

# DATA REPORT

## What influences salesmen's goal orientations?

Chi Zhang   Wei Zhou

### PART 1   Research Introduction

The existing research has concluded that relationship between supervisors' and sales reps' goal orientations can alter reps' role stress. Moreover, results reveal that firms can magnify the effects of supervisor-sales rep fit on innovation sales success by tying sales reps' variable compensation more closely to innovation sales.

What we want to do in this research is to step backward to estimate goal orientations with secondary data at hand, such as resumes or performance reports. It could be more efficient for the firms to combine the two methods together. To some extent, they could predict the performance from the resume or some other information that could be found easily.

To be specific, we want to find the relationship between the goal orientations and individual education, experience, gender, etc., through analyzing the resume and some survey data. The information of the resume has plenty of values for each variable. How to categorize and organize the information will be the key point. Therefore, the data cleaning will be the first step and very essential.

### PART 2   Description of Data

#### 1. Dependent Variables and Independent Variable

The dataset was collected in Germany, we translated the German information into English. The original information comes from resumes and some survey results of 175 sales people in Germany. We categorize the information in to the variables as listed below.

In this data set, we have three dependent variables.

Dependent Variables	Explanation	Extra Information
LEOR	Learning Orientation	Numeric scaled between 1-7
PAPO	Performance Orientation	Numeric scaled between 1-7
FAVO	Failure Avoidance Orientation	Numeric scaled between 1-7

We also have plenty of independent variables which could be separated into two categories. The first type of variables based on resumes or personal history. We call these variables Type 1, which is from Column 2-21 in the final dataset.

Independent Variables	Explanation	Extra Information
L_BE	Learned business or economics in education experience	Binary, yes-1, no-0
L_TE	Learned tech in education experience	Binary, yes-1, no-0
E_LEV	English level	Native-4, fluent-3, well-2, could speak but no info-1, couldn't-0
L_NUM	Number of languages he/she could speak	Integer with range 1-5
TR_Q	Has sales training in qualification	Binary, yes-1, no-0
TECH_Q	Has Tech as qualification	Binary, yes-1, no-0
THI_SK	whether consider thinking as a skill	Binary, yes-1, no-0
TEC_SK	whether consider tech as a skill	Binary, yes-1, no-0
INT_23	Interested in arts, literature and culture	Binary, yes-1, no-0

INT_14	Interested in sports or family life	Binary, yes-1, no-0
INT_56	Interested in science or technology	Binary, yes-1, no-0
INT_789	Interested in business, communication or learning	Binary, yes-1, no-0
WS	How long have you been working in sales?	Numeric with range 1-34
CCS	How long have you been working for the current company in sales	Numeric with range 0-32
WCC	How long have you been working for the current company	Numeric with range 0-174
TOP	Generally, I am among the TOP __% of the sales employees in my company	Numeric with range 1-6
J_LEV	Current job level	Numeric with range 1-5
GEN	Gender	Binary, 1 or 2?
GPA	GPA of highest education	Numeric with range -9-38

The second type of variables are related to selling behavior or consciousness which was collected from survey. We call these variables Type 2, which is from Column 25-44 in the final dataset.

