首先清洗359个单词的数据：有几个异常值，2022/11/26给出的word为clen不满足五字母要求，2022/4/29的tash同样如此；2022/11/12给出的单词为naïve，不符合英文单词格式。在考虑常见的五字母英文单词及上述三个单词可能的搭配后，决定不删去这些词，而是替换为clean，trash，naive

然后根据观察和对英文单词特征的总结，得出了如下几个可能和困难模式下结果相关的单词的变量：

单词的26个字母组成；

单词在英文中的使用频率；

单词中含有的信息熵（比如说连续单词，一些重要字母等等）

difficulty表示的是一个单词的难度系数，有以下公式得到：

难度得分函数：（贡献法）↵

按猜测成功获得 1 分。↵

1 次猜测就成功则每次猜测的贡献值为 1 分↵

2 次猜测就成功则每次猜测的贡献值为 1/2 分↵

3 次猜测就成功则每次猜测的贡献值为 1/3 分↵

。。。↵

6 次都失败则不得分，每次猜测的贡献值为 0 ↵

↵

所以统计所有猜测次数的贡献值的平均数，则得到难度得分函数↵

首先，我们考虑单词的26个字母组成，即每个单词分别含有从a到z的哪些字母，分别有几个；然后再列出每个单词含有字母的表，在stata中将含有26的个数设置为26个虚拟变量，从而进行多元线性回归。

	word	diffic...	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
1	manly	.2395623	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	
2	molar	.2545455	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	1	0	
3	havoc	.2383333	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	
4	impel	.2525252	0	0	0	0	1	0	0	0	1	0	0	1	1	0	0	1	0	0	0	
5	condo	.2355	0	0	1	1	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	
6	judge	.1855219	0	0	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	
7	extra	.2618333	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
8	poise	.3052805	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	1	0	0	1	
9	aorta	.2632013	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	
10	excel	.2286195	0	0	1	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
11	lunar	.2778912	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	
12	third	.3148515	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	
13	slate	.3452862	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	
14	taper	.2666667	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	
15	chord	.291	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	
16	probe	.2563333	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	
17	rival	.2698333	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	
18	usual	.2498333	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	
19	spoke	.2774411	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	1	
20	apply	.2663366	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	0	0	0	
21	naïve	.2713333	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
22	knock	.2498333	0	0	1	0	0	0	0	0	0	0	2	0	0	1	1	0	0	0	0	
23	braid	.2964286	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	
24	infer	.2409864	0	0	0	0	1	1	0	0	1	0	0	0	0	1	0	0	0	1	0	
25	joust	.267	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	
26	amber	.2538333	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	
27	woken	.2161667	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	
28	adore	.3193069	1	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	
29	torso	.2734007	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	1	1	
30	chafe	.2681818	1	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	
31	eject	.2295	0	0	1	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	
32	study	.2671667	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
33	undue	.2430693	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
34	tepid	.2981482	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	
35	happy	.2715488	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	
36	clean	.3215	1	0	1	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	
37	itchy	.2776667	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	
38	feast	.3173267	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	
39	drive	.2976667	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	
40	prime	.2836667	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0	1	0	
41	axiom	.2465	1	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	
42	brave	.2457096	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
43	avert	.2681667	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	
44	glyph	.2575	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	1	0	0	0	
45	there	.3038333	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	1	0	
46	baker	.2205	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	
47	snarl	.2455	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	
48	maple	.2747475	1	0	0	0	1	0	0	0	0	0	0	1	1	0	0	1	0	0	0	
49	inane	.262	1	0	0	0	1	0	0	0	1	0	0	0	0	2	0	0	0	0	0	
50	valet	.2558922	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
51	medal	.2626667	1	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	

