标签大多是汉化过的，不过数据比较杂，可能会有很多非数值特征，使用起来大多需要处理（如何选择和处理见下面两小节）。不过好的是这里的检索系统还是挺靠谱的，你可以在此检索一些想要的数 据，也可以搜索像“价格”、“销售”之类的字眼，会更加利于回归任务。

基本上这两个 source 就可以让大家的数据不重样了。你可别小看这两个数据集，一个 Air Quality 就能有 9000 多条数据，你选个 100 条出来就差不多了，这要是也能有哪位兄弟能跟你选重样了，那可真是……

## 2.2 数据选择

有了数据源，就要学会如何选择，这样才能更好地完成大作业。在数据的选择上，我认为需要注意两点。

一、数据一定要是**数值特征**，否则没法做回归。

举个例子，我想做笔记本电脑价格的多元线性回归，可以预见到，内存越大、频率越高对价格肯定是正增益的，屏幕分辨率越高、重量越轻、厚度越薄等因素肯定也是正增益的，这些数据都是**数值的**。这样回归出来的效果好吗？估计是不好的。就算好，也可能回归出来那个常量很大，因为影响电脑的主要因素在于 CPU、显卡这些指标，而像 i3、i5、i7，AMD 还是 Intel，6800XT 还是 3080，这些东西是很难以数值去表示的。如果你的数据集在这些数据上波动很大（比如你拿一个 i5 的 960 和一个 i9 的 3080 去比，就算他们内存一样、薄厚重量差不多，价位肯定也是天差地别，而你把这些因素忽略掉了），那么回归的效果肯定是不好的，遇到这种情况的话我给两条建议：

1. 放弃整个数据集，正所谓 there's plenty of fish in the sea，天涯何处无芳草嘛。
2. 放弃这项非数值特征，反正按照孙老师的说法，点到为止，回归出来线性程度不好就大胆拒绝，有结论就行。
3. 对这些非数值特征做数值化。比如 CPU 和 GPU 的型号可以对应地换成他们在某个 benchmark 上的评分，然后再去做回归即可，当然了，处理数据会有一些工作量，这就要看你的时间精力，以及对这项工作的喜爱程度了。

二、数据要尽量有实际意义，一是方便你写论文，二也算是让自己做的东西有些意义。

这是一个数据量爆炸的时代，但并不一定所有数据都有用。比如我手头有近些年英国足球超级联赛的数据，我该怎么入手分析呢？也就是一个核心的问题：如何确定因变量，以及**回归到底有什么用**？这才是这个大作业需要把握的核心问题。

我认为，回归就是为了确定是不是线性关系，以及能够为将来的数据做出预测。

比如射门数对于进球数的贡献是不是线性增益？是我射门数越多进球概率就会越大，还是说呈指数性地，我就算射几十上百次也不一定进了几个球，这是「确定线性关系」。

如果我做完回归后，得到了进球数与射门数、角球数、犯规数等诸多指标的线性拟合式，我作为一个教练是不是就可以更好地布置战术，比如犯规对进球数是正增益我就要多让球员犯规；我作为一个赌狗，是不是就可以更好地根据比赛数据预测比赛结果，进行相应地投注指导；我作为一个显卡爱好者，是不是就可以根据这些计算单元数量、吞吐量等指标来预测显卡“香不香”，这些都是所谓「预测」。

带着这些目的，就可以更好地选择数据和选择相关变量，至于你要确定什么关系，要做何种预测，就是自己的事情了。

## 2.3 数据处理

这里的数据处理主要是针对于非数值特征而言的。我建议大家使用 excel 去处理数据，还是比较便捷的，使用一些基本的逻辑语句即可。在这里我以我自己的大作业为例，教大家一些简单的数据处理方法（如果不需要处理，或 excel 技巧高超，此节内容跳过即可）。

