STATISTICAL STUDY ON NORTHWIND DB

EXECUTIVE SUMMARY

September 2019, Nuran Kairakbayev

Research content. Statistical Hypotheses

A. Quantity - Discount

Does <u>DISCOUNT</u> amount have a significant effect on the <u>QUANTITY</u> of a product in an order?

B. Sales - Discount

Does <u>DISCOUNT</u> amount have a significant effect on the <u>CASH VALUE (sales)</u> of a product in an order?

C. Shipper – Delivery

Does <u>SHIPPER</u> have a significant effect on the <u>DELAY</u> of product delivery in an order?

D. Discount predictors

What <u>FACTORS</u> have a significant effect on <u>DISCOUNT</u> levels?

E. DISCOUNT level classifier model

Is it possible to create a <u>CLASSIFIER</u> model on <u>DISCOUNT</u> level with robust level of prediction accuracy?
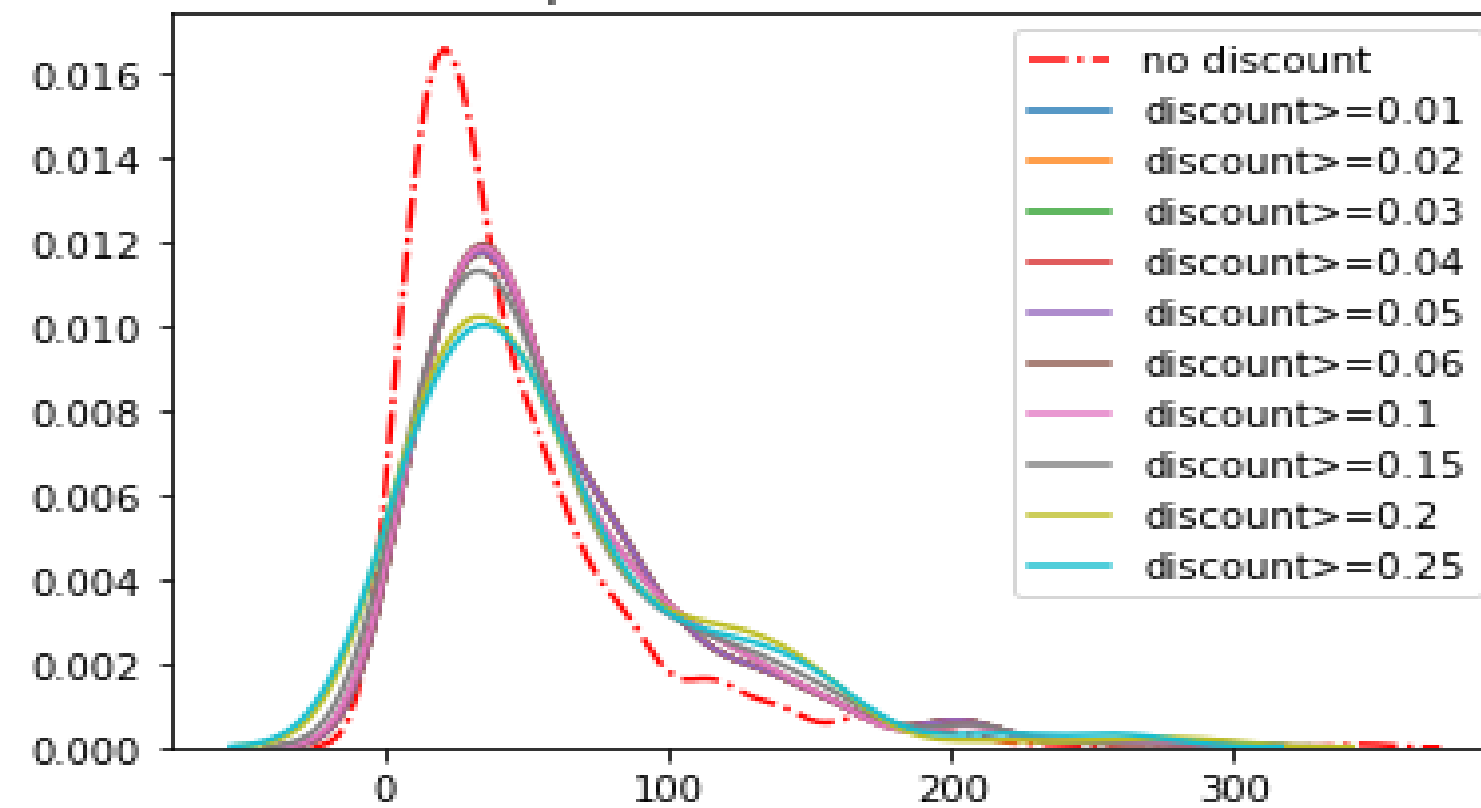
A. QUANTITY – DISCOUNT (1/5)

?

Key Question

Does **DISCOUNT** amount have a significant effect on the **QUANTITY** of a product in an order?