Independent Variables	Explanation	Extra Information
ROLAM	role ambiguity with respect to innovation selling	Numeric with range -9-7
ROLCONF	role conflict with respect to innovation selling	Numeric with range -9-7
esROLAM	role ambiguity with respect to selling established solutions	Numeric with range -9-7
esROLCONF	role conflict with respect to selling established solutions	Numeric with range -9-7
custselect	I systematically select the customers I approach to sell innovations.	Numeric with range -9-7
escustselect	I systematically select the customers I approach to sell established solutions.	Numeric with range -9-7
hugecust	I focus on approaching large customers for selling our innovations.	Numeric with range -9-7
eshugecust	I focus on approaching large customers for selling our established solutions.	Numeric with range -9-7
newcust	I generally approach new customers for selling our innovations.	Numeric with range -9-7
esnewcust	I generally approach new customers for selling our established solutions.	Numeric with range -9-7
estabcust	I focus on approaching customers that I have a long relationship for selling our innovations.	Numeric with range -9-7
esestabcust	I focus on approaching customers that I have a long relationship for selling our established solutions.	Numeric with range -9-7
informalcust	? (no information)	Numeric with range -9-7
esinformalcust	? (no information)	Numeric with range -9-7
formalcust	? (no information)	Numeric with range -9-7
esformalcust	? (no information)	Numeric with range -9-7
MSSUCC	Average of: (1)Relative to competing products how successful are the innovations in terms of market share? (2)Relative to your company's objectives how successful are the innovations in terms of market share?	Numeric with range -9-7
PROSUCC	Average of: (1)Relative to competing products how successful are the innovations in terms of profits? (2)Relative to your company's objectives how successful are the innovations in terms of profits?	Numeric with range -9-7
esMSSUCC	Average of:	Numeric with range -9-7

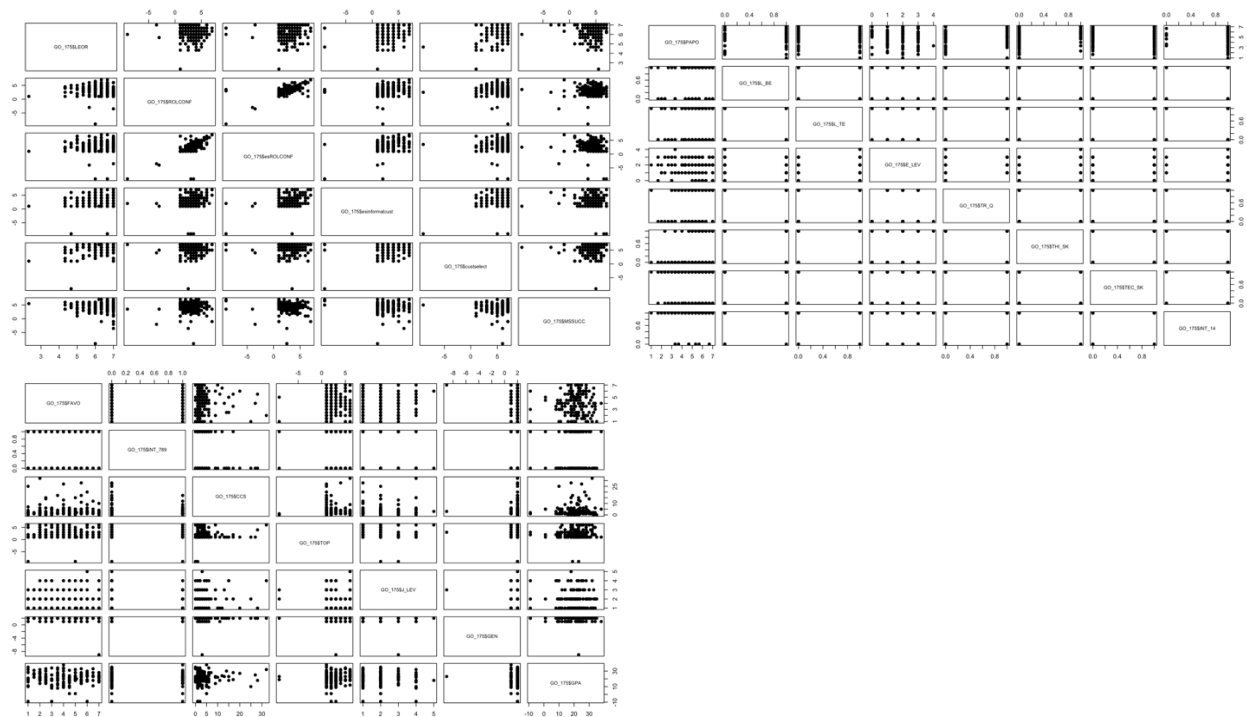
	(1)Relative to competing products how successful are the established solutions in terms of market share? (2)Relative to your company's objectives how successful are the established solutions in terms of market share?	
esPROSUCC	Average of : (1)Relative to competing products how successful are the established solutions in terms of profits? (2)Relative to your company's objectives how successful are the established solutions in terms of profits?	Numeric with range -9-7

## 2. Basic Description of Relationships between Variables

Firstly, because there are three dependent variables, we want to explore the correlation between each other. The conclusion is that the relationships are not strong enough in the result listed below.

