Name: 林成鍇 Julius Ling Sing Kai

Department: Psychology 112

Student ID: D84085360

*** This assignment is performed with *Python* coding. For detailed display in *IPython Notebook*, click the link below.

https://colab.research.google.com/drive/1Y8EWqtfmqU0QhwzhK6rEfeb9Y5VZlmwE?usp=sharing

Homework 4 - Logistic Regression, Polynomial Regression & Surface Analysis

Dataset: 2010 World Values Survey, Taiwan （2010 世界價值觀調查-台灣）

Source: 中研院 SRDA https://srda.sinica.edu.tw/datasearch_detail.php?id=1220

The dataset contains 1238 observations and a total of 308 columns with the respective response on the questionnaire.

Outline

I. Logistic Regression (with Ridge Regularization) Analysis on the how *Personal Values* 個人價值觀 influence one's *Perceived Importance of Certain Qualities Held by the Future Generation* 認爲下一代所需具備品質之重要性.

II. Polynomial Regression and Surface Analysis on the interaction of one's Faith in Religion 宗教情懷 and Faith in Science 科學精神 in affecting one's *Frequency of Existentialistic Pondering* 思考人生意義之頻率.

Going on the extra miles with the result of the previous analyses, I question on the mechanism of Faith in Religion in providing "Reassurance" to the individuals by conducting Sentiment Analysis on the religious text, shown as below:

a) *Meditations* by Marcus Aurelius
b) *The Book of Mormon*
c) *The Gospel of Buddha*
d) *The King James Bible*
e) *The Quran*

Logistic Regression (with Ridge Regularization)

Sample size = 828

Variables of Interest:

Control variable

Edu 學歷: min = 1, max = 5. The higher the higher educational background.

Independent variable

Val_Creativity 重視創造力的程度: min = 1, max = 6. The higher the greater value of creativity.

Val_Tradition 重視傳統價值的程度: min = 1, max = 6. The higher the greater value of tradition.

Val_Achievement 重視成就的程度: min = 1, max = 6. The higher the greater value of achievement.

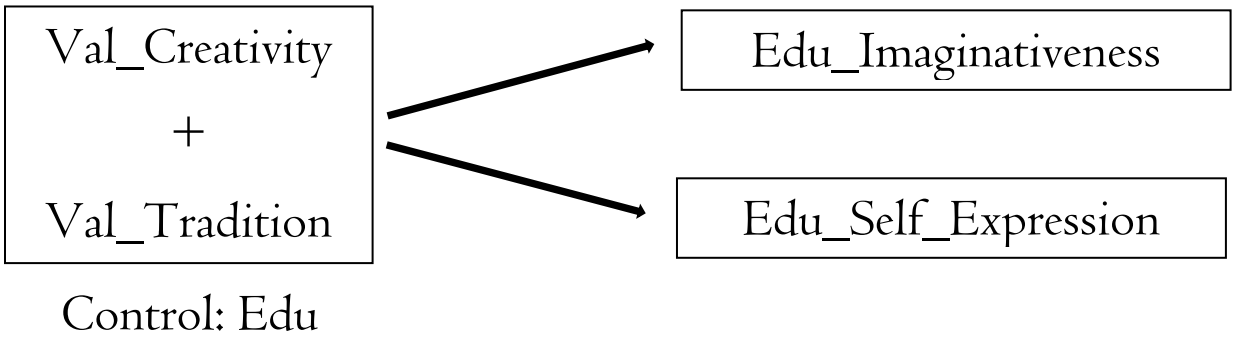
Val_Adventure 重視冒險精神的程度: min = 1, max = 6. The higher the greater value of adventure.

Dependent variable

Edu_Imaginativeness 認爲下一代想象力之重要性: min = 0, max = 1. A '1' represents a check.

Edu_Self_Expression 認爲下一代自我表達之重要性: min = 0, max = 1. A '1' represents a check.

Edu_Deligence 認爲下一代勤奮之重要性: min = 0, max = 1. A '1' represents a check.



Control: Edu

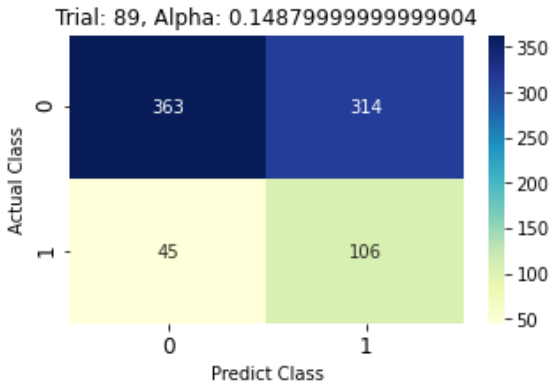
Hyperparameter alpha of the Ridge Regularization = 0.1488

Logit Regression Results

Dep. Variable:	Edu_Imaginativeness	No. Observations:	828
Model:	Logit	Df Residuals:	825
Method:	MLE	Df Model:	2
Date:	Wed, 18 Nov 2020	Pseudo R-squ.:	-0.4462
Time:	16:59:46	Log-Likelihood:	-568.74
converged:	True	LL-Null:	-393.27
Covariance Type:	nonrobust	LLR p-value:	1.000

