Jean Pockets and Gender*

subtitle

Hari Lee Robledo

April 24, 2024

Fashion has long reflected and reinforced gender inequalities, with women's clothing often lacking functional pockets found in men's garments. A 2018 study by Jan Diehm & Amber Thomas on The Pudding revealed this discrepancy, highlighting inferior pocket functionality in women's jeans. This paper aims to reproduce these findings and explore if gender disparities extend to pricing. Using the same data, we analyze the correlation between prices of men's and women's jeans. Results suggest that women's jeans not only have smaller pockets but also tend to be more expensive. This study sheds light on gender inequalities in everyday products and their broader implications. [UPDATE according to findingsThe findings revealed that pocket area alone may not be a strong predictor of jeans prices for either gender, and there may be other factors influencing the pricing of jeans.]

Table of contents

*Code and data are available at: https://github.com/hari-lr/pockets-and-gender

1 Introduction

Throughout history, women have faced numerous restrictions and limitations, often ingrained within societal norms and cultural practices. Among these limitations, the realm of fashion has served as both a reflection and an enforcer of gender-based inequalities. From ancient times to the modern era, clothing has been a means of expressing social status, cultural identity, and gender roles. According to design expert, Hannah Carlson, pocket inequality stems from the way clothes are designed and made, she explains that: "From a very early age, I think we sort of agree as a culture that womenswear, girlswear is meant to be pretty," she said. "And menswear, boyswear is meant to be utilitarian" (Alwahaidi 2018). Men's garments historically incorporated functional pockets, allowing them to carry essentials such as money, keys, and tools, women's clothing has often lacked this practical feature. Instead, women's garments have frequently been designed with form-fitting silhouettes, delicate fabrics, and ornamental details, prioritizing aesthetics over utility.However, the 1880s, women began to be vocal about pocket inequality, as the "demand for the vote and the demand for pockets were made together" (Alwahaidi 2018).

In 2018, Jan Diehm & Amber Thomas published Womens´s Pockets are Inferior on *The Pudding*, a data-centric digital publication. In this paper, they "measured pockets in both men´s and women's pants in 20 of the US' most popular blue jeans brands" (Jan Diehm 2018). They "programmatically determined whether various everyday items could fit in an otherwise empty pocket in jeans that aren't being worn" (Jan Diehm 2018)., and concluded that women´s pockets were less functional than men´s.

Using the data from the previously described article, this paper will graphically reproduce the results and examine whether there is a correlation between the prices of men's jeans and women's jeans. The study seeks to determine if women's jeans not only have smaller and less functional pockets but also tend to be more expensive. This will help better understand if gender inequality is only reflected in the jean's pockets or in pricing as well. A linear regression analysis revealed divergent findings regarding the relationship between pocket area and price in men's and women's jeans. For men's jeans, no significant correlation was found between pocket area and price, indicating that changes in pocket size do not significantly affect the price of men's jeans. In contrast, although a negative correlation was observed in women's jeans, it was not statistically significant, suggesting that while larger pocket areas might be associated with slightly lower prices, other factors likely play a more influential role in determining the price of women's jeans.

This paper is structured into the following sections: Data, which explains the collection and cleaning process; Results, presenting trends and correlations found in the data; Discussion, comparing and evaluating the data; and Conclusion, summarizing the findings.

2 Data

The data utilized in this paper was retrieved from The Pudding GitHub Site Portal, specifically the data collected by Jan Diehm and Amber Thomas (Jan Diehm 2018). Data was collected, cleaned, and analyzed using the open-source statistical programming software R (R Core Team 2023). This process involved various packages within R, including `tidyverse` (Wickham et al. 2019), `jsonlite` (Ooms 2014), `readr` (Wickham, Hester, and Bryan 2024), `janitor` (Firke 2021), `dplyr` (Wickham et al. 2023), `tibble` (Müller and Wickham 2023), `ggplot2` (Wickham 2016), `knitr` (Xie 2024), and `kableExtra` (Zhu 2021). A comprehensive description of the data gathering and cleaning process is provided in the following subsections.