	LEOR	PAP0	FAVO
LEOR	1.00000000	0.19676781	-0.08540722
PAP0	0.19676781	1.00000000	0.06208148
FAVO	-0.08540722	0.06208148	1.00000000

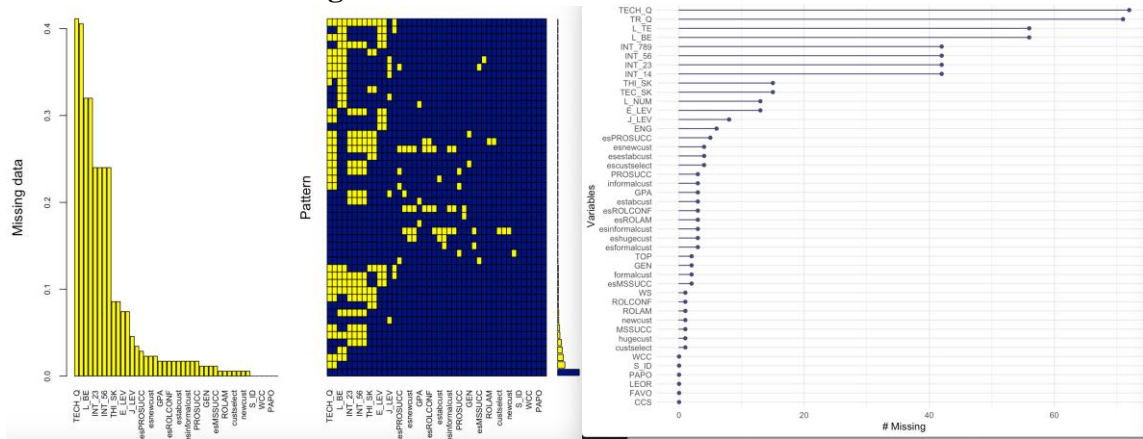
In order to get rough relationships of the variables, we make some plot through pairs function in R. Some independent variables and three dependent variables are described in the following graphs. However, those are not exhaustive.



In these graphs, there seems to be no obvious patterns between three dependent variables and independent variables. However, are they the total story of this dataset? We try to do further exploration through some analysis, such as PCA, SUR model, OLS.

## PART 3 Missing Data Problem

### 1. Visualize the Missing Data



### 2. Check the Mechanism of Missing Data

I use the classification system to check the missing data mechanism, by **splitting the data into two groups**. Then the **means** of three DV's are **compared between 118 complete observations (Group 1) and 57 uncomplete observations (Group 2)**. The differences, as below in the table, are not so big. Therefore, we could assume the missing data mechanism is **MCAR** (missing data complete at random).

mean (Group 1) – mean (Group 2)	-0.1198434
mean (Group 1) – mean (Group 2)	0.01233025
mean (Group 1) – mean (Group 2)	0.3133809

