SESUG Paper 142-2025

Investigation of the Wage Gap between Sexes in Agriculture

Jordyne Lewis, Elon University

ABSTRACT

The pay gap between different genders is a prevalent issue on a global scale. It is especially pronounced in developing and third-world countries, where economies are agricultural-based rather than industrial, further exacerbating this problem. The resolution of the gender wage gap is crucial to the interdependence of the global market. The purpose of this paper is to study whether there is a possible difference between the average income of male small-scale food producers and female small-scale food producers in certain countries. This paper examines the gender pay gap through data pulled from the United Nations SDG codebook and Database and analyzes it through hypothesis tests, confidence intervals, and graphs that are created through SAS® Studio through SAS® OnDemand for Academics. This study explores the potential wage gap for small-scale food producers who depend on these practices and where more resources may possibly need to be funneled in terms of assistance.

INTRODUCTION

The United Nations Sustainable Development Goals are a collection of action calls for the world to partner together on different initiatives. There are 17 in total, and each focus on a different topic, from Life Below Water to Good Health and Well-Being. Each Goal has different Targets, and within those are Indicators. The Targets list the different missions that fit into the Goals, while the indicators list how the progress on those missions will be measured. The Goals are a part of the 2030 Agenda for Sustainable Development, adopted in September 2015. Every year, a progress report on the Goals is released, and there is also an SDG Indicators Database available online (https://unstats.un.org/sdgs/dataportal) to access data on the Goals.

The focus of this paper is on Goal 2, which is about ending hunger and increasing sustainable agriculture, food security, and nutrition. The specific Target is 2.3, which homes in on small-scale food producers. Indicator 2.3.2 uses the average income of small-scale food producers by sex and indigenous status to measure.

Wage gaps between sexes are a common issue in today's world in many disciplines. In 2022, according to the Pew Research Center, women in America usually earned 82 cents per every dollar made by men. This was only up 2 cents from 2002. The US is a developed, first world country, and still struggles with this issue. You can only imagine that it is more exacerbated in Third World and developing countries, understanding the lack of resources compared to First World countries. This problem not only affects specific countries but can have an impact on the global system as well. A paper written for The Food and Agriculture Organization of the UN did a cross-country examination focusing on gender pay gaps among agricultural workers and stated that "addressing the gender wage gap is…of pivotal importance in the international development agenda" (Benali et al., 2024).

This paper will go through the processes of a hypothesis test and confidence interval in SAS® Studio to examine data on small-scale food producers and the possible wage gap between male and females. The outcome of this paper will allow you to see if there is a difference at all, as well as next steps needing to be taken in terms of the small-scale food producer field for research.

METHODOLOGY

DATA COLLECTION

To research this question, data points were pulled from the United Nations SDG codebook and Database. The study was limited to only a few years' worth of data. The year 2014 was chosen as it seemed there was a good amount of income data for developing and rural countries. To help give more specific recommendations, the countries chosen were narrowed to come from the continents of Africa and South America. Data from multiple different countries were examined to get an adequate sample size of data points. These countries were Cameroon, Ecuador, Ethiopia, Guatemala, Mali, Nicaragua, Niger, Peru, Rwanda, Uganda, and Burkina Faso. One variable was sex, recorded as either male or female, to allow for analysis based on these two groups. Average income was also included, which was measured purchasing power parity (PPP) USD, allowing for comparison between different countries and currencies.

LOADING IN AND FILTERING THE DATASET

In SAS® Studio, you start off with the base Excel data from the UN SDG Database. You have data for all of Goal 2 and need to load it in for use. Due to it being an Excel file, you need to copy the file path for use through both the FILENAME statement and the IMPORT procedure:

```
 FILENAME REFILE '/home/u64148480/my_shared_file_links/u49358361/data/UN
SDG Data/Goal2.xlsx';
proc import DATAFILE=REFILE
DBMS=XSLX
  OUT=goal2_import;
  GETNAMES=YES;
run;
```

