Misha O'Keeffe
Econ 310: Statistics in Economics

Justification:

It's a man's world. With women accounting for 4% of Fortune 500 CEO's[1], 19% of Congress[1], and 38% of the world's wealth[2], there is no debate: the economy is male dominated. The question remains, however, how deeply ingrained is this gender inequality? Using publicly available data[3], this report will examine the 2021 fiscal year salaries of University of Wisconsin-Madison professors to determine if there is weight to this wage gap assumption.

Method:

Data was obtained using an online database[3] of 2021 fiscal-year UW-Madison faculty compensation. Observations were restricted to faculty whose title included "professor" and whose salary was above $28,100 (the minimum pay grade for 12-month full-time academic staff). Total observations were n=3,272. The relevant population is all academic staff at UW-Madison. To clean the data, several measures were performed. First, total pay was converted to a standard numerical format for analysis purposes. Additionally, for each participant, time spent with the University was calculated using their position start date (provided in database). Finally, a machine-learning python algorithm, "gender-guesser", assigned a value of either one (males) or two (females) to each observation by analyzing name length, character sequence, and frequency of occurrence against a known database. The names of self-identifying victims of violence and those with gender neutral names were excluded, bring the final sample to n=3,080.

Results:

Since later results depend on the accuracy of the python "gender-guesser" algorithm, an additional hypothesis test was hand coded on a simple random sample of n=30 to determine the generalizability of the criterion. Since the algorithm was found to be accurate in n=29 (97%), the decision was made to proceed with caution. Additionally, in the case that this algorithm is not accurate, it is suspected that no statistical differences would be found, since any potential differences are randomized due to error—this is not the case, however.

First, the dataset was summarized by fiscal year salary: male professors (n=1542) had an average salary of $142,181 and standard deviation of $76,385 whereas female professors (n=1539) had an average salary of $127,778 and standard deviation of $61,107. Preliminary hypothesis testing revealed that the difference in mean salary between the two groups was

significant, t(3079) = 5.75, p < .001, with male faculty having a significantly higher mean salary ($14,903.35, 95% CI [$9,823.66, $19,983.05]) than female faculty.

      Since salary is affected by time spent teaching, there could be a confounding relationship in drawing a correlation between male and female salaries, if males tend to have worked longer on average. To test whether the time spent teaching is different, I will perform a second independent samples ttest since there are two populations, a scale dependent variable and a scale independent variable. Results of the ttest revealed that on average, males worked significantly longer (M = 17.92 years, SD = 11.40) than females (M = 14.10 years, SD = 9.99), t(3079) = 9.92, p < .001, Cohen's d = 0.35 (95% CI [0.27, 0.43]). The effect size suggests a moderate difference in length of work between males and females.

      To account for time spent teaching as a confounding variable, a third ttest was conducted which restricted data with time spent at the University to within the range of [0,14] years. Within this range, if there still remains a significant difference, one can conclude that, even when controlling for time spent teaching, male and female professors have different salaries. A two-sample t-test was conducted to compare the mean total pay between males (M = $119,113.5, SD = $71,285.87) and females (M = $110,773.8, SD = $57,637.36) who worked at the university for 14 years or less. Results showed that males earned significantly more than females, t(1541) = 2.54, p < .01, d = 0.14.

General Discussion:

The findings of this investigation reveal a significant difference in the remuneration of male and female faculty members, even after controlling for the duration of their service at the University of Wisconsin-Madison. Notably, female faculty are expected to receive a total pay that is $8,339 less than their male counterparts. These results unequivocally demonstrate the existence of a gender pay gap at the university. Despite controlling for the variable of time spent teaching, the gender disparity in salary persists. These results prompt reflection on the possible causes of gender bias in the recruitment, promotion, or salary negotiation processes at the University of Wisconsin-Madison. While we cannot infer causation from the data, the results of this study indicate the potential presence of gender inequity. Consequently, this essay underlines the necessity of employing data-driven approaches to investigate and confront gender inequality in the workplace. Future investigations should consider controlling for additional confounding variables and employ more reliable gender classification algorithms, such as those using social media data, to obtain more accurate results.