2.1 Dataset

In the original study, researchers examined jean pocket sizes across 20 brands, encompassing both men's and women's jeans, including skinny and straight styles. The assessment involved scrutinizing four pairs of jeans per brand, all tailored with a 32-inch waistband, despite falling below the median waist size for women in the US. Measurements were meticulously taken in physical stores located in Nashville, New York, and Seattle, with pocket dimensions recorded while jeans were unbuttoned and unzipped, from the inside (Jan Diehm 2018). The study incorporated computer-generated images to illustrate pocket curvature. Various objects, such as standard pen sizes and a front pocket wallet, were utilized to evaluate pocket compatibility. Additionally, area measurements were computed using the `d3.polygonArea()` function in d3.v4, offering valuable insights into the spatial characteristics of the jean pockets. These methods were intended to give a complete understanding of jean pocket sizes and how well they can hold everyday items. (Jan Diehm 2018).

In this paper, the dataset used for analysis was obtained from `measurementRectangles.json`. This dataset mirrors the measurements found in `measurements.csv`, including brand, style, gender specification (men or women), product name, brand size, waist size, fabric composition, price, and height and width measurements of jean pockets (see Table 1). Notably, the dataset contains a "pocketArea" column representing the area of the polygon generated from a pocket's measurements. However, it's important to note that the units of measurement for this area are not explicitly specified in the provided metadata. To determine the units, one must examine how the measurements were collected and processed. Since the measurements were processed using computer-generated calculations and intended for web display, it is possible that the pocket area is denoted in square centimeters (px^2). However, without explicit confirmation from the data documentation or creators of the dataset, absolute certainty regarding the units of measurement remains elusive.

Table 1: Sample of Raw Jean Pockets Data

Brand	Style	Men/Women	Name	Fabric	Price
7 for All Mankind	straight	men	The Straight	98% cotton, 2% spandex	179.00
7 for All Mankind	skinny	men	Paxtyn	92% cotton, 6% polyester, 2% spandex	209.00
Abercrombie	slim	men	Langdon Slim	95% cotton, 4% polyester, 1% elastane	78.00
Abercrombie	straight	men	Kennan Straight	95% cotton, 4% polyester, 1% elastane	78.00
American Eagle	straight	men	Straight	85% cotton, 13% polyester, 2% elastane	49.95

2.2 Data Cleaning

Initially, the code read the JSON data containing pocket area measurements, standardized the column names using the `clean_names()` function, and stored the cleaned data in the `cleaned_area_data` dataframe. Subsequently, the data was filtered based on gender, separating jeans intended for women and men using the `filter()` function. Only the "price" and "pocket_area" columns were retained for further analysis in the resulting dataframes `men_jeans_data` and `women_jeans_data` (see Table 2), respectively. In the final step, the cleaned data was saved into CSV files. The `write.csv()` function was utilized to export the data, with separate CSV files generated for women's and men's jeans.

Table 2: Sample of Cleaned Men's and Women´s Data

(a) Men's Data

Price	Pocket Area
179.00	11222.73
209.00	10841.94
78.00	10624.65
78.00	10134.15
49.95	10555.40

(b) Women's Data

Price	Pocket Area
199.00	5557.218
159.00	5811.112
78.00	8757.578
88.00	9688.483
39.95	6408.587

To visually explore the relationship between pocket area and price for both men's and women's jeans scatter plots with regression lines were created (see Figure 1 and 2).

