Name: 林成鍇 Julius Ling

Student ID: D84085360

Department: Psychology 112

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

https://colab.research.google.com/drive/1hXTv63QZHoVKjbu54YpSGPkqqvir-HXr?usp=sharing

Homework 5 - Confirmatory Factor Analysis (CFA)

The project aims to demolish (while it's assumes an enlightenment to me) what I've built up in

Homework 2 – Exploratory Factor Analysis (EFA) &

Homework - Logistic Regression, Polynomial Regression & Surface Analysis

Outline of such demolition, piece by piece

I. Homework 2 Revisit – A Comparison Between Exploratory & Confirmatory Factor Analysis.

II. Homework 4 Revisit – Justification of the Validity of Items Integration with CFA

Homework 2 Revisit – A Comparison Between Exploratory & Confirmatory Factor Analysis.

Dataset: Airline Passenger Satisfaction

This dataset contains an airline passenger satisfaction survey with the variables of interest shown as below:

- Gender: Gender of the passengers (Female, Male)
- Customer Type: The customer type (Loyal customer, disloyal customer)
- Age: The actual age of the passengers
- Type of Travel: Purpose of the flight of the passengers (Personal Travel, Business Travel)
- Class: Travel class in the plane of the passengers (Business, Eco, Eco Plus)
- Flight distance: The flight distance of this journey
- Inflight wifi service: Satisfaction level of the inflight wifi service (0: Not Applicable;1-5)
- Departure/Arrival time convenient: Satisfaction level of Departure/Arrival time convenient
- Ease of Online booking: Satisfaction level of online booking
- Gate location: Satisfaction level of Gate location
- Food and drink: Satisfaction level of Food and drink
- Online boarding: Satisfaction level of online boarding
- Seat comfort: Satisfaction level of Seat comfort
- Inflight entertainment: Satisfaction level of inflight entertainment
- On-board service: Satisfaction level of On-board service
- Leg room service: Satisfaction level of Leg room service
- Baggage handling: Satisfaction level of baggage handling
- Check-in service: Satisfaction level of Check-in service
- Inflight service: Satisfaction level of inflight service
- Cleanliness: Satisfaction level of Cleanliness
- Departure Delay in Minutes: Minutes delayed when departure
- Arrival Delay in Minutes: Minutes delayed when Arrival
- Satisfaction: Airline satisfaction level (Satisfaction, neutral or dissatisfaction)

It has 103,904 observations and 25 columns, with 14 of those representing customers responses, on a scale of 1 to 5, to a survey evaluating different aspects of the flights (Inflight wifi service, food and drink, online boarding, seat comfort, etc). These 14 columns will be very important for our upcoming factor analysis.

Recall in the first homework on the Exploratory Factor Analysis, the final result shown as below:

3 factors under promax (oblique) rotation (loading = 0.5 as a cutoff point):

	COMFORT	CONVENIENCE	SERVICE	TOTAL_COMMUNALITIES
INFLIGHT WIFI SERVICE	0.116614	0.715414	0.060798	0.529112
DEPARTURE/ARRIVAL TIME CONVENIENT	-0.04504	0.543998	0.035553	0.299226
EASE OF ONLINE BOOKING	-0.01174	0.89807	-0.03511	0.8079
GATE LOCATION	-0.02918	0.562047	-0.05571	0.319852
FOOD AND DRINK	0.792072	0.007425	-0.11396	0.640419
SEAT COMFORT	0.774135	0.000262	-0.05337	0.602132
INFLIGHT ENTERTAINMENT	0.747417	-0.00681	0.364857	0.691799
ON-BOARD SERVICE	0.022285	-0.0019	0.686404	0.471651
BAGGAGE HANDLING	-0.05069	0.0049	0.772885	0.599945
INFLIGHT SERVICE	-0.06249	-0.00506	0.826092	0.68636
CLEANLINESS	0.876884	-0.01212	-0.05935	0.772595
SS_LOADING	2.577808	1.930478	1.912705	
PROPORTIONAL_VARIANCE	0.234346	0.175498	0.173882	
CUMULATIVE_VARIANCE	0.234346	0.409844	0.583726	

At the end the oblique rotation method is applied to optimize the result.

