Hypothesis Testing in the NorthWind Database

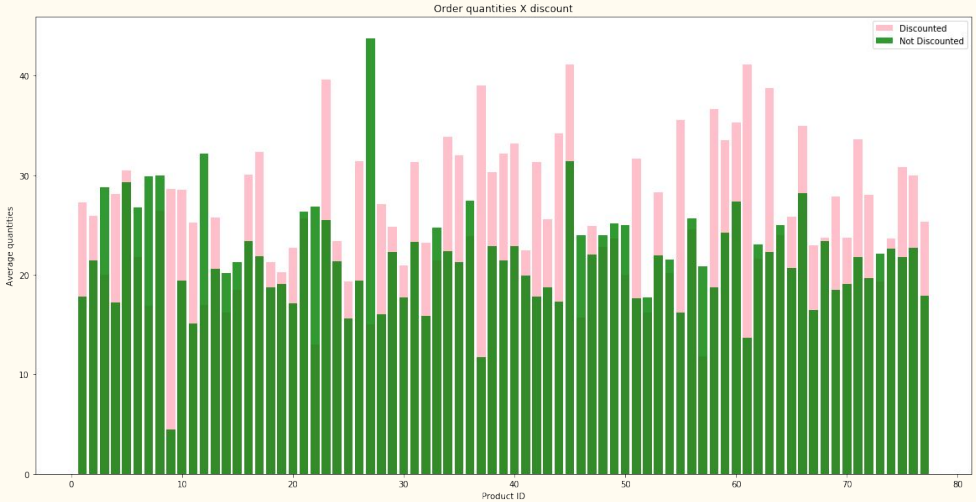
—

Module 2 Final Project

Emily Pfeifer

Q1a: Do discounts have a statistically significant effect on the number of products customers order?

- Split data in discount variable by (discount = 0) & (discount =/= 0) & grouped by average quantity of order
- Created visualization and cleaned data
- Welch's T-Test - "two-sided test for the null hypothesis that 2 independent samples have identical average (expected) values"
- Results determined we reject our null - meaning **there IS a difference between the average quantities of products ordered with no discount vs. products ordered with discounts!**



...If so, at what level(s) of discount? (Q1b)

- Q1b: ANOVA followed by Tukey test (pictured) to compare levels of discount (indicated by group1 and group2)
- Fail to reject null - **NO difference across levels of discount!**
- Conclusion: It does not matter whether the discount is 5% off or 25% off, **the important thing to the consumer is the existence of a discount at all.**

Multiple Comparison of Means - Tukey HSD,FWER=0.05

group1	group2	meandiff	lower	upper	reject
0.0	0.05	6.2955	2.0814	10.5097	True
0.0	0.1	3.5217	-0.8187	7.8622	False
0.0	0.15	6.6669	2.1352	11.1986	True
0.0	0.2	5.3096	0.8285	9.7907	True
0.0	0.25	6.525	1.954	11.096	True
0.05	0.1	-2.7738	-8.4504	2.9028	False
0.05	0.15	0.3714	-5.4528	6.1955	False
0.05	0.2	-0.986	-6.7708	4.7989	False
0.05	0.25	0.2294	-5.6253	6.0842	False
0.1	0.15	3.1452	-2.771	9.0613	False
0.1	0.2	1.7879	-4.0896	7.6653	False
0.1	0.25	3.0033	-2.943	8.9496	False
0.15	0.2	-1.3573	-7.3775	4.6628	False
0.15	0.25	-0.1419	-6.2292	5.9454	False
0.2	0.25	1.2154	-4.8343	7.2652	False

Q2: Will there be an effect on sales if Customers are from the same region as the Employees that selling to them?