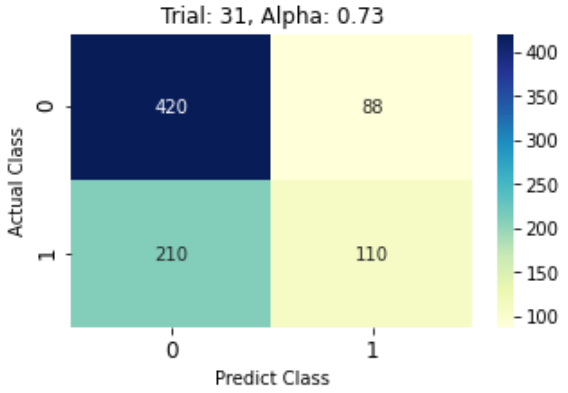
	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	0	nan	nan	nan	nan	nan
edu	0.2820	0.076	3.712	0.000	0.133	0.431
Val_Creativity	0.2531	0.074	3.397	0.001	0.107	0.399
Val_Tradition	-0.2462	0.075	-3.285	0.001	-0.393	-0.099

AUC score: 0.6190879121954083
Accuracy: 0.5664251207729468
Cohen's Kappa: 0.14076103923132943



Hyperparameter alpha of the Ridge Regularization = 0.73

```
                           Logit Regression Results
==============================================================================
Dep. Variable:     Edu_Self_Expression   No. Observations:                  828
Model:                           Logit   Df Residuals:                      824
Method:                            MLE   Df Model:                            3
Date:                 Wed, 18 Nov 2020   Pseudo R-squ.:                 0.06368
Time:                         17:21:30   Log-Likelihood:                -517.22
converged:                       True   LL-Null:                       -552.40
Covariance Type:             nonrobust   LLR p-value:                 3.587e-15
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.5023      0.075     -6.678      0.000      -0.650      -0.355
edu            0.4613      0.081      5.667      0.000       0.302       0.621
Val_Creativity 0.1384      0.077      1.789      0.074      -0.013       0.290
Val_Tradition -0.2553      0.076     -3.364      0.001      -0.404      -0.107
==============================================================================
AUC score:  0.5852608267716535
Accuracy:  0.6400966183574879
Cohen's Kappa:  0.18346437931856086
```

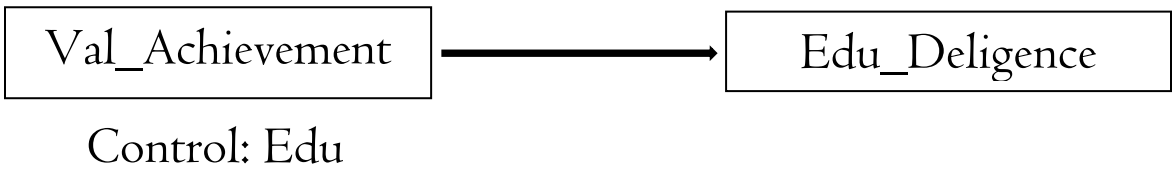


Here I'm facing a tradeoff in model selection between the pseudo-R-square and prediction accuracy. In order to obtain a well performance and more generalized model fitted, a ridge penalty is introduced and the highest-Cohen's-kappa model selection criterion is to have selected the model best fitted.

The models both indicate kappa higher than 0, indicating a tendency in terms of probability greater than mere random assignment. Hence, it's statistically valid to interpret the regression coefficients of both models as a relative impact in affecting the subjects' *Perceived Importance of Certain Qualities Held by the Future Generation.*

And in the first model we are to see the *Val_Creativity* contributes to higher chance of a tick in *Edu_Imaginativeness* while *Val_Tradition* acts on the opposite direction. Both are statistically significant. On the other hand, *Val_Tradition* seems statistically significant in contributing a higher chance of a tick in *Edu_Self_Expression* while *Val_Creativity* alleviates, however not reaching a statistical significance.

The control variable *Edu* shows positive impacts in *Edu_Imaginativeness* & *Edu_Self_Expression*.



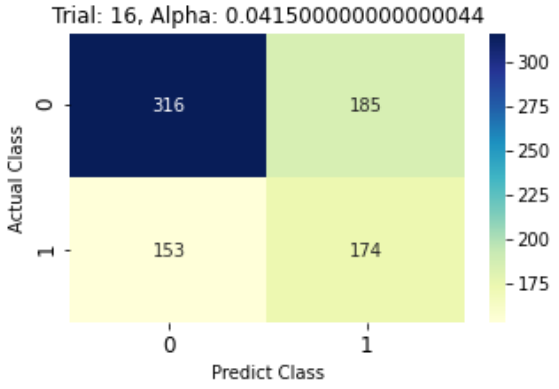
Hyperparameter alpha of the Ridge Regularization = 0.0415

Logit Regression Results

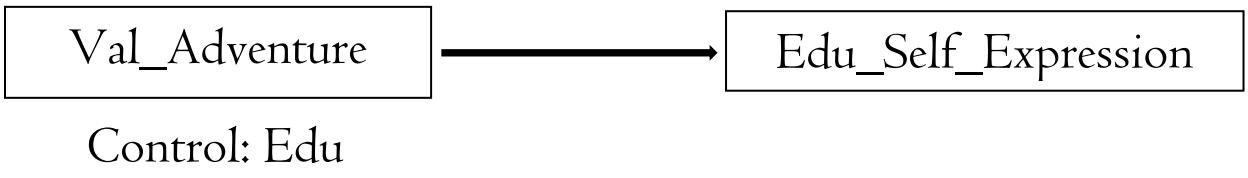
Dep. Variable:	Edu_Deligence	No. Observations:	828
Model:	Logit	Df Residuals:	826
Method:	MLE	Df Model:	1
Date:	Wed, 18 Nov 2020	Pseudo R-squ.:	0.003056
Time:	17:20:56	Log-Likelihood:	-553.81
converged:	True	LL-Null:	-555.51
Covariance Type:	nonrobust	LLR p-value:	0.06537

