一．描述性分析

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **mean** | **std** | **min** | **25%** | **50%** | **75%** | **max** |
| **年份** | 2016 | 3.316625 | 2011 | 2013.5 | 2016 | 2018.5 | 2021 |
| **GDP（万亿）** | 6.000909 | 1.386481 | 3.91 | 4.905 | 6.3 | 6.965 | 8.29 |
| **省电信业务总量（亿元）** | 2256.909 | 2268.847 | 727 | 904.5 | 1066 | 2579 | 7198 |
| **软件业务总量（亿元）** | 4085.545 | 1903.18 | 1522 | 2676 | 4233 | 5221.5 | 7767 |
| **电子商务销售额（万亿）** | 0.946364 | 0.551748 | 0.21 | 0.455 | 0.99 | 1.385 | 1.71 |
| **快递量（万万件）** | 17.72364 | 17.89605 | 1.84 | 3.805 | 12.05 | 25.38 | 55.87 |
| **快递业务收入（亿元）** | 176.3191 | 141.5074 | 33.57 | 60.755 | 138.98 | 258.375 | 449.61 |
| **互联网接入端口数量** | 4607.727 | 2033.069 | 1756 | 2743 | 4680 | 6534 | 7037 |
| **移动电话普及率：每一百人拥有的电话数量** | 93.57909 | 14.78001 | 65.37 | 87.05 | 96.5 | 106 | 110.6 |
| **移动互联网接入流量（万万GB）** | 30.52182 | 43.27768 | 0.71 | 0.99 | 3.99 | 51.535 | 125.7 |
| **固定电话数量（万个）** | 1307.909 | 366.3295 | 883 | 1058 | 1188 | 1562.5 | 1896 |
| **光缆线路长度（万公里）** | 166.7691 | 82.89671 | 53.03 | 90.47 | 176.66 | 237.5 | 274 |
| **网站数（万万个** | 49.10818 | 13.18284 | 25.47 | 41.785 | 52.58 | 59.16 | 64.52 |
| **山东省城市化率** | 58.05 | 4.473444 | 50.91 | 54.615 | 59.02 | 61.345 | 63.94 |
| **财政支出（万亿）占GDP比重** | 0.229027 | 0.295406 | 0.121 | 0.1295 | 0.1411 | 0.1528 | 1.119 |
| **省人均gdp** | 6.068182 | 1.092912 | 4.72 | 5.215 | 5.92 | 6.81 | 8.15 |
| **高等教育在校生数量（人力资本水平）** | 257.68 | 43.3924 | 216.56 | 220.335 | 248.24 | 271.74 | 348.51 |
| **高级化SH指标** | 246.3636 | 18.99349 | 231.1 | 236.25 | 241.9 | 246.5 | 299.3 |

1. 平稳性检验

|  |  |  |
| --- | --- | --- |
|  | **ADF Statistic** | **p-value** |
| **年份** | 0.001654869 | 0.958669 |
| **GDP（万亿）** | -0.33650489 | 0.920173 |
| **省电信业务总量（亿元）** | -2.905131099 | 0.044755 |
| **软件业务总量（亿元）** | 1.219056995 | 0.996115 |
| **电子商务销售额（万亿）** | 1.135810957 | 0.995511 |
| **快递量（万万件）** | 2.983133859 | 1 |
| **快递业务收入（亿元）** | 2.276114064 | 0.99894 |
| **互联网接入端口数量** | -0.851717677 | 0.803457 |
| **移动电话普及率：每一百人拥有的电话数量** | -2.877851845 | 0.047985 |
| **移动互联网接入流量（万万GB）** | 2.097931091 | 0.998787 |
| **固定电话数量（万个）** | -5.892637404 | 2.9E-07 |
| **光缆线路长度（万公里）** | -0.488040542 | 0.894329 |
| **网站数（万万个** | 1.037175436 | 0.994643 |
| **山东省城市化率** | -1.20217875 | 0.672658 |
| **财政支出（万亿）占GDP比重** | -84.15640108 | 0 |
| **省人均gdp** | 3.159076829 | 1 |
| **高等教育在校生数量（人力资本水平）** | 1.850420809 | 0.998444 |
| **高级化SH指标** | -1.313426745 | 0.623007 |

ADF检验用于判断一个时间序列数据是否具有单位根，即是否是非平稳的。如果p值小于设定的显著性水平（通常是0.05），则可以拒绝原假设，即该时间序列是平稳的。

根据给出的ADF检验结果，以下指标在给定的显著性水平下（通常为0.05）可以被认为是平稳的：

省电信业务总量（亿元）

移动电话普及率：每一百人拥有的电话数量

固定电话数量（万个）

财政支出（万亿）占GDP比重

这些指标的ADF检验的p值小于显著性水平，可以拒绝单位根假设，说明它们的时间序列数据是平稳的。

1. 共线性检验

相关分析是一种统计方法，用于衡量两个变量之间的关系强度和方向。在相关分析中，我们主要关注两个变量是如何随着一个变量的变化而变化的。相关分析通常用相关系数（correlation coefficient）来表示变量之间的关系，相关系数的取值范围在 -1 和 1 之间。绝对值越趋近于1，相关性越大