Misha O'Keeffe
Econ 310: Statistics in Economics

Works Cited:

[1]

https://www.worldbank.org/en/topic/gender/publication/unrealized-potential-the-high-cost-of-gender-inequality-in-earnings#:~:text=Globally%2C%20women%20account%20for%20only,of%20%2423%2C620%20per%20person%20globally.

[2]

https://news.harvard.edu/gazette/story/2016/03/the-costs-of-inequality-for-women-progress-until-they-get-near-power/

[3]

https://madison.com/uw-salary-database/html_e37ff025-9a87-5a31-91ea-b2eb50aba4cb.html

Error! Bookmark not defined.
https://www.collegefactual.com/colleges/university-of-wisconsin-madison/student-life/sports/#:~:text=The%20sports%20at%20UW%20-%20Madison%20are%20led,salary%20of%20women%E2%80%99s%20team%20head%20coaches%20is%20%24271%2C489.

[4]

https://badgerherald.com/sports/2022/10/06/the-cost-of-success-a-look-into-badger-coaches-salaries/

[5]

https://www.analyticsvidhya.com/blog/2023/03/name-based-gender-identification-using-nlp-and-python/

Misha O'Keeffe
Econ 310: Statistics in Economics

Appendix:

Fiscal year	Name	Title	Dept Description	Campus	Start Date	Total Pay	Details	gender_value	Current Date	Time at University (years)
2021	Ruben Alexanian	Asst Professor (Chs)	Smph/Medicine/Cardiology	UW Madison	2020-04-06 00:00:00	28364	view	1	4/10/23	3.010958904
2021	Andrew Thliveris	Professor (Chs)	Smph/Ophthal&Vis Sci/Gen	UW Madison	2000-09-01 00:00:00	29125.02	view	1	4/10/23	22.61917808
2021	Stephen Sauer	Assoc Professor (Chs)	Smph/Ophthal&Vis Sci/Gen	UW Madison	2001-07-01 00:00:00	29407.2	view	1	4/10/23	21.7890411
2021	Dinesh Shah	Professor Emer	Smph/Ob-Gyn/Fetal Med	UW Madison	2005-03-11 00:00:00	30797.1	view	1	4/10/23	18.09315068
2021	Samir Sultan	Asst Professor (Chs)	Smph/Medicine/Pulmon Med	UW Madison	2019-10-01 00:00:00	31229.78	view	1	4/10/23	3.526027397
2021	Herbert Wang	Professor Emer	L&S/Geoscience/Geoscience	UW Madison	1972-08-21 00:00:00	31498.59	view	1	4/10/23	50.66849315
2021	Michael Ward	Asst Professor (Chs)	Smph/Emerg Med	UW Madison	2017-11-01 00:00:00	31739.44	view	1	4/10/23	5.44109589
2021	Stephen Carpenter	Professor Emer	L&S/Ctr For Limnology	UW Madison	1989-08-28 00:00:00	31890.21	view	1	4/10/23	33.63835616
2021	Jan Kucher	Adjunct Professor	Engr/Civil & Environ Engr	UW Madison	2019-01-01 00:00:00	31985.24	view	1	4/10/23	4.273972603
2021	Mark Oleinik	Adjunct Professor	Engr/Civil & Environ Engr	UW Madison	2015-05-25 00:00:00	31985.24	view	1	4/10/23	7.882191781
2021	Matthew Kalscheur	Asst Professor (Chs)	Smph/Medicine/Cardiology	UW Madison	2017-07-17 00:00:00	32154.6	view	1	4/10/23	5.734246575
2021	David Bohnhoff	Professor Emer	Cals/Bio Sys Engr	UW Madison	1988-01-08 00:00:00	32347.44	view	1	4/10/23	35.27671233
2021	James Zagzebski	Professor Emer	Smph/Med Physics/Med Physic	UW Madison	2020-06-01 00:00:00	34282.92	view	1	4/10/23	2.857534247
2021	Rajan Suri	Professor Emer	Engr/Industrial & Systems Engr	UW Madison	1985-08-26 00:00:00	34486.82	view	1	4/10/23	37.64657534