A. QUANTITY – DISCOUNT (2/5)

Does <u>DISCOUNT</u> amount have a significant effect on the <u>QUANTITY</u> of a product in an order?
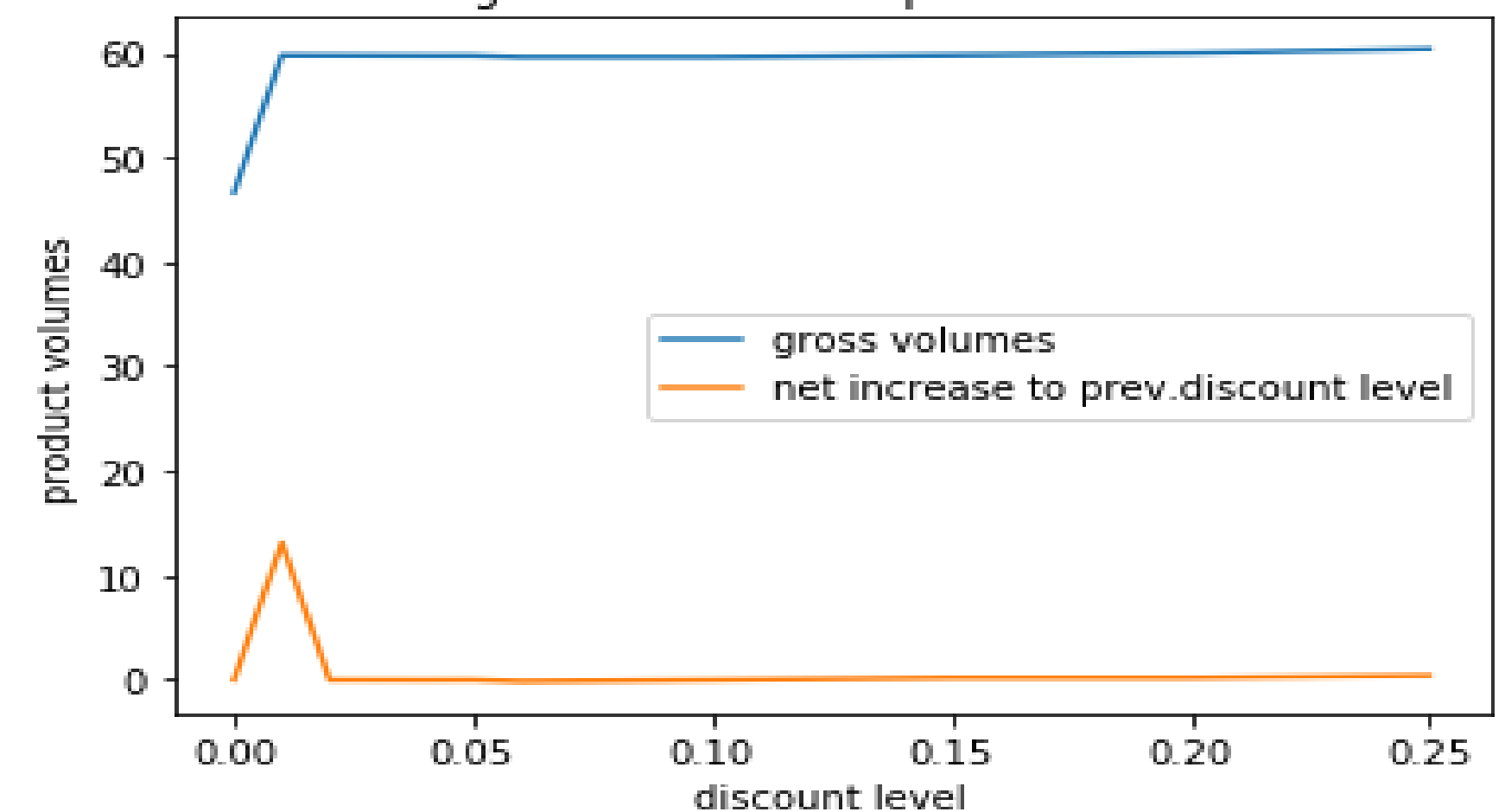
Visual analysis

Product Quantities per Order less for Products without discounts

Marginal increase of Product Quantities:

0% - 2% discount rates

Total Effect and Marginal Effect will be tested in statistical analysis

A. QUANTITY – DISCOUNT (3/5)

Does <u>DISCOUNT</u> amount have a significant effect on the <u>QUANTITY</u> of a product in an order?

|	**Test Version 1: testing gross difference**	**Test Version 2: testing marginal difference**
? Methodology	• Ho: **<u>E{Quantity | no discount} = E{Quantity | discount}</u>** • Testing method: **<u>Welch-test (2-tail)</u>** • Loop: varying **<u>discount groups</u>** by discount level	• Ho: **<u>E{Quantity | prev.discount} = E{Quantity | next.discount}</u>** • Testing method: **<u>Welch-test (2-tail)</u>** • Loop: varying **<u>intra-discount groups</u>** by discount level

