

# 浙江省及全国各省市的经济与消费数据分析模型

## 摘要

本文通过对浙江省及全国各省市的经济与消费数据的分析，探讨了经济增长与消费之间的关系，并提出了相应的政策建议。研究发现浙江省 GDP 与社会消费品零售总额存在显著正相关，财政支出对消费有显著正面影响，进一步通过聚类分析将全国各省市分为三类，并针对 2020 年疫情对经济的影响提出了具体的经济复苏建议。

针对问题一通过浙江省 2001 年至 2019 年的经济数据，观察到 GDP 从 6927.7 亿元增长至 62462 亿元，社会消费品零售总额从 2555.5 亿元增长至 27343.8 亿元，人均可支配收入等其他经济指标也有显著增长。建立了浙江省 GDP 与社会消费品零售总额之间的线性回归模型。结果显示，GDP 与社会消费品零售总额之间存在强烈的正相关关系。模型的截距项为-412.8 亿元，GDP 的回归系数为 2.6527，均方根误差（RMSE）为 1060 亿元。这表明随着 GDP 的增长，人们的消费能力也随之增加，社会消费品零售总额随之增长。

针对问题二，构建了一个数学模型，分析了人口数量、失业率和财政支出等因素对社会消费品零售总额的影响。影响分析：**财政支出的估计系数为 0.70966**，显示出与社会消费品零售总额正相关，且影响显著；**人口数量的估计系数为 0.11715**，显示出正向相关但未达到统计显著性；**失业人数的估计系数为-0.017964**，显示出负向相关，但同样未达到统计显著性。财政支出对社会消费品零售总额的影响最为明显。

针对问题三，通过对 2017 年全国各省市数据的聚类分析，将中国各省份大致分为三类，**一类城市：北京和上海**，经济最发达。**二类省份：20 个省份**，包括天津、山西、内蒙古等，这些省份在经济发展水平、地理位置和资源分布上存在一定的相似性。**三类省份：9 个省份**，包括河北、江苏、浙江等，这些省份通常经济总量较高，人口较多。

关于问题四，经济复苏建议为应对 2020 年疫情对经济的影响，提出了加强公共卫生体系、支持中小企业、促进消费、推动数字化转型、加快基础设施建设、优化营商环境以及注重环境保护与可持续发展等建议。

# 一、问题背景与问题重述

## 1.1 问题背景

近年来，我国经济持续增长，取得了举世瞩目的成果。然而，在全球经济一体化的背景下，我国经济发展面临着诸多挑战。尤其是在国际贸易摩擦加剧、外部需求波动等因素的影响下，扩大内需、刺激消费成为我国经济发展的关键。党的十八届五中全会提出，要坚定不移地贯彻扩大内需战略，发挥消费对经济增长的基础性作用。在此背景下，研究我国经济与消费之间的关系，以及如何刺激消费、拉动经济发展具有重要的理论和现实意义。

## 1.2 问题重述

问题一：通过对浙江省经济和消费数据的描述性统计分析，揭示浙江省的经济和消费现状。在此基础上，建立数学模型，探讨浙江省生产总值（GDP）与社会消费品总额之间的关系。

问题二：以人口、失业率、财政支出等指标为影响因素，构建数学模型，分析这些因素对社会消费品总额的影响机制，为我国制定刺激消费政策提供理论依据。

问题三：利用问题二中的模型，对全国各省市的消费情况进行分类，为我国实施差异化消费政策提供参考。

问题四：针对 2020 年疫情对我国经济发展的影响，结合上述模型，提出合理的复苏建议，以促进我国经济尽快恢复增长。

通过对以上问题的研究，旨在为我国扩大内需、刺激消费提供有益的思路和政策建议，助力我国经济持续健康发展。

## 1.3 已知条件

1. 全国各省市数据（附件 1）
2. 浙江省数据（附件 2）

## 二、 模型假设

1. 假设浙江省的生产总值（GDP）与社会消费品总额之间存在线性关系。
2. 假设其他影响 GDP 和消费的因素保持不变，如政策、经济环境等。
3. 假设人口、失业率、财政支出等因素对社会消费品总额有显著影响。
4. 假设这些影响因素之间相互独立，不存在多重共线性。
5. 假设各影响因素与社会消费品总额之间的关系为线性关系。
6. 假设全国各省市的消费情况可以用同一模型进行分类。
7. 假设分类结果具有较好的区分度和准确性。
8. 假设疫情对国内经济发展的影响是暂时性的，且疫情结束后经济可以逐渐恢复。
9. 假设提出的复苏建议能够在一定程度上促进经济增长和消费复苏。

