MODULE 3 PROJECT

STATISTICAL STUDY ON NORTHWIND DB

EXECUTIVE SUMMARY

September 2019, Nuran Kairakbayev

Research content. Statistical Hypotheses

A. Quantity - Discount

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

B. Sales - Discount

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

C. Shipper – Delivery

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

D. Discount predictors

What FACTORS have a significant effect on DISCOUNT levels?

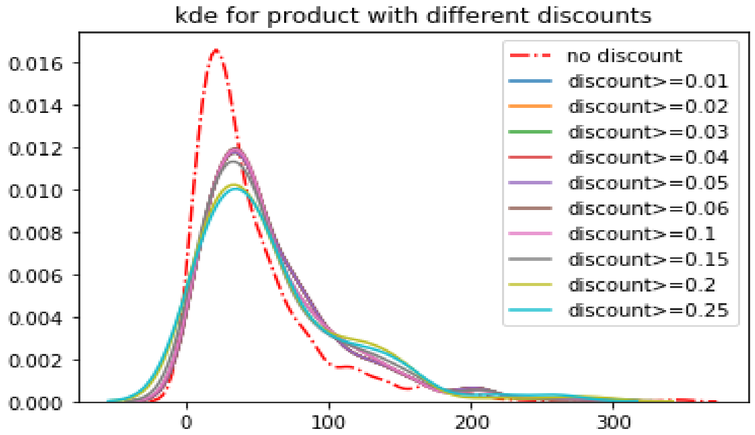
E. DISCOUNT level classifier model

Is it possible to create a classifier model on DISCOUNT level with robust level of prediction accuracy?

A. QUANTITY – DISCOUNT (1/2)

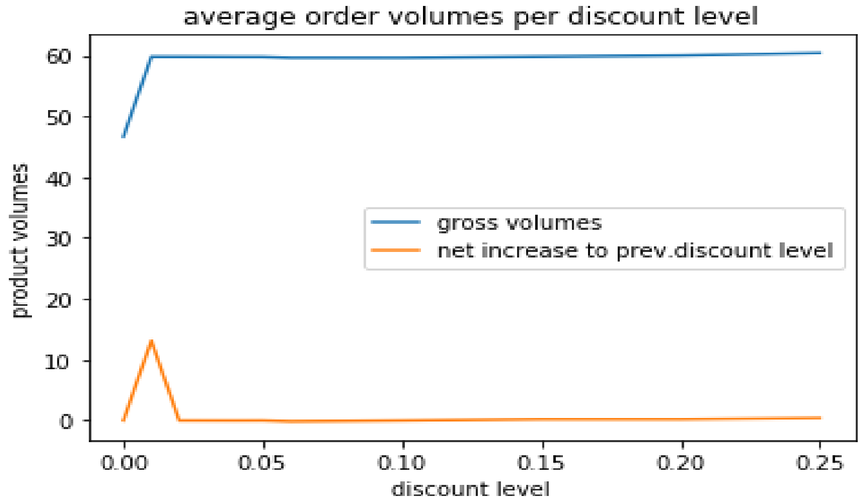
Does DISCOUNT amount have a significant effect on the QUANTITY of a product in an order?

Visual analysis



Product Quantities per Order less for Products without discounts

Marginal increase of Product Quantities: 0% - 2% discount rates



For statistical testing, 2 version of Hypothesis will be tested: on total effect and on marginal effect

A. QUANTITY – DISCOUNT (2/2)

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

Methods and Ho

Test Version 1: testing gross difference

- Ho: E{Quantity | no discount} = E{Quantity | discount}
- Testing method: Welch-test (2-tail)
- Loop: varying discount groups by discount level

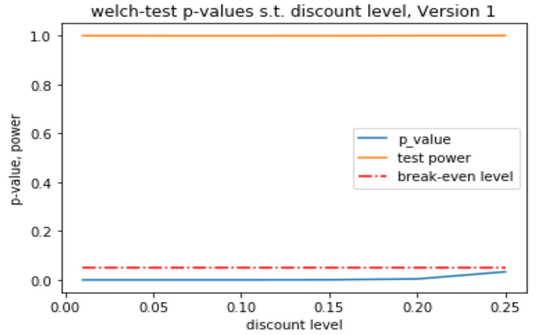
Test Version 2: testing marginal difference