With Loading = 0.5 as a cutoff point, we have the three factors nicely created:

1) Comfort: *Food and Drink, Seat Comfort, Inflight Entertainment, Cleanliness*
2) Service: *On-board Service, Baggage Handling, Inflight Service*
3) Convenience: *Inflight Wifi Service, Departure/Arrival Time Convenient, Ease of Online Booking, Gate Location*

3 variables are left out: *Online Boarding, Leg Room Service, Check In Service* (Note that their respective communalities are minute as well)

Though the total communalities of *Departure/Arrival time convenient* and *Gate location* are below 0.3, I think they are still meant to be preserved as their loading on the 3rd factor are quite big, big enough to lay my eyes on them. I think it's reasonable to cluster them under Factor 3.

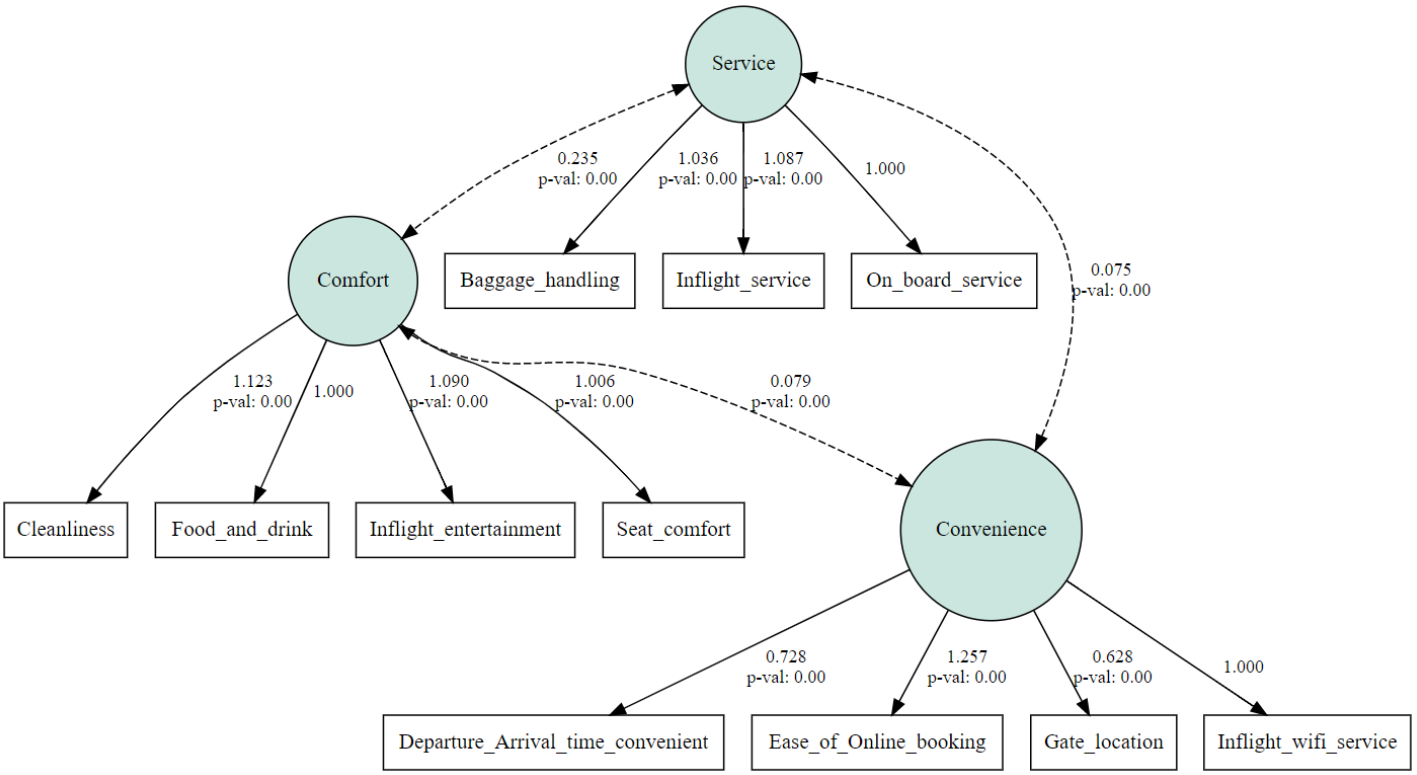
Cronbach's alpha of Factor 1 = `0.8762877916624099`

Cronbach's alpha of Factor 2 = `0.7942916933090214`

Cronbach's alpha of Factor 3 = `0.7679754211110683`

A Confirmatory Factor Analysis is conducted as a revisit on the case.