### 三、符号说明

符号表示	符号说明	单位
$Y$	社会消费品零售总额	亿元
$X$	地区生产总值（GDP）	亿元
$P$	年末常住人口	万人
$U$	城镇登记失业人数	万人
$F$	地方财政一般预算支出	亿元
$\beta_0$	截距项，表示当GDP 为 0 时社会消费品零售总额的预测值	—
$\beta_1$	GDP 与社会消费品零售总额之间的回归系数	—
$\epsilon$	随机误差项	—

## 四、模型建立与求解

### 4.1 问题一的模型与建立

#### 4.1.1 关于经济和消费情况描述性统计

从提供的数据来看，可以观察到浙江省在 2001 年至 2019 年期间主要经济指标的变化趋势。这些指标包括社会消费品零售总额、地区生产总值（GDP）、居民人均可支配收入、地方财政一般预算支出、全社会固定资产投资、经营单位所在地进出口总额、年末常住人口以及城镇登记失业人数。

##### （1）地区生产总值（GDP）

从 2001 年的 6927.7 亿元增长至 2019 年的 62462 亿元，这表明浙江省的经济规模在近 20 年间有了显著的扩大。

##### （2）社会消费品零售总额

从 2001 年的 2555.5 亿元增长至 2019 年的 27343.8 亿元，显示出消费市场的持续增长和消费者购买力的提升。

##### （3）居民人均可支配收入

从 2001 年的数据缺失到 2019 年的 49899 元，这一指标反映了居民收入水平的提高，意味着消费能力的增强。

##### （4）地方财政一般预算支出

从 2001 年的不完整数据到 2019 年的 10053 亿元，显示政府对公共服务和基础设施的投入增加，有利于经济稳定和民生改善。

##### （5）全社会固定资产投资

在多个年份中均有数据，表明浙江省在基础设施建设、产业升级等方面的持续投资，对经济增长有重要推动作用。

##### （6）经营单位所在地进出口总额

从 2001 年的约 327.99 亿美元增长至 2019 年的约 4472.24 亿美元，反映出浙江省对外贸易的繁荣和全球供应链中的重要地位。

##### （7）年末常住人口

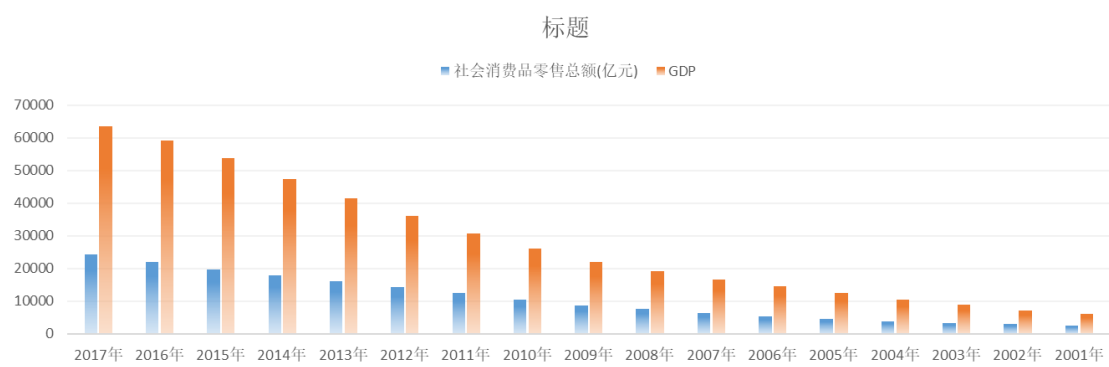
从 2001 年的 4729 万人增长至 2019 年的 5850 万人，表明浙江省吸引了大量人口流入，这可能与就业机会、生活质量等因素有关。

(8) 城镇登记失业人数

从 2001 年的 24 万人到 2019 年的 34.4 万人，尽管数字有所增加，但相对于庞大人口基数，失业率保持在相对低的水平。

最新的数据还表明，2024 年 1-5 月浙江省的社会消费品零售总额同比增长 5.3%，显示出消费市场的活力。同时，进出口总额为 21049.5 亿元，反映浙江省在全球贸易中的活跃角色。