2021	James Maloney	Professor (Chs)	Smph/Surgery/Cardiothoracic	UW Madison	2006-07-01 00:00:00	35181.24	view	1	4/10/23	16.78630137
2021	Luke Hillman	Asst Professor (Chs)	Smph/Medicine/Gastroent	UW Madison	2020-08-17 00:00:00	35513.25	view	1	4/10/23	2.646575342
2021	Christopher Demarco	Professor Emer	Engr/Electrical & Computer Eng	UW Madison	1985-01-01 00:00:00	36079.52	view	1	4/10/23	38.29589041
2021	Matthew Wolff	Professor (Chs)	Smph/Medicine/Cardiology	UW Madison	2018-03-01 00:00:00	36326.69	view	1	4/10/23	5.112328767
2021	Robert Lasseter	Professor Emer	Engr/Wisperc	UW Madison	1980-07-01 00:00:00	36510.8	view	1	4/10/23	42.80273973
2021	Tuncer Edil	Professor Emer	Engr/Civil & Environ Engr	UW Madison	1974-04-01 00:00:00	36594.99	view	1	4/10/23	49.05753425
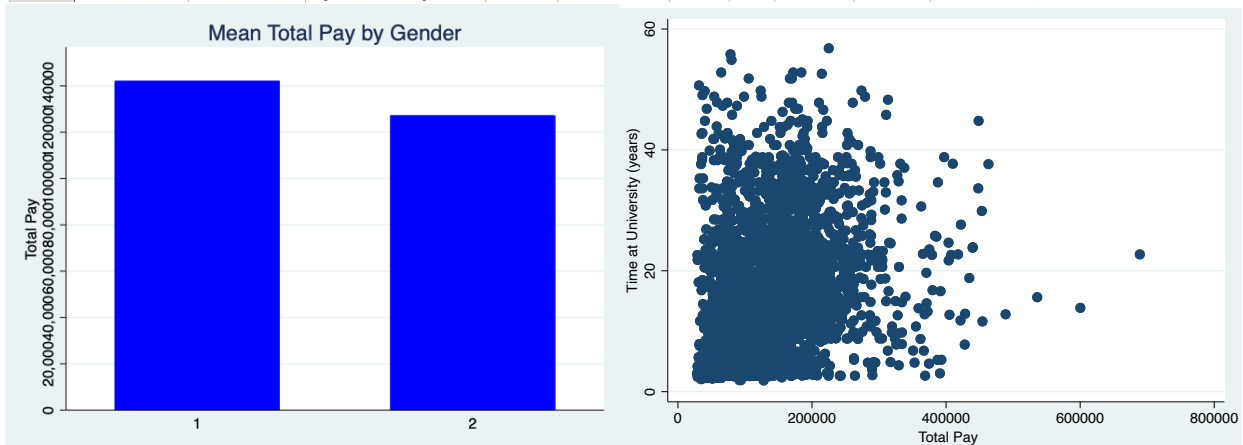

```
. ttest timeatuniversityyears, by(gender_value)

Two-sample t test with equal variances

Group      Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]

    1     1,542    17.92444    .2901874    11.39517    17.35524    18.49365
    2     1,539    14.09564    .2546364    9.989408    13.59617    14.59511

Combined  3,081    16.01191    .1960851    10.88405    15.62743    16.39638

 diff              3.828803    .3861171                3.07173     4.585876

    diff = mean(1) - mean(2)                              t =   9.9162
H0: diff = 0                             Degrees of freedom =     3079

  Ha: diff < 0              Ha: diff != 0                  Ha: diff > 0
Pr(T < t) = 1.0000      Pr(|T| > |t|) = 0.0000        Pr(T > t) = 0.0000
```

```
. ttest totalpay, by(gender_value)

Two-sample t test with equal variances

Group      Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]

    1     1,542    142181.9     1945.22    76385.5     138366.4    145997.5
    2     1,539    127278.6    1710.616    67107.64    123923.2    130633.9

Combined  3,081    134737.5    1302.085    72274.51    132184.4    137290.5

 diff              14903.35    2590.709                9823.661    19983.05

    diff = mean(1) - mean(2)                              t =   5.7526
H0: diff = 0                             Degrees of freedom =     3079

  Ha: diff < 0              Ha: diff != 0                  Ha: diff > 0
Pr(T < t) = 1.0000      Pr(|T| > |t|) = 0.0000        Pr(T > t) = 0.0000
```

```
Two-sample t test with equal variances

Group      Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]

    1       669    119113.5    2756.071    71285.87    113701.9    124525.1
    2       874    110773.8    1949.613    57637.36    106947.3    114600.2

Combined  1,543    114389.6    1629.936    64025.56    111192.5    117586.7

 diff              8339.691    3283.229                1899.622    14779.76

    diff = mean(1) - mean(2)                              t =   2.5401
H0: diff = 0                             Degrees of freedom =     1541

  Ha: diff < 0              Ha: diff != 0                  Ha: diff > 0
Pr(T < t) = 0.9944      Pr(|T| > |t|) = 0.0112        Pr(T > t) = 0.0056
```