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	0	nan	nan	nan	nan	nan
edu	-0.4497	0.075	-6.019	0.000	-0.596	-0.303
Val_Achievement	0.2375	0.074	3.214	0.001	0.093	0.382

AUC score: 0.5814243073486055
Accuracy: 0.5917874396135265
Cohen's Kappa: 0.1601274825791814

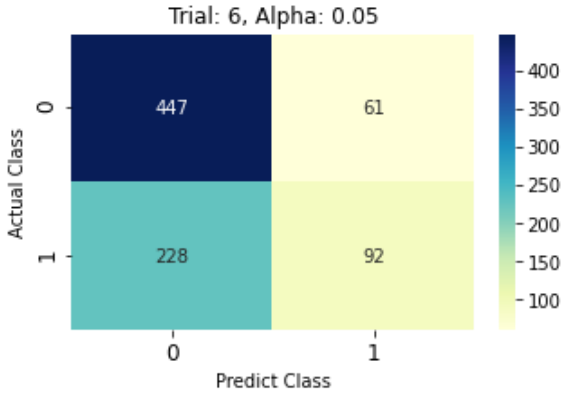


The model reveals that *Val_Achievement* contributes to higher chance of a tick in *Edu_Deligence*.. while the control variable *Edu* seems negatively correlate with *Edu_Deligence*, which is reasonable just as the old adage holds: Deligence is the cornerstone of Achievement.



Hyperparameter alpha of the Ridge Regularization = 0.0415

```
                           Logit Regression Results
==============================================================================
Dep. Variable:     Edu_Self_Expression   No. Observations:                  828
Model:                           Logit   Df Residuals:                      825
Method:                            MLE   Df Model:                            2
Date:                 Wed, 18 Nov 2020   Pseudo R-squ.:                0.05747
Time:                         17:31:16   Log-Likelihood:                -520.65
converged:                        True   LL-Null:                       -552.40
Covariance Type:            nonrobust   LLR p-value:                 1.630e-14
==============================================================================
                    coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept        -0.5040      0.075     -6.719      0.000      -0.651      -0.357
edu               0.5207      0.079      6.591      0.000       0.366       0.676
Val_Adventure     0.2166      0.074      2.932      0.003       0.072       0.361
==============================================================================
AUC score:  0.5837106299212599
Accuracy:  0.6509661835748792
Cohen's Kappa:  0.18531682804265237
```



The model reveals that *Val_Adventure* contributes to higher chance of a tick in *Edu_Self_Expression*.. while the control variable *Edu* seems positively correlate with *Edu_Self_Expression*. Reasonable as well, venturing out by daring to express oneself.

Surprisingly, *Edu_Imaginativeness*, and *Edu_Self_Expression* possess their own elitist nature as they're positively correlated with *Edu* (sounds a bit melancholic, at least to me).

Polynomial Regression & Surface Analysis

Sample size = 967

Variables of Interest:

Independent variable

Faith_Religion 宗教情懷: min = 1, max = 6. The higher the greater value of creativity.

Faith_Science 科學精神: min = 1, max = 6. The higher the greater value of tradition.

Dependent variable

Existentialistic_Ponder 思考人生意義之頻率: min = 0, max = 1. A '1' represents a check.

Model Setting:

Existentialism = β_0 + β_1 * Faith_Religion + β_2 * Faith_Science

+ β_3 * Faith_Religion^2 + β_4 * Faith_Religion*Faith_Science + β_5 * Faith_Science^2,

which β_0, β_1, β_2, β_3, β_4, β_5 are the regression coefficients.

OLS Regression Results

Dep. Variable:	Existentialism	R-squared:	0.031
Model:	OLS	Adj. R-squared:	0.026
Method:	Least Squares	F-statistic:	6.110
Date:	Mon, 16 Nov 2020	Prob (F-statistic):	1.40e-05
Time:	03:29:08	Log-Likelihood:	-1357.0
No. Observations:	967	AIC:	2726.
Df Residuals:	961	BIC:	2755.
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-4.142e-17	0.032	-1.3e-15	1.000	-0.062	0.062
Faith_Religion	0.1464	0.342	0.428	0.669	-0.526	0.818
Faith_Science	0.9018	0.315	2.860	0.004	0.283	1.520
Faith_Religion**2	0.3929	0.328	1.198	0.231	-0.251	1.037
Faith_Religion:Faith_Science	-0.7876	0.317	-2.482	0.013	-1.410	-0.165
Faith_Science**2	-0.2407	0.199	-1.212	0.226	-0.630	0.149

Omnibus:	48.305	Durbin-Watson:	2.086
Prob(Omnibus):	0.000	Jarque-Bera (JB):	52.997
Skew:	-0.556	Prob(JB):	3.10e-12
Kurtosis:	2.719	Cond. No.	26.4

95% confidence interval of a1, a2, a3 and a4

	lower_boundary	upper_boundary	
a1	0.122328	1.966503	(Statistically Significant)
a2	-1.440803	0.167392	
a3	-1.477964	-0.076844	(Statistically Significant)
a4	0.099155	1.839747	(Statistically Significant)

