What Really Went Wrong? A Comparison of Two Approaches to Penalty Analysis

#49 **Mike Gasho, Chrisly Philip, Alexa Williams** | PepsiCo, Inc, USA

PEPSICO Sensory and Consumer Product Insights

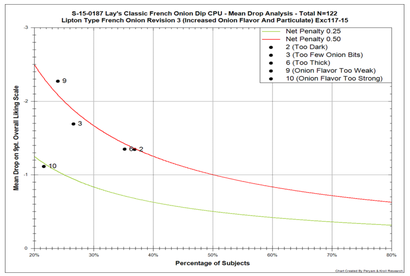
INTRODUCTION

Penalty Analysis, at its core, is an analysis strategy which guides the prioritization of product attributes developers might correct when optimizing future iterations of their product design. Higher priority is given to higher calculated penalties. Penalties are calculated using the relationship between liking data and JAR (Just About Right) data. In this research, two approaches to Penalty Analysis were compared to understand their relative strengths and weaknesses as a guidance method for product development. The first approach, Traditional, is arithmetical with graphical representation. On a per attribute ("one-at-a-time") basis, mean liking is compared between consumers scoring the product "just right" and those scoring it "not just right". The second approach is regression based. Multiple linear regression is conducted on the raw consumer data set to produce a simultaneous solution ("all attributes") to the penalty question.

METHODS

Traditional Penalty Analysis

(Assesses Penalties Arithmetically; Conclusions Displayed Graphically)



Total Penalty = (Mean Drop)(% Not-JAR)

Mean Drop is calculated by subtracting the average Overall Liking score of people who found a trait "Just About Right" from the OA of those finding the trait "Not Just about Right".

This is done on each attribute ("one-at-a-time")