这是每个单词含有字母的表的一部分（变量中每个字母数字n表现该字母在单词中出现且仅仅出现n-1次，比如A1表示A在单词中出现0次是否为真，当然每个单词在）

多元线性回归分析的结果如下：

```
. regress difficulty A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3 F4 G1
>  G2 G3 H1 H2 H3 I1 I2 I3 J1 J2 K1 K2 K3 L1 L2 L3 M1 M2 M3 M4 N1 N2 N3 O1 O2 O3
>  P1 P2 P3 Q1 Q2 R1 R2 R3 S1 S2 S3 T1 T2 T3 U1 U2 U3 V1 V2 V3 W1 W2 X1 X2 Y1 Y2
>  Y3 Z1 Z2
note: A2 omitted because of collinearity.
note: B2 omitted because of collinearity.
note: C3 omitted because of collinearity.
note: D2 omitted because of collinearity.
note: E2 omitted because of collinearity.
note: F4 omitted because of collinearity.
note: G3 omitted because of collinearity.
note: H3 omitted because of collinearity.
note: I2 omitted because of collinearity.
note: J2 omitted because of collinearity.
note: K3 omitted because of collinearity.
note: L3 omitted because of collinearity.
note: M4 omitted because of collinearity.
note: N3 omitted because of collinearity.
note: O3 omitted because of collinearity.
note: P2 omitted because of collinearity.
note: Q2 omitted because of collinearity.
note: R2 omitted because of collinearity.
note: S2 omitted because of collinearity.
note: T3 omitted because of collinearity.
note: U2 omitted because of collinearity.
note: V3 omitted because of collinearity.
note: W1 omitted because of collinearity.
note: X2 omitted because of collinearity.
note: Y3 omitted because of collinearity.
note: Z2 omitted because of collinearity.
```

Source	SS	df	MS			
				Number of obs	=	359
				F(49, 309)	=	7.71
Model	.195513403	49	.003990069	Prob > F	=	0.0000
Residual	.15988158	309	.000517416	R-squared	=	0.5501
				Adj R-squared	=	0.4788
Total	.355394983	358	.000992723	Root MSE	=	.02275

Number of obs 指样本观测值数量，共359条数据观测值；F(49,309)=MSR/MSE，显示的是F检验-方差检验结果，为模型的全局检验，来表明拟合方程是否有意义，其中，49是回归自由度，即回归模型中没有误差的个数；309是残差自由度，即回归模型中有误差的个数；下面一条表示F检验的显著性，即F(49,309)在对应X值为7.71时的概率密度函数值，用Matlab计算得到约为2.8932e-30，基本为零，显然小于5%，因此F显然大于F(0.05表示下标)（49,309），表示显著的线性关系；R^2=SSR/SST 为相关系数R的平方，值在0-1之间，表示模型的拟合优度，越大说明模型预测越准；Adj R^2则表示的为调整后的拟合优度，同样越大表示模型预测越准确；这里R^2大于0.5，调整R方略小于0.5，表现这一多元线性回归分析有一定准确度；Root MSE表示残差标准差。