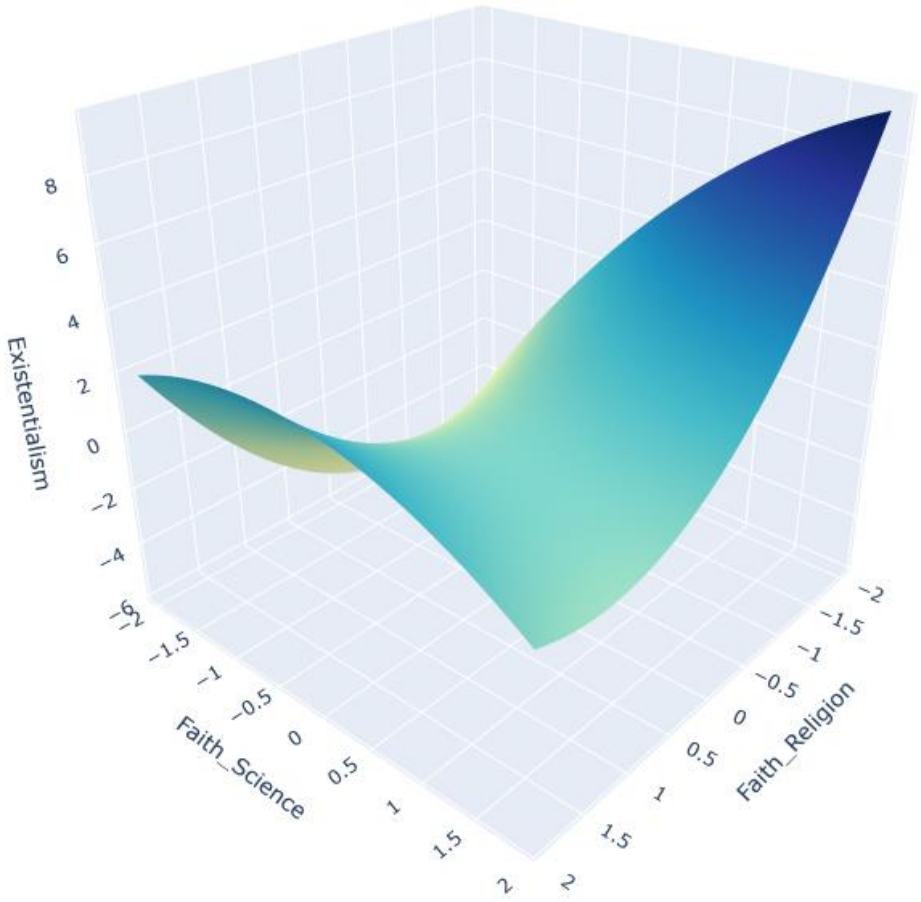
*** a1: linear effect along the line of congruency LOC (X = Y),

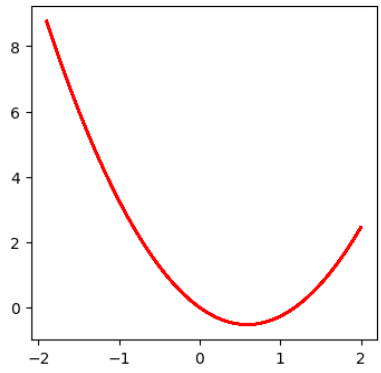
a2: curvilinear effect along the line of congruency LOC (X = Y),

a3: linear effect along the line of incongruency LOIC (X = -Y),

a4: curvilinear effect along the line of incongruency LOIC (X = -Y)

Surface Plot:





`Blue curve: Projection along the line of congruency LOC (X = Y)`

`Red curve: Projection along the line of incongruency LOIC (X = -Y)`

The linear effect along the LOC, linear effect along the LOIC & curvilinear effect along the LOIC is significant.

The surface plot itself tells an intriguing story: Faith_Religion positively correlates to Existentialistic_Ponder. In average, those with a relative match in *Faith_Science* & *Faith_Religion* have lower frequency of *Existentialistic_Ponder*, in comparison with those with a mismatch. Higher matching in both increases the frequency of *Existentialistic_Ponder* while the story gets complicated in case of mismatch in both. Both higher mismatch in both direction, whether high-*Faith_Science*-low-*Faith_Religion* or low-*Faith_Science*-high-*Faith_Religion*, contributes to higher frequency of *Exiatentialistic_Ponder*. A 'sweet spot' lies in between as the red curve shown.

Contemporary Existential Psychologist and Economist:

Clay Routledge

What makes life meaningful for theists and atheists? *Psychology of Religion and Spirituality.*

- Nelson, T. A., Abeyta, A. A., & Routledge, C. (2019).

Abstract

Meaning in life is a predictor of psychological and physical health, and religiosity has long been associated with meaning. However, little is known about how theists and atheists compare regarding their views on what gives life meaning. We conducted a study to explore similarities and differences between atheists and theists on perceived sources of meaning in life as well as overall perceptions of meaning. Participants responded to a writing prompt, and then their responses were analyzed by trained coders. Although both groups mentioned interpersonal relationships more than any other source of meaning, theists were more likely to reference social (e.g., relationships, parenting) and religious sources of meaning, whereas atheists were more likely to describe their lives as having no meaning and to mention sources of meaning that were not captured by the coded categories. Finally, theists scored significantly higher than atheists on the presence of meaning in life and the need for meaning but not on the search for meaning.

Faith_Religion seems to provide a sense of reassurance and meaning of life to one and hence reducing one's frequency in search of meaning. The story is consistent with the one of the previous literature.

*** The results and probable explanation of such phenomenon may sound thrilling at first. However a revisit in *Homework 5 – Confirmatory Factor Analysis (CFA)* invalidate such analyses. It's a mirage!

Below are my playland regarding the topic.

Enter if you please.

Sentiment Analysis of the Religious Text with NRC Emotional Lexicon

About NRC Emotional Lexicon:

The NRC Emotion Lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive).