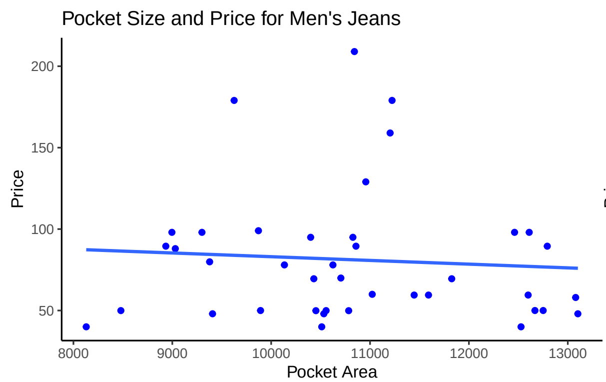

Figure 1: Jean Price vs Pocket Size

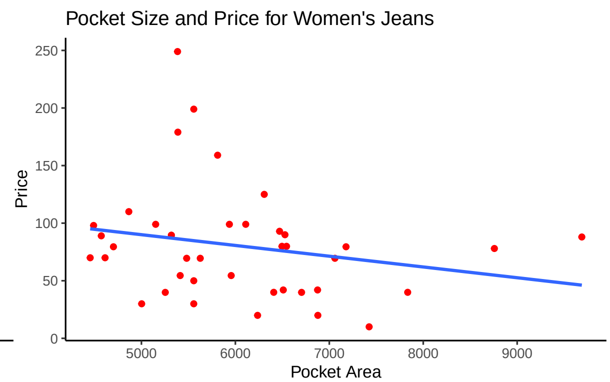

Figure 2: Jean Price vs Pocket Size

These plots visually depict the relationship between pocket area and price for men's and women's jeans. Additionally, to compute the standard deviation of price for both men's and women's jeans, the `sd()` function can be applied to the respective columns in the `men_jeans_data` and `women_jeans_data` dataframes. The standard deviation of price for women's jeans is 48.79, while for men's jeans, it is 40.49, indicating the variability in prices within each gender category. This measure helps assess the spread of prices around the mean and provides insight into the degree of dispersion in price values among the observed jeans.

3 Model

To gain further insights and make predictions about jeans pricing, a linear regression model was implemented.

The final model is displayed here.

$$L = \beta_0 + \beta_1 S + \epsilon$$

Where:

- L represents the dependent variable, which is the overall price of jeans
- S represents the independent variable, which is the pocket area
- β_0 represents the intercept of the regression line, which is the predicted value of L when S is equal to zero
- β_1 represents the slope of the regression line, which is the change in L for a one-unit increase in S
- ϵ represents the random error term, which accounts for variability in L that is not explained by the relationship with S

The linear regression seeked to determine the optimal parameters β_0 and β_1 that minimize the discrepancies between predicted and observed prices across varying pocket areas in the dataset. Through this minimization process, the model constructs a line that most accurately represents the association between pocket area and price. Consequently, we can predict price values based on known pocket area measurements. This regression analysis sheds light on how pocket area influences the pricing of jeans for both men and women, offering valuable insights into the pricing dynamics within each gender category.

4 Results

A merged scatterplot was generated to compare pocket area and price for both men's and women's jeans. Each gender category was represented by a distinct color in the plot. Additionally, linear regression lines were overlaid for each gender group to observe potential differences in the relationship between pocket area and price (see Figure 3).

This visualization facilitated the comparison of how pocket area related to price for men's and women's jeans separately. By including separate regression lines for each gender, we could assess the direction and strength of the relationship between these variables within each group. The use of different colors for men and women aided in distinguishing between the two groups, enabling easy interpretation of the plot.

After visualizing the linear regression for both men's and women's jeans, the results obtained from the regression models were analyzed.