A. QUANTITY – DISCOUNT (4/5)

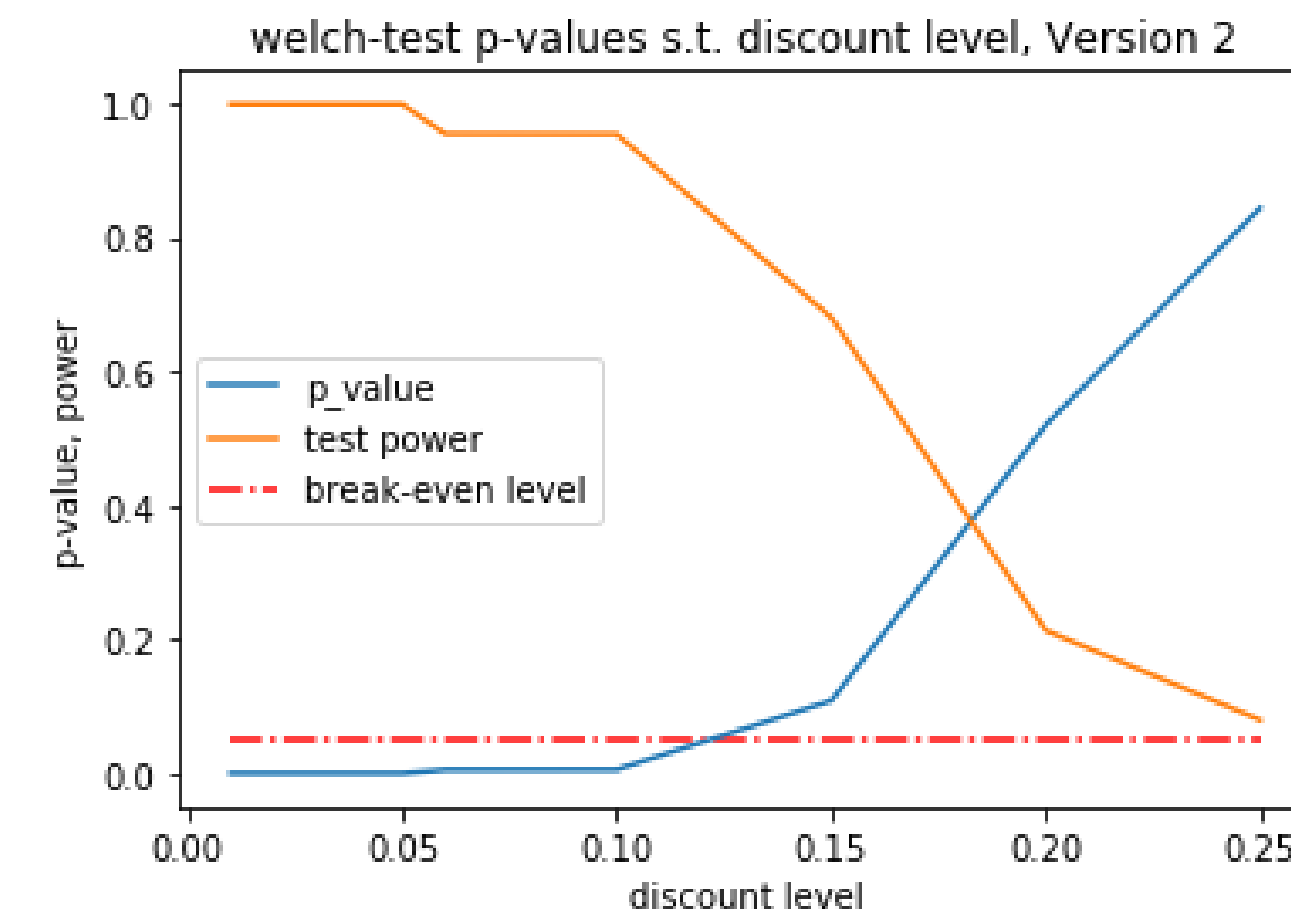**Does <u>DISCOUNT</u> amount have a significant effect on the <u>QUANTITY</u> of a product in an order?**