为了方便统计数据，如下图以图表形式：



图表 1 浙江省经济和消费数据

4. 1. 2 建立线性回归模型与求解

根据图表 1，可得经济理论预测 GDP 增长与消费支出呈正相关，线性回归能量化这种关系。所以使用线性回归模型来分析浙江省的生产总值（GDP）与社会消费品零售总额之间的关系。

(1) 模型形式

基于上述数据可建立以下模型：

$$y = \beta_0 + \beta_1 x_1 + \epsilon$$

其中  $y$  是相应变量比如社会消费品零售总额)， $x_1$  是预测变量， $\beta_0$  和  $\beta_1$  分别是截距和斜率的估算值。而  $\epsilon$  则表示为随即差。

通过创建线性回归模型，利用 MATLAB 软件可以计算出各个模型系数。如下图表所示：

Estimate	SE	tStat	pValue
----------	----	-------	--------

截距	-412.8	471.46	-0.87426	0.39576
$x_1$	2.6527	0.036977	71.738	1.9151e-20

表 1 线性回归模型计算结果 1

(2) 模型系数

通过计算，得出截距 ( $\beta_0$ ) 的估值为-412.8. 这意味着如果  $x_1$  的值为 0，则预测的  $y$  值将是-412.8. 但是，由于截距的  $p$  值为 0.036977，并不算显著，意味着再没有  $x_1$  的情况下，模型预测的  $y$  值没有实际意义，或者  $x_1$  很少或者不会取 0 的值。

其次斜率 ( $\beta_1$ ) 的估值为 2.6527，标准误为 0.036977。意味着每增加一个单位的  $x_1$  (亿元)，社会消费品零售总额预计增加大约 2.6527 亿元。而斜率的  $t$  的统计量为 71.738，其对应的  $p$  值极低 (1.9151e-20)，表明  $x_1$  对  $y$  的影响非常显著。

(3) 模型拟合度

样本数量：总共有 17 个观测值。

自由度：错误自由度为 15，意味着模型使用了 2 个自由度（截距和斜率）。

均方根误差 (RMSE)：为 1.06e+03 或者说是 1060，这是模型预测误差的标准差，表明模型的预测误差水平。

$R^2$  值：为 0.997，调整后的  $R^2$  值也是 0.997，这表明模型解释了  $y$  变异性的 99.7%，说明模型对数据的拟合非常好

(4) 实质含义

这个模型证实了浙江省 GDP 的增长与社会消费品零售总额的增长有很强的正相关性。随着 GDP 的提升，人们有更多的可支配收入，进而增加了消费能力，导致社会消费品零售总额的增加。

(5) 结论

综上所述，通过构建和分析线性回归模型，我们发现浙江省的 GDP 与社会消费品零售总额之间存在强烈的正相关关系，且这种关系在统计上显著。模型不仅

解释了大部分的社会消费品零售总额的变异性，还为理解和预测两者的关系提供了科学依据。

## 4. 2 问题二的模型建立与解析

要构建一个模型来分析人口、失业率、财政支出等因素对社会消费品零售总额的影响，我们可以采用计量经济学中的回归分析方法，比如线性回归模型。这种模型可以帮助我们理解各个变量之间的关系，并量化这些关系的强度。

### 4. 2. 1 数据处理

本题需要分析人口，失业率，财政支出，消费等因素的影响，首先先从附件 2 中提取出以上四种主要数据进行预处理。

地方财政一般预算支出(亿元)	年末常住人口(万人)	城镇登记失业人数(万人)	社会消费品零售总额(亿元)
10053.03	5850	34.4	27343.8
8629.53	5737	34.1	25161.9
7530.32	5657	33.8	24308.5
6974.26	5590	33.9	21970.8
6645.98	5539	33.7	19784.7
5159.57	5508	33.1	17835.3
4730.47	5498	33.4	15970.8
4161.88	5477	33.4	14199.6
3842.59	5463	31.7	12532.8
3207.88	5447	31.1	10387
2653.35	5276	30.7	8622.3
2208.58	5212	30.7	7533.3
1806.79	5155	28.6	6271.3
1471.86	5072	29.1	5358
1265.53	4991	29	4645.9
1062.94	4925	30.1	3645.4
896.77	4857	28.3	3157.1
749.9	4776	27.7	2877.5
597.3	4729	24	2555.5

