Зависимость производства ликеро-водочных изделий

Содержание

Постановка задачи. Первичный анализ исходных данных.

Корреляционно-регрессионный анализ.

Способ 1.

Способ 2.

метод пресс

метод исключения

метод главных компонент

прогнозирование

заключение

# Постановка задачи.

Определить существует ли зависимость между производством ликеро-водочных изделей (Y) и :

1. валовый сбор зерна (X1);
2. валовый сбор сахарной свеклы (X2);
3. потребление пива (X3);
4. население России (X4);
5. потребление водки (X5).

В случае обнаружения зависимости построить оптимальную модель, котороя могла бы быть пригодной для прогноза.

# Первичный анализ исходных данных.

Анализ динамики производства ликеро-водочных изделий (Y) показывает, что за период наблюдения (N=21) минимальное производство был равно 138.1, а максимальным 209.2, тем самым изменение величины Y было в пределах 71.1. Вариация равная 12.2126% свидетельствует об однородности величины Y (<33%). Отклонение от среднего значения (176.5905) в среднем не превышало 17.5814 (среднее абсолютное отклонение), эксцесс (-1.1554) и асимметрия (-0.1873) утверждает, что распределение величины Y имеет незначительный сдвиг влево и достаточно выраженную плосковершинность.   
 Величина Y имеет тенденцию к увеличению, средний темп прироста составляет -0.981% .

Анализ динамики валового сбора зерна (X1) показывает, что за период наблюдения (N=21) минимальный сбор был равен 248.1, а максимальным 356.3, тем самым изменение величины X1 было в пределах 108.2. Вариация равная 10.6046% свидетельствует об однородности величины X1 (<33%). Отклонение от среднего значения (313.5953) в среднем не превышало 33.2555 (среднее абсолютное отклонение), эксцесс (-0.9713) и асимметрия (-0.5517) утверждает, что распределение величины X1 имеет незначительный сдвиг влево и достаточно выраженную плосковершинность.

Величина X1 имеет тенденцию к увеличению, т.к. средний темп прироста составляет 1.0741% или на 0.0254 единиц измерения (% от номинала в миллионах тонн). Сбор до 16 наблюдения имеет тенденцию к увеличению, в период от 16 до 21 наблюдается падение сбора.

Анализ динамики валового сбора сахарной свеклы (X2) показывает, что за период наблюдения (N=21) минимальный сбор был равен 20812, а максимальный 33177, тем самым изменение величины X2 было в пределах 12365. Вариация равная 13.9157% свидетельствует об однородности величины X2 (<33%). Отклонение от среднего значения (26846.0952) в среднем не превышало 3735.8119 (среднее абсолютное отклонение), эксцесс (-1.1144) и асимметрия (0.324) утверждает, что распределение величины X2 имеет незначительный сдвиг вправо и плосковершинность.

Величина X2 имеет тенденцию к увеличению, т.к. средний темп прироста составляет 0.9409%.

Анализ динамики потребление пива (X3) показывает, что за период наблюдения (N=21) минимальное потребление пива было 92.4, а максимальная 106.1, тем самым изменение величины X3 было в пределах 13.7. Вариация равная 3.8059% свидетельствует об однородности величины X3 (<33%). Отклонение от среднего значения (99.5857) в среднем не превышало 3.7902 (среднее абсолютное отклонение), эксцесс (5.6717) и асимметрия (1.4085) утверждает, что распределение величины X3 имеет незначительный сдвиг вправо и достаточно выраженную островершинность.

Величина X3 имеет тенденцию к росту, т.к. средний темп прироста составляет 0.0821% . Потребление пива во время 9 наблюдения имеет резкое падение.

Анализ динамики населения России (X4) показывает, что за период наблюдения (N=21) минимальное население было 130.1, а максимальное 147.4, тем самым изменение величины X4 было в пределах 17.3. Вариация равная 3.6811% свидетельствует об однородности величины X4 (<33%). Отклонение от среднего значения (138.7) в среднем не превышало 5.1057 (среднее абсолютное отклонение), эксцесс (-1.2575) и асимметрия (0.1499) утверждает, что распределение величины X4 имеет незначительный сдвиг вправо и незначительную плосковершинность.

Величина X4 имеет тенденцию к возрастанию, т.к. средний темп прироста составляет 0.6262% .Кривая распределения величины Х4 имеет небольшой подъем вверх.

Анализ динамики потребления водки (X5) показывает, что за период наблюдения (N=21) минимальное потребление было 133.5, а максимальное 208.5, тем самым изменение величины X5 было в пределах 75. Вариация равная 11.4207% свидетельствует о однородности величины X5 (<33%). Отклонение от среднего значения (175.9905) в среднем не превышало 20.0993 (среднее абсолютное отклонение), эксцесс (-0.7625) и асимметрия (-0.1934) утверждает, что распределение величины X5 имеет незначительный сдвиг влево и достаточно выраженную плосковершинность.

Величина X5 имеет тенденцию к уменьшению, т.к. средний темп прироста составляет -1.1457% . Потребление до 13 наблюдения возрастает, затем последовал медленный спад до 21 наблюдения.