Results

Any discount rate – affect on Product Volumes

- P-values: <5% for all discount levels
- Test power: good for discount levels

welch-test p-values s.t. discount level, Version 1
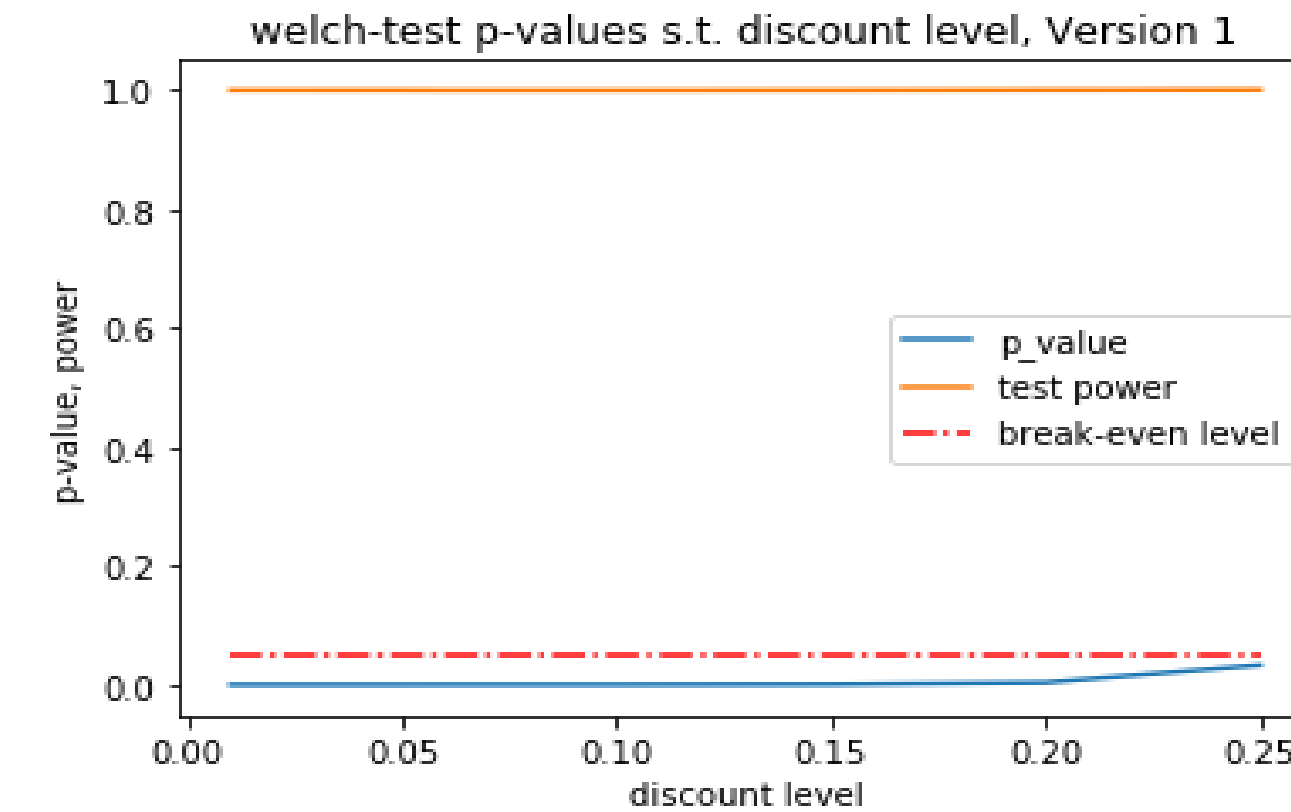
Positive marginal effect on Volumes – for 1%-10% discount rates

- P-values: <5% for discount levels in range 1%-10%
- Test power: good for discount level in range 1%-10%

welch-test p-values s.t. discount level, Version 2

A. QUANTITY – DISCOUNT (5/5)

Does <u>DISCOUNT</u> amount have a significant effect on the <u>QUANTITY</u> of a product in an order?

Issues and Recommendations

ISSUES:

- **Discounts on prices increases product quantities in an order, but only for 1%-10% range of discount levels.**

RECOMMENDATIONS:

- **Use minimum discount rates (1%-10%) for product volumes improvement**

?

Key Question

Does DISCOUNT amount have a significant effect on the CASH VALUE (sales) of a product in an order?

B. SALES – DISCOUNT (2/4)

Does <u>DISCOUNT</u> amount have a significant effect on the <u>CASH VALUE (sales)</u> of a product in an order?

|	**Test Version 1: testing gross difference**	**Test Version 2: testing marginal difference**
? **Methodology**	• Ho: <u>**E{Quantity | no discount} = E{Quantity | discount}**</u> • Testing method: <u>**Welch-test (2-tail)**</u> • Loop: varying **<u>discount groups</u>** by discount level • Sales – **<u>undiscounted original price</u>**	• Ho: <u>**E{Quantity | prev.discount} = E{Quantity | next.discount}**</u> • Testing method: <u>**Welch-test (2-tail)**</u> • Loop: varying **<u>intra-discount groups</u>** by discount level