The CFA model setup shown as below:



Parameters estimation:

Factor	Item	Estimate	Std. Err	z-value	p-value
Comfort	Food_and_drink	1	-	-	-
	Seat_comfort	1.005532	0.004137	243.0597	0
	Inflight_entertainment	1.089933	0.004174	261.12	0
	Cleanliness	1.122604	0.004136	271.4113	0
Service	On_board_service	1	-	-	-
	Baggage_handling	1.035904	0.005341	193.9584	0
	Inflight_service	1.08708	0.005604	193.9843	0
Convenience	Inflight_wifi_service	1	-	-	-
	Departure_Arrival_time_convenient	0.728487	0.004792	152.0101	0
	Ease_of_Online_booking	1.256779	0.00568	221.2621	0
	Gate_location	0.628464	0.004009	156.7752	0

Various fit indices:

	Value
DoF	41.000
DoF Baseline	55.000
chi2	61011.147
chi2 p-value	0.000
chi2 Baseline	498306.944
CFI	0.878
GFI	0.878
AGFI	0.836
NFI	0.878
TLI	0.836
RMSEA	0.120
AIC	48.826
BIC	287.606
LogLik	0.587

Average Variance Extracted (AVE) & Construct Reliability (CR):

	Average Variance Extracted (AVE)	Construct Reliability (CR)
Comfort	0.642	0.877
Service	0.572	0.800
Convenience	0.483 < 0.5	0.777

It's notable that the overall fit of such measurement model is undesirable, though the construct reliability itself isn't bad at all.

Though the comparison between EFA & CFA is not fair in this case due to the disparity between the presumption and practical implication of two models. It's still notable that the result of CFA is not necessarily converges with EFA's as CFA we're using is based on the framework of Classical Test Theory (CTT), & such 'misfit' of CFA probably indicates the violation of the presumption in CTT (eg. Unidimensionality, as shown in EFA's that certain items have a cross-loading on more than one factor), in other words, such items maybe have assessed on more than one construct, which is my probable explanation on such CFA's misfit.

I'm some how reluctant to modify my CFA model in the effort obtaining a better fit (it makes no sense to me in terms of making a comparison with EFA). Personally, I'm feeling that exploration in such confirmatory analyses is more or less contradictory.

As proposed in *Exploratory Data Analysis – Chong Ho Yu* :

> *"EDA is not "fishing" or "torturing" the data set until it confesses."*

Thus, I'm resting my case in this revisit.

Homework 4 Revisit – Justification of the Validity of Items Integration with CFA

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.

Items of interest:

1. Religion_Insistence: 當科學和宗教衝突時候，宗教總是對的。
2. Religion_Only_Acceptable: 我信仰的宗教是唯一可接受的。
3. Religion_Edu_Diversity: 我們的公立學校應該教導各種宗教。
4. Religion_Relativity: 信仰其他宗教的，可能和信仰我的宗教的人一樣有道德。
5. Sci_Convenience: 科技是我們的生活更健康。方便，舒適。
6. Sci_Future_Generation_Dvlp: 因爲科技發達，下一代將會有更多發展機會。
7. Sci_Overdependence: 我們太過依賴科學而不夠依賴信仰。
8. Sci_Distortion_Val_Jgmt: 科學破壞人對於是非的看法。
9. Sci_Irrelevance: 瞭解科學對我的日常生活並不重要。

It's expected that the item (1)-(4) correspond to a factor —— Religious Excludability while item (5)-(9) correspond to another —— Scientific Prospect.