Regression Penalty Analysis

(Liking Modelled via Multiple Linear Regression; Conclusions based on classical statistics rules)

```
The regression equation is
Overall Liking = 7.30 - 1.65 Overall Flavor TOO MUCH
                      - 0.994 Overall Flavor NOT ENOUGH
                      - 0.825 Flavor X NOT ENOUGH
                      + 0.148 Flavor Y NOT ENOUGH
                      - 0.716 Sweetness NOT ENOUGH
                      - 0.471 Saltiness NOT ENOUGH
                      + 0.508 Tartness NOT ENOUGH
                      - 0.105 Mouthfeel NOT ENOUGH
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	7.3048	0.2607	28.02	0.000	
Overall Flavor TOO MUCH	-1.6511	0.4096	-4.03	0.000	1.436
Overall Flavor NOT ENOUGH	-0.9936	0.4780	-2.09	0.040	1.558
Flavor X NOT ENOUGH	-0.8247	0.5376	-1.53	0.128	2.649
Flavor Y NOT ENOUGH	0.1478	0.5205	0.28	0.777	2.547
Sweetness NOT ENOUGH	-0.7157	0.3826	-1.87	0.064	1.180
Saltiness NOT ENOUGH	-0.4711	0.4256	-1.11	0.271	1.326
Tartness NOT ENOUGH	0.5076	0.4079	1.24	0.216	1.384
Mouthfeel NOT ENOUGH	-0.1054	0.3833	-0.28	0.784	1.371

Penalties are the coefficients from a dummy variable multiple linear regression of JAR attributes against Overall Liking

Dummy Variable Assignment

for i = 1, …, k, where k = number of JAR attributes, we have

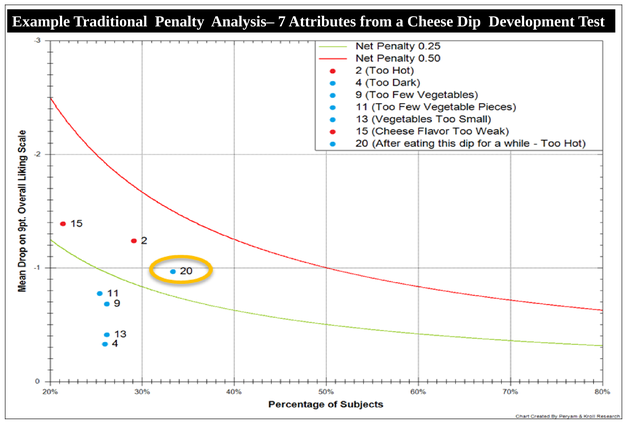
TOO MUCH(i) = 1, if JAR(***i***) response is 4 or 5; **0**, otherwise

NOT ENOUGH(i) = 1, if JAR(***i***) response is 1 or 2; **0**, otherwise

- For both approaches, attributes with less than 20% not-just-right scores were removed from consideration as penalties
- In total, 62 attributes from 14 sequential monadic test cells were analyzed using two penalty analysis approaches

RESULTS

Example Traditional Penalty Analysis– 7 Attributes from a Cheese Dip Development Test



Example Regression Penalty Analysis– Same Data Set As Above

	Regression Penalty Output					
	Coefficients	Std. Error	t Stat	P Value	V I F	Percent
Intercept	8.08	0.23	35.59	<0.0001		
For a Spicy or Hot Dip, the HEAT LEVEL of this dip is ... Too Hot	-0.83	0.40	-2.22	0.03	1.86	30%
The COLOR is... Too Dark	-0.01	0.33	-0.03	0.97	1.19	25%
The AMOUNT of VEGETABLES is ... Too Few	0.26	0.99	0.26	0.80	9.79	27%
The AMOUNT of VEGETABLE PIECES is ... Too Few	-0.74	0.94	-0.80	0.43	8.48	26%
The SIZE of the VEGETABLES is ... Too Small	-0.06	0.43	-0.15	0.88	1.97	27%
The CHEESE FLAVOR is ... Too Weak	-1.17	0.35	-3.35	0.00	1.14	22%
After eating this dip for a while, the HEAT LEVEL of this dip is ... Too Hot	-0.33	0.38	-0.86	0.39	1.93	34%

To compare the interpretation of penalty severity, the following definitions were devised.

	Traditional	Regression
Primary Penalty	Net Penalty ≥ 0.50	Coefficient significant with 95% C.L.
Secondary Penalty	Net Penalty between 0.25 and 0.50	Coefficient significant with 80% C.L.
Not Meaningful	Net Penalty < 0.25	Coefficient not significant

Cross Classification of Attribute Severity by Penalty Analysis Method

		Traditional Penalty Analysis		
		Primary	Secondary	Not Meaningful
Regression Penalty Analysis	Primary	3 (5%)	14 (23%)	5 (8%)
	Secondary	1 (1%)	6 (10%)	4 (6%)
	Not Meaningful	1 (1%)	9 (15%)	19 (31%)

46%	of the time characterizations agree
24%	of the time the severity of the penalty shifted (Secondary to Primary or Vice Versa)
14%	of the time Results that were "Not Meaningful" according to the Traditional Method are deemed as Primary or Secondary penalties according to Regression Penalty
16%	of the time attributes deemed as Primary or Secondary in the Traditional Method are 'Not Meaningful' according to Regression Penalty

Acknowledgements: Consumer studies were analyzed and cross tabulated by Nancy Kramer, Patti Wojnicz, and Juan Martinez of P&K Research

Key Takeaways

- Fewer numbers of meaningful penalties were expected for regression penalty analysis due to its consideration of attribute JAR correlation (significant terms in MLR solution are impacted by multicolinearity among attributes). However, the total number of attributes with meaningful penalties remained roughly the same between the two analyses.
- Variance Inflation Factors, a measure of much the variance of the estimated regression coefficients ("penalties") are inflated due to attribute correlation. This metric is unique to regression penalty analysis.
- A higher than expected percentage of disagreement in penalty severity between the two methods was observed. Nearly 30% of attributes with >20% Not-JAR scores were classified as "Meaningful" (Primary or Secondary) by one method and "Not Meaningful" by the other. This begged the question, 'which method is correct'?
- Comparison of client guidance (i.e. which attributes to prioritize for improvement) suggested from both quantitative penalty methods and open ended suggestions provided by consumers in the same test reveal a slightly better match for traditional penalty analysis. However, the advantage was not consistent enough to be definitive.

CONCLUSIONS

- Both methods to penalty analysis have value but can lead to different results, with neither method reflecting "best" post test guidance .
- However, as a result of these conclusions, an appreciation for the simplicity of traditional penalty analysis (ease of explanation to clients, graphical presentation of results) was gained. These "beyond analysis technique" features may make it advantaged for many Sensory Scientists (including the author of this poster).
- High Variance Inflation Factors calculated as part of regression penalty analysis (because they reflects attribute correlation) provides an interesting post-test diagnostic tool for questionnaires. (i.e. Were there superfluous questions? Were questions providing unique information?)