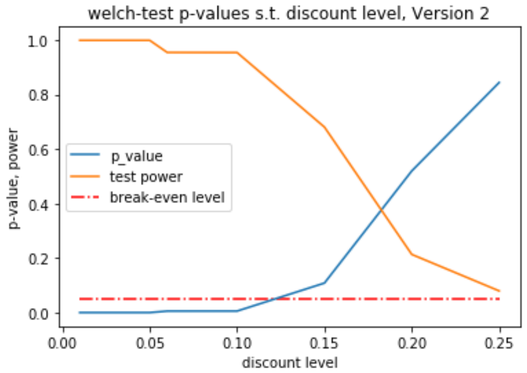
- Ho: E{Quantity | prev.discount} = E{Quantity | next.discount}
- Testing method: Welch-test (2-tail)
- Loop: varying intra-discount groups by discount level

Results

Any discount rate – affect on Product Volumes

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





Marginal effect on Volumes – for 1%-10% discount rates

Issues and Recommendations

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

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

B. SALES – DISCOUNT

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

Methods and Ho

Test Version 1: testing gross difference

- Ho: E{Sales | no discount} = E{Sales | discount}
- Testing method: Welch-test (2-tail)
- Loop: varying discount groups by discount level
- Sales – based on undiscounted original price

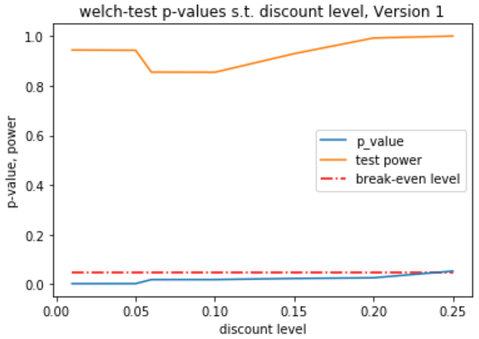
Test Version 2: testing marginal difference

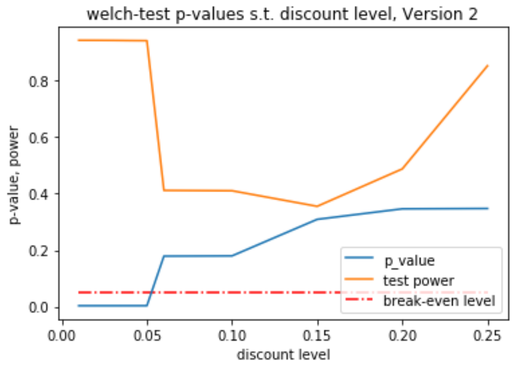
- Ho: E{Sales | prev.discount} = E{Sales | next.discount}
- Testing method: Welch-test (2-tail)
- Loop: varying intra-discount groups by discount level

Results

Any discount rate – affect on Sales in cash terms

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





Marginal effect on Sales – for 1%-5% discount rates

Issues and Recommendations

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

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

C. DELAY – SHIPPERS

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

Methods and Ho

Test Method 1: multiple AB-testing

- Delay = ShippedDate – RequiredDate
- Ho: E{Delay | Shipper A} = E{Delay | Shipper B}
- Testing method: Welch-test (2-tail)
- Shippers: SpeedyExpress, United Package, Federal Shipping
- Loop: varying in all combinations of groups by shipper pairs

Test Method 2: ANOVA

- Testing method: ANOVA
- Formula: Delay ~ C(Shipper)

Results

No significant difference of delays between all three shippers, but test power - low

- P-values: >5% for all pair-groups
- Low test power for all pair-groups
- Test figures:

AB-test group	P-value	Cohen's D	Test power
Federal Shipping – Speedy Express	**0.55**	**0.05**	**0.09**
Federal Shipping – United Package	**0.10**	**0.14**	**0.34**
Speedy Express– United Package	**0.36**	**0.08**	**0.14**

No significant difference of delays between all three shippers

- P-value (F-stat) = 0.27

Issues and Recommendations

I: There is no significant effect of shipper on level of shipping delay

R: Analyze and compare shipping tariffs for all shippers with same service quality

D. DISCOUNT PREDICTORS

What FACTORS have a significant effect on DISCOUNT levels?