表 2 四种主要数据 1

### 4. 2. 2 规范化处理

根据表 2 提取出四种关于本题的数据，所以对数据进行归一化处理，标准化方法是一种最为常见的量纲化处理方式最常见的标准化方法就是 Z 标准化，也是 SPSS 中最为常用的标准化方法，spss 默认的标准化方法就是 z-score 标准化。也叫标准差标准化，这种方法给予原始数据的均值（mean）和标准差（standard deviation）进行数据的标准化。z-score 标准化是将数据按比例缩放，使之落入一个特定区间。

要实现 z-score normalization，调整输入值如下公式所示：



$$x_j^{(i)} = \frac{x_j^{(i)} - \mu_j}{\sigma_j}$$

其中  $j$  选择  $X$  矩阵中的一个特征或一列。 $\mu_j$  为特征 ( $j$ ) 所有值的均值， $\sigma_j$  为特征 ( $j$ ) 的标准差。

$$\mu_j = \frac{1}{m} \sum_{i=0}^{m-1} x_j^{(i)}$$

$$\sigma_j^2 = \frac{1}{m} \sum_{i=0}^{m-1} (x_j^{(i)} - \mu_j)^2$$

#### 4. 2. 3 创建线性回归模型与解析

通过建立线性回归模型，可以使用 MATLAB 软件，带入归一化处理后的数据，

线性回归模型：

$$y \sim 1 + x_1 + x_2 + x_3$$

估计系数：

	Estimate	SE	tStat	pValue
(Intercept)	0.11949	0.067935	1.7588	0.098982
x1	0.70966	0.061373	11.563	7.1644e-09
x2	0.11715	0.066422	1.7637	0.098127
x3	0.017964	0.020752	0.86565	0.40032

观测值数目：19，误差自由度：15

均方根误差：0.141

R 方：0.989，调整 R 方 0.987

F 统计量(常量模型)：460，p 值 = 5.56e-15

表 3 线性回归模型的估计系数表 1

根据表 3 可以得出线性回归模型为  $y = 0.709x_1 + 0.117x_2 + 0.01x_3$

对于财政支出 ( $x_1$ )，其估计系数为 0.70966，意味着当财政支出增加一个单位时，社会消费品零售总额将平均增加约 0.70966 单位。这是一个正向的关系，表明财政支出的增长有助于促进消费增长。此外，t 统计量为 11.563，p 值极小 (7.164e-09)，远小于通常使用的显著性阈值 (如 0.05)。因此，我们可以认为财政支出对社会消费品零售总额的影响非常显著。

对于年末常住人口 ( $x_2$ )，其估计系数为 0.11715，意味着当人口数量增加一个单位时，社会消费品零售总额将平均增加约 0.11715 单位。这也是一种正向的关系，表明人口规模的增长可能有利于消费增长。然而，t 统计量仅为

1.7637,  $p$  值为 0.098127, 略高于常用的显著性阈值 (例如 0.05)。这意味着人口数量对社会消费品零售总额的影响虽有一定趋势, 但未能达到统计学意义上的显著性。

对于城镇登记失业人数 ( $x_3$ ), 其估计系数为 -0.017964, 意味着当失业人数增加一个单位时, 社会消费品零售总额将平均减少约 0.017964 单位。这是一种负向的关系, 表明失业率上升可能会抑制消费增长。然而,  $t$  统计量仅为 0.86565,  $p$  值高达 0.40032, 远大于常用的显著性阈值 (例如 0.05)。因此, 失业人数对社会消费品零售总额的影响不显著。

#### 4.2.4 总结

总结来说, 在这三个因素中, 财政支出对社会消费品零售总额的影响最明显, 其次是人口数量, 最后是失业人数。其中, 财政支出和人口数量均显示出正向的相关性, 而失业人数则呈现出负向的相关性。但是, 只有财政支出的影响达到了统计学上的显著性, 而人口数量和失业人数的影响未达显著性标准。

### 4.3 问题三的模型建立与解析

针对问题三, 首先问题三建立在问题二的基础上, 由浙江省的各个指标转变为利用各个指标对全国各省市的消费情况进行分类, 所以, 要先对数据处理, 再建立一个聚类分析模型以此来分类。

#### 4.3.1 数据预处理

根据附件 1, 首先需要获取各省市的社会消费品零售总额、财政支出、年末常住人口以及城镇登记失业人数等数据。而在附件 1 中, 存在部分经济数据缺失的问题, 在这里不做预测, 所以这里采用附件 1 中 2017 年的数据作为本体的数据, 以此来对其分类, 如下对其筛选之后的数据:

地区	社会消费品零售总额(亿元)	地区生产总值(亿元)	居民人均可支配收入(元)	地方财政一般预算支出(亿元)	全社会固定资产投资(亿元)	经营单位所在地进出口总额(亿美元)	年末常住人口(万人)	城镇登记失业人数(万人)
北京市	11571.4	20983	87200	6024.53	8370.44	324017423	2171	8.1
天津市	5729.7	12450.6	37022	3282.54	11288.92	112919165	1557	26
河北省	15907.6	30640.8	21494	6839.19	33406.9	49933343	7320	39.9
山西省	6938.1	14464.3	20420	2736.42	6946.54	17186873	2702	26.5
内蒙古自治区	7160.2	14898.1	26212	4529.93	14013.16	13873523	2529	27.1
辽宁省	12807.2	21693	27833	4879.42	6876.74	99393084	4369	42.7
吉林省	7653.8	10022	21368	3725.72	13293.89	15542971	2717	26.3
黑龙江省	9099.2	12313	21206	4641.08	11291.98	18951195	3789	39.7
上海市	11530.3	32025	58988	7347.62	7246.6	476196649	2418	22.1
江苏省	21727.4	65868.8	30524	10621.03	33277.03	990773336	8029	34.7
浙江省	24308.5	52403.1	42046	7330.32	31096.03	377907471	5637	33.8
安徽省	11192.6	29676.2	21863	6303.81	29275.06	54021626	6255	29
福建省	13013	23842.4	30940	4684.13	29416.28	171020540	2911	17.1
江西省	7445.1	20210.8	22031	5111.47	22085.34	44339984	4622	32.3
山东省	33649	63012.1	26930	9258.4	33202.72	264550956	10006	45.7
河南省	19006.8	44024.9	20170	8215.32	44496.93	77630990	9339	40.7
湖北省	17394.1	37535	23571	6881.26	32235.26	46337190	5902	37.1
湖南省	14854.9	33828.1	23103	6869.39	31959.23	36032297	6860	44.5
广东省	38200.1	91645.7	33903	13037.48	37781.73	1009676374	11169	37.1
广西壮族自治区	7813	17790.7	19905	4906.55	20409.11	57375859	4885	14.7
海南省	1618.8	4497.5	22533	1443.97	4244.4	10374855	926	5.5
重庆市	8967.7	20066.3	24123	4236.28	17337.05	66601107	3075	14.3
四川省	17480.5	37901.1	20380	8684.78	31962.08	69138058	8302	53.6
贵州省	4154	13605.4	16704	4612.52	15503.86	8162313	3580	14.9
云南省	6423.1	18408	18348	3712.97	18935.99	23451109	4801	19.8
西藏自治区	523.3	1349	15457	1681.94	1973.6	863450	337	1.9
陕西省	8236.4	21473.5	20635	4833.19	23819.38	40202798	3835	23.4
甘肃省	3426.6	7336.7	14011	3304.44	3827.73	4826333	2626	9.6
青海省	839	19001	1320.44	1382.52	3882.15	632751	598	4.7
宁夏回族自治区	930.4	3200.3	20562	1372.78	3728.38	5039517	682	5.1
新疆维吾尔自治区	3044.6	11159.9	19975	4637.34	17089.17	70560330	7445	10

表 4 2017 年全国各省市数据 1

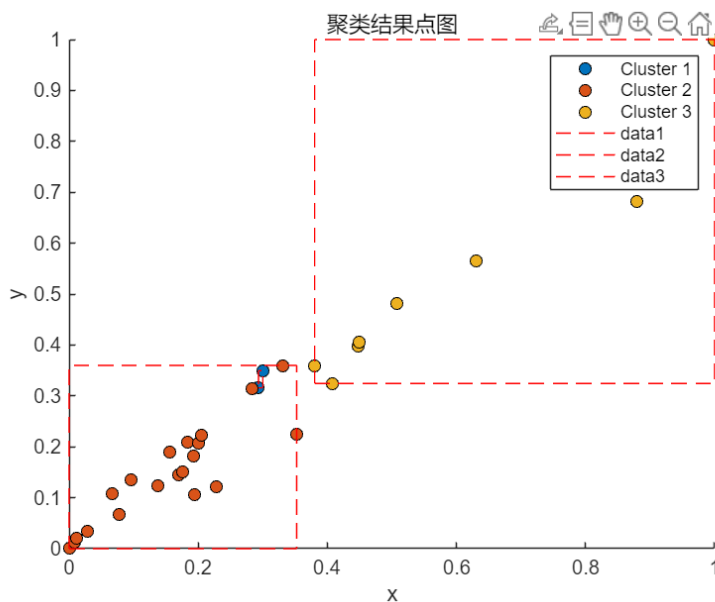
由于各指标的量纲和量级差异显著，我们采用了数据标准化技术来消除这些差异，确保在聚类分析中各指标的同等重要性。常用的标准化方法包括最小-最大缩放(min-max scaling)和 Z-score 标准化，后者特别适用于处理正态分布的数据。