B. SALES – DISCOUNT (3/4)

Does <u>DISCOUNT</u> amount have a significant effect on the <u>CASH VALUE (sales)</u> of a product in an order?
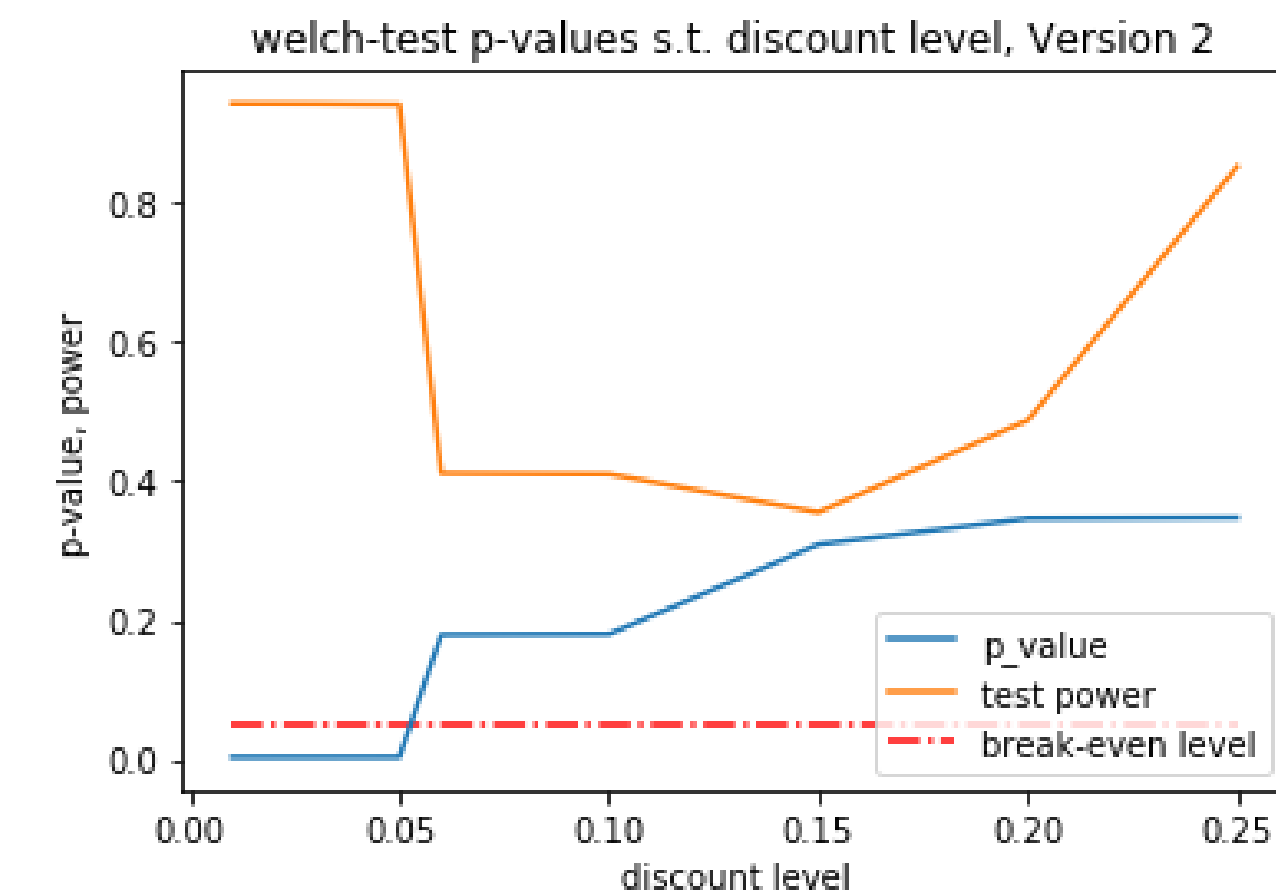
Results

Any discount rate – affect on Product Sales

- P-values: <5% for all discount levels
- Test power: good for discount levels
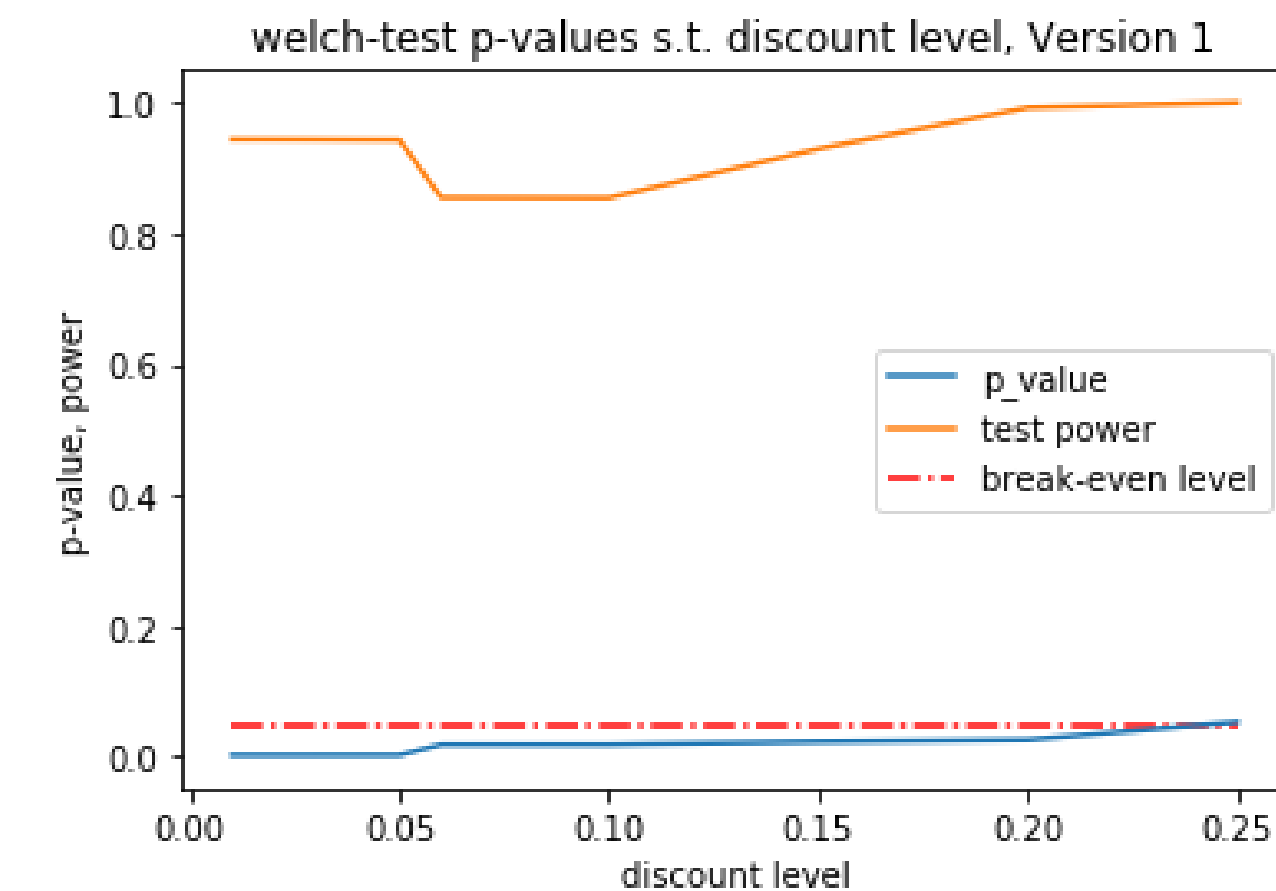
Positive marginal effect on Sales – for 1%-5% discount rates