比如以下是我拿到的原始数据：

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	Div	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	Referee	HS	AS	HST	AST	HF	AF	HC	/
2	E0	2009/8/15 0:00	Aston Villa	Wigan	0	2	A	0	1	A	M Clattenk	11	14	5	7	15	14	4	
3	E0	2009/8/15 0:00	Blackburn	Man City	0	2	A	0	1	A	M Dean	17	8	9	5	12	9	5	
4	E0	2009/8/15 0:00	Bolton	Sunderland	0	1	A	0	1	A	A Marriner	11	20	3	13	16	10	4	
5	E0	2009/8/15 0:00	Chelsea	Hull	2	1	H	1	1	D	A Wiley	26	7	12	3	13	15	12	
6	E0	2009/8/15 0:00	Everton	Arsenal	1	6	A	0	3	A	M Halsey	8	15	5	9	11	13	4	
7	E0	2009/8/15 0:00	Portsmouth	Fulham	0	1	A	0	1	A	M Atkinson	16	9	4	3	11	18	6	
8	E0	2009/8/15 0:00	Stoke	Burnley	2	0	H	2	0	H	S Bennett	12	9	5	5	15	10	3	
9	E0	2009/8/15 0:00	Wolves	West Ham	0	2	A	0	1	A	C Foy	19	16	11	13	9	5	8	
10	E0	2009/8/16 0:00	Man United	Birmingham	1	0	H	1	0	H	L Mason	26	6	17	4	13	7	13	
11	E0	2009/8/16 0:00	Tottenham	Liverpool	2	1	H	1	0	H	P Dowd	17	6	11	3	14	16	6	
12	E0	2009/8/18 0:00	Sunderland	Chelsea	1	3	A	1	0	H	S Bennett	4	20	3	9	14	10	1	
13	E0	2009/8/18 0:00	Wigan	Wolves	0	1	A	0	1	A	M Jones	18	9	5	3	8	21	4	
14	E0	2009/8/19 0:00	Birmingham	Portsmouth	1	0	H	0	0	D	L Probert	9	9	5	4	11	20	4	
15	E0	2009/8/19 0:00	Burnley	Man United	1	0	H	1	0	H	A Wiley	8	18	2	9	8	12	1	
16	E0	2009/8/19 0:00	Hull	Tottenham	1	5	A	1	3	A	C Foy	9	18	7	12	23	13	6	
17	E0	2009/8/19 0:00	Liverpool	Stoke	4	0	H	2	0	H	P Walton	18	6	13	3	10	9	11	
18	E0	2009/8/22 0:00	Arsenal	Portsmouth	4	1	H	2	1	H	S Bennett	19	9	16	4	9	10	8	
19	E0	2009/8/22 0:00	Birmingham	Stoke	0	0	D	0	0	D	C Foy	8	13	6	3	8	13	3	
20	E0	2009/8/22 0:00	Hull	Bolton	1	0	H	0	0	D	M Jones	12	20	5	10	19	19	4	
21	E0	2009/8/22 0:00	Man City	Wolves	1	0	H	1	0	H	L Mason	16	10	11	2	7	11	5	
22	E0	2009/8/22 0:00	Sunderland	Blackburn	2	1	H	1	1	D	A Wiley	8	18	5	5	12	14	0	
23	E0	2009/8/22 0:00	Wigan	Man United	0	5	A	0	0	D	H Webb	16	16	7	13	11	8	3	
24	E0	2009/8/23 0:00	Burnley	Everton	1	0	H	1	0	H	P Dowd	8	17	4	11	11	13	3	
25	E0	2009/8/23 0:00	Fulham	Chelsea	0	2	A	0	1	A	A Marriner	4	12	1	4	11	10	4	
26	E0	2009/8/23 0:00	West Ham	Tottenham	1	2	A	0	0	D	M Clattenk	17	17	8	10	13	11	4	
27	E0	2009/8/24 0:00	Liverpool	Aston Villa	1	3	A	0	2	A	M Atkinson	21	7	12	4	15	11	10	