SS列对应误差平方和，第一行回归平方和SSR，第二行剩余平方和SST；df列指自由度；MS指均方差，第一行MSR回归均方差，第二行剩余均方差MSE

difficulty	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
A1	.0205611	.0171782	1.20	0.232	-.0132399	.0543621
A2	0	(omitted)				
A3	-.0543102	.0181821	-2.99	0.003	-.0900866	-.0185339
B1	.0347145	.0172881	2.01	0.046	.0006972	.0687317
B2	0	(omitted)				
B3	-.0537927	.0288082	-1.87	0.063	-.1104777	.0028924
C1	.1123758	.0351156	3.20	0.002	.0432799	.1814718
C2	.0837235	.0205567	4.07	0.000	.0432747	.1241724
C3	0	(omitted)				
D1	.0301092	.0169606	1.78	0.077	-.0032637	.0634821
D2	0	(omitted)				
D3	-.0727669	.0235333	-3.09	0.002	-.1190727	-.026461
E1	.0317639	.0168697	1.88	0.061	-.0014301	.0649579
E2	0	(omitted)				
E3	-.0377803	.0176095	-2.15	0.033	-.07243	-.0031307
F1	.1542821	.0553019	2.79	0.006	.0454661	.2630981
F2	.1061687	.0409187	2.59	0.010	.0256541	.1866833
F3	.0728812	.03652	2.00	0.047	.0010218	.1447405
F4	0	(omitted)				
G1	.0867166	.0355053	2.44	0.015	.0168539	.1565794
G2	.0384305	.0209419	1.84	0.067	-.0027763	.0796373
G3	0	(omitted)				
H1	.1295868	.0408335	3.17	0.002	.0492399	.2099336
H2	.1005918	.0286923	3.51	0.001	.0441348	.1570488
H3	0	(omitted)				
I1	.019116	.0167843	1.14	0.256	-.0139099	.0521419
I2	0	(omitted)				
I3	-.101152	.0236016	-4.29	0.000	-.1475922	-.0547118
J1	.0677914	.0204779	3.31	0.001	.0274977	.1080851
J2	0	(omitted)				
K1	.1022619	.0369531	2.77	0.006	.0295504	.1749734
K2	.0583985	.0235535	2.48	0.014	.0120531	.104744
K3	0	(omitted)				
L1	.0910643	.0339	2.69	0.008	.0243603	.1577683
L2	.0629823	.018167	3.47	0.001	.0272356	.098729
L3	0	(omitted)				
M1	.1730772	.0552429	3.13	0.002	.0643774	.281777
M2	.1339015	.0409488	3.27	0.001	.0533277	.2144753
M3	.0754666	.0332795	2.27	0.024	.0099835	.1409496
M4	0	(omitted)				
N1	.0952367	.0371337	2.56	0.011	.0221699	.1683036
N2	.06663	.0233044	2.86	0.005	.0207747	.1124854
N3	0	(omitted)				
O1	.068346	.034037	2.01	0.046	.0013724	.1353196
O2	.0435462	.0181786	2.40	0.017	.0077766	.0793158
O3	0	(omitted)				
P1	.0342507	.0169931	2.02	0.045	.0008138	.0676876
P2	0	(omitted)				
P3	-.0220408	.0235297	-0.94	0.350	-.0683395	.0242579
Q1	.0508564	.0200383	2.54	0.012	.0114276	.0902852
Q2	0	(omitted)				
R1	.028336	.0167142	1.70	0.091	-.0045521	.0612241
R2	0	(omitted)				
R3	-.0900102	.0204867	-4.39	0.000	-.1303212	-.0496991
S1	.0270878	.0170838	1.59	0.114	-.0065276	.0607031
S2	0	(omitted)				
S3	-.0612114	.0203015	-3.02	0.003	-.1011581	-.0212648
T1	.0841784	.0342911	2.45	0.015	.0167048	.1516521
T2	.0659675	.0186427	3.54	0.000	.0292848	.1026502
T3	0	(omitted)				
U1	.0339299	.0169972	2.00	0.047	.0004849	.0673749
U2	0	(omitted)				
U3	-.0509212	.0215338	-2.36	0.019	-.0932926	-.0085499
V1	.0412502	.0439191	0.94	0.348	-.0451682	.1276685
V2	-.0200514	.0330763	-0.61	0.545	-.0851347	.045032
V3	0	(omitted)				
W1	0	(omitted)				
W2	-.0533145	.0170922	-3.12	0.002	-.0869463	-.0196827
X1	.0608612	.0187244	3.25	0.001	.0240178	.0977046
X2	0	(omitted)				
Y1	.1650411	.0410069	4.02	0.000	.0843531	.2457291
Y2	.1209142	.0291277	4.15	0.000	.0636005	.1782279
Y3	0	(omitted)				
Z1	.0780264	.019703	3.96	0.000	.0392574	.1167954
Z2	0	(omitted)				
_cons	-1.402361	.5711775	-2.46	0.015	-2.52625	-.2784716

P>|t|的值小于0.05表示在95%置信水平下回归系数显著异于零，Coefficient为该项在回归方程中对应的系数，Std.error表示该项的标准差

之后采用BP检验的方法对该回归做异方差假设检验，得到BP检验结果如下

```
    chi2(49) =  88.14
Prob > chi2 = 0.0005
```