You now must filter it appropriately to only include the correct data points for small-scale producers, Indicator 2.3.2, 2014, and the correct countries. Additionally, in the Excel file there are data points that aggregate both male and female incomes together. In order to keep SAS® Studio from using these, you must filter accordingly. These steps involve the SET statement, the IF statement, and the KEEP statement:

```
   Data goal2 (KEEP = Indicator SeriesDescription GeoAreaName TimePeriod
Sex Value Units Average_Income);
set goal2_import;
Average_Income = input(Value,5.2);
if Indicator='2.3.2';
if SeriesDescription = "Average income of small-scale food producers, PPP
(constant 2017 international $)";
if GeoAreaName = "Cameroon" or GeoAreaName="Ecuador" or
GeoAreaName="Ethiopia"
or GeoAreaName="Guatemala" or GeoAreaName="Mali"
or GeoAreaName="Nicaragua" or GeoAreaName="Niger" or GeoAreaName="Peru" or
GeoAreaName="Rwanda"
or GeoAreaName="Uganda" or GeoAreaName="Burkina Faso";
if Sex="MALE" or Sex="FEMALE";
if TimePeriod="2014";
run;
```

As you may have noticed, there was also an INPUT statement included in the code. At the start, SAS® Studio recognizes our Average Income variable (named “Value”) as a character rather than a number, as seen below in Table 1**.** You can check how your variables are being defined through the CONTENTS procedure, meant to describe the dataset and its variables.

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
16	Age	Char	6	$6.	$6.	Age
8	BasePeriod	Char	4	$4.	$4.	BasePeriod
9	FootNote	Char	573	$573.	$573.	FootNote
5	GeoAreaName	Char	92	$92.	$92.	GeoAreaName
1	Goal	Num	8	BEST.		Goal
3	Indicator	Char	5	$5.	$5.	Indicator
12	Location	Char	7	$7.	$7.	Location
17	Nature	Char	2	$2.	$2.	Nature
14	Observation Status	Char	1	$1.	$1.	Observation Status
4	SeriesDescription	Char	140	$140.	$140.	SeriesDescription
10	Severity of price levels	Char	8	$8.	$8.	Severity of price levels
11	Sex	Char	7	$7.	$7.	Sex
2	Target	Char	3	$3.	$3.	Target
6	TimePeriod	Num	8	BEST.		TimePeriod
13	Type of product	Char	3	$3.	$3.	Type of product
15	Units	Char	11	$11.	$11.	Units
7	Value	Char	22	$22.	$22.	Value

Table 1. Output from PROC CONTENTS before INPUT

You can create a new variable named Average_Income using the INPUT statement to keep the same numbers from VALUE but change the format to have it recognized as a number rather than a character string. Here is the line of code from before that makes this change:

```
Average_Income = input(Value,5.2);
```

Below, you can see in Table 2 how that INPUT statement changes the code output of PROC CONTENTS:

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
8	Average_Income	Num	8			
3	GeoAreaName	Char	92	$92.	$92.	GeoAreaName
1	Indicator	Char	5	$5.	$5.	Indicator
2	SeriesDescription	Char	140	$140.	$140.	SeriesDescription
6	Sex	Char	7	$7.	$7.	Sex
4	TimePeriod	Num	8	BEST.		TimePeriod
7	Units	Char	11	$11.	$11.	Units
5	Value	Char	22	$22.	$22.	Value

Table 2. Output from PROC CONTENTS after INPUT

PROCEDURES AND VISUAL ANALYSIS

DESCRIPTIVE STATISTICS

After filtering the data, you can run some descriptive statistics to get a feel for your data and understand the characteristics of the data you're working with. You can utilize the MEANS procedure for this, to analyze your Average_Income variable by sex. This procedure will generate the Average Income mean, standard deviation, minimum, and maximum split by male and female, seen below in Output 1:

The MEANS Procedure

Analysis Variable : Average_Income						
Sex	N Obs	N	Mean	Std Dev	Minimum	Maximum
FEMALE	11	11	913.3363636	761.6210767	104.6000000	2105.00
MALE	11	11	1409.05	1048.25	148.5000000	3041.00

Output 1. Output from PROC MEANS

You can edit the output for clarity in a Word document, as seen below in

Table 3:

Sex	N	Mean	Std. Dev.	Minimum	Maximum
Female	11	913.34	761.62	104.60	2105.00
Male	11	1409.05	1048.25	148.50	3041.00

Table 3. Statistics for Average Income, classified by Sex

As seen above, females had a much lower standard deviation of 761.6, while males were 1048.3. Males' income values are more spread out over a wider range, and they also have a higher average income of $1409. However, this doesn't tell the whole story. There could be outliers affecting the means. In order to view the spread in a visual way, you can generate boxplots and histograms.