图 3: 原始数据

原始的数据是从 data fountain 上获取的（去查了原始网站），是每轮比赛的对阵情况，以及对应的一些比赛数据。那么首先要做的就是去除冗余信息，比如每轮的对阵情况太过杂乱，要么我就单取出几轮，去分析整个联赛的情况；要么我就单独取出一支队伍，然后去考察这支队伍的情况。

这里我是通过查找操作，手工高亮出某支球队（Arsenal）的比赛数据（毕竟一个赛季也就 38 场，我觉得手工操作量不大）。然后将这些英文的特征对应原始网站提供的 notes.text（大部分数据集也都会提供类似的说明文件）汉化一下，并去除冗余信息后大概是这个样子：

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	Div	Date	Time	HomeTeam	AwayTeam	全场主队进球数	全场客队进球数	全场结果	半场主队进球数	半场客队进球数	半场结果	主裁判	主队射门数	客队射门数	主队射正数	客队射正数	主队犯规数	客队犯规数	主队角球数	客队角球数
2	E0	11/08/2011	14:00	Newcastle	Arsenal	0	1	A	0	0	D	M Atkinson	9	8	2	2	12	7	5	
3	E0	17/08/2011	12:30	Arsenal	Burnley	2	1	H	1	1	D	M Dean	16	18	9	5	13	11	10	
4	E0	24/08/2011	17:30	Liverpool	Arsenal	3	1	H	1	0	H	A Taylor	25	9	5	3	8	5	6	
5	E0	01/09/2011	16:30	Arsenal	Tottenham	2	2	D	1	2	A	M Atkinson	26	13	8	9	13	19	11	
6	E0	15/09/2011	16:30	Watford	Arsenal	2	2	D	0	2	A	A Taylor	31	10	7	4	14	4	7	
7	E0	22/09/2011	16:30	Arsenal	Aston Villa	3	2	H	0	1	A	J Moss	21	14	6	9	13	15	9	
8	E0	30/09/2011	20:00	Man United	Arsenal	1	1	D	1	0	H	K Friend	16	10	4	5	18	13	8	
9	E0	06/10/2011	14:00	Arsenal	Bournemouth	1	0	H	1	0	H	M Atkinson	12	10	2	2	12	6	14	
10	E0	21/10/2011	20:00	Sheffield U	Arsenal	1	0	H	1	0	H	M Dean	8	9	2	3	10	12	7	
11	E0	27/10/2011	16:30	Arsenal	Crystal Palace	2	2	D	2	1	H	M Atkinson	15	10	6	4	18	9	12	
12	E0	02/11/2011	15:00	Arsenal	Wolves	1	1	D	1	0	H	M Oliver	10	25	4	8	6	15	8	
13	E0	09/11/2011	17:30	Leicester	Arsenal	2	0	H	0	0	D	C Kavanagh	19	8	7	1	10	10	5	
14	E0	23/11/2011	15:00	Arsenal	Southampton	2	2	D	1	1	D	S Attwell	12	21	5	6	13	19	6	
15	E0	01/12/2011	14:00	Norwich	Arsenal	2	2	D	2	1	H	P Tierney	15	16	8	7	8	10	7	
16	E0	05/12/2011	20:15	Arsenal	Brighton	1	2	A	0	1	A	G Scott	12	20	5	9	10	11	9	
17	E0	09/12/2011	20:00	West Ham	Arsenal	1	3	A	1	0	H	M Dean	11	10	4	3	12	6	4	
18	E0	15/12/2011	16:30	Arsenal	Man City	0	3	A	0	3	A	P Tierney	6	14	1	7	9	24	3	
19	E0	21/12/2011	12:30	Everton	Arsenal	0	0	D	0	0	D	K Friend	9	6	0	2	10	11	5	
20	E0	26/12/2011	15:00	Bournemouth	Arsenal	1	1	D	1	0	H	S Attwell	12	17	4	2	5	13	8	
21	E0	29/12/2011	14:00	Arsenal	Chelsea	1	2	A	1	0	H	C Pawson	7	13	2	4	13	19	2	
22	E0	01/01/2012	20:00	Arsenal	Man United	2	0	H	2	0	H	C Kavanagh	10	10	4	4	11	15	1	
23	E0	11/01/2012	12:30	Crystal Palace	Arsenal	1	1	D	0	1	A	P Tierney	6	7	3	4	14	22	1	
24	E0	18/01/2012	15:00	Arsenal	Sheffield U	1	1	D	1	0	H	M Dean	11	12	4	4	9	13	4	
25	E0	21/01/2012	20:15	Chelsea	Arsenal	2	2	D	1	0	H	S Attwell	19	2	8	2	11	6	17	
26	E0	02/02/2012	14:00	Burnley	Arsenal	0	0	D	0	0	D	C Kavanagh	15	13	2	2	8	11	5	
27	E0	16/02/2012	16:30	Arsenal	Newcastle	4	0	H	0	0	D	L Mason	15	10	7	2	15	9	5	