原假设为扰动项不存在异方差。而假设检验的P值小于0.05，说明在95%的置信水平下拒绝原假设，即我们认为扰动项存在异方差。

因此我们需要使用“OLS + 稳健的标准误”方法来解决异方差问题，再对上述数据做回归

OLS+稳健的标准误方法得到的结果如下：

```
.inear regression                               Number of obs     =        359
                                                F(42, 309)        =          .
                                                Prob > F          =          .
                                                R-squared         =     0.5501
                                                Root MSE          =     .02275
```

可以看到和普通OLS相比R方和RMSE几乎没有变化

下面是对于每个参数在回归方程中的值，稳健的标准误，t检验，P>|t|的值和95%置信区间

difficulty Coefficient std. err. t P>t [95% conf. interval]

A1 .0205611 .0072743 2.83 0.005 .0062477 .0348746
A2 0 (omitted)
A3 -.0543102 .0080993 -6.71 0.000 -.070247 -.0383734
B1 .0347145 .0063509 5.47 0.000 .022218 .047211
B2 0 (omitted)
B3 -.0537927 .0077149 -6.97 0.000 -.0689732 -.0386122
C1 .1123758 .0219704 5.11 0.000 .0691452 .1556064
C2 .0837235 .0202657 4.13 0.000 .0438473 .1235998
C3 0 (omitted)
D1 .0301092 .0065085 4.63 0.000 .0173026 .0429158
D2 0 (omitted)
D3 -.0727669 .0169356 -4.30 0.000 -.1060906 -.0394431
E1 .0317639 .0061704 5.15 0.000 .0196225 .0439052
E2 0 (omitted)
E3 -.0377803 .0072819 -5.19 0.000 -.0521087 -.023452
F1 .1542821 .0178135 8.66 0.000 .1192311 .1893332
F2 .1061687 .0128015 8.29 0.000 .0809796 .1313578
F3 .0728812 .0063509 11.48 0.000 .0603847 .0853777
F4 0 (omitted)
G1 .0867166 .014244 6.09 0.000 .0586891 .1147442
G2 .0384305 .0103799 3.70 0.000 .0180063 .0588547
G3 0 (omitted)
H1 .1295868 .013025 9.95 0.000 .1039579 .1552156
H2 .1005918 .0079592 12.64 0.000 .0849307 .1162529
H3 0 (omitted)
I1 .019116 .0059603 3.21 0.001 .0073881 .0308439
I2 0 (omitted)
I3 -.101152 .0100087 -10.11 0.000 -.1208458 -.0814582
J1 .0677914 .0106184 6.38 0.000 .0468979 .0886848
J2 0 (omitted)
K1 .1022619 .0162611 6.29 0.000 .0702654 .1342584
K2 .0583985 .013651 4.28 0.000 .0315379 .0852592
K3 0 (omitted)
L1 .0910643 .0120744 7.54 0.000 .0673058 .1148228

