Appendix: Identifying salary ranges for jobs relevant to the data cycle

Lars Vilhuber

2019-11-12

In the main text, we identify certain job descriptions, with associated salary ranges, from L (lowest) to VH (highest). This appendix identifies possible job titles and associated salary ranges observed in workplace and occupational surveys conducted by the Bureau of Labor Statistics (Bureau of Labor Statistics 2019b).

1 Data used

1.1 Occupational Employment Statistics

Collection methods, estimation methodology, and coverage are described in Bureau of Labor Statistics (2019c). We downloaded the data from https://www.bls.gov/oes/special.requests/oesm18nat.zip on 2019-10-30. From the downloaded data, we used `national_M2018_dl.xlsx`.

1.2 Occupational Information Network (O*NET)

The Occupational Information Network (O*NET) database comprises worker attributes and job characteristics. Information is collected using a two-stage design in which:

- a statistically random sample of businesses expected to employ workers in the targeted occupations is identified and
- a random sample of workers in those occupations within those businesses are selected. Data is collected by surveying job incumbents using a randomly assigned standardized questionnaire on occupation characteristics, out of three questionnaires. Additional questions cover tasks and demographic information.
- Abilities and Skills information is developed by occupational analysts using the updated information from incumbent workers (National Center for O*NET Development 2019a).

The resulting data covers several thousand occupations. A data dictionary (National Center for O*NET Development 2019b) provides additional information.

We downloaded version 23_2 of the data (National Center for O*NET Development 2019c) from https://www.onetcenter.org/dl_files/database/db_23_2_excel.zip on 2019-10-30. From the downloaded data, we used both Occupation Data.xlsx and Alternate Titles.xlsx.

2 Methods

2.1 Mapping job titles to SOC

O*Net is structured around Standard Occupational Categories (SOC) (Bureau of Labor Statistics 2019a). Our main text has a normative list of job description based on data management practiced at university libraries. These may not match reported standard occupation titles exactly. The O*Net data provides a long but not exhaustive list of alternate mentions of job titles for specific occupations (`Alternate Titles.xlsx`). Using both the standard occupation title as well as the alternate mention, we match the normative job title

via probabilistic matching, using the Jaro-Winkler distance (Winkler 1990) as implemented in the R package `fuzzyjoin` (Robinson 2019). We keep all *reasonable* matches ($d < 0.05$) to obtain a list of similar occupations and their SOC codes.

2.2 Mapping SOC into salary ranges

OES computes for each SOC code a salary range, comprised of annual salary and hourly wages, and characterized by the 25th and 75th percentile, as well as the median. We attach the annual salary distributions to each of the identified occupations (Table 1), and then collapse these statistics to a triplet of information for each normative job description (Table 2). To do so, we chose to use the minimum of all observed 25th percentiles, the median of all observed medians, and the maximum of all observed 75th percentiles. No weights were applied. An alternative implementation might use the employment shares to create weighted statistics. We do not attempt to compute reliability statistics, as the resulting table is meant to be indicative, not precise. Finally, Table 3 reports

3 Results

Table 1 (nlm.extract.csv) lists the annual salaries, as of 2018, by job title (median, and the 25% and 75% percentile), for all occupations identified as having similar names as the normative description. Blank salaries ("NA") indicate that no occupation code could be found on O*Net based on the normative description.