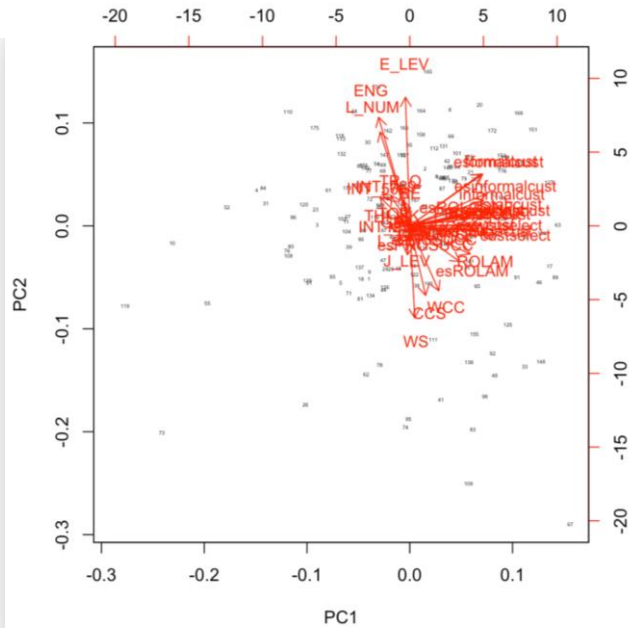
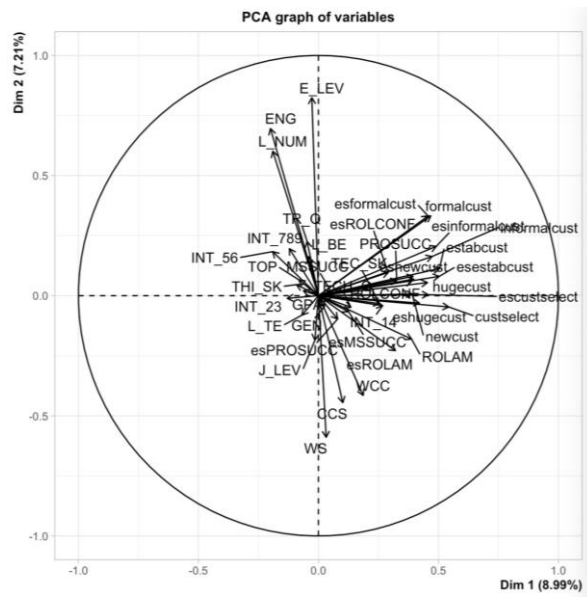
### 3. Multiple Imputation of Missing Data

MICE function in R is used to impute the missing data, whose logic is based on MAR. The condition of MCAR is stricter than MAR, thus the imputation still works here. Then I get a complete data set.

## PART 4 Data Analysis

### 1. Step 1: PCA Model

Based on the new complete data set, we do the PCA model to reduce the dimension of the dataset. The pictures as below are two results from prcomp function and PCA function.



Based on the results, we could ignore GPA, GEN, THI\_SK, L\_TE, TOP, TECH\_Q, esPROSUCC, esMSSUCC, MSSUCC, INT\_14, TEC\_SK, INT\_23, because of the low weight.

We could also ignore L\_NUM, TR\_Q, esformalcust, PROSUCC, esnewcust, esROLCONF, ROLCONF, hugecust, eshugecust, newcust, L\_BE, because the same direction with some IDV's but lower weight.

The IDV's for analysis are **E\_LEV, ENG, TEC\_SK, INT\_56, INT\_789, WS, CCS, WCC, J\_LEV, ROLAM, esROLAM, custselect, escustselect, estabcust, esestabcust, informalcust, esinformalcust, formalcust**

## 2. Step 2: SUR Model

We want to observe if there is any relationship between the equations, so we use SUR model. Two SUR models were conducted, but we only keep the better one based on the AIC BIC estimation. The three equations are not much related with each other. In addition, the SUR results confirm with the intuition to some extent.

### (1) Correlation between three equations

**The correlations of the residuals:**

	EQ11 (LEAO)	EQ22(PAPO)	EQ33(FAVO)
EQ11 (LEAO)	1.0000000	0.1999762	-0.0834598
EQ22(PAPO)	0.1999762	1.0000000	0.0759663
EQ33(FAVO)	-0.0834598	0.0759663	1.0000000

### (2) SUR Equation 1: Learning Orientation

SUR estimates for 'EQ11' (equation 1)

Model Formula: Y11 **LEAO** ~ X11

IDV's	Interpretation	Category	Estimate	Std. Error	t value	Pr(> t )
(Intercept)			4.18472661	0.40099546	10.43585	< 2e-16 ***
<b>X114 completeData\$ INT_56</b>	<b>Interested in science or technology</b>	Yes-1, No-0	<b>0.33233540</b>	<b>0.14014991</b>	<b>2.37129</b>	<b>0.018947</b> *
<b>X1112 completeData\$ custselect</b>	<b>I systematically select the customers I approach to sell innovations.</b>	Numeric with range 1-7	<b>0.08852874</b>	<b>0.03937295</b>	<b>2.24847</b>	<b>0.025947</b> *
Other variables are not significant, therefore, we don't put them in the table.						