# Корреляционно-регрессионный анализ.

Анализ коэффициентов парной корреляции говорит о наличии интенсивной связи Y с Х5 (0.9834), средней с Х4 (-0.5315) -знак минус указывает на обратную зависимость- и Х3 ( -0.4266), слабой с Х2 (-0.1890) и Х1 (0.1176). Значит в модель стоит включить факторы Х3, Х4,Х5.

Следующим этапом идет проверка на мультиколлениарность, существует несколько способов данной проверки.

## Способ 1.

При проверке на мультиколлениарность (коэффициенты частной корреляции и t-статистика) видно, что существует взаимосвязь между:

|  |  |  |  |
| --- | --- | --- | --- |
| x1 | x2 | x3 | x4 |
| x2 | x1 |  | x1 |
| x4 | x4 |  | x2 |
|  |  |  |  |

следовательно в модель включается Х5 и Х4, т.к. коэффициент парной корреляции Y-X4 (-0.5315) больше, чем коэффициенты парной корреляции Y-X1 (0.1170) и Y-X3 (-0.4266) и Y-Х2(-0.1890).

## Способ 2.

Этот метод основан на анализе распределения корреляционной матрицы. Идея метода заключается в том что вводятся некоторые критерии на основе которого можно проверить о значимости отклонения корреляционной матрицы от ортогональной, для этого вводится величина:

Х^2= N-1-1/6(2\*n+5)\*ln|R|

по расчетам ХИ квадрат равно 80.469 больше табличного, значит между переменными существует мультиколлениарность. Для определения степени мультиколлениарности вводим величину:

W=(Cii-1)-(N-n)/(n-1)

где Сii - диагональный элемент матрицы обратной корреляционной.

|  |  |  |
| --- | --- | --- |
| Wii | Wii | f-критерий |
| W11 | 3.622 | 0.0139 |
| W22 | 1.93 | 0.12648 |
| W33 | 6.18 | 0.00081 |
| W44 | 2.181 | 0.08999 |
| W55 | 6.225 | 0.00077 |

Данная таблица указывает, что наиболее коллениарна Х2, затем Х4 и можно сказать что Х3 и Х5 вовсе не коллениарны. Следовательно в модель лучше включить Х3 и Х5, но проведенный последующий регрессионный анализ указывает что лучше включать в модель Х2 и Х3, т.е. производство ликеро-водочных изделий (Y) зависит от валового сбора сахарной свеклы (X2) и потребления пива (X3).

Анализ уравнения регрессии говорит, что при росте Х5 на 1 единицу в своих единицах измерения увеличит Y на 1.0552 единицы в своих единицах измерения, Отклонения основного тренда носят случайный характер, а данная модель определяет Y на 96.71% ( R-квадрат). Относительная ошибка апроксимации указывает об адекватности математической модели. Степень рассеянности Y мала (дисперсия=3.909). Распределение Y является нормальным, в ряду нет автокорреляции нельзя , а проверка на стационарность случайного компонента с помощью Х^2 (Х^2=10.04) указывает что коэффициенты корреляции неоднородны.

# Метод пресс.

Основан на выборе наилучшего уравнения регрессии для этого рассчитывают значения сумм квадратов расхождения:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Хi | отклонение | Хi | отклонение | Хi | отклонение | Хi | отклонение | Хi | отклонение |
| 1 | 9174.74 | 12 | 5598.67 | 123 | 5589.96 | 1234 | 538.735 | 12345 | 185.547 |
| 2 | 8969.93 | 13 | 7329.06 | 124 | 545.654 | 1235 | 217.694 |  |  |
| 3 | 7608.97 | 14 | 2226.17 | ***125*** | ***217.86*** | ***1245*** | ***185.690*** |  |  |
| 4 | 6674.29 | 15 | 256.857 | 134 | 1176.13 | 1345 | 236.652 |  |  |
| ***5*** | ***305.611*** | 23 | 7607.95 | 135 | 240.845 | 2345 | 224.784 |  |  |
|  |  | 24 | 256.856 | 145 | 256.53 |  |  |  |  |
|  |  | ***25*** | ***227.26*** | 234 | 3506.0 |  |  |  |  |
|  |  | 34 | 5628.28 | 235 | 224.949 |  |  |  |  |
|  |  | 35 | 275.868 | 245 | 226.924 |  |  |  |  |
|  |  | 45 | 266.522 | 345 | 236.662 |  |  |  |  |

Из таблицы видно лучше всего взять модель 25 или 125.

|  |  |  |
| --- | --- | --- |
| модель | R2 | дисперсия |
| 25 | 0.9756 | 3.3709 |
| 125 | 0.9766 | 3.3005 |

Последующая проверка говорит, что модель 25 наиболее выгодна. Значит

производство ликеро-водочных изделий (Y) зависит от 2- валового сбора сахарной свеклы (X2), 5- потребления водки (X5) на 97.66%.

