# Exploring Pocket Size and Jean Prices for Men and Women*

**An Analysis of the 2018 *Women´s Pockets are Inferior* Data**

Hari Lee Robledo

April 24, 2024

This paper explores the intersection of gender, fashion, and pricing dynamics, focusing on pocket inequality in jeans. The study investigates whether there is a correlation between pocket size and jean prices for men and women. Utilizing data collected by Diehm & Thomas (2018) from 20 popular US jeans brands, the analysis employs linear regression models, standard deviation and averages to examine the relationship between pocket area and price. The findings reveal that pocket area is not a significant predictor of price for either men's or women's jeans, indicating complex pricing dynamics beyond pocket considerations. The implications of these findings underscore the need for further discussions surrounding the pink tax and pocket inequality by providing empirical evidence of pricing disparities. While the average price of women's jeans appears slightly lower than men's, the considerable variability in prices within each gender category suggests deeper complexities.

## 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 examines 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).

Additionally,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. These area measurements were computed using the `d3.polygonArea()` function in d3.v4 (Jan Diehm 2018).

For this paper, the dataset used was obtained from `measurementRectangles.json`. This dataset mirrors the measurements found in `measurements.csv`, including pixelsrand, 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 pixels ($px^2$). 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 and save the data.

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).

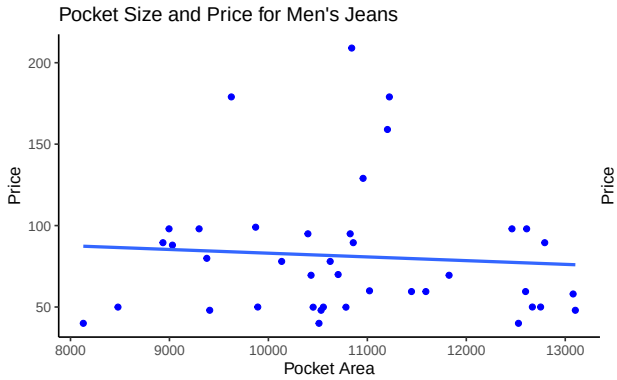


Figure 1: Jean Price vs Pocket Size

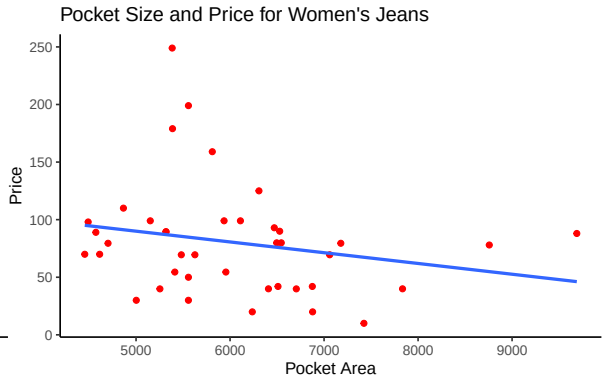


Figure 2: Jean Price vs Pocket Size

## 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
- $\beta_0$ represents the intercept of the regression line, which is the predicted value of $L$ when $S$ is equal to zero
- $\beta_1$ represents the slope of the regression line, which is the change in $L$ for a one-unit increase in $S$
- $\epsilon$ 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 $\beta_0$ and $\beta_1$ that minimize the discrepancies between predicted and observed prices across varying pocket areas in the dataset. 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).

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

For men's jeans, the coefficient estimate for pocket area ($\beta_1$) 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,