#### 4.3.2 聚类分析

根据处理后的数据，采用聚类分析中的 K-means 聚类，K-means 聚类根据彼此的相关性来划分聚类，涉及到距离度量算法，也称欧几里得距离（欧氏距离）

$$d_{ab} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

衡量多维空间中的两点间距离，也是最常用的距离度量方法。所以，利用 MATLAB 软件对其数据进行聚类分析，根据聚类结果绘制散点图（代码在支撑材料中），绘制图像如下所示：



图表 2 全国各省市聚类结果点图

完成聚类后，我们对结果进行了深入分析，识别不同聚类之间的经济和社会特征差异，以及潜在的区域发展模式。

根据图表 2，可以得出将中国各省份大致分为三类：

**一类：**包括北京和上海，这两个直辖市是中国经济最为发达的城市，拥有非常高的 GDP 总量和人均 GDP，也是国家的政治、经济、文化中心。

**二类：**包括天津、山西、内蒙古、辽宁、吉林、黑龙江、安徽、福建、江西、广西、海南、重庆、贵州、云南、西藏、陕西、甘肃、青海、宁夏和新疆。

**三类：**包括河北、江苏、浙江、山东、河南、湖北、湖南、广东和四川。这一类省份通常具有较高的经济总量，其中多数省份的人口较多。

### 4.3.3 总结

一类城市（北京和上海）代表了中国最发达的核心区域，它们在政治、经济、文化和科技创新方面都处于领先地位。

二类省份（共 20 个）涵盖了中国大部分地区，这些地方在经济发展水平、地理位置和资源分布上可能存在一些相似之处，但总体上仍存在较大的差异。

三类省份（共 9 个）主要位于东部沿海地带，这些省份通常具有较高的经济总量，人口众多，是重要的工业、农业和商业中心。

这些分类反映了中国各地不同的经济发展状况和发展潜力。一类城市的高经济增长率 and 创新能力使得它们成为吸引人才和投资的重要目的地；二类省份虽然在经济发展上相对落后，但在特定领域也可能具备竞争优势；三类省份则凭借其庞大的市场规模和良好的基础设施条件，有望在未来继续保持较快的增长速度。