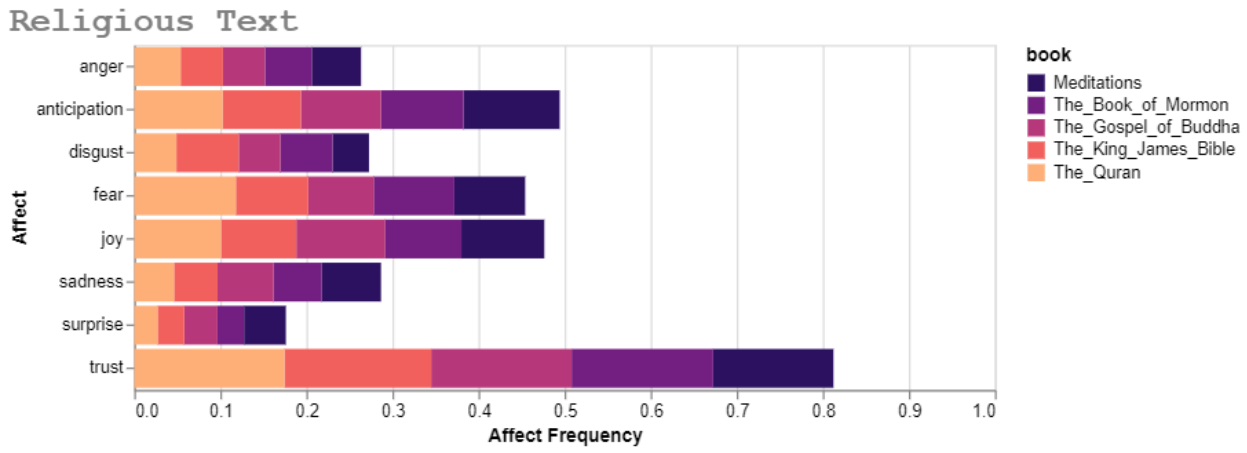
Eg. Sampling of the lexicons associated with 'Trust' within the dictionary

```
[['respectful'],
 ['victory'],
 ['cement'],
 ['armor'],
 ['antidote'],
 ['explain'],
 ['impenetrable'],
 ['seal'],
 ['thoughtful'],
 ['steward']]
```

```
[['expert']
 ['helper']
 ['credential']
 ['governess']
 ['guarded']
 ['sermon']
 ['enliven']
 ['generous']
 ['episcopal']
 ['reinforcement']]
```

```
[['elucidate']
 ['unity']
 ['fain']
 ['connoisseur']
 ['convince']
 ['morals']
 ['secrecy']
 ['institute']
 ['chaplain']
 ['substantiate']]
```

Sentiment Analysis:



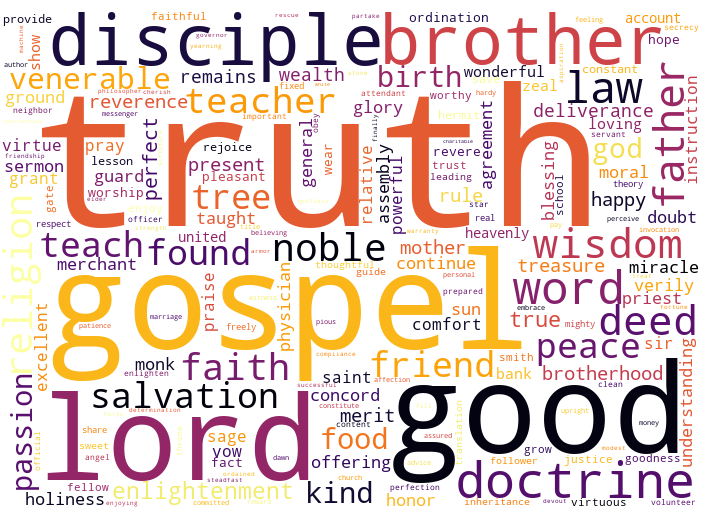
The score of "Trust" is notably high.

And there a curiosity to know the context of each religious text with respect to each affect of interest (Trust, Fear and Anticipation).

Word cloud visualization of each religious text:

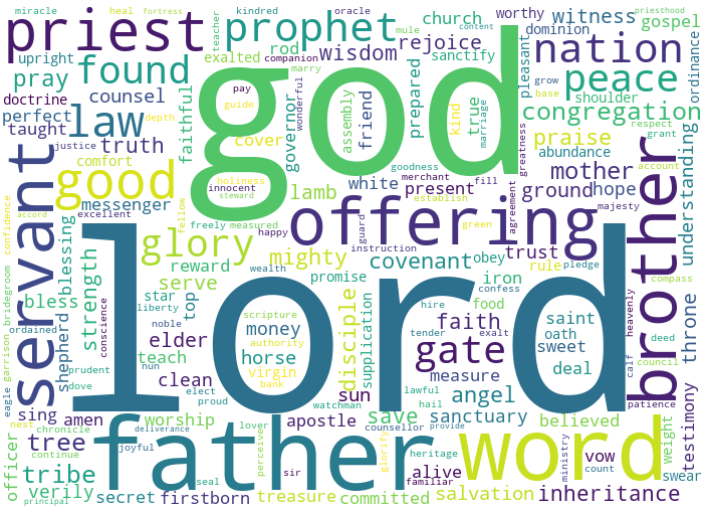
Under the context of Trust

Meditations

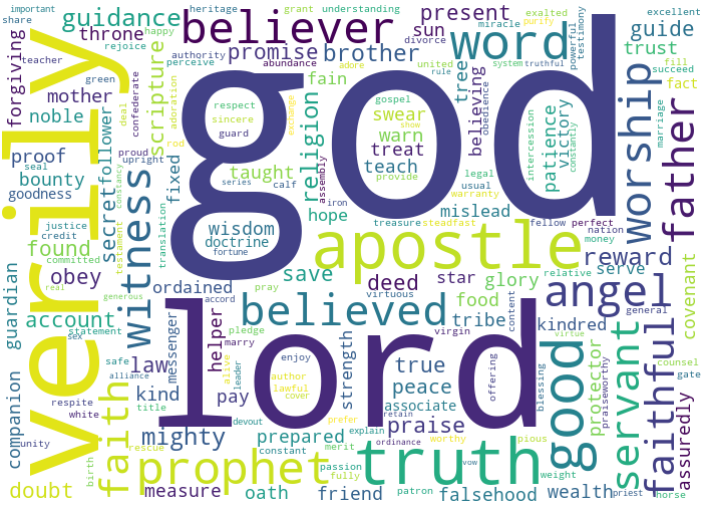


The Gospel of Buddha

The Book of Mormon



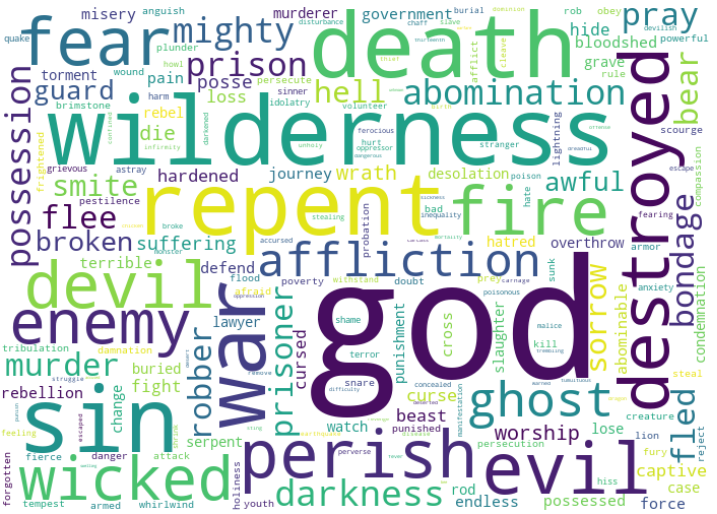
The King James Bible



The Quran

Under the context of Fear:

Meditations



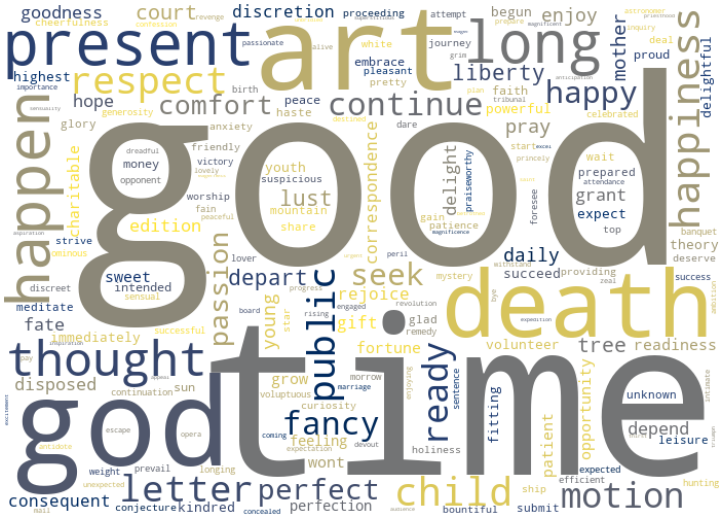
The Book of Mormon

The King James Bible

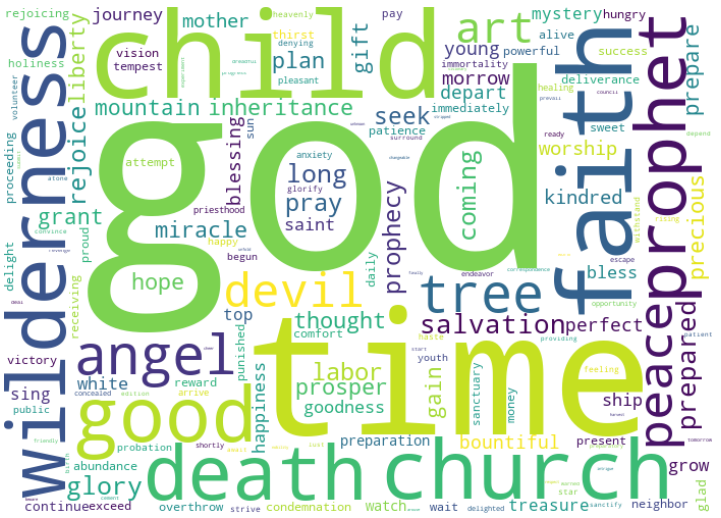
The Gospel of Buddha

The Quran

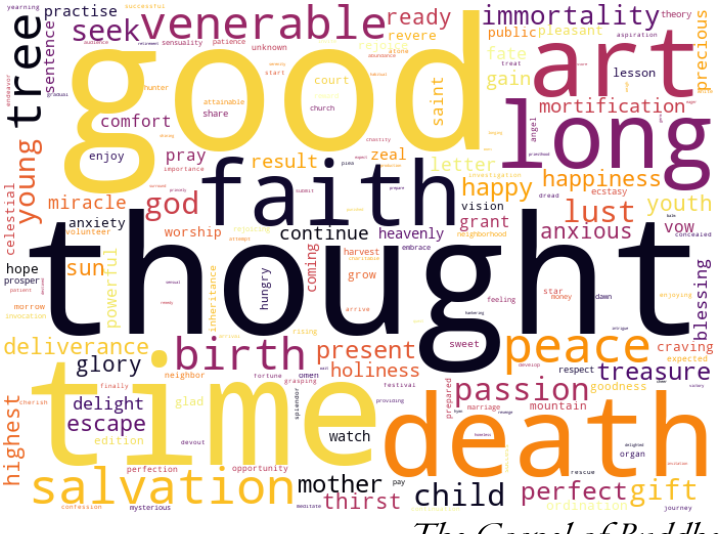
Under the context of Anticipation:



Meditations

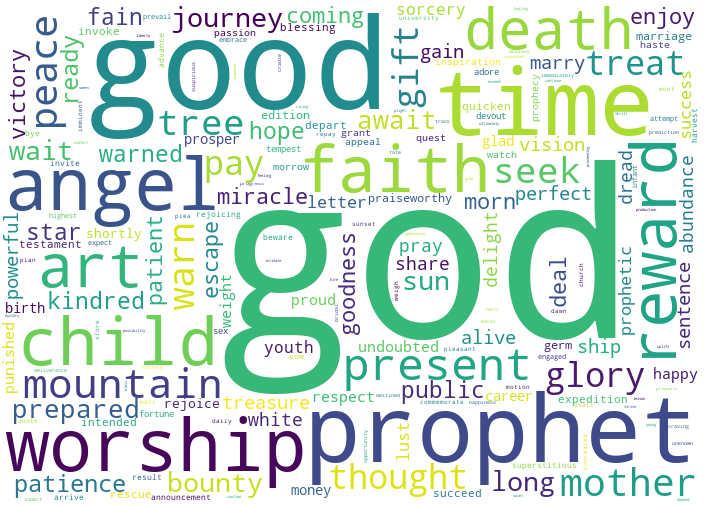


The Book of Mormon



The Gospel of Buddha

The King James Bible

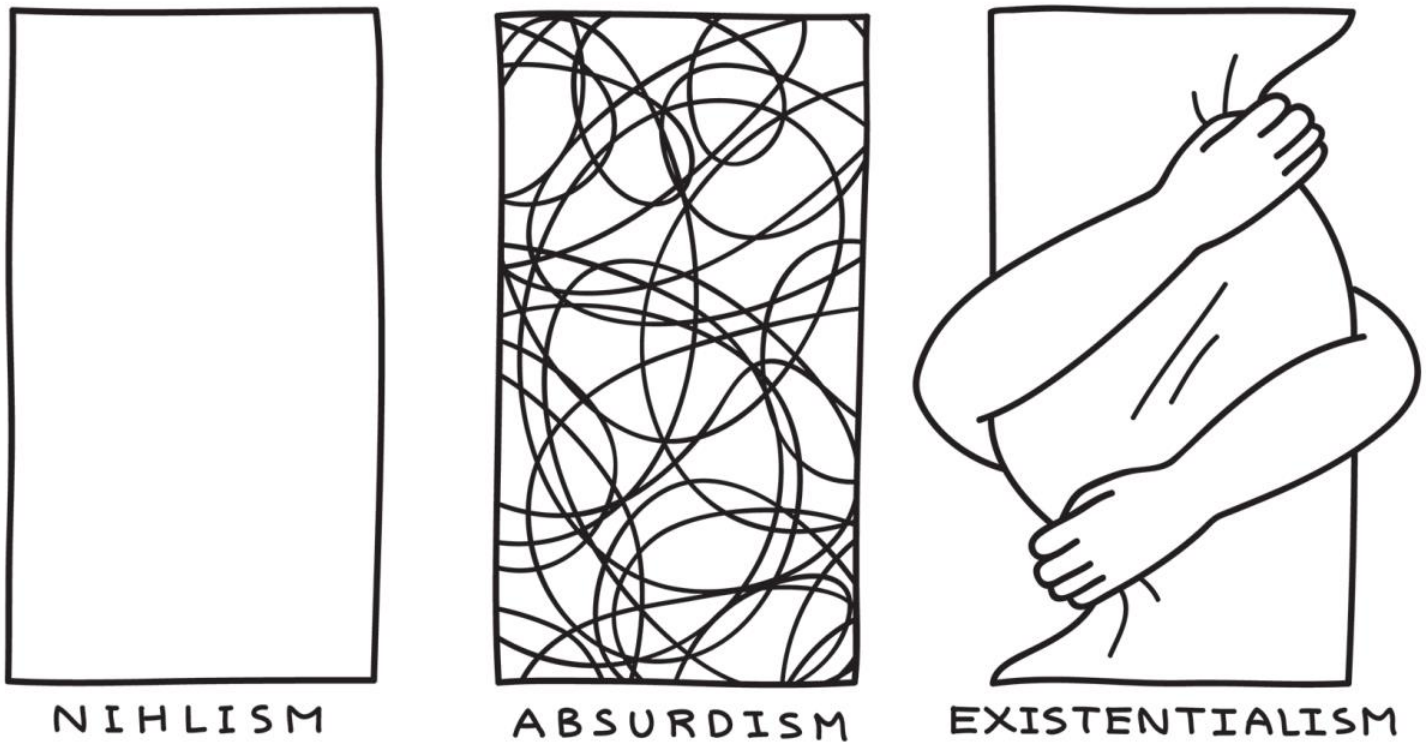


The Quran

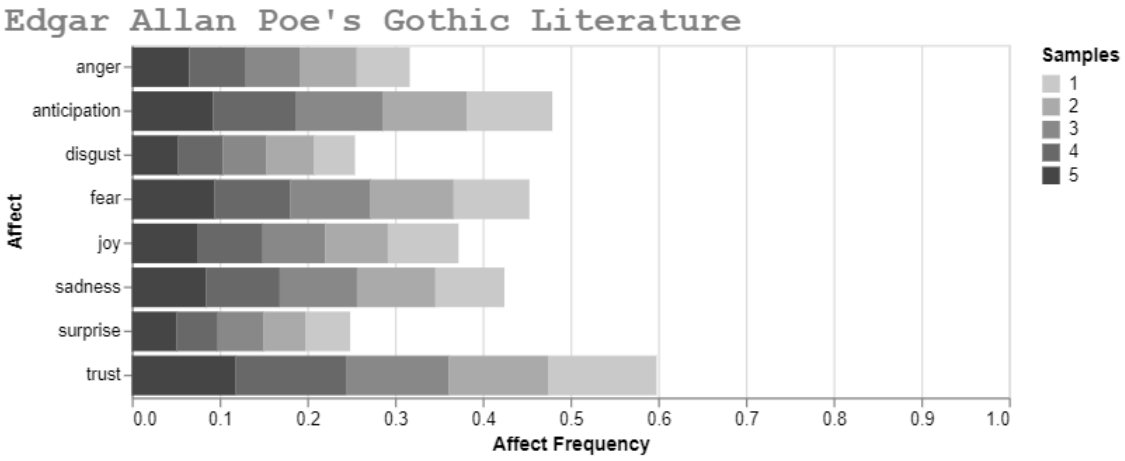
To be continued:

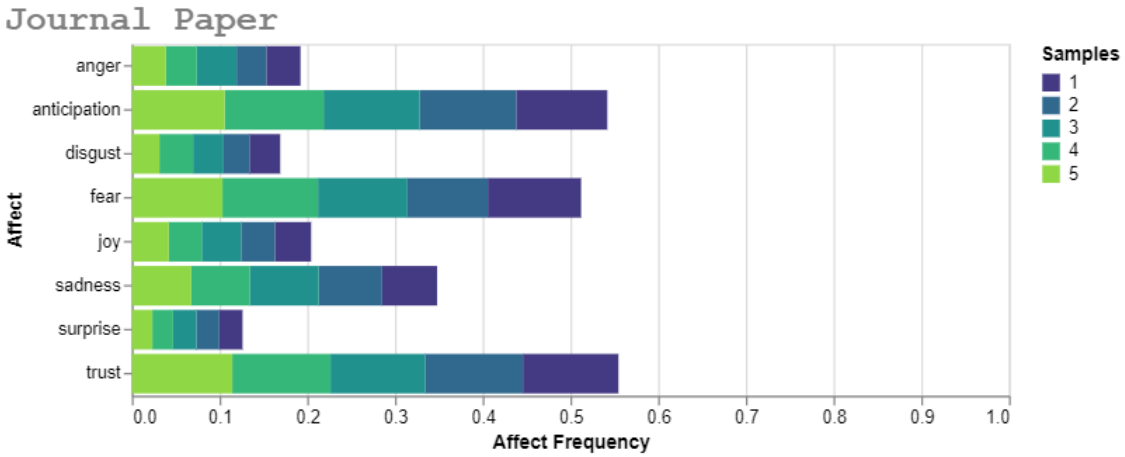
a. Polynomial Regression and Surface Analysis on the interaction of one's Faith in Religion 宗教情懷 and Faith in Science 科學精神 in affecting one's *Judgement 'bout a Certain Issue* 對某個議題的評斷 (Suicide 自殺, Euthanasia 安樂死, Homosexuality 同性戀, Sex Trade 性交易, Abortion 墮胎, Divorce 離婚, Premarital Sex 婚前性行爲
Controlled by different Religion (Christian, Buddhism, Folk Belief etc.)

