Boomers, Kiss My Hass!

An Analysis of Avocado Consumption in Relation to Millennial Populations
miniscule data energy: as130, clam14, mdong5, schen96

Hypothesis

Recently, avocados have been tied to the millennial generation due to its increasing presence in social media and public platforms. While this connection is widely acknowledged, we wanted to see whether it could be statistically supported that **cities in the United States with a higher population of millenials consume more avocados, despite price increases in recent years.**

Data

We collected avocado data for years 2016 to 2018 from the Hass Avocado Board. These datasets contained information about prices and total avocado consumption for 42 major US cities and 8 major regions (each region is defined by a collection of states).

We use two datasets for demographic data:

- US Census Bureau, with data regarding population and age per state from 2016-2018
- National Association of Realtors, with data on proportion of millennials in cities across all states in 2017.

We joined these data sources into two tables: One with region-level data that spanned three years and the other with city-level data for 2017. We computed the proportion of millennials for the region-level data and calculated the annual per-capita avocado consumption for both datasets.

Findings

Claim #1: On a regional level, we cannot conclude that cities with a higher proportion of millennials consume more avocados.

Support for Claim #1: Since we had region-level data for three distinct years, we ran a panel data regression with entity fixed effects to see if there's a correlation between per capita avocado consumption and the percentage of millennials in a region. We believed prices are closely related to demand, so we also included the average selling price of an avocado as a regressor. We found that the coefficients on percentage of millennials and average selling price were significant with values of -6.7 and -3.3, respectively. Figure 1 shows our full regression results. This means that, all else being equal, a one unit increase in the proportion of millennials or average selling price was associated with a decrease in mean annual per capita avocado consumption by -6.7 and -3.3 units, respectively. However, we realized there may also be time fixed effects. After running a panel data regression with both entity and time fixed effects, the results suggested that the coefficients on percentage of millennials and average selling price were both not significant at the 5% level. Figure 2 shows our full regression results. In fact, the only significant variable was

the dummy variable for the year 2018 ($p < 0.001$). This indicated that avocados became more popular across all regions in 2018, regardless of millennial populations or price.

Given the limited amount of data points in our region-level analysis, we decided to conduct more analysis on finer-grained city-level data.

```
. areg per_cap_avo pctg_millenial avg_unit_price, absorb(region) robust cluster(region)

Linear regression, absorbing indicators         Number of obs     =         24
                                                F(   2,      7)   =      25.09
                                                Prob > F          =     0.0006
                                                R-squared         =     0.9716
                                                Adj R-squared     =     0.9533
                                                Root MSE          =     0.4580

                                   (Std. Err. adjusted for 8 clusters in region)
------------------------------------------------------------------------------
             |               Robust
   per_cap_avo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
pctg_millenial |  -6.741106   1.362739    -4.95   0.002    -9.963472   -3.518739
avg_unit_price |  -3.29708   .4979285    -6.62   0.000    -4.474494   -2.119666
         _cons |   157.5037   30.07173     5.24   0.001     86.39539    228.6121
-------------+----------------------------------------------------------------
        region |  absorbed                                     (8 categories)
```

```
. areg per_cap_avo pctg_millenial avg_unit_price i.year, absorb(region) robust cluster(region)

Linear regression, absorbing indicators         Number of obs     =         24
                                                F(   4,      7)   =      31.79
                                                Prob > F          =     0.0001
                                                R-squared         =     0.9957
                                                Adj R-squared     =     0.9918
                                                Root MSE          =     0.1920

                                   (Std. Err. adjusted for 8 clusters in region)
------------------------------------------------------------------------------
             |               Robust
   per_cap_avo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
pctg_millenial |  -.9396341   1.838343    -0.51   0.625    -5.286625    3.407356
avg_unit_price |   3.077384    2.20973     1.39   0.206    -2.147797    8.302565
             |
        year |
       2017  |  -.5493844   .3143707    -1.75   0.124    -1.292753    .1939842
       2018  |   1.357968   .2216598     6.13   0.000     .8338264    1.882111
             |
         _cons |   22.63105   41.06019     0.55   0.599    -74.46088     119.723
-------------+----------------------------------------------------------------
        region |  absorbed                                     (8 categories)
```

Figures 1 (entity fixed effects only) & 2 (entity & time fixed effects)

Claim #2: On a city level, we also cannot conclude that cities with a higher proportion of millennials consume more avocados.

Support for Claim #2: Since we only had city-level data for one year (2017), we ran robust multiple regression to measure the relationship between changes in per capita avocado consumption and the percentage of millennials in a city. Once again, we controlled for average selling price by including it as a regressor. We found that the coefficients on percentage of millennials and average selling price were not significant at the 5% level, with p-values of 0.079 and 0.540, respectively. A robust regression serves as an alternative to a least squares regression when the data contains outliers or influential observations, which we believe applied to our city-level data since we saw two outliers (with suspiciously high avocado consumption) in our exploratory scatter plots. Manually removing the two outliers and running regular OLS did not yield any major changes in coefficients or p-values. Figure 3 shows our full regression results.

```
. reg per_cap_avo pctg_millenial avg_unit_price, robust

Linear regression                               Number of obs     =         41
                                                F(2, 38)          =       1.63
                                                Prob > F          =     0.2098
                                                R-squared         =     0.0683
                                                Root MSE          =      21.72

------------------------------------------------------------------------------
             |               Robust
   per_cap_avo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
pctg_millenial | -2.416985   1.341163    -1.80   0.079    -5.132028    .2980577
avg_unit_price | -9.227515   14.93325    -0.62   0.540    -39.45829    21.00326
         _cons |  110.6269   43.85572     2.52   0.016     21.84568    199.4082
------------------------------------------------------------------------------
```

Figure 3 (robust multiple regression)

Claim #3: Avocado prices have decreased over the years of observation (2016 to 2018)
Support for Claim #3: Our intuition that avocado prices have increased in recent years is incorrect. In fact, prices have decreased over the studied three-year period, possibly due to larger supplies from growers. We gathered the average selling price for each year across all the regions and saw that although there were increases in price from 2016 to 2017, there was a more significant decrease from 2017 to 2018 causing the overall percentage change to be negative. In our analysis above, there is no evidence to suggest that price is correlated with consumption. Regional price data and changes across years are shown in Figure 4.

	Avg Unit Price ($/Avocado)			Annual Percent Change		2-Yr Change
Region	**2016**	**2017**	**2018**	**2016-17**	**2017-18**	**2016-18**
California	1.17	1.34	1.16	14.81%	-13.70%	-0.91%
Great Lakes	1.23	1.42	1.15	15.14%	-18.77%	-6.47%
Midsouth	1.21	1.36	1.18	12.61%	-13.82%	-2.95%
Northeast	1.31	1.49	1.28	14.14%	-14.60%	-2.53%
Plains	1.16	1.35	1.10	16.49%	-18.41%	-4.95%
South Central	0.89	1.01	0.84	12.86%	-16.72%	-6.01%
Southeast	1.16	1.37	1.12	18.93%	-18.81%	-3.45%
West	1.08	1.21	1.08	11.59%	-10.90%	-0.57%

Figure 4 (heatmap of unit price changes, 2016-18)