The result of CFA presented as below:

Factor	Item	Est. Std	Std. Err	z-value	p-value
Religious Excludability	Religion_Insistence	0.901385	-	-	-
	Religion_Only_Acceptable	0.501883	0.145649	3.822749	0.000132
	Religion_Edu_Diversity	-0.19521	0.063882	-3.39062	0.000697
	Religion_Relativity	-0.01559	0.03848	-0.44963	0.65298
Scientific Prospect	Sci_Convenience	1	-	-	-
	Sci_Future_Generation_Dvlp	0.559538	0.1159	4.827732	1.38E-06
	Sci_Overdependence	-0.11183	0.038455	-2.90808	0.003637
	Sci_Distortion_Val_Jgmt	0.001971	0.03133	0.062909	0.949839
	Sci_Irrelevance	0.177277	0.047237	3.753001	0.000175

Note that certain items have undesirable factor loading. Thus they are dropped.

CFA result with the unwanted items dropped:

Factor	Item	Est. Std	Std. Err	z-value	p-value
Religious Excludability	Religion_Insistence	0.733098	-	-	-
	Religion_Only_Acceptable	0.617125	0.39269	2.143822	0.032047
Scientific Prospect	Sci_Convenience	1	-	-	-
	Sci_Future_Generation_Dvlp	0.559604	0.306496	1.82593	0.067861

	Average Variance Extracted (AVE)	Construct Reliability (CR)
Religious Excludability	0.459	0.628
Scientific Prospect	0.657	0.780

	Value
DoF	1.000
DoF Baseline	6.000
chi2	11.883
chi2 p-value	0.001
chi2 Baseline	635.326
CFI	0.983
GFI	0.981
AGFI	0.888
NFI	0.981
TLI	0.896
RMSEA	0.103
AIC	17.977
BIC	62.316
LogLik	0.012

It's procedural invalid as too few items are included in each factors, and the overall fit of such CFA model isn't satisfying. However, here's a will to carry on conducting a polynomial regression analysis in spite of the procedural injustice.

Model Setting:

Existentialistic Pondering $= \beta_0 + \beta_1$ * Religious Excludability $+ \beta_2$ * Scientific Prospect

$+ \beta_3$ * Religious Excludability ^2

$+ \beta_4$ * Religious Excludability * Scientific Prospect

$+ \beta_5$ * Scientific Prospect ^2,

which $\beta_0,\ \beta_1,\ \beta_2, \beta_3,\ \beta_4,\ \beta_5$ are the regression coefficients.

OLS Regression Results

Dep. Variable:	Existentialistic_Pondering	R-squared:	0.025
Model:	OLS	Adj. R-squared:	0.020
Method:	Least Squares	F-statistic:	5.089
Date:	Wed, 30 Dec 2020	Prob (F-statistic):	0.000129
Time:	20:06:49	Log-Likelihood:	-1433.3
No. Observations:	1019	AIC:	2879.
Df Residuals:	1013	BIC:	2908.
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	8.197e-17	0.031	2.64e-15	1.000	-0.061	0.061
Religious_Excludability	-0.1644	0.033	-4.957	0.000	-0.230	-0.099
Scientific_Prospect	0.0020	0.034	0.059	0.953	-0.064	0.068
Religious_Excludability_Scientific_Prospect	0.0078	0.031	0.250	0.803	-0.053	0.069
Religious_Excludability_sq	0.0835	0.033	2.511	0.012	0.018	0.149
Scientific_Prospect_sq	-0.0001	0.034	-0.004	0.997	-0.066	0.066

Omnibus:	51.579	Durbin-Watson:	2.109
Prob(Omnibus):	0.000	Jarque-Bera (JB):	57.523
Skew:	-0.569	Prob(JB):	3.23e-13
Kurtosis:	2.759	Cond. No.	1.57

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

	lower_boundary	upper_boundary
a1	-0.245177	-0.081361
a2	-0.020899	0.204416
a3	-0.246110	-0.086411
a4	-0.192924	0.044160

*** 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)

And note that the curvilinear effect is no longer significant. In this case only the linear effect of Religious Excludability is significant.

The thrill is gone and the mirage fades away.

This homework is a regurgitation and correction on my progression within the courses. Such and such puzzling are intriguing. Hope my echo reverberates.

Guess I did much response within my homework. I'm resting my case this time.

Thank you & thousands appreciation.