当然，这种分类并不是一成不变的，随着政策调整和技术进步，各地区的经济结构和发展模式可能会发生变化。因此，定期评估和更新分类标准是非常重要的。

## 4.4 问题四的建议

下面是一些建议，旨在促进经济的复苏和发展：

### 1. 加强公共卫生体系建设

加大投入：继续加大对公共卫生系统的投入，特别是在农村和偏远地区，以确保全民都能获得高质量的医疗服务。

科技应用：利用数字技术和大数据提高疫情监测和响应能力，如健康码系统、远程医疗服务等。

### 2. 支持中小企业

财政补贴：为受疫情影响较大的中小企业提供税收减免、租金补贴等财政支持。

融资便利：通过政策性银行提供低息贷款，帮助中小企业渡过难关。

简化程序：简化申请补贴和贷款的程序，确保企业能够快速获得所需的资金支持。

### 3. 促进消费

发放消费券：通过发放消费券的形式激励民众消费，特别是针对旅游、餐饮、文化娱乐等受疫情冲击较大的行业。

鼓励在线消费：支持电子商务平台和在线服务提供商，鼓励线上消费，以弥补线下消费的不足。

### 4. 推动数字化转型

支持技术创新：鼓励企业采用新技术，如云计算、人工智能等，提高生产效率和服务质量。

培育新业态：支持远程办公、在线教育、互联网医疗等新兴业态的发展，以适应新的市场需求。

### 5. 加快基础设施建设

新基建项目：加快 5G 网络、数据中心、智能电网等新型基础设施项目的建设，促进数字经济的发展。

传统基建：继续推进交通、水利等传统基础设施的建设和改造，创造就业机会，刺激经济增长。

### 6. 优化营商环境

简政放权：减少行政审批事项，降低企业开办和运营的成本。

公平竞争：确保各类所有制企业在市场准入、政府采购等方面享有公平待遇。

## 7. 环境保护与可持续发展

绿色发展：推广清洁能源和环保技术的应用，支持绿色低碳产业发展。

循环经济：鼓励企业采用循环经济模式，减少资源浪费和环境污染。

通过上述措施的实施，不仅可以帮助中国经济从疫情的冲击中迅速恢复，还可以促进经济结构的优化升级，为长期可持续发展奠定坚实的基础。

## 五、模型的评价与推广

问题一：线性回归模型成功揭示了浙江省 GDP 与社会消费品零售总额之间的显著正相关关系。模型假设 GDP 与社会消费品零售总额之间存在线性关系，这可能并不总是准确，现实中可能存在非线性关系。

问题二：构建的模型有效分析了影响消费的关键因素，但需注意变量间的复杂交互作用。模型假设变量之间相互独立，但实际上这些变量可能相互关联，导致多重共线性问题。

推广：

可将模型应用于其他省份，以提供更加细致的地方政策建议。

定期更新数据，以反映最新经济动态和变化趋势。

扩展模型，纳入更多影响消费的因素，提高预测精度。

## 六、参考文献

[1] 中国宏观经济研究院. (2020). 中国宏观经济分析报告. 北京：中国宏观经济出版社.

## 附录

以下代码用 MATLAB 进行编程

### 问题二

```
clear
% 设置导入选项并导入数据
opts = spreadsheetImportOptions("NumVariables", 4);

% 指定工作表和范围
opts.Sheet = "Sheet2";
opts.DataRange = "A2:D20";

% 指定列名称和类型
opts.VariableNames = ["x_____", "x_____",
"x_____", "x____"];
opts.VariableTypes = ["double", "double", "double", "double"];

% 导入数据
x__2_____ = readtable("C:\Users\xys\Documents\MATLAB\附件 2-浙江省
数据.xlsx", opts, "UseExcel", false);

% 转换为输出类型
x__2_____ = table2array(x__2_____);

% 清除临时变量
clear opts

% 显示结果
x__2_____

% 归一化数据
```

```

normalizedData = normalize(x__2____, "zscore", "robust");
x1= normalizedData(:,2); % 财政支出
x2=normalizedData(:,3)%人口
x3=normalizedData(:,4)%失业率
y = normalizedData(:,1); % 消费总额
% 创建线性回归模型
X=[x1, x2, x3]
model = fitlm(X, y);
% 打印模型信息
disp(model)

```

### 问题三

```

% 读取数据
[X, textdata] = xlsread('附件 1-2017 全国各省市数据.xlsx')

% 提取数值列
numeric_columns = textdata(2:32,1)
numeric_data =X

% 归一化处理
normalized_data = (numeric_data - min(numeric_data)) ./
(max(numeric_data) - min(numeric_data));

% 进行聚类
z = linkage(normalized_data, 'ward');
clusters = cluster(z, 'maxclust', 3) % 设置聚类阈值

% 绘制点图
figure;
hold on;

```



```

% 根据聚类结果绘制点图
colors = lines(max(clusters));
for i = 1:max(clusters)
cluster_points = normalized_data(clusters == i, :);
scatter(cluster_points(:, 1), cluster_points(:, 2), 36,
'MarkerEdgeColor', 'k', ...
'MarkerFaceColor', colors(i, :), 'DisplayName',
sprintf('Cluster %d', i));
end

xlabel('x');
ylabel('y');
title('聚类结果点图');
legend('show');

% 画出各类范围
for i = 1:max(clusters)
cluster_points = normalized_data(clusters == i, :);
x_min = min(cluster_points(:, 1));
x_max = max(cluster_points(:, 1));
y_min = min(cluster_points(:, 2));
y_max = max(cluster_points(:, 2));
plot([x_min, x_max, x_max, x_min, x_min], [y_min, y_min, y_max,
y_max, y_min], 'r--');
end

hold off;
numeric_columns (clusters == 1) % 查看第 1 类所包含的城市

```

```
numeric_columns (clusters == 2) % 查看第 2 类所包含的城市  
numeric_columns (clusters == 3) % 查看第 3 类所包含的城市
```