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.715245 on 156 degrees of freedom

Number of observations: 175 Degrees of Freedom: 156

SSR: 79.805735 MSE: 0.511575 Root MSE: 0.715245

Multiple R-Squared: 0.21606 Adjusted R-Squared: 0.125605

### (3) SUR Equation 2: Performance Orientation

SUR estimates for 'EQ22' (equation 2)

Model Formula: Y22 **PAPO** ~ X22

IDV's	Interpretation	Category	Estimate	Std. Error	t value	Pr(> t )
(Intercept)			4.53383030	0.79382439	5.71138	5.4634e-08 ***
<b>X223 completeData\$ TEC_SK</b>	<b>whether consider tech as a skill</b>	Yes-1, No-0	<b>-0.48284249</b>	<b>0.22644052</b>	<b>-2.13231</b>	<b>0.034536</b> *
<b>X224 completeData\$ INT_56</b>	<b>Interested in science or technology</b>	Yes-1, No-0	<b>0.61347060</b>	<b>0.27709792</b>	<b>2.21391</b>	<b>0.028276</b> *
Other variables are not significant. Therefore, we don't put them in the table.						

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.415862 on 157 degrees of freedom

Number of observations: 175 Degrees of Freedom: 157

SSR: 314.732514 MSE: 2.004666 Root MSE: 1.415862

Multiple R-Squared: 0.091451 Adjusted R-Squared: -0.006927

### (4) SUR Equation 3: Failure Avoidance Orientation

SUR estimates for 'EQ33' (equation 3)

Model Formula: Y33(**FAVO**) ~ X33

IDV's	Interpretation	Category	Estimate	Std. Error	t value	Pr(> t )
-------	----------------	----------	----------	------------	---------	----------

(Intercept)			5.11237124	0.95848710	5.33379	3.3026e-07 ***
<b>X335 completeData\$INT_789</b>	<b>Interested in business, communication or learning</b>	Binary, yes-1, no-0	<b>-0.58087296</b>	<b>0.28384386</b>	<b>-2.04645</b>	<b>0.0423798 *</b>
<b>X336 completeData\$WS</b>	<b>How long have you been working in sales?</b>	Numeric with range 1-34	<b>-0.04613277</b>	<b>0.01729736</b>	<b>-2.66704</b>	<b>0.0084542 **</b>
Other variables are not significant. Therefore, we don't put them in the table.						

## PART 5 Interpretation and Limitation

### 1. Result Explanation

#### (1) For Learning Orientation:

If the salesman is interested in science or technology, there will be a significant 33.2% increase in the learning orientation scale value.

The level on which the sales person systematically select the customers to sell innovations increase by 1, there will be a significant 8.9% increase in the learning orientation scale value.

#### (2) For Performance Orientation:

If a sales person considers tech as a skill, there will be a significant 48.3% decrease in the performance orientation scale value.

If the salesman is interested in science or technology, there will be a significant 61.3% increase in the learning orientation scale value.

#### (3) The Failure Avoidance Orientation:

If a sales person is interested in business, communication or learning, there will be a significant 58.1% decrease in the failure avoidance orientation scale value.

If a person has one-year longer working experience in sales, there will be a significant 4.6% decrease in the failure avoidance orientation scale value.

### 2. Model and Analysis Limitation

The sample size is not big enough. In addition, there are too much missing data that might have negative effect on results and the model power. The multiple imputation usually works well for different variables of the same features, instead of broad different aspects. Thus, the multiple imputation might not accurate enough, and large number of data is needed. However, more content could be possibly teased out by using different methods based on the existing raw data.