Misha O'Keeffe
Econ 310: Statistics in Economics

Command Line Code and Results:

```
 1   * 1. Set up Environment
 2   clear all
 3   set more off
 4
 5   * Describe Variables
 6   describe
 7
 8   * Summarize mean of totalpay for females, males
 9   summarize totalpay if gender_value == 1
10   summarize totalpay if gender_value == 2
11
12   * ttest of totalpay for females, males
13   ttest totalpay, by(gender_value)
14
15   * ttest of timeatuniversityyears for females, males
16   ttest timeatuniversityyears, by(gender_value)
17
18   * Effect Sizes
19   estat esize, test(ttest) over(totalpay) by(gender_value)
20
21   * ttest for totalpay for females, males while controlling for
     timeatuniversityyears
22   ttest totalpay if timeatuniversityyears <= 14, by(gender_value)
23
24   * Find mean of time spent working for females, males
25   summarize timeatuniversityyears if gender_value == 1
26   summarize timeatuniversityyears if gender_value == 2
27
28   * Confidence Intervals
29   ci mean totalpay if gender_value == 1, level(99)
30   ci mean totalpay if gender_value == 2, level(99)
31
32   * Scatterplot and correlation between timeatuniversityyears and
     totalpay
33   scatter timeatuniversityyears totalpay
34   correlate timeatuniversityyears totalpay
35
36   * Regression analysis w/ totalpay as dependent variable
37   regress totalpay timeatuniversityyears i.gender_value
38   test 2.gender_value = 1.gender_value
```

```
1 . do "/var/folders/1t/0_3m3jb16cjcty9jpldbb_gc0000gn/T//SD36017.000000"

2 . * Describe Variables
3 . describe

  Contains data
   Observations:         3,272
      Variables:            11

  Variable      Storage   Display    Value
    name         type     format     label      Variable label

  fiscalyear     int      %8.0g                 Fiscal year
  name           str35    %35s                  Name
  title          str21    %21s                  Title
  deptdescription str30    %30s                  Dept Description
  campus         str10    %10s                  Campus
  startdate      str19    %19s                  Start Date
  totalpay       float    %9.0g                  Total Pay
  details        str4     %9s                   Details
  gender_value   byte     %8.0g
  currentdate    str7     %9s                   Current Date
  timeatunivers~s float   %9.0g                 Time at University (years)

  Sorted by:
       Note: Dataset has changed since last saved.

4 .
5 . * Summarize mean of totalpay for females, males
6 . summarize totalpay if gender_value == 1

      Variable |        Obs        Mean    Std. dev.       Min        Max

      totalpay |      1,542    142181.9     76385.5      28364     600270

7 . summarize totalpay if gender_value == 2

      Variable |        Obs        Mean    Std. dev.       Min        Max

      totalpay |      1,539    127278.6    67107.64    29215.9   688963.3

8 .
9 . * ttest of totalpay for females, males
10 . ttest totalpay, by(gender_value)

  Two-sample t test with equal variances

     Group |     Obs       Mean    Std. err.   Std. dev.   [95% conf. interval]

         1 |   1,542   142181.9     1945.22     76385.5    138366.4    145997.5
         2 |   1,539   127278.6    1710.616    67107.64    123923.2    130633.9

  Combined |   3,081   134737.5    1302.085    72274.51    132184.4    137290.5

      diff |            14903.35    2590.709                9823.661    19983.05

      diff = mean(1) - mean(2)                              t =    5.7526
```

```
HO: diff = 0                                           Degrees of freedom =     3079

    Ha: diff < 0                 Ha: diff != 0                   Ha: diff > 0
 Pr(T < t) = 1.0000        Pr(|T| > |t|) = 0.0000        Pr(T > t) = 0.0000

11 .
12 . * ttest of timeatuniversityyears for females, males
13 . ttest timeatuniversityyears, by(gender_value)

   Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
1	1,542	17.92444	.2901874	11.39517	17.35524	18.49365
2	1,539	14.09564	.2546364	9.989408	13.59617	14.59511
Combined	3,081	16.01191	.1960851	10.88405	15.62743	16.39638
diff		3.828803	.3861171		3.07173	4.585876