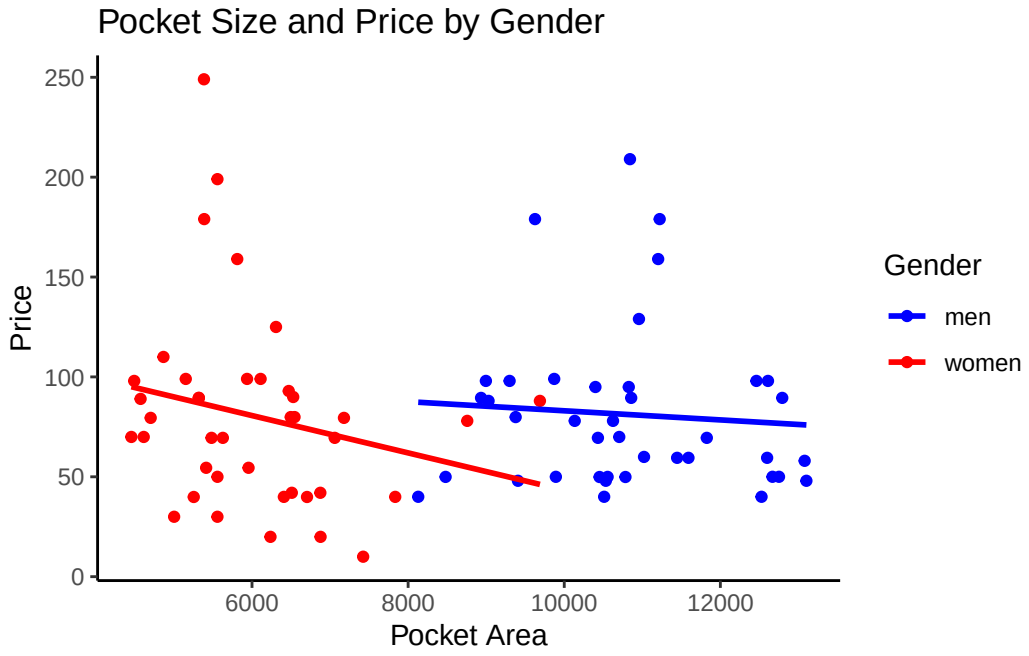


Figure 3: Jean Price vs. Pocket Size per Gender

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).

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.

Furthermore, to quantify the central tendency of prices, the average price for both men's and women's jeans was calculated. Using the `mean()` function, it was determined that the average price for men's jeans was $81.184, while for women's jeans, it was $80.316. These average values offer insight into the typical price range within each gender category.

In addition to the average price, the standard deviation of priceswas calculated using the `sd()` function. The standard deviation measures the degree of dispersion or variability in prices

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 |

within each gender category. For women's jeans, the standard deviation of price was found to be \$48.79, indicating a considerable variability in prices. Similarly, for men's jeans, the standard deviation was \$40.49, reflecting variability within this gender category as well.

# 5 Discussion

## 5.1 Price Disparity in Clothing

The pink tax refers to the phenomenon where products marketed towards girls and women are priced higher than similar products targeted towards boys and men.It represents a type of pricing gender discrimination, leading to women being charged more for items like clothing, toys, and personal care products in comparison to men (Taylor (2023)).

An example direclty applicable to clothes where the "Levi's 501 jeans for women [that] are on average 46% more expensive than the men's version" (Worstall (2016)). The same pair of jeans for women are priced at \$98, while the men's 501 original fit jeans are \$89.50 (McCormick (2023)).

However, when considering the specific aspect of pocket size in jeans, the analysis reveals an interesting but inconclusive finding. While there appears to be a small negative association between pocket area and price for women's jeans, indicating that larger pocket sizes may be associated with slightly lower prices, this relationship is statistically insignificant. The analysis also demonstrates that pocket area explains only a small proportion of the variance in price for women's jeans, suggesting that other factors likely play a more significant role in determining pricing.

Therefore, while pricing disparities according to gender remain a relevant financial phenomenon, the specific relationship between pocket size and price in women's jeans remains weak and inconclusive, which underscores the complexity of pricing dynamics in the fashion industry.

Moreover, while the average price for women's jeans ($80.316) is slightly lower than that for men's jeans ($81.184), this finding may seem contradictory to the general notion of the pink tax, which suggests that women are typically charged more for comparable products. However, when considering the standard deviations calculated for both men's and women's jeans, variability in prices within each gender category becomes evident. The considerable standard deviation for women's jeans ($48.79) implies a wide range of prices, aligning with the idea of pricing gender discrimination and the existence of the pink tax. Yet, the standard deviation for men's jeans ($40.49) also indicates variability, suggesting that pricing disparities may not be exclusive to women's products.

## 5.2 Weaknesses and next steps

The analysis has several limitations. Firstly, it narrowly focuses on the relationship between pocket area and price, neglecting other potential pricing determinants in the fashion industry. The dataset's limited scope, derived from measurements of specific brands and styles, may not be representative of the broader jeans market. Additionally, the regression analysis yielded statistically insignificant results for pocket area coefficients in both men's and women's jeans, indicating a weak relationship. Moreover, the unknown units of pocket area measurement limits accurate interpretation.

To address these limitations, future research could incorporate additional pricing determinants like brand reputation and market demand. Diversifying the dataset to include jeans from additional brands and price points would enhance representativeness. Qualitative research methods could provide deeper insights into consumer behavior, offering a more comprehensive understanding of purchasing decisions beyond pocket considerations. Considering cultural and societal factors influencing fashion trends would offer valuable context for interpreting the results. Moreover, it would be important to increment the number of jeans measured to improve the robustness and generalizability of the findings.

## 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*.

McCormick, Moira. 2023. *Why Women Pay More Than Men for Products*.

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/.

Taylor, Brianna. 2023. *WHAT IS THE PINK TAX? A GUIDE FOR TEENS*.

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.

Worstall, Tim. 2016. *Why Do Retailers Charge Women More Than Men? Because They Can*.

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*.