Methods and Ho

- Ho: E{Discount | Factor = A} = E{Discount | Factor = B}
- Testing methods: Welch-test (2-tail) and ANOVA
- Factors:
  - Not-ordered product in Stock
  - Product Category (only ANOVA)
  - Supplier (only ANOVA)

Results

No effect on Discount from Unordered product , product category and supplier

FACTOR	P-value (AB)	P-value (ANOVA)
Not-ordered product in Stock	**0.14**	**0.75**
Product Category	**n.a.**	**0.15**
Supplier	**n.a.**	**0.06**

Issues and Recommendations

I: There is no significant effect of not-ordered stock, category and Supplier on Discount Rate

E. DISCOUNT CLASSIFIER

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

Model parameters

- Model type: Naive Bayes Classifier
- Target: Discount rate
- Predictors: Product Price (before discount), Quantity and Supplier
- Model options (Discount levels):
  - A. 2-level (Discount / No discount)
  - B. Rounded levels (0-5-10-15-20-25)
- Train - Test: 0.75 - 0.25

Results

FACTOR	TRAIN	TEST
Model with 2-level discount groups	**0.62**	**0.61**
Model with 6-level discount groups	**0.60**	**0.60**

Issues and Recommendations

I: Accuracy level of classification model is low on given factors (~60%)

Business Recommendations

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

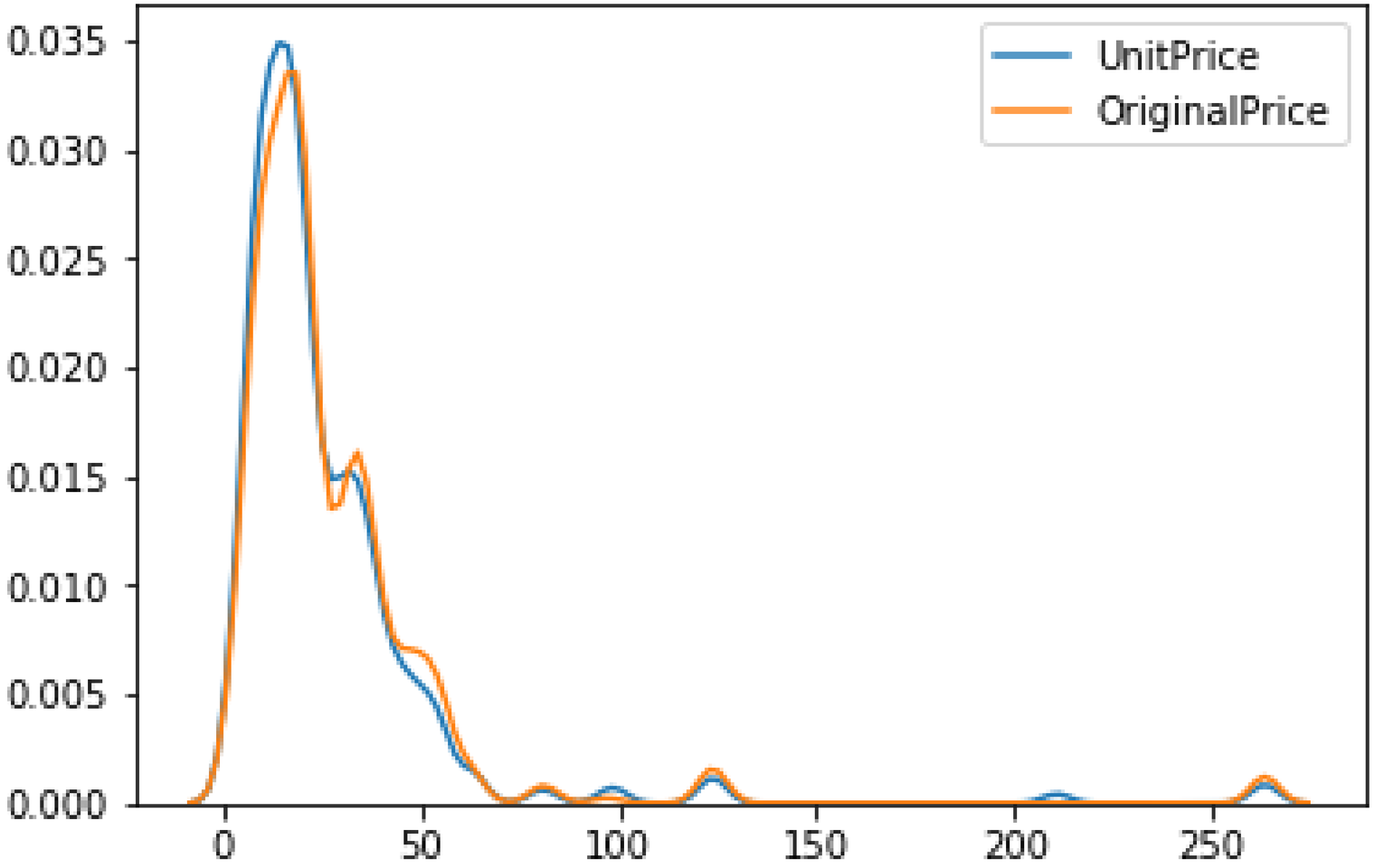
2. Use discounts on product in smaller range (1%-5%) for increasing gross revenues of product sales