GENERATING BOXPLOTS

To create a boxplot in SAS® Studio, you can use the SGPLOT procedure. This will allow you to define the dataset you want to use, what kind of plot you want to use, such as a vertical or horizontal boxplot, histogram, or scatterplot. You can also input labels and footnotes under the procedure line. Below is code to generate a horizontal boxplot:

```
   title'Ranges of Average Income of Small-Scale Food Producers by Sex for
a collection of Countries, 2017';
proc sgplot data=goal2;
hbox Average_Income / group=Sex;
xaxis label='Average Income';
footnote "Note: All average incomes are measured in PPP USD.";
run;
title;
```

In this section of code, as seen above, you can add a footnote, axis label, and title. You can also give the boxplot a title, through the TITLE statement. Just make sure to add “title;” after the RUN line – this prevents your title from being applied to the next graph. After inputting all this code, you create a boxplot that graphs Average income classified by sex. See below in Figure 1:

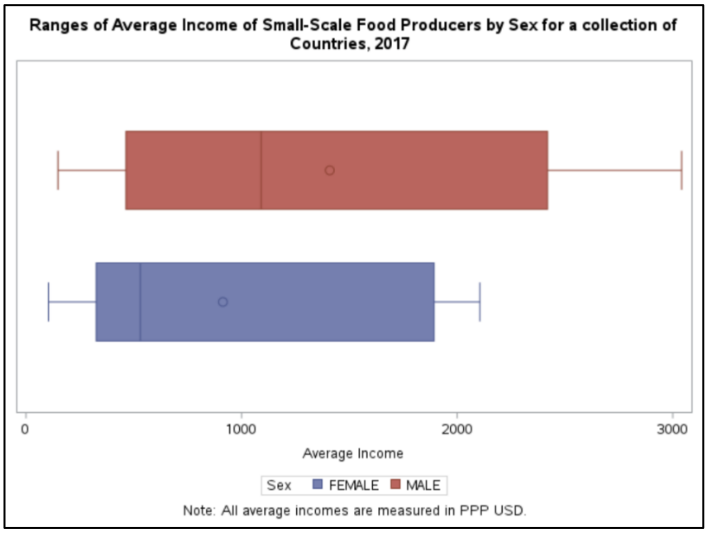


Figure 1. Boxplots, by Sex

Boxplot Analysis

The minimum value for male producers is greater than the minimum value for female producers. This is important to note because not only was the average income bigger, but males are still making more money on average than all small-scale females even at the minimum value. Circles in both the boxplots represent the mean value, furthering the point made prior about how males have a significantly larger mean value than females. The median average income for males is also significantly higher than females. It’s also important to note that there aren’t any outliers.

GENERATING HISTOGRAMS

Histograms are similar to generate, but because you’re classifying by sex, it requires a few extra steps, outlined below:

1. You must add a line detailing the SORT procedure before PROC SGPLOT, letting Studio know what variable to split the plots by.
2. Inside your PROC SGPLOT, reiterate to Studio what you are splitting the plots by.

By following these steps, Studio should generate you two histograms, split by the sex variable. The finished code looks like this:

```
   proc sort data=goal2;
by Sex;

proc sgplot data=goal2;
title'Histograms of Average Income of Small-Scale Food Producers by Sex for
a collection of Countries, 2017';
histogram Average_Income;
by Sex;
xaxis label='Average Income';
yaxis label='Percentage of Sample';
footnote "Note: All average incomes are measured in PPP USD.";
run;
title;
```

From this code, the histograms generate and look like Figure 2 below:

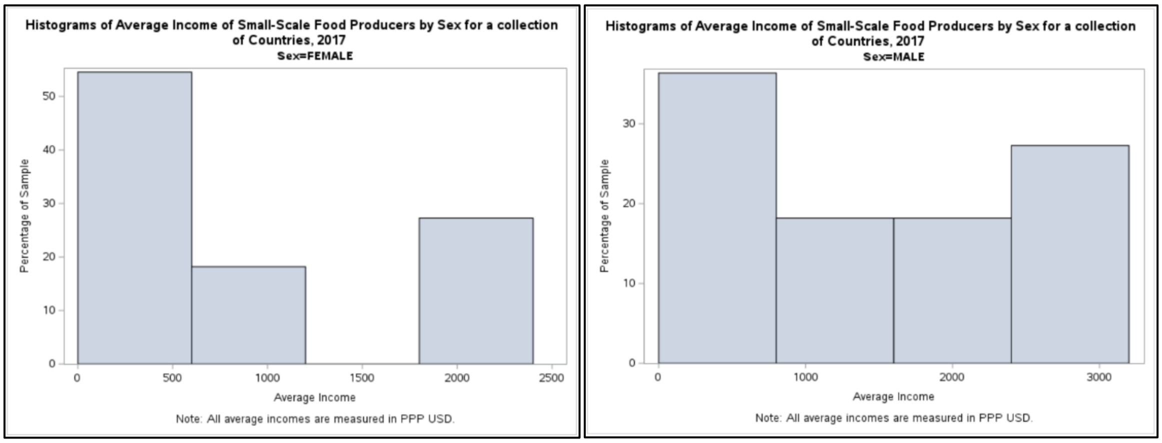


Figure 2. Female and Male Histograms, Default

As seen from this first round of histograms, it looks like female average income data is heavily skewed right, while males' data looks to be bimodal. However, the bins are a bit large, and so it's hard to tell. To make it easier to view, you can adjust where the bins start and how wide you want them to be inside the HISTOGRAM statement by adding a forward slash after defining what variable you want your Histograms to describe:

```
histogram Average_Income / binstart=0 binwidth=500 showbins;
```

Look at the impact that a small addition to your code can have on your histograms' spread in Figure 3:

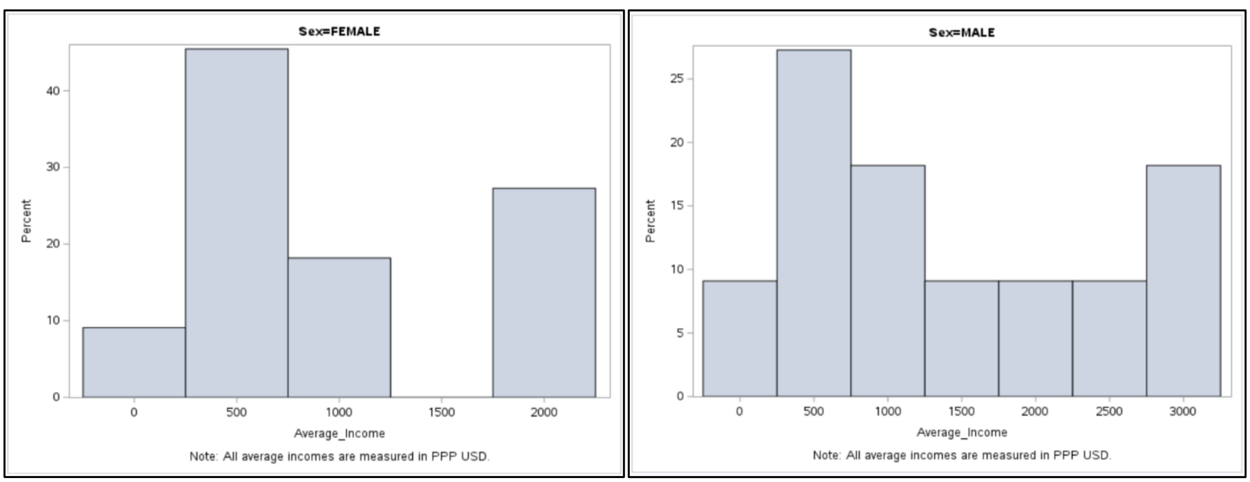


Figure 3. Female and Male Histograms, Modified

Histogram Analysis

With females, the highest percentage of the average income data is concentrated within the 0 to about 500-dollar interval, with over 50% of our data. There are significantly lower percentages of female average incomes as the interval amount increases. The data is right skewed, with more data concentrated on the lower end of income ranges. For males, there's more of an even distribution across incomes. About 50% of the values from the male group are from the 0-1000 (approximate) dollar interval. Like the boxplot graph, these graphs help demonstrate the difference between female and male income by showing the wider distribution males have compared to females, as well as the fact that females are more likely to make less. Additionally, the added code allows for a more in-depth analysis of the spread.

STATISTICIAL ANALYSIS

While there was a visual difference between the two sexes, you still must investigate whether this difference was statistically significant. To figure this out, you can use a one-sided, two-independent sample t-test. Though you may be used to doing this by hand or on a calculator, SAS® Studio can perform the calculations instantly.

Your hypotheses are as follows (1=Female, Male=2):

- H_o: $\mu_1 - \mu_2 = 0$ → The difference between the population average income for female small-scale food producers and the population average income for male small-scale food producers is equal to 0.
- H_a: $\mu_1 - \mu_2 < 0$ The difference between the population average income for female small-scale food producers and the population average income for male small-scale food producers is less than 0.