- P-values: <5% for discount levels in range 1%-5%
- Test power: good for discount level in range 1%-5%

welch-test p-values s.t. discount level, Version 1

welch-test p-values s.t. discount level, Version 2

B. SALES – DISCOUNT (4/4)

Does <u>DISCOUNT</u> amount have a significant effect on the <u>CASH VALUE (sales)</u> of a product in an order?

Issues and Recommendations

ISSUES:

- **Discounts on prices increases product Sales in an order, but only for 1%-5% range of discount levels.**

RECOMMENDATIONS:

- **Use minimum discount rates (1%-5%) for sales volumes improvement**

C. DELAY – SHIPPERS (1/4)

?

Key Question

Does <u>SHIPPER</u> have a significant effect on the <u>DELAY</u> of product delivery in an order?

C. DELAY – SHIPPERS (2/4)

Does <u>SHIPPER</u> have a significant effect on the <u>DELAY</u> of product delivery in an order?

	Test Version 1: Multiple AB-testing	**Test Version 2: ANOVA model**
? **Methodology**	• **Delay** = ShippedDate – RequiredDate • Ho: **E{Delay | Shipper 1} = E{Delay | Shipper 2}** • Testing method: **Welch-test (2-tail)** • Loop: varying in **all combinations** of groups by shipper pairs	• Formula: **Delay ~ C(Shipper)**

C. DELAY – SHIPPERS (3/4)

Does <u>SHIPPER</u> have a significant effect on the <u>DELAY</u> of product delivery in an order?

Results

Multiple AB-testing:

- **No significant difference between shippers for delay**

- P-values: <5% for 3 AB-tests (0,10 – 0,55)
- Test power: weak for all 3 AB-tests (0,09 – 0,34)

ANOVA-model

- **No significant difference between shippers for delay**

- P-value (F-stat) = 0,27

C. DELAY – SHIPPERS (4/4)

Does SHIPPER have a significant effect on the DELAY of product delivery in an order?

Issues and Recommendations

ISSUES:

- **There is no significant effect of shipper on level of shipping delay**

RECOMMENDATIONS:

- **Analyze and compare shipping tariffs for all shippers with same service quality**

?

Key Question

Are there any significant PREDICTORS for DISCOUNT rate?

D. DISCOUNT PREDICTORS (2/4)

Are there any significant <u>Predictors</u> for <u>Discount</u> rate?

Candidates for testing	**Testing parameters**

• **Not-ordered product in Stock**
 	
• **Product Category (only ANOVA)**
 	
• **Supplier (only ANOVA)**
	• Methodology:
	
	   • **Single AB-test –** for 2-class predictor
	   • **ANOVA –** for multi-class predictor
	
	• **Ho: E{Discount | Factor = A} = E{Discount | Factor = B}**
	
	• Statistics for AB-test: **Welch-test (2-tail)**

?

Methodology

D. DISCOUNT PREDICTORS (3/4)

Are there any significant <u>Predictors</u> for <u>Discount</u> rate?

Results

No significant predictors for Discount level:

- **<u>Not-ordered product in Stock</u>**:
  - p-value (AB) = 0,14
  - p-value (F-stat) = 0,75

- **<u>Product Category:</u>** p-value (F-stat) = 0,15

- **<u>Product Supplier:</u>** p-value (F-stat) = 0,06 (weak significance)

D. DISCOUNT PREDICTORS (4/4)

Are there any significant <u>Predictors</u> for <u>Discount</u> rate?