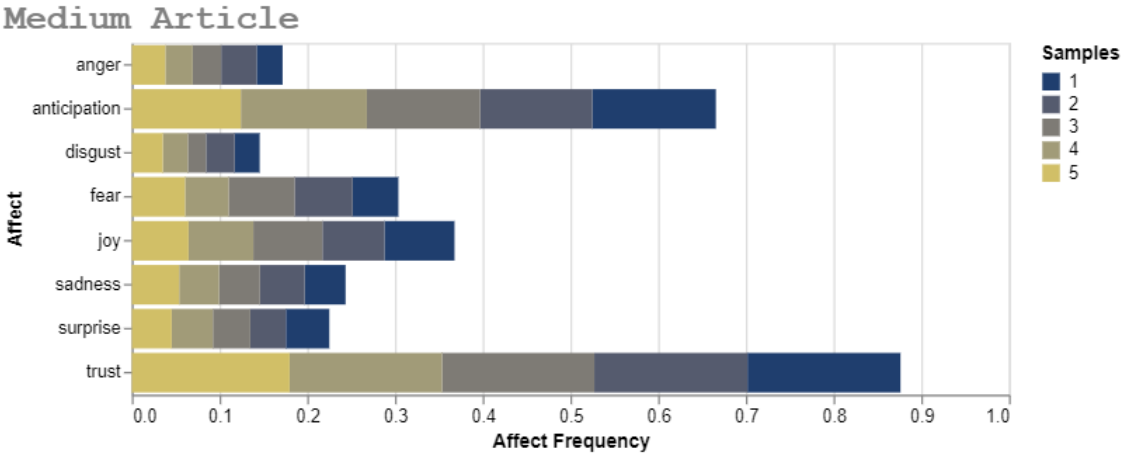
b. Text Mining on the Religious Texts used by ISIS, which play a key role in spreading ISIS ideology, propaganda, and recruitment. (Sample size = 2685)



Appendix: Sentiment Analysis with different corpus







It's hard to compare the results of each sentiment analyses with mere absolute value of affect frequency between each corpus as we believe that these corpuses are sampling from heterogenous population. I'm concerning more on the change and contrast between each affect within a corpus. Thus, this section is conducted aimlessly.

It's fun though, isn't it?