L2	.0629823	.0077912	8.08	0.000	.0476517	.0783129
L3	0 (omitted)					
M1	.1730772	.0183338	9.44	0.000	.1370023	.2091521
M2	.1339015	.0135026	9.92	0.000	.1073327	.1604702
M3	.0754666	.0170013	4.44	0.000	.0420135	.1089196
M4	0 (omitted)					
N1	.0952367	.0142119	6.70	0.000	.0672723	.1232011
N2	.06663	.0101114	6.59	0.000	.046734	.086526
N3	0 (omitted)					
O1	.068346	.0123693	5.53	0.000	.0440073	.0926847
O2	.0435462	.0079	5.51	0.000	.0280016	.0590908
O3	0 (omitted)					
P1	.0342507	.0069628	4.92	0.000	.0205502	.0479512
P2	0 (omitted)					
P3	-.0220408	.0071304	-3.09	0.002	-.036071	-.0080106
Q1	.0508564	.0091991	5.53	0.000	.0327557	.0689572
Q2	0 (omitted)					
R1	.028336	.0059398	4.77	0.000	.0166483	.0400237
R2	0 (omitted)					
R3	-.0900102	.0347781	-2.59	0.010	-.158442	-.0215783
S1	.0270878	.0066941	4.05	0.000	.013916	.0402596
S2	0 (omitted)					
S3	-.0612114	.0084174	-7.27	0.000	-.0777741	-.0446488
T1	.0841784	.0174219	4.83	0.000	.0498979	.118459
T2	.0659675	.0146789	4.49	0.000	.0370842	.0948508
T3	0 (omitted)					
U1	.0339299	.0062752	5.41	0.000	.0215824	.0462774
U2	0 (omitted)					
U3	-.0509212	.0073824	-6.90	0.000	-.0654473	-.0363952
V1	.0412502	.0150458	2.74	0.006	.0116451	.0708553
V2	-.0200514	.0116073	-1.73	0.085	-.0428906	.0027879
V3	0 (omitted)					
W1	0 (omitted)					
W2	-.0533145	.0067034	-7.95	0.000	-.0665045	-.0401245
X1	.0608612	.0097988	6.21	0.000	.0415803	.0801421
X2	0 (omitted)					
Y1	.1650411	.0134511	12.27	0.000	.1385736	.1915085
Y2	.1209142	.0097975	12.34	0.000	.101636	.1401925
Y3	0 (omitted)					
Z1	.0780264	.0063426	12.30	0.000	.0655463	.0905065
Z2	0 (omitted)					
_cons	-1.402361	.1995849	-7.03	0.000	-1.795078	-1.009644

然后对该回归做多重共线性检验，通过计算每个变量的方差膨胀因子（VIF）实现，下表为结果

M1	258.07	0.003875
T1	183.04	0.005463
H1	179.62	0.005567
O1	178.13	0.005614
N1	174.29	0.005738
F1	162.50	0.006154
Y1	162.41	0.006157
L1	158.25	0.006319
M2	134.76	0.007421
C1	129.87	0.007700
G1	99.52	0.010048
H2	87.70	0.011403
F2	83.54	0.011970
V1	83.49	0.011978
Y2	80.84	0.012369
K1	79.09	0.012645
N2	67.54	0.014806
T2	52.26	0.019136
A1	49.71	0.020118
E1	49.04	0.020391
O2	47.45	0.021076
V2	45.52	0.021970
R1	44.62	0.022410
C2	42.42	0.023572
L2	41.04	0.024364
I1	38.79	0.025779
S1	35.99	0.027783
G2	32.08	0.031171
K2	30.37	0.032930
P1	28.63	0.034933
U1	28.27	0.035370
D1	24.72	0.040449
B1	15.88	0.062969
W2	15.52	0.064420
E3	10.78	0.092728
A3	6.21	0.160998
X1	5.30	0.188679
M3	4.26	0.234900
Q1	3.83	0.261360
Z1	3.70	0.270331
R3	3.21	0.311675

J1	3.21	0.311944
I3	3.20	0.312234
S3	3.15	0.317387
U3	2.67	0.375080
F3	2.57	0.389036
D3	2.13	0.469753
P3	2.13	0.469898
B3	1.60	0.625200

平均VIF为60.88，可以看到一大半的变量的VIF都大于10，有严重的多重共线性这些变量单个对总的估计是不准的

尝试使用向前和向后逐步回归分析解决这一问题

向前及结果：

```
regress difficulty A1 A3 B1 B3 C1 C2 D1 D3 E1 E3 F1 F2 F3 G1 G2 H1 H2 I1 I3 J1 K1 K2 L1 L2 M1 M2 M3 N1 N2 O1 O2 P1 P3 Q1 R1 R3 S1 S3 T1 T2 U1
U3 V1 V2 W2 X1 Y1 Y2 Z1 r pe(0.05)
```