Job Title	Title	SC
Researcher	Industrial Ecologists	19-
Researcher	Anthropologists	19-
Researcher	Historians	19-
Researcher	Biofuels/Biodiesel Technology and Product Development Managers	11-
Researcher	Mathematicians	15-
Researcher	Chemical Engineers	17-
Researcher	Nanosystems Engineers	17-
Researcher	Manufacturing Engineering Technologists	17-
Researcher	Biologists	19-
Researcher	Biochemists and Biophysicists	19-
Researcher	Bioinformatics Scientists	19-
Researcher	Medical Scientists, Except Epidemiologists	19-
Researcher	Chemists	19-
Researcher	Hydrologists	19-
Researcher	Remote Sensing Scientists and Technologists	19-
Researcher	Geographers	19-
Data Librarian	Librarians	25-
Data Librarian	Library Science Teachers, Postsecondary	25-
Data Librarian	Archivists	25-
Metadata Librarian	Librarians	25-
Metadata Librarian	Library Science Teachers, Postsecondary	25-
Metadata Librarian	Archivists	25-
Records Management Specialist	Librarians	25-
Records Management Specialist	Library Science Teachers, Postsecondary	25-
Records Management Specialist	Archivists	25-
Curator	Curators	25-
Curator	Archivists	25-
Curator	Archeologists	19-
Research Domain Curator	Biofuels/Biodiesel Technology and Product Development Managers	11-

Job Title	Title	SC
Research Domain Curator	Mathematicians	15-
Research Domain Curator	Chemical Engineers	17-
Research Domain Curator	Nanosystems Engineers	17-
Research Domain Curator	Manufacturing Engineering Technologists	17-
Research Domain Curator	Biologists	19-
Research Domain Curator	Biochemists and Biophysicists	19-
Research Domain Curator	Bioinformatics Scientists	19-
Research Domain Curator	Medical Scientists, Except Epidemiologists	19-
Research Domain Curator	Chemists	19-
Research Domain Curator	Climate Change Analysts	19-
Research Domain Curator	Hydrologists	19-
Research Domain Curator	Remote Sensing Scientists and Technologists	19-
Research Domain Curator	Anthropologists	19-
Research Domain Curator	Geographers	19-
Research Domain Project Manager	Biofuels/Biodiesel Technology and Product Development Managers	11-
Research Domain Project Manager	Mathematicians	15-
Research Domain Project Manager	Chemical Engineers	17-
Research Domain Project Manager	Nanosystems Engineers	17-
Research Domain Project Manager	Manufacturing Engineering Technologists	17-
Research Domain Project Manager	Biologists	19-
Research Domain Project Manager	Biochemists and Biophysicists	19-
Research Domain Project Manager	Bioinformatics Scientists	19-
Research Domain Project Manager	Medical Scientists, Except Epidemiologists	19-
Research Domain Project Manager	Chemists	19-
Research Domain Project Manager	Climate Change Analysts	19-
Research Domain Project Manager	Hydrologists	19-
Research Domain Project Manager	Remote Sensing Scientists and Technologists	19-
Research Domain Project Manager	Anthropologists	19-
Research Domain Project Manager	Geographers	19-
Informatician	Computer Systems Analysts	15-
Informatician	Information Technology Project Managers	15-
Data Wrangler	Information Technology Project Managers	15-
Education Specialist	Health Educators	21-
Education Specialist	Special Education Teachers, Secondary School	25-
Education Specialist	Instructional Coordinators	25-
Communication Specialist	Public Relations Specialists	27-
Software Engineer	Computer and Information Research Scientists	15-
Software Engineer	Software Developers, Applications	15-
Software Engineer	Software Developers, Systems Software	15-
IT Security Specialist	Security Management Specialists	13-
IT Systems Engineer	Computer and Information Systems Managers	11-
IT Systems Engineer	Information Technology Project Managers	15-
IT Project Manager	Computer and Information Systems Managers	11-
IT Project Manager	Information Technology Project Managers	15-
Project Manager	Construction Managers	11-
Project Manager	Architectural and Engineering Managers	11-
Project Manager	Managers, All Other	11-
Project Manager	Information Technology Project Managers	15-
Project Manager	Environmental Engineers	17-
Project Manager	Wind Energy Engineers	17-
Project Manager	Environmental Restoration Planners	19-
Project Manager	Social Science Research Assistants	19-

