The San Francisco city government pays out a wide range of salaries to workers who perform many different jobs, including those who work part-time. This paper investigates the question "what factors determine how much a San Francisco city employee makes?" by using data from the SF Salaries records to examine how salaries vary for part-time and full-time workers as well as how they have changed from year to year.

The mean total salary for any San Francisco city employee in a given year is $74,768.32 with a 95 percent confidence interval of -$24,244.68 to $173,781.32 (1).[1] This confidence interval is clearly improbable; no employees are paying the San Francisco government thousands of dollars for their jobs (though Joe Lopez, a "Log Cabin Ranch Counselor," apparently did pay the city $618.13 in 2014) (2). The salary distribution underlying this distorted confidence interval is strongly skewed to the right due to a few employees who make especially large salaries (like Metropolitan Transit Authority General Manager Nataniel Ford, whose $567,595.43 2011 salary was the highest on record) and due to the inclusion of many workers, most of whom are likely part-time, who made less than $10,000 (2).[2]

The salary distribution for only full-time employees is more realistic though still skewed for the right, yielding a mean total salary for full-time employees of $103,505.76 with a 95 percent confidence interval of $23,690.61 to $183,320.92 (3). If only part-time employees are examined, the distribution is even more strongly skewed than that for all employees because of a large amount who make very little being counted alongside a handful who still make over

[1] #confidenceintervals: I correctly calculate a 95 percent confidence interval and explain why the underlying distribution makes it biased.
[2] #distributions: I recognize that the distribution is not normal, explain its effects the application of the confidence interval, and identify why it is skewed to the right.

$100,000. The total mean salary, therefore, is $35,806.99 with a 95 percent confidence interval of -$38,095.08 to $109,709.05 (4).

Both of these calculations, however, are only relevant for 2014; in the previous three years, employee status was not recorded (5). It is impossible, therefore, to represent the data for all given years completely accurately, since any distribution of total pay in 2011, 2012, or 2013 would misleadingly place those who work one day for the city alongside those who work 365. This also precludes including distinctions between part-time and full-time employees in an analysis of how mean salaries have changed over time, which I perform in the following paragraphs.[3]

Over the four years included in the data, the mean total salary for any employee has been steadily increasing, at a rate of approximately $1,239.94 a year (6). Between 2011 and 2014, therefore, the mean total salary for any employee increased by $3,719.81 (6). A difference of sample means hypothesis test (with a null hypothesis that there is no change between the two years and an alternative hypothesis that there is any change) found that this difference is statistically significant, with a p-value of approximately 0, far below the 0.05 alpha level (8). The effect size for this difference, however, is only 0.07 (8), suggesting that the increase from 2011 to 2014 has little practical significance (it likely merely adjusts for the rising cost of living in San Francisco).[4]

The yearly change in salary is slightly different, however, when it is separately examined for the employees who made the most and least. For the top ten percent of total salaries (those

[3] #induction: I identify a shortcoming of the data (employees were only split into part-time/full-time in 2014) and explain how it limits my analyses.
[4] #significance: I perform a difference of sample means hypothesis test and interpret the results, which reject my null hypothesis. I also calculate the effect size for the two means, which shows that the difference is not very practically significant, and suggest a practical reason for the difference.

above $142,489.28) the difference of sample means hypothesis test between the mean salary for 2011 and 2014 found another significant increase, with a p-value of approximately 0 (9). The effect size, however, was 0.17, making it larger than that for all salaries but still small. The bottom ten percent of total salaries (below $7,297.90), however, did not change significantly between 2011 and 2014 (9). The difference of sample means hypothesis test resulted in a p-value of 0.44, far above the 0.05 alpha cutoff (9). In fact, the lowest mean salaries actually decreased slightly, from $3,021.88 in 2011 to $2,984.45 in 2014 (9). This analysis suggests that the increase in mean total salary over the years was driven by those who are paid the most by the San Francisco government and rely on it as their main source of income. Those who do part-time work for the city have had no such adjustment in their salaries, and would presumably have to make up for rising costs of living with income from other sources.

The difference between these two income bands is further reflected in the conditional probability for San Francisco government employee being in the top or bottom 10 percent of earners given the year that they received their salary. An employee who received their salary in 2014 is very slightly more likely to be in the upper 10 percent of all salaries (11 percent (10)) than someone who was paid in 2011 (8 percent (11)). The conditional probabilities for an employee being in the bottom 10 percent of earners is almost the same for both years (approximately 10 percent (12, 13)).[5]

This analysis of how salaries have changed over time is limited, however, because only four years of data were recorded, making it difficult to infer general trends that could hold true across decades. Furthermore, there is no data for 2015, 2016, or 2017, preventing the

[5] #probability: I calculate conditional probabilities for a salary being in either the top or bottom 10% given that they are from 2011 or 2014 and explain how these probabilities relate to the broader analysis.

calculations from reflecting more recent changes and limiting the potential of generalizing trends to a contemporary context. A more fundamental shortcoming of the data is that it does not include information about demographic characteristics (age, race, gender, etc.) or even what department an employee worked for, both of which would have allowed for much richer analyses.[6] Despite these drawbacks, however, this analysis showed how salary distributions for all types of employees are skewed to the right to different degrees, as well as that there has been a significant increase in mean salaries over the four years provided, though higher salaries have increased more than lower ones.

Word Count: 991

[6] #induction: I discuss the factors that limit the relevance of my inferences, mainly the limited amount of data available, which considerably weaken any inferences I could draw.