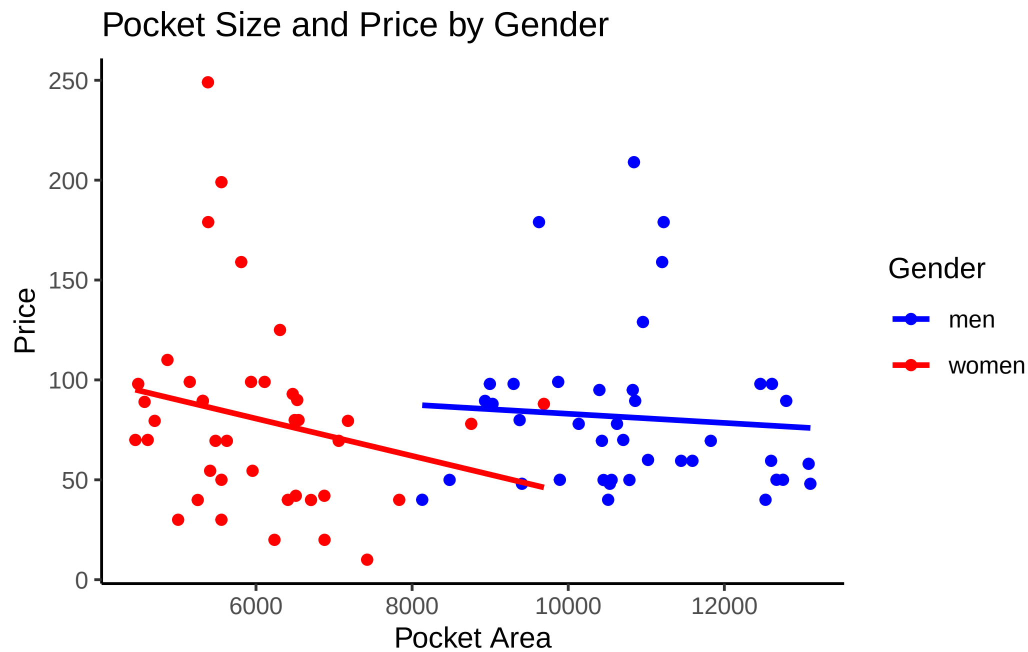

Figure 3: Jean Price vs. Pocket Size per Gender

For men's jeans, the coefficient estimate for pocket area (β_1) is -0.002286, indicating a very small and statistically insignificant negative association between pocket area and price. This suggests that, on average, changes in pocket area do not have a significant impact on the price of men's jeans. The p-value associated with the pocket area coefficient is 0.6445, which is much greater than the typical significance level of 0.05. Additionally, the (R^2) value, which measures the proportion of variance in the dependent variable explained by the independent variable, is very low at 0.005662. Overall, these results suggest that there is little evidence to suggest a meaningful relationship between pocket area and price for men's jeans (see Table 3).

In contrast, for women's jeans, the coefficient estimate for pocket area is -0.009345, also indicating a small negative association with price, although it is slightly larger in magnitude compared to men's jeans. However, like men's jeans, this coefficient is statistically insignificant with a p-value of 0.1804. The (R^2) value for women's jeans is slightly higher at 0.04671, indicating that pocket area explains a slightly larger proportion of the variance in price compared to men's jeans. Despite this, the overall relationship between pocket area and price remains weak for women's jeans (see Table 4).

In summary, the results suggest that pocket area is not a significant predictor of price for either men's or women's jeans in the dataset analyzed. Other factors not considered in this analysis are likely to have a greater influence on the pricing of jeans for both genders.

Table 3: Summary of Linear Regression Model for Men's Jeans

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	105.9002168	53.5254438	1.9785024	0.0551543
pocket_area	-0.0022856	0.0049134	-0.4651738	0.6444619

Table 4: Summary of Linear Regression Model for Women's Jeans

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	136.7159600	42.0324855	3.252626	0.0024018
pocket_area	-0.0093445	0.0068483	-1.364495	0.1804384

5 Discussion

5.1 First discussion point

5.2 Second discussion point

5.3 Third discussion point

##Weaknesses and next steps

References

Alwahaidi, Keena. 2018. *Women's Apparel Doesn't Have Enough Pockets. This Expert Says That Has to Change*.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

Jan Diehm, Amber Thomas. 2018. *Womens´s Pockets Are Inferior*.

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*. https://tibble.tidyverse.org/.

Ooms, Jeroen. 2014. "The Jsonlite Package: A Practical and Consistent Mapping Between JSON Data and r Objects." *arXiv:1403.2805 [Stat.CO]*. https://arxiv.org/abs/1403.2805.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://readr.tidyverse.org.

Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*.