Job Title	Title	SC
Project Manager	Remote Sensing Technicians	19-
Project Manager	Technical Directors/Managers	27-
Project Manager	Intelligence Analysts	33-
Senior Staff	NA	NA
Policy Specialist	NA	NA
Administrative Staff	First-Line Supervisors of Office and Administrative Support Workers	43-
Administrative Staff	Executive Secretaries and Executive Administrative Assistants	43-
Administrative Staff	Secretaries and Administrative Assistants, Except Legal, Medical, and Executive	43-
Administrative Staff	Business Operations Specialists, All Other	13-
Administrative Staff	Billing and Posting Clerks	43-
Administrative Staff	New Accounts Clerks	43-
Administrative Staff	Medical Secretaries	43-
Facilities Manager	General and Operations Managers	11-
Facilities Manager	Administrative Services Managers	11-
Facilities Manager	Property, Real Estate, and Community Association Managers	11-
Facilities Manager	First-Line Supervisors of Housekeeping and Janitorial Workers	37-
Facilities Manager	First-Line Supervisors of Office and Administrative Support Workers	43-
Facilities Manager	First-Line Supervisors of Mechanics, Installers, and Repairers	49-
Facilities Manager	Maintenance and Repair Workers, General	49-
Data Scientist	Computer and Information Research Scientists	15-

Table 2 (nlm.collapsed.csv)lists the ranges, as defined above, for each of the normative description, based on the underlying occupations identified.

Job Title	PCT25	MEDIAN	PCT75
Administrative Staff	28930	37800	94890
Communication Specialist	44490	60000	81550
Curator	38090	53780	80230
Data Librarian	38090	59050	90550
Data Scientist	91650	118370	149470
Data Wrangler	66410	90270	117070
Education Specialist	39800	60600	82860
Facilities Manager	29560	58340	157120
Informatician	66410	89505	117070
IT Project Manager	66410	116400	180190
IT Security Specialist	52200	70530	94890
IT Systems Engineer	66410	116400	180190
Metadata Librarian	38090	59050	90550
Policy Specialist	Inf	NA	-Inf
Project Manager	35450	87620	173180
Records Management Specialist	38090	59050	90550
Research Domain Curator	47500	80300	173180
Research Domain Project Manager	47500	80300	173180
Researcher	40670	79945	173180
Senior Staff	Inf	NA	-Inf
Software Engineer	79340	110000	149470

Table 3 (nlm.categories.csv)lists the statistics associated with each of the L-VH categories. While we defined the categories based on our own experience, ex ante, they match up well with observed median salaries in 2018.

Relative Salary	PCT25	MEDIAN	PCT75	N	Missing
L	28930	37800	94890	7	0
M	29560	61505	173180	34	0
H	40670	80300	180190	50	1
VH	52200	103620	180190	10	1

4 Full code and data

The code and data underlying this chapter, including an exhaustive list of our own edits (inclusions and exclusions) to the list of the occupations, is available at https://github.com/labordynamicsinstitute/job-description-and-wages and (DOI TBD).

References

Bureau of Labor Statistics. 2019a. "2018 Standard Occupational Classification System." https://www.bls.gov/soc/2018/major_groups.htm.

———. 2019b. "Occupational Employment Statistics." Dataset. Bureau of Labor Statistics, OES Program. https://www.bls.gov/oes/home.htm.

———. 2019c. "Survey Methods and Reliability Statement for the May 2018 Occupational Employment Statistics Survey." Bureau of Labor Statistics, OES Program. https://www.bls.gov/oes/current/methods_statement.pdf.

National Center for O*NET Development. 2019a. "O*NET Data Collection Overview." https://www.onetcenter.org/dataCollection.html.

———. 2019b. "O*NET® 23.2 Database." Data Dictionary. O*NET Resource Center. https://www.onetcenter.org/dl_files/database/db_23_2_dictionary.pdf.

———. 2019c. "O*NET® Database Release 23.2." Dataset. O*NET Resource Center. https://www.onetcenter.org/db_releases.html.

Robinson, David. 2019. *Fuzzyjoin: Join Tables Together on Inexact Matching*. https://github.com/dgrtwo/fuzzyjoin.

Winkler, William E. 1990. "String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage." *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 354–59. https://files.eric.ed.gov/fulltext/ED325505.pdf.