```
    diff = mean(1) - mean(2)                                  t =    9.9162
HO: diff = 0                                         Degrees of freedom =     3079

    Ha: diff < 0                 Ha: diff != 0                   Ha: diff > 0
 Pr(T < t) = 1.0000        Pr(|T| > |t|) = 0.0000        Pr(T > t) = 0.0000

14 .
   end of do-file

15 . do "/var/folders/1t/0_3m3jb16cjcty9jp1dbb_gc0000gn/T//SD36017.000000"

16 . * ttest for totalpay for females, males while controlling for timeatuniversityyears
17 . ttest totalpay if timeatuniversityyears <= 14, by(gender_value)

   Two-sample t test with equal variances
```

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf. interval]	
1	669	119113.5	2756.071	71285.87	113701.9	124525.1
2	874	110773.8	1949.613	57637.36	106947.3	114600.2
Combined	1,543	114389.6	1629.936	64025.56	111192.5	117586.7
diff		8339.691	3283.229		1899.622	14779.76

```
    diff = mean(1) - mean(2)                                  t =    2.5401
HO: diff = 0                                         Degrees of freedom =     1541

    Ha: diff < 0                 Ha: diff != 0                   Ha: diff > 0
 Pr(T < t) = 0.9944        Pr(|T| > |t|) = 0.0112        Pr(T > t) = 0.0056

18 .
19 . * Find mean of time spent working for females, males
20 . summarize timeatuniversityyears if gender_value == 1
```

Variable	Obs	Mean	Std. dev.	Min	Max
timeatuniv~s	1,542	17.92444	11.39517	1.857534	56.8137

```
21 . summarize timeatuniversityyears if gender_value == 2
```

Variable	Obs	Mean	Std. dev.	Min	Max
timeatuniv~s	1,539	14.09564	9.989408	1.89863	52.81096

```
22 .
23 . * Confidence Intervals
24 . ci mean totalpay if gender_value == 1, level(99)
```

Variable	Obs	Mean	Std. err.	[99% conf. interval]	
totalpay	1,542	142181.9	1945.22	137165.1	147198.7

```
25 . ci mean totalpay if gender_value == 2, level(99)
```

Variable	Obs	Mean	Std. err.	[99% conf. interval]	
totalpay	1,539	127278.6	1710.616	122866.8	131690.3

```
26 .
27 . * Scatterplot and correlation between timeatuniversityyears and totalpay
28 . scatter timeatuniversityyears totalpay
```

```
29 . correlate timeatuniversityyears totalpay
   (obs=3,081)
```

	timeat~s	totalpay
timeatuniv~s	1.0000	
totalpay	0.2830	1.0000

```
30 .
31 . * Regression analysis w/ totalpay as dependent variable
32 . regress totalpay timeatuniversityyears i.gender_value
```

Source	SS	df	MS			
				Number of obs	=	3,081
				F(2, 3078)	=	139.32
Model	1.3356e+12	2	6.6778e+11	Prob > F	=	0.0000
Residual	1.4753e+13	3,078	4.7931e+09	R-squared	=	0.0830
				Adj R-squared	=	0.0824
Total	1.6089e+13	3,080	5.2236e+09	Root MSE	=	69232

totalpay	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
timeatuniversityyears	1814.783	116.431	15.59	0.000	1586.493	2043.073
2.gender_value	-7954.908	2534.071	-3.14	0.002	-12923.55	-2986.265
_cons	109652.9	2731.99	40.14	0.000	104296.2	115009.7

```
33 . test 2.gender_value = 1.gender_value

   ( 1)  - 1b.gender_value + 2.gender_value = 0

       F(  1,  3078) =    9.85
```