# Метод исключения.

Метод исключения основан на анализе коэффициентов регрессионного уравнения при условии, что переменная при этом коэффициенте в модель была включена последней.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| переменные в моделе | f-кри-  терий | переменные в моделе | f-кри-  терий | переменные в моделе | f-кри-  терий | переменные в моделе | f-кри-  терий | переменные в моделе | f-кри-  терий |
| Х1 | 3.1719 | Х1 | 0.5331 | Х1 | 0.7335 |  |  |  |  |
| Х2 | 4.1314 | Х2 | 1.7014 | Х2 | 3.0429 | Х2 | 1.8365 |  |  |
| Х3 | 0.0115 | Х3 | 0.0121 |  |  |  |  |  |  |
| Х4 | 2.5988 |  |  |  |  | Х4 | 8.6594 |  |  |
| Х5 | 28.553 | Х5 | 394.844 | Х5 | 419.872 |  |  | Х5 | 23.6498 |
| Fкр | 4.4100 | Fкр | 4.4100 | Fкр | 4.4100 | Fкр | 4.4100 | Fкр | 4.4100 |

Следовательно в модель включается только Х5. Данная модель определяет Y на 96.71%, значит потребление водки (X5) значительно влияет на производство ликеро-водочных изделий (Y).

# Метод главных компонент.

Метод главных компонент был предложен К. Пирсоном в 1901 году, а в дальнейшем развит и доработан. Метод основан на стандартизации переменных для чего используют следующие формулы:

Zij=(Xij-Xiсред)Si ;

Si=[1/(n-1)\*сумма(Xij-Xiсред)^2]^(1/2) ;

где Zij стандартизованные переменные;

Si стандартизированное отклонение.

В модели участвуют главные компоненты Wj, которые представляют собой следующее:

Wj=V1Z1+V2Z2+...+VrZr

где Vj собственный вектор, который удовлетворяет системе уравнений:

(Z’z-KI)\*Vj=0

где Z’z корреляционная матрица;

КI характеристические корни уравнения | Z’z-KI|=0 .

Корреляция главных компонент показывает тесноту связи Хi с главными компонентами. Переменные Х1,Х2,Х4 имеют интенсивную связь с первой главной компонентой, а Х3 среднюю, вторая главная компонента интенсивно связана с переменной Х5. Следовательно валовый сбор зерна (X1), валовый сбор сахарной свеклы (X2), население России (X4), потребление пива (X5) имеют некоторую гипотетическую величину, зависимую от них. Модель полученная по методу главных компонент определяет величину Y на 87.43% ( R квадрат).

# Прогнозирование.

Проведем прогнозы по полученным моделям и сделаем оценки прогнозов.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| прогноз | Gt | Dср | Eпр-сред | K | KH | KH1 | V | Vмю | Vs | Vl |
| регрессия от факторов | **2.5273** | **1.552086** | **0.843786** | **0.13734** | **0.015911** | **0.0164** | **0.1373** | **0.008** | **0.009699** | **169.4348** |
| регрессия от главных компонент | *6.633742* | *4.78329* | *2.587049* | *0.360434* | *0.041764* | *0.0432* | *0.3604* | *0.002* | *0.076127* | *124.1527* |
| экспоненциальное сглаживание | *11.42036* | *7.739524* | *3.974608* | *0.62061* | *0.071899* | *0.0744* | *0.6206* | *0.006* | *0.169182* | *168.1134* |
| метод гармонических весов | *8.637442* | *3.711905* | *2.035688* | *0.46938* | *0.054378* | *0.0563* | *0.4693* | 0.018 | *0.074788* | *157.9697* |
| регрессия от времени | *16.61707* | *11.85095* | *6.213912* | *0.903012* | *0.104615* | *0.1083* | *0.903* | *0.012* | *0.169182* | *263.5587* |

Из данной таблицы видно, что наиболее точной моделью прогноза считается регрессия от факторов, т.к. Gt=2.5273. Eпр-сред указывает о точности высокой точности прогноза, К - о том что данная модель довольно сильно близка к эталонной (простая экстрополяция), КН - модель близка к совершенной, а КН1 - что модель лучше чем модель на уровне средней, V - что модель близка к простой экстрополяции, Vмю - что центральная тенденция определена точно, Vs - что отклонения фактических и прогнозных достаточно точно совпадают, Vl - слабая связь между прогнозными и фактическими значениями.

**Заключение.**

Основными выводами по проведенной работе можно считать следующее:

1- производство ликеро-водочных изделий (Y) имеет тенденцию к постоянному росту;

2 - наиболее сильно оно зависит от потребления водки (Х5) и от валового сбора сахарной свеклы (X2) ;

3 - наиболее лучшей моделью для проведения прогноза служит модель полученная по корреляционно-регрессионному методу , которая на 97,66% описывает

производство ликеро-водочных изделий (Y);

4 - прогноз следует проводить по модели регрессии от факторов, характеристики которой наиболее достоверные;

5 - для построения наиболее точной модели следует рассмотреть большее количество факторов, влияние которых в большей мере бы определяло производство ликеро-водочных изделий (Y);

6 - влияние валового сбора зерна (X1), потребления пива (Х3) и населения России (Х4)фактически не существенно сказывается на изменение производства ликеро-водочных изделий (Y);

7 - полученная модель пригодна для прогноза лишь на краткосрочный период.