图 4: 初步去除后

根据我作为一个足球狗的先验知识，客场作战肯定影响很大，我不想放弃掉这一信息，所以就单开了一栏表示是否客场。这里我将这个非数值特征二值化了——虽然武断了些，但也有好处：如果我们将客场以 1 表示，主场以 0 表示，那么最后分析出来的那个系数就应该是直接增益！倘若我的因变量是胜率，那么最后就能直接回归出，如果是客场就可以直接带来多少的胜率增加/减少，这是非常直观的。

这里可以使用 Excel 中的 IF 逻辑判断。选中某一个块，在其上输入判断语句：

=IF(E2="Arsenal",1,0)

其中 IF 语句的第一个分量表示判断条件，第二个分量为判断为真时此块的 value，第三个分量为判断为假时的 value（多说一句，所以这其实是个三目表达式啊），这里 E2 表示 E 列的第二个块，在我的数据中表示客场的球队。

=IF(E2="Arsenal",1,0)									
U	V	W	X	Y	Z	AA	AB	AC	AD
主队黄牌数	客队黄牌数	主队红牌数	客队红牌数	AvgH	AvgD	AvgA	是否客场		
1	3	0	0	4.49	3.82	1.79	=IF(E2="Arsenal",1,0)		
2	1	0	0	1.33	5.49	9.48	IF(logical_test, [value_if_true], [value_if_false])		
1	1	0	0	1.5	4.76	6.09			
3	5	0	0	2.38	3.63	2.9			
3	3	0	0	3.64	3.74	2.01			

图 5: IF 语句的使用

如果一切顺利的话，就会输出 1 或者 0，接下来要做的就是大家在 Excel 中都会的，将鼠标指针放到块的右下角，变成加号后按住左键往下拉，就可以神奇地自动逐行生成判断了。由此，我们便完成了将一个非数值特征二值化的操作。

此后还有一些对数据的运算，基本就是在 Excel 中敲公式了，如果不会敲公式的话，可以去查阅我的大作业以及文件夹中的.xlsx 文件，在此不再赘述。

再就是尽量选一些数字大的东西去回归（比如胜率能用 50% 就不用 0.5），这样回归出来的系数不至于太难看。就像我也放弃了进球数作为因变量，而是用博彩胜率这样一个阈值更大的数据作为因变量。