WHY:

- People connect better with someone they have something in common
- Common region could establish rapport or help foster effective communication

HOW:

- Created a variable that indicates whether consumer & employee helping them share a region
- Grouped with average order quantity as well as average order monetary value

```
Same Region Price Stats:

 count    140.000000
mean      26.919064
std       14.954526
min        2.500000
25%       16.562500
50%       23.825000
75%       32.162500
max       81.000000
Name: UnitPrice, dtype: float64

Different Region Price Stats:

 count    660.000000
mean      23.855547
std       14.217361
min        2.000000
25%       15.237500
50%       20.750000
75%       28.000000
max      123.790000
Name: UnitPrice, dtype: float64
```

```
Same Region Quantity Stats:

 count    140.000000
mean      24.983548
std       14.476155
min        2.000000
25%       14.125000
50%       24.000000
75%       34.150000
max       75.000000
Name: Quantity, dtype: float64

Different Region Quantity Stats:

 count    661.000000
mean      22.538427
std       14.281011
min        1.000000
25%       12.666667
50%       20.000000
75%       28.333333
max       86.500000
Name: Quantity, dtype: float64
```

Q2: Results

- *H*0 : There is no difference in the average monetary value for orders made by customers who live in the same region as the employee who helps them - Reject
  - When customers are living in the same region as the employee responsible for their order, they will on average spend more money on their order.

- *H*0 : There is no difference in the average quantity of orders made by customers who live in the same region as the employee who helps them - Fail to Reject
  - Living in a common region with an employee does not have an effect on the average quantity of a customer's order.

Conclusion: Trying to connect employees and consumers who are located in the same area may help employees upsell customers on certain products, but it will not aid an employee in selling larger amounts of a product.

Q3: Will products that reference a "Chef" in their name sell better than the products that do not reference Chefs in their title?

WHY:

- Noticed portion of product names included the name of a "Chef", i.e. 'Sir Rodney's Scones' or 'Gnocchi di Nonna Alice'.
- Establishes ethos/credibility with customer?
- Product name is key part of marketing

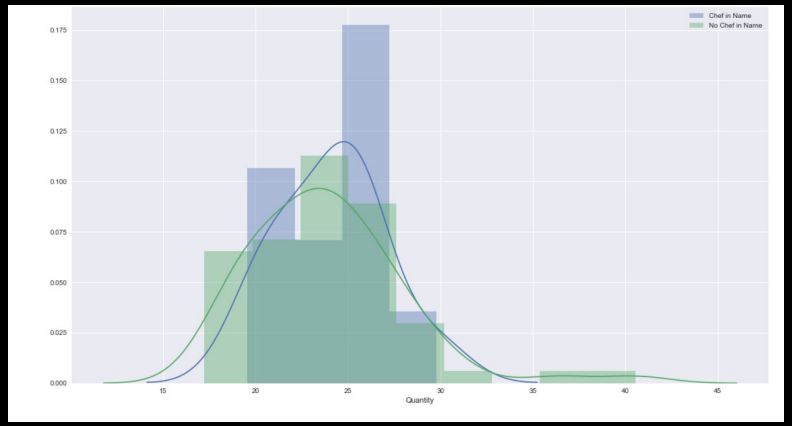
HOW:

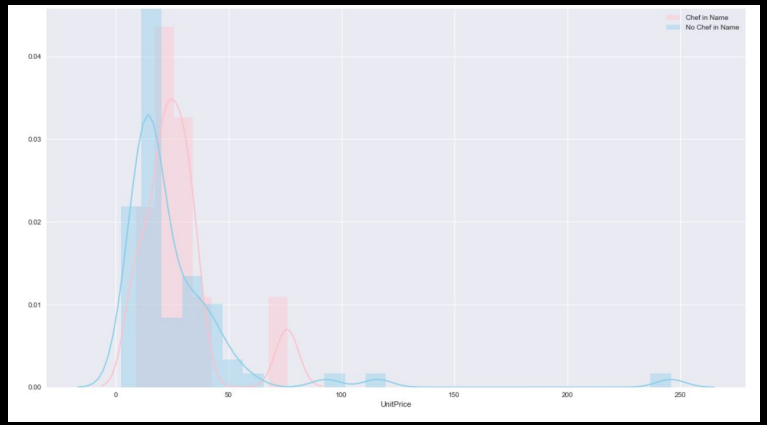
- Created a variable that indicates whether product name includes chef
- Grouped with average order quantity as well as average order monetary value

```
Out[229]: array(['Queso Cabrales', 'Singaporean Hokkien Fried Mee',
                 'Mozzarella di Giovanni', 'Tofu', 'Manjimup Dried Apples',
                 "Jack's New England Clam Chowder",
                 'Louisiana Fiery Hot Pepper Sauce', "Gustaf's Knäckebröd",
                 'Ravioli Angelo', "Sir Rodney's Marmalade", 'Geitost',
                 'Camembert Pierrot', 'Gorgonzola Telino', 'Chartreuse verte',
                 'Maxilaku', 'Guaraná Fantástica', 'Pâté chinois', 'Longlife Tofu',
                 'Chang', 'Pavlova', 'Inlagd Sill', 'Raclette Courdavault',
                 'Perth Pasties', 'Original Frankfurter grüne Soße',
                 'Schoggi Schokolade', "Chef Anton's Gumbo Mix",
                 'Mascarpone Fabioli', "Sir Rodney's Scones", 'Gravad lax',
                 'Tarte au sucre', 'Outback Lager', 'Steeleye Stout',
                 "Uncle Bob's Organic Dried Pears", 'Gnocchi di nonna Alice',
                 'Nord-Ost Matjeshering', 'Alice Mutton',
                 'Queso Manchego La Pastora', 'Boston Crab Meat', 'Lakkalikööri',
                 'Thüringer Rostbratwurst', 'Ipoh Coffee', 'Ikura', 'Flotemysost',
                 'Konbu', 'Rössle Sauerkraut', 'Gula Malacca', 'Vegie-spread',
                 'Röd Kaviar', 'Rhönbräu Klosterbier', 'Teatime Chocolate Biscuits',
                 'Genen Shouyu', 'Laughing Lumberjack Lager', 'Chai',
                 'Sasquatch Ale', 'Spegesild', 'Tourtière', 'Scottish Longbreads',
                 'Aniseed Syrup', 'Wimmers gute Semmelknödel', 'Carnarvon Tigers',
                 'Gudbrandsdalsost', 'Louisiana Hot Spiced Okra',
                 "Chef Anton's Cajun Seasoning", "Grandma's Boysenberry Spread",
                 'Escargots de Bourgogne', 'Filo Mix', 'NuNuCa Nuß-Nougat-Creme',
                 'Côte de Blaye', 'Gumbär Gummibärchen', 'Zaanse koeken',
                 'Tunnbröd', 'Northwoods Cranberry Sauce', 'Valkoinen suklaa',
                 'Rogede sild', 'Chocolade', "Sirop d'érable", 'Mishi Kobe Niku'],
                dtype=object)
```

Q3: Results

- *H*0 : There is no difference in the average quantity sold of products that reference a chef in their name versus products that do not reference chefs in their title - Fail to Reject
- *H*0 : On average, customers will not pay more for products that reference a "Chef" in their name than the products that do not reference Chefs in their title - Reject
- Conclusion: Referencing a chef in a product name does not increase demand for the product. However, a chef-inspired name may help a consumer justify spending more money on the product.





Q4: Will the average re-order level differ between shipping companies?

WHY:

- Reorder Level is important to product success
- Consumers appreciate when their orders are shipped in a timely fashion and their products arrive undamaged
- Associate successful shipping with product & be motivated to reorder product

HOW:

- Used existing variable indicating shipping company (Speedy Express, Federal Shipping, or United Package)
- Grouped with average re-order level per product

Results:

- *H*0 : There is no difference in the average re-order levels between shipping companies **- Fail to Reject**
- The Shipping company used to deliver an order has no effect on the reorder levels of a product.
- The shipping companies had no difference in performance, so if one company is difficult to work with or increases their prices, they can be easily replaced by one of the other two shipping companies.