Issues and Recommendations

ISSUES:

- **There are no Predictors with significant relationship with Discount rate**

RECOMMENDATIONS:

- **Continue to analyze possible Discount rate factors**

E. DISCOUNT CLASSIFIER (1/4)

?

Key Question

Is it possible to create robust <u>CLASSIFIER MODEL</u> for <u>DISCOUNT</u> rate?

E. DISCOUNT CLASSIFIER (2/4)

Is it possible to create robust classifier model for discount rate?

Methodology

Model Parameters

- Model type: **Naïve Bayes Classifier**

- Target: **Discount class**

- Model options:
  - **2-class** (discounted vs no-discounted)
  - **6-class** (no-discount, 1-5, 5-10, 10-15, 15-20, 20+)

- Model predictors:
  - **Product Price (before discount)**
  - **Quantity**
  - **Supplier**

E. DISCOUNT CLASSIFIER (3/4)

Is it possible to create robust classifier model for discount rate?

Results

Classifier model accuracy: moderate level

- 2-class model: accuracy score ~ 61%

- 6-class model: accuracy score ~ 60%

E. DISCOUNT CLASSIFIER (4/4)

Is it possible to create robust classifier model for discount rate?

Issues and Recommendations

ISSUES:

- **Accuracy level of classification model is low on given factors (~60%)**

RECOMMENDATIONS:

- **Continue to analyze possible predictors for discount rate**

Business Recommendations

1. Use discounts on product in small range (1%-10%) for increasing volumes of product sales

2. Use discounts on product in smaller range (1%-5%) for increasing gross revenues of product sales

3. Shipping services for all shippers are similar, so check and compare shippers' tariffs on inequality

4. Discount rate for a product could be predicted, but with low level of accuracy

Future Work

1. Analyze and test other predictors for Sales and Volumes

2. Find and understand significant factors on Discount volumes

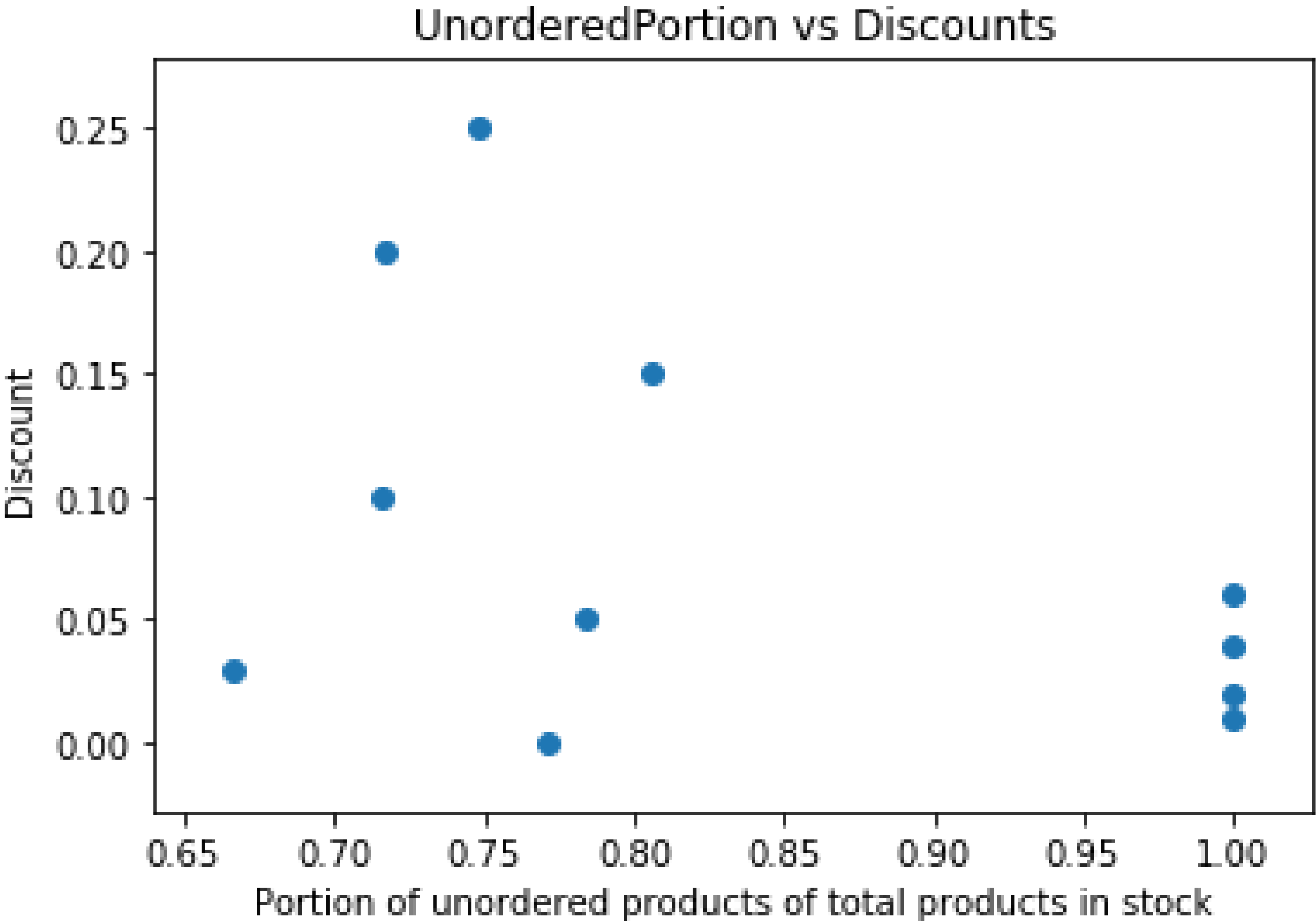3. Based on above, create classifier model on Discount rates