由此我们便得到了最终的数据：

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	场次	是否客场	阿森纳射门数	阿森纳射正数	阿森纳犯规数	阿森纳角球数	阿森纳黄牌数	阿森纳红牌数	阿森纳获胜概率	阿森纳打平概率	阿森纳不败概率	阿森纳输球概率	获胜概率（百分之）
2	1	1	8	2	7	3	3	0	0.445	0.378	0.823	0.177	44.455
3	2	0	16	9	13	10	2	0	0.582	0.337	0.918	0.082	58.160
4	3	1	9	3	5	4	1	0	0.121	0.385	0.507	0.493	12.146
5	4	0	26	8	13	11	3	0	0.325	0.407	0.733	0.267	32.548
6	5	1	10	4	4	1	3	0	0.388	0.398	0.786	0.214	38.765
7	6	0	21	6	13	9	5	1	0.542	0.355	0.897	0.103	54.236
8	7	1	10	5	13	7	2	0	0.266	0.398	0.663	0.337	26.577
9	8	0	12	2	12	14	1	0	0.510	0.379	0.889	0.111	50.992
10	9	1	9	3	12	12	4	0	0.404	0.395	0.799	0.201	40.417
11	10	0	15	6	18	12	2	0	0.515	0.367	0.882	0.118	51.467
12	11	0	10	4	6	8	0	0	0.455	0.371	0.826	0.174	45.455
13	12	1	8	1	10	4	1	0	0.212	0.403	0.614	0.386	21.193
14	13	0	12	5	13	6	6	0	0.513	0.372	0.885	0.115	51.292
15	14	1	16	7	10	12	1	0	0.409	0.417	0.825	0.175	40.869
16	15	0	12	5	10	9	3	0	0.481	0.376	0.857	0.143	48.089
17	16	1	10	3	6	3	0	0	0.364	0.426	0.790	0.210	36.373
18	17	0	6	2	9	3	1	0	0.113	0.416	0.529	0.471	11.319
19	18	1	6	2	11	4	3	0	0.259	0.418	0.677	0.323	25.947
20	19	1	17	2	13	3	4	0	0.374	0.411	0.785	0.215	37.393
21	20	0	7	2	13	2	5	0	0.263	0.417	0.680	0.320	26.286
22	21	0	10	4	11	1	2	0	0.296	0.412	0.708	0.292	29.558
23	22	1	7	4	22	4	3	1	0.424	0.391	0.815	0.185	42.366
24	23	0	11	4	9	4	1	0	0.450	0.364	0.814	0.186	44.960
25	24	1	2	2	6	5	1	1	0.167	0.380	0.547	0.453	16.731
26	25	1	13	2	11	7	3	0	0.403	0.380	0.783	0.217	40.323
27	26	0	15	7	15	5	2	0	0.561	0.335	0.896	0.104	56.145
28	27	0	9	4	12	6	0	0	0.392	0.388	0.780	0.220	39.159
29	28	0	9	2	11	6	1	0	0.494	0.378	0.867	0.133	49.360
30	29	1	3	0	7	2	1	1	0.090	0.368	0.458	0.542	9.024
31	30	1	13	6	8	7	1	0	0.345	0.383	0.728	0.272	34.545
32	31	1	10	5	14	6	2	0	0.301	0.396	0.697	0.303	30.137

图 6: 最终数据

其中打平、不败、输球这三栏数据我是做备用的，如果胜率做因变量的效果不好，可以换这些试试。

### 3 使用 SPSS 完成回归任务

将最终数据存好在.xls/.xlsx 文档之后，便可以导入 SPSS 完成回归任务了。这里有许多细节需要注意，我在这里先给出完整的操作，再补充解释。

1. 导入数据，可以通过欢迎页的打开导入，也可以通过菜单栏的文件-> 打开-> 数据打开，只是注意要在文件类型处选择 excel 表格，才能找到你的数据。

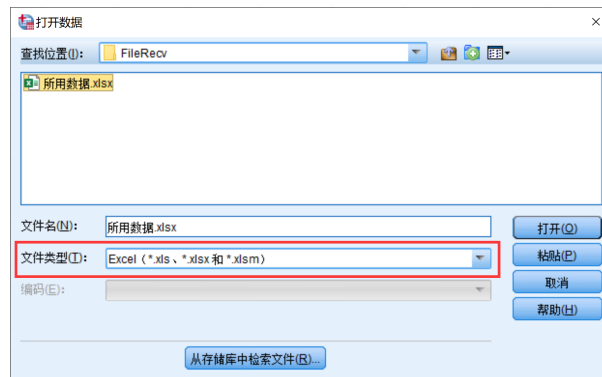


图 7: 导入数据

2. 点击打开后会弹出一个对话框。如果工作表有多个 sheet，选择你使用数据的那个 sheet；然后默认会有三个 tick，第二个 tick “用于确定……” 建议反选掉，虽然选上好像也没啥事儿。然后点击确定。
3. 这时应该会看到数据被导入。在菜单栏选择：分析-> 回归-> 线性，会弹出一个对话框，此时将因变量选中，点击添加按钮添加到因变量区域，再选中自变量，然后添加到自变量区域，如下图所示。友情提示，这里选中自变量的时候可以巧用 shift 或者 ctrl 操作，不用一个个点。