（已经剔除完全多线性的变量）

Linear regression

Number of obs	359
F(25, 329)	.
Prob > F	.
R-squared	0.4371
Root MSE	.02466

difficulty	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
B3	-.0145594	.0045854	-3.18	0.002	-.0235797	-.005539
Z1	.0418259	.002821	14.83	0.000	.0362765	.0473753
F3	.0381667	4.37e-09	8.7e+06	0.000	.0381667	.0381667
M3	-.0235394	.0153794	-1.53	0.127	-.0537937	.006715
T2	.0178318	.0030405	5.86	0.000	.0118504	.0238131
I3	-.0657696	.0055542	-11.84	0.000	-.0766959	-.0548433
P3	.0177687	.0058707	3.03	0.003	.0062198	.0293177
Y1	.0917019	.0036391	25.20	0.000	.084543	.0988608
Y2	.0786059	.0044934	17.49	0.000	.0697665	.0874454
D3	-.0333746	.0138896	-2.40	0.017	-.0606982	-.006051
S3	-.0268284	.0055025	-4.88	0.000	-.037653	-.0160038
W2	-.0182123	.0040631	-4.48	0.000	-.0262053	-.0102193
G1	.0115603	.0029078	3.98	0.000	.0058401	.0172805
U3	-.0135587	.0063818	-2.12	0.034	-.026113	-.0010043
J1	.0315567	.0069909	4.51	0.000	.0178042	.0453091
V2	-.0514399	.0072782	-7.07	0.000	-.0657576	-.0371221
V1	-.0302668	.0054623	-5.54	0.000	-.0410123	-.0195213
X1	.0229313	.0075749	3.03	0.003	.0080299	.0378326
I1	-.0150586	.0033052	-4.56	0.000	-.0215606	-.0085566
A1	-.0138796	.0031255	-4.44	0.000	-.020028	-.0077312
A3	-.0226132	.007746	-2.92	0.004	-.0378512	-.0073753
O2	.0100841	.0028417	3.55	0.000	.004494	.0156743
S1	-.008885	.0032192	-2.76	0.006	-.0152178	-.0025523
N2	.0272166	.0068195	3.99	0.000	.0138013	.040632
N1	.021305	.0067688	3.15	0.002	.0079895	.0346206
F1	.0513366	.0036223	14.17	0.000	.0442109	.0584624
F2	.0407341	.0047435	8.59	0.000	.0314026	.0500656
L2	.0291774	.0047047	6.20	0.000	.0199223	.0384324
L1	.0229516	.0045558	5.04	0.000	.0139893	.0319138
_cons	.0201982	.0217834	0.93	0.354	-.0226541	.0630505

向后逐步线性回归的结果：

p = 0.0851 >= 0.0500 removing V2

Linear regression

Number of obs	359
F(43, 310)	.
Prob > F	.
R-squared	0.5496
Root MSE	.02272

difficulty	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]