However, before you can test, you need to check our assumptions – independence, equal variances, and normality.

INDEPENDENCE AND VARIANCES

For simplicity purposes, you can assume that the two groups (male and female) were independent of each other and that each observation in the sample is independent of the other, as well as the randomness of our sample. When it came to variances, you can perform the test for equal variances – making sure the only potential differences could be in the means rather than variances or anything else. Through the TTEST procedure that you'll use to run your hypothesis test; you can also check the variances:

```
proc ttest h0=0 data=goal2 sides=l;

class Sex;

var Average_Income;

run;
```

In the code, you'll identify what your two groups you're trying to compare, as well as your hypothesis and null value. It is critical that you understand Studio chooses group 1 and group 2 based on *alphabetical* order. Therefore, when Studio takes the difference, it will take the female data less the male data. Since we are trying to test whether male data is greater, you will look for the side lower than 0, representing negative values, because you're trying to see if there's evidence that group 2 (the males) are bigger!

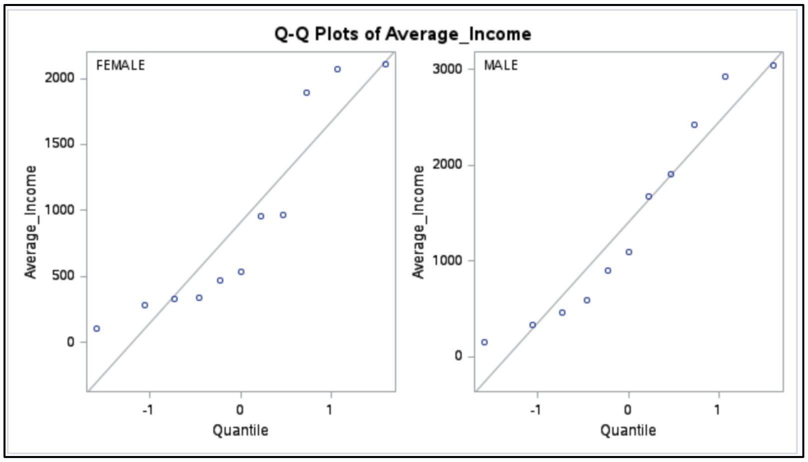
For this analysis, you look at the Equality of Variances table and see that the p-value is equal to 0.3284. Since p is so large, there is insufficient evidence to support the claim that the variances are unequal. Your code from above generates multiple tables, but one specifically looks like Output 2 below:

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	10	10	1.89	0.3284

Output 2. TTEST Variance

NORMALITY

When normality is tested, you receive an unexpected result. To assess the normality condition, a Q-Q plot **IS** used that was also generated with PROC TTEST. For the sample (and population) to be considered normal, the data points should fall mainly on a 45-degree line, meaning it follows the specific normal probability distribution. However, when you perform this test, neither the male nor female data points really followed the Normal line exactly, although the males more than female. You can see this below in Output 3. Q-Q Plots:



Output 3. Q-Q Plots

Though this is not desired, there are many possible reasons for the data's orientation. Both results could be attributed to the fact that neither data sample followed the normal shape. The male data was bimodal, while the female data was strongly skewed right, as we saw from the histograms. This is a limitation of your t-test. While all the other conditions were met, it can't really be said that the populations were normally distributed. If given another chance, a different test would probably be performed, one that could give more accurate results based on the distributions given. You can continue with your analysis, but understand and acknowledge the constraint.

HYPOTHESIS TEST RESULTS

T-test Results

Although assumptions are not met perfectly, you can move forward and perform the test. When you input the code for a PROC TTEST, aside from the Q-Q plot and equal variance tests you also generate data tables for your chosen variable containing your p-value and other descriptive statistics. As a note, make sure you have the class statement since you're using two independent samples.