图 8: 选择变量

4. 将方法从默认的「输入」改成「步进」，这样就是逐步回归法了。
5. 随后点击「图」按钮，勾选标准化残差图区域中的两个框框，这东西用于分析数据的正态性。



图 9: 「图」按钮对话框



6. 继续后点击「选项」按钮，“使用 F 的概率”那一条中，可以自己设置进入和除去值。一般来讲 0.05 和 0.10 的组合都得出不了啥结论，可以调成 0.2+0.25 或是 0.4+0.5 的组合分别看看效果，都试试。

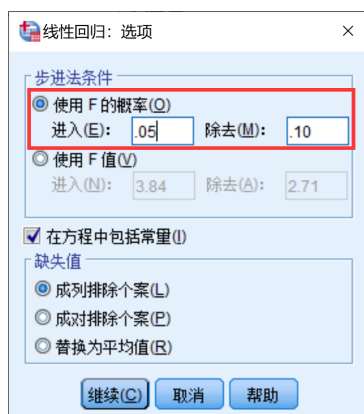


图 10: 「选项」按钮对话框

7. 其余选项都默认不要管，然后确定，即可生成各个表格。
8. 接下来要做的就是把这些表格对应地替换报告模板中的表格即可。

这里稍微对分析结果进行解释一下。

1. 输入/除去的变量中，表示了在你设置的这个 F 值组合中，按照显著性的程度，步进获得的变量。也就是说表格中越靠上的变量显著性越小（置信度越高）。另外就是，通常来讲是没有除去变量的。
2. 根据输入/除去的变量顺序，因为其逐步性，可以分别得到几个模型，这些模型都记录在了模型摘要表中。
3. ANOVA 不用太在意，照搬即可，一般就看最后那个显著性一栏，不太离谱就行。
4. 系数表比较重要，你最终会选择其中一个模型作为最优模型（你看着选，毕竟点到为止，有结构就行），那么最终多元回归的表达式就应该是  $y = b + \sum c_i x_i$  的形式，这里的  $c_i$  就都在这张表里了。你的报告里也应该有对应的完整表达式。
5. 系数表的显著性其实比较重要，一般来讲要小于 0.05 才算是比较能接受模型，但是你也得多点变量嘛，所以这里可以睁一只眼闭一只眼。
6. 排出的变量这张表没啥好说的，我认为不用贴；参差统计和后面的两个正态性的图都要贴上去，这可是老师点名要的图。

## 4 完成报告

这一部分相信大家 can 照葫芦画瓢完成，应该没什么太多要说的。至于报告各部分的内容，韦老师和孙老师班的要求略有差异，但我认为我这样对于两个班的要求来讲应该都是 ok 的，大家还是要把重心放在学习知识和考试上，大作业这种东西，大家体验体验就是了哈哈哈。

格式方面，我在上传的.docx 报告文件中也调整了样式库，大家如果需要自行加标题的话记得选样式、敲公式的话记得选公式样式然后使用制表位，参考文献换行会自动生成编号，最后也别忘了更新目录。

内容方面，大家在替换表格的时候可以从 SPSS 中复制，然后粘贴的时候选择保留源格式，然后选中后设置居中及边框全满，大概看起来会好看些吧。总之这些大家随意吧，原理的那些死的东西也都不用

改，涉及数据和分析的活的东西要自己写一写，大概就是这样，如有任何问题欢迎在 GitHub 上提 Issue，我会随缘解答。