A1	.0256613	.0073379	3.50	0.001	.0112228	.0400997
A3	-.0591343	.0087234	-6.78	0.000	-.0762988	-.0419698
B1	.039774	.0068638	5.79	0.000	.0262684	.0532796
B3	-.0583056	.0086404	-6.75	0.000	-.0753068	-.0413043
C1	.121974	.02303	5.30	0.000	.0766592	.1672889
C2	.0883842	.0206325	4.28	0.000	.0477868	.1289815
D1	.0347053	.0076946	4.51	0.000	.0195651	.0498454
D3	-.0777734	.0169684	-4.58	0.000	-.1111613	-.0443856
E1	.036713	.006718	5.46	0.000	.0234944	.0499315
E3	-.0428022	.0076564	-5.59	0.000	-.0578673	-.0277372
F1	.1690626	.0196099	8.62	0.000	.1304773	.207648
F2	.1161206	.0137698	8.43	0.000	.0890266	.1432146
F3	.0779407	.0068638	11.36	0.000	.0644351	.0914463
G1	.0964321	.0152787	6.31	0.000	.0663691	.1264952
G2	.0432982	.0106219	4.08	0.000	.022398	.0641985
H1	.1393506	.0143485	9.71	0.000	.1111178	.1675833
H2	.1054052	.0086019	12.25	0.000	.0884797	.1223307
I1	.0240472	.0066144	3.64	0.000	.0110323	.037062
I3	-.1009611	.0111745	-9.03	0.000	-.1229486	-.0789736
J1	.0726231	.0111583	6.51	0.000	.0506676	.0945786
K1	.1117874	.017593	6.35	0.000	.0771707	.1464041
K2	.06316	.0140667	4.49	0.000	.0354816	.0908384
L1	.1007703	.0135038	7.46	0.000	.0741997	.127341
L2	.0678582	.0083636	8.11	0.000	.0514016	.0843148
M1	.1878199	.0200208	9.38	0.000	.1484262	.2272137
M2	.1438289	.0142263	10.11	0.000	.1158366	.1718211
M3	.0805073	.0168956	4.76	0.000	.0472627	.1137519
N1	.1047833	.0157631	6.65	0.000	.0737672	.1357994
N2	.0711847	.0109911	6.48	0.000	.0495582	.0928112
O1	.0782774	.0133124	5.88	0.000	.0520833	.1044716
O2	.0484605	.0082346	5.89	0.000	.0322578	.0646633
P1	.0390819	.0075122	5.20	0.000	.0243006	.0538632
P3	-.0268408	.0080251	-3.34	0.001	-.0426315	-.0110502
Q1	.0557767	.0095884	5.82	0.000	.0369102	.0746432
R1	.0331508	.0067362	4.92	0.000	.0198964	.0464052
R3	-.0949135	.0345265	-2.75	0.006	-.1628495	-.0269776
S1	.0319981	.0073646	4.34	0.000	.0175071	.0464891
S3	-.0659402	.009117	-7.23	0.000	-.0838792	-.0480012
T1	.0938575	.018495	5.07	0.000	.0574658	.1302492
T2	.0707373	.0151134	4.68	0.000	.0409996	.1004751
U1	.0388088	.0069761	5.56	0.000	.0250823	.0525354
U3	-.0558873	.0077253	-7.23	0.000	-.0710878	-.0406867
V1	.0658864	.0093391	7.05	0.000	.0475103	.0842626
Z1	.0828152	.0072159	11.48	0.000	.0686168	.0970136
W2	-.0581001	.0073928	-7.86	0.000	-.0726466	-.0435536
X1	.06567	.0103507	6.34	0.000	.0453034	.0860366
Y1	.1747458	.0147022	11.89	0.000	.1458172	.2036745
Y2	.1257511	.010183	12.35	0.000	.1057146	.1457875
_cons	-1.582623	.2197729	-7.20	0.000	-2.015058	-1.150187

可以看到向前线性回归的R方明显变小，准确性降低；而向后线性回归的结果几乎没有变化，严重的多重共线性对所有变量的整体效应本身影响也并不算太大。因此采用OLS+稳健的标准误得到的结果作为26个字母在单词中出现的次数和单词本身困难程度的线性回归结果

所以说线性回归结果为每个单词所贡献的难度系数 Difficulty = A1×0.0205611+A3×0.0543102
B1×0.0347145
B3×0.0537927
C1×0.1123758
C2×0.0837235

D1×0.0301092
D3×-0.0727669
E1×0.0317639
E3×-0.0377803
F1×0.1542821
F2×0.1061687
F3×0.0728812
G1×0.0867166
G2×0.0384305
H1×0.1295868
H2×0.1005918
I1×0.019116
I3×-0.101152
J1×0.0677914
K1×0.1022619
K2×0.0583985
L1×0.0910643
L2×0.0629823
M1×0.1730772
M2×0.1339015
M3×0.0754666
N1×0.0952367
N2×0.06663
O1×0.068346
O2×0.0435462
P1×0.0342507
P3×-0.0220408
Q1×0.0508564
R1×0.028336
R3×-0.0900102
S1×0.0270878
S3×-0.0612114
T1×0.0841784
T2×0.0659675
U1×0.0339299
U3×-0.0509212
V1×0.0412502
V2×-0.0200514
W2×-0.0533145
X1×0.0608612
Y1×0.1650411
Y2×0.1209142
Z1×0.0780264
-1.402361

但是……常数都负一点几了，所有的五字母单词带进去都是负的啊……这……这模型废了……