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

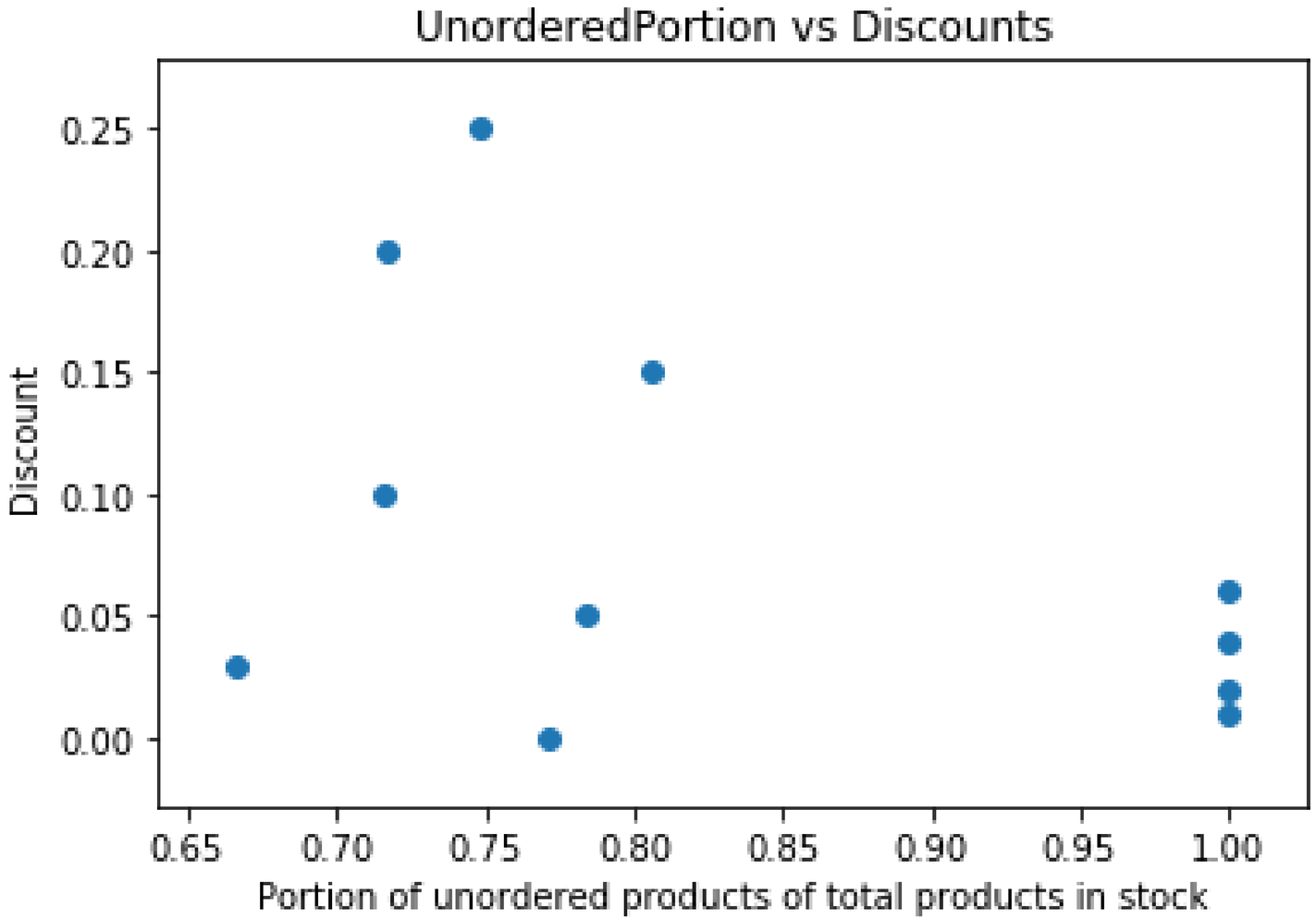
4. Discount rate for a product could be predicted, but with low level of accuracy

APPENDIX

A1. Original price VS Discounted price



A2. Avg.Discount per Unordered product portion



A3. Avg.Discount per product category