APPENDIX

A1. Original price VS Discounted price
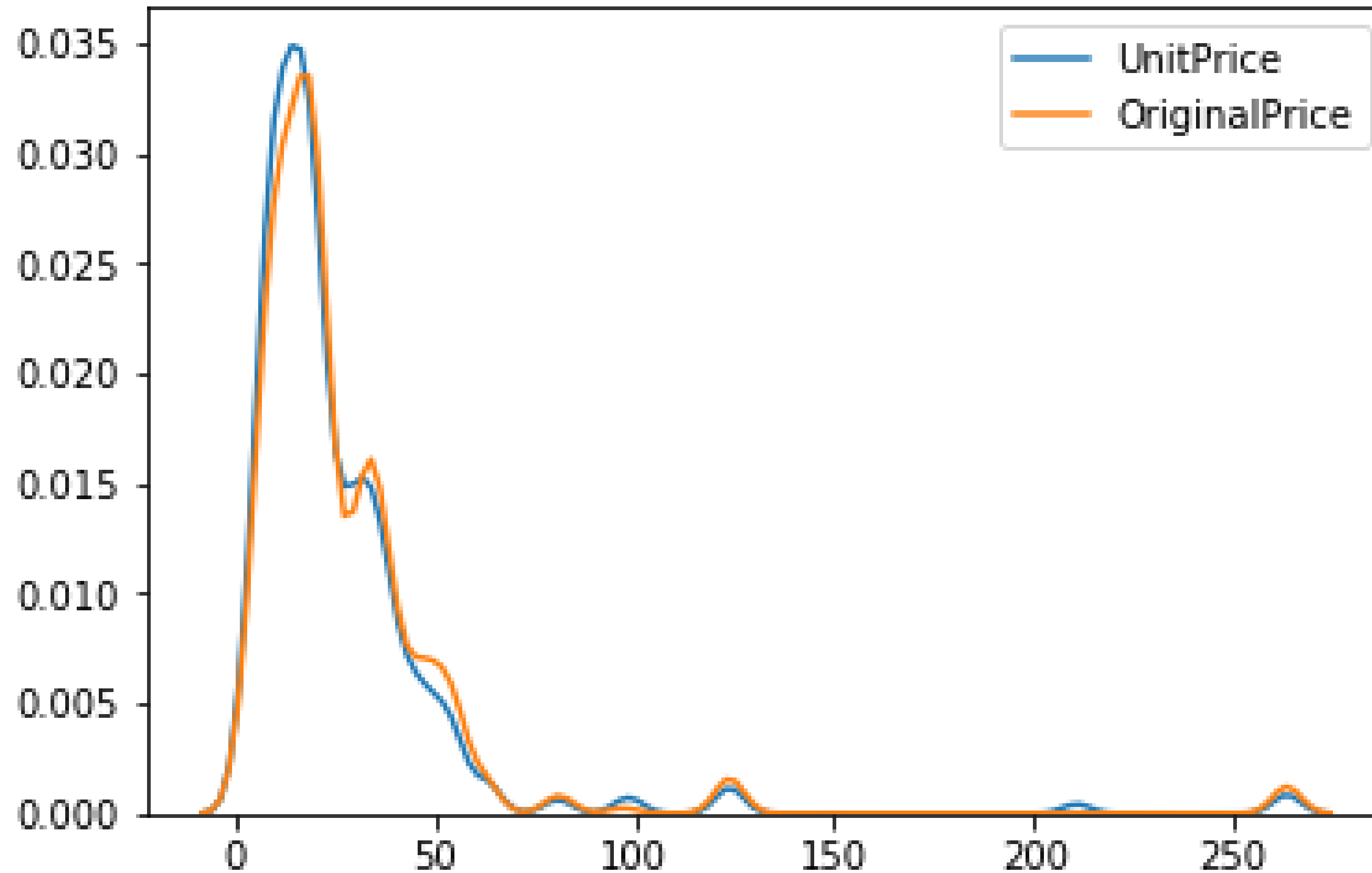
A2. Avg.Discount per Unordered product portion

UnorderedPortion vs Discounts

A3. Avg.Discount per product category