

ANALYZE RESPONSIBLY

Welcome to Cask Studies, where you can properly age your skills without getting old. Even sour grapes can become fine wines here.

October 17, 2014

Gregory Taketa's



CASK STUDIES

Pay Discrimination (Regression – Binary, Control Variables)

A fictional case study by Gregory Taketa. Non-data managers can briefly read this document to see how data analysis helps in hidden ways. Meanwhile, regression practitioners can hone using binary and control variables with a MS Excel Data Set to approach a realistic problem. Babs, an experienced office worker, believes that she and her female counterparts are victims of gross pay discrimination based on gender.

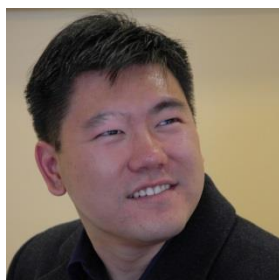
Babs: “I’m telling you, Gregory, there’s blatant gender discrimination at my workplace!”

Me: “What is your evidence of that, Babs?”

Babs: “I’ve been hearing about so many men at the office who are paid over \$50,000 a year while a number of experienced women are still being paid in the \$40,000s.”

Me: “I can understand that you perceive an injustice at first glance. Have you ruled out other factors for differences in pay, including education, hours worked, and certain results achieved?”

Babs: “Well, nobody gets paid higher for being more educated, since our office work does not require formal education to be excellent. Some have claimed greater experience as a justification



Gregory Taketa is the Data Decanter, serving refined, well-breathed data analysis while keeping out the sediments.

E-mail: greq@gregariousconsulting.com

See more of The Data Decanter at: <http://gregariousconsulting.com>

Copyright 2014 Gregory Taketa. All rights reserved.

to be paid more. We also have a scorecard filled out by the manager during our performance reviews.”

Eventually, Babs and I discuss possible key factors the management uses for pay. The manager agrees that these are meaningful factors in deciding salary, and a random sample of 30 employees (14 men & 16 women) is interviewed. The factors and some statistics are shown below:

Variable	The Logic Behind the Variable	Men's Average	Women's Average
Salary	This is the output variable and is the pre-tax total compensation for this year.	\$53,529	\$46,513
Merit (Composite Performance Score)	Higher weighted-average scores suggest more valuable results achieved in the eyes of management. The score assesses performance in terms of: <ul style="list-style-type: none"> • Efficiency: savings costs while meeting objectives. • Innovation: finding new ways to make a profit. • Interpersonal: being a “team player” in the department to sustain morale and overall efficiency. • Interdepartmental: working with other departments to help grow company profits. 	5.52 (out of 10)	5.16 (out of 10)
Hours/Week	Whether you are highly talented or almost highly talented, hard work is valued.	42.4	40.3
Years in This Position at the Company	Employee commitment and experience.	5.43	5.31
Years in That Profession	Experience and Skills overall.	8.43	8.38
# Raises	The more raises you were awarded, the higher your salary. Normally, the employee asks for a raise.	2.71	2.06

Manager: “Babs, I think it’s quite clear that there is no gender discrimination here. As you can see, the men on average have been achieving better results, working harder, holding more experience, and asking for more raises. Our company’s business analysts and general counsel have confirmed this after seeing these statistics.”

Babs: “Gregory, is he right? Or did you interview the wrong people?”

Me: “The sample is fine, Babs. Although these averages are as your manager and advisors say, I have a feeling you still have a case.”

Very quickly, I demonstrated a \$4,500 shortfall for each woman in the sample and gave the management a couple of easy suggestions to implement, including encouraging women to ask for more raises. Babs and her co-workers were thrilled to quickly receive a well-deserved \$72,000 collectively.

[Click Here to Download Data \(MS Excel 2007+\)](#)

Cask Questions:

- 1) Most data sets in real life tend to report qualitative data such as gender in text format (e.g. “Female,” “Male”). How would you change these data to work under a mathematical model such as regression?
- 2) Many data analysts would be satisfied to have only 1 input variable for their regression. In this case, they would simply use the gender variable. Why is this a poor practice for this case?
- 3) Although the “Merit” variable comes from subjective, ordinal data (ranked scores of 1-10), we often accept this as a necessity. What is the real problem behind the Merit score, and how could you quickly get management to mitigate that problem?
- 4) Do any of those variables seem redundant or conflicting with each other? What can you do to make the analysis easier?
- 5) Run a regression model using your own judgment.
 - a. What sorts of statistical diagnostics can you use to check that your model has satisfactory support?
 - b. What do you infer from your own findings?
 - c. What advice would you give this company based on your findings?

My own approach is provided in the latter red tabs of the Excel file. There is more than 1 way to legitimately approach the problem, and I do not claim to have the “perfect” or “the absolutely right” method. However, we can examine ourselves and determine whether we have a satisfactory model and inferences. Did you learn anything new from my own example?

- 6) How likely would your own business analysts agree with the manager upon seeing the statistics? What new value does the data analyst provide?

Skol!