Sex	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
FEMALE		11	913.3	761.6	229.6	104.6	2105.0
MALE		11	1409.1	1048.3	316.1	148.5	3041.0
Diff (1-2)	Pooled		-495.7	916.2	390.7		
Diff (1-2)	Satterthwaite		-495.7		390.7		

Sex	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
FEMALE		913.3	401.7	1425.0	761.6	532.2	1336.6
MALE		1409.1	704.8	2113.3	1048.3	732.4	1839.6
Diff (1-2)	Pooled	-495.7	-Infty	178.1	916.2	701.0	1323.1
Diff (1-2)	Satterthwaite	-495.7	-Infty	181.2			

Method	Variances	DF	t Value	Pr < t
Pooled	Equal	20	-1.27	0.1095
Satterthwaite	Unequal	18.257	-1.27	0.1102

Output 4. PROC TTEST, with p circled

As shown by the table, you generate a p-value of 0.1095. You use the Pooled method p-value, since you are assuming our variances are equal between our two samples. This value meant there was a 10%

chance of seeing data as extreme or more extreme than the observed data if there is no true difference in the population means between the two sexes. Due to this p-value of .1095, there is insufficient evidence to say there is evidence of a difference in the average incomes of male and female small scale food producers. This is to say that while the male average income is undeniably higher than the female value, it is more so attributed to the random chance of observing income values that happened to be different rather than a true income difference between sexes.

Confidence Interval Results

Though there was insufficient evidence, your exploration of the data does not have to stop here – you can also run a confidence interval to garner more insight from the data. In order to generate one, all you need to do is modify your previous PROC TTEST. You can add the alpha option next to TTEST and input an alpha value of whatever you choose. Additionally, make sure to change your sides to equal 2 to generate an appropriate result. Your code now looks like this:

```
proc ttest alpha=.1 data=goal2 sides=2;
```

Our code states an alpha value of .1, to generate a 90% confidence interval. Now, you may have noticed that you have generated a confidence interval, when you ran PROC TTEST the first time. SAS® Studio generates a confidence interval automatically when you run the procedure but uses an alpha value of .05 *unless* you specify otherwise. Your new output for the 90% Confidence Interval should look similar to Output 4 above, but your CL in the second table now reads 90%. You get a table like Output 5:

Sex	Method	Mean	90% CL Mean		Std Dev	90% CL Std Dev	
FEMALE		913.3	497.1	1329.5	761.6	562.9	1213.3
MALE		1409.1	836.2	1981.9	1048.3	774.7	1669.9
Diff (1-2)	Pooled	-495.7	-1169.5	178.1	916.2	731.1	1243.9
Diff (1-2)	Satterthwaite	-495.7	-1172.7	181.2			

Output 5. 90% Confidence Interval

As seen above, the interval for the male population mean minus the female population mean came out to be –1169.5 dollars to 178.1 dollars. Therefore, we are 90% confident that the true difference between female and male average incomes falls between those numbers. Although the interval was heavier on the negative side, possibly giving evidence that males do statistically make more than females, 0 is contained in our interval. This leaves no difference in average incomes between the sexes as a possibility, meaning it's not certain if there's any statistically significant evidence that there is a disparity between males and females based on the data.

CONCLUSION

Throughout this paper, you have run your own statistical analysis from a UN dataset. You were able to import an Excel file from their database, filter to get to the data you needed, and generate descriptive statistics for your data. You also generated visuals, ran your own hypothesis test, created a confidence interval, and checked for assumptions being met.

An important lesson was also learned – just because there may be a difference when looking at descriptive statistics or visuals, this doesn't mean the difference is statically significant. Though you could not prove anything in your analysis, you can still take the possibility of there being a difference into consideration and use this as a spring-off point for more intensive research into the topic. There's no telling what could be true until a more extensive investigation is done on this topic, possibly considering other countries and years.

REFERENCES

Benali, M., Slavchevska, V., Piedrahita, N., Davis, B., Sitko, N., Nico, G., & Azzarri, C. 2023. *Gender pay gaps among agricultural and non-agricultural wage workers: A cross-country examination - Background paper for The status of women in agrifood systems.* Rome, Italy: FAO.

Kochhar, Rakesh. “The Enduring Grip of the Gender Pay Gap.” Pew Research Center. Accessed August 20, 2024. Available at https://www.pewresearch.org/social-trends/2023/03/01/the-enduring-grip-of-the-gender-pay-gap/.

United Nations: Department of Economic and Social Affairs, Statistics. “SDG Indicators Database.” Accessed August 22, 2025. Available at https://unstats.un.org/sdgs/dataportal.

United Nations. THE 17 GOALS.” Accessed August 20, 2025. Available at https://sdgs.un.org/goals.

ACKNOWLEDGMENTS

I would like to thank Dr. Kirstie Doehler for her support and help as I formulated my topic, wrote my paper, and for introducing me to this Conference. Her belief in me has been invaluable. Any errors in this paper are entirely my own.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jordyne Lewis
Elon University
9802711619
jlewis55@elon.edu