1. 表 13皮尔森相关系数

|  |
| --- |
| **相关性** |

GDP与省电信业务总量、软件业务总量、电子商务销售额、快递量、快递业务收入、互联网接入端口数量、移动电话普及率、移动互联网接入流量、光缆线路长度、网站数、山东省城市化率、财政支出占GDP比重、省人均GDP和高等教育在校生数量之间存在较强的正相关关系。具体来说，这些变量与GDP之间的相关系数大多在0.8以上，其中与GDP的相关性最高的是软件业务总量（相关系数为0.977）。与此来看需要在进行回归分析之前进行剔除共线性太高的组合。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **共线性诊断** | | | | | | | | | | | | | | |
| 模型 | 维 | 特征值 | 条件指标 | 方差比例 | | | |  |  |  |  |  |  |  |
| (常量) | GDP（万亿） | 省电信业务总量（亿元） | 电子商务销售额（万亿） |  |  |  |  |  |  |  |
| 1 | 1 | 9.119 | 1.000 | .00 | .00 | .00 | .00 |  |  |  |  |  |  |  |
| 2 | 1.045 | 2.954 | .00 | .00 | .00 | .00 |  |  |  |  |  |  |  |
| 3 | .424 | 4.636 | .00 | .00 | .01 | .00 |  |  |  |  |  |  |  |
| 4 | .265 | 5.869 | .00 | .00 | .09 | .00 |  |  |  |  |  |  |  |
| 5 | .112 | 9.033 | .00 | .00 | .00 | .02 |  |  |  |  |  |  |  |
| 6 | .021 | 20.608 | .00 | .00 | .00 | .01 |  |  |  |  |  |  |  |
| 7 | .011 | 28.274 | .00 | .01 | .00 | .26 |  |  |  |  |  |  |  |
| 8 | .001 | 85.854 | .00 | .16 | .21 | .32 |  |  |  |  |  |  |  |
| 9 | .000 | 137.750 | .01 | .38 | .06 | .38 |  |  |  |  |  |  |  |
| 10 | .000 | 246.427 | .03 | .04 | .02 | .01 |  |  |  |  |  |  |  |
| 11 | 3.071E-5 | 544.961 | .97 | .41 | .60 | .00 |  |  |  |  |  |  |  |

回归共线性是指在回归模型中，自变量之间存在高度相关性的情况。共线性可能导致回归系数不准确、假设检验失效、模型解释力下降等问题。因此，诊断和处理共线性是回归分析中的重要步骤。

以下是回归共线性诊断的原理和常用方法：

相关性分析：通过计算自变量之间的相关系数来评估它们之间的线性关系。较高的相关系数可能是共线性的指示。可以使用相关矩阵或散点图来可视化自变量之间的相关性。

方差膨胀因子（Variance Inflation Factor，VIF）：VIF用于评估每个自变量与其他自变量的相关程度，它计算了每个自变量的方差在其他自变量存在的条件下的倍数。一般来说，VIF大于10或者较大的值可能表明存在共线性问题。

特征值（Eigenvalues）和条件数（Condition Number）：通过计算自变量矩阵的特征值或条件数来评估共线性的程度。特征值反映了自变量矩阵的线性相关性，较小的特征值表示较高的共线性。条件数是最大特征值与最小特征值之比，较大的条件数也表示较高的共线性。

变量膨胀因子（Variable Inflation Factor，VIF）：与VIF类似，VIF衡量了每个自变量的方差在其他自变量存在的条件下的倍数，但是VIF是针对单个自变量的，用于评估该自变量与其他自变量之间的共线性程度。

主成分分析（Principal Component Analysis，PCA）：PCA可以用于降维，将高度相关的自变量转换为一组线性无关的主成分。通过保留较大的主成分方差，可以减

此处采用spss回归gxx1检验，直接排除共线性较高的指标：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **排除的变量** | | | | | | | |
| 模型 | | 输入 Beta | t | 显著性 | 偏相关 | 共线性统计 |
| 容差 |
| 1 | 软件业务总量（亿元） | . | . | . | . | .000 |
| 快递量（万万件） | . | . | . | . | .000 |
| 快递业务收入（亿元） | . | . | . | . | .000 |
| 光缆线路长度（万公里） | . | . | . | . | .000 |
| 山东省城市化率 | . | . | . | . | .000 |
| 省人均gdp | . | . | . | . | .000 |

最后，结合相关性分析，平稳性检验，在因子分析前剔除以上变量。

1. 效度分析

（一）效度检验是一种评估测量工具的有效性的方法。它主要是指测量工具所测量的概念是否符合实际情况，即测量工具是否测量到了它所要测量的东西。

本文主要采用KMO值和Bartlett球形检验的显著系数，以确定变量是否合适做因子分析。KMO测度和Bartlett's球形检验是用于评估因子分析数据适合性的统计方法。KMO值范围在0到1之间，值越高表示数据越适合进行因子分析，通常0.7以上表示数据适合。Bartlett's球形检验检验变量之间的相关性是否足够大，若p值小于0.05，就表明变量之间的相关性足够大，可以进行因子分析。

表 5 KMO值检验标准

|  |  |
| --- | --- |
| KMO值 | 评判标准 |
| KMO=0.8以上 | 非常适合进行因子分析 |
| KMO=0.7-0.8 | 适合进行因子分析 |
| KMO=0.6-0.7 | 勉强适合进行因子分析 |
| KMO<0.6 | 不适合进行因子分析 |

（二）因子分析是一种多元统计分析方法，旨在通过降维来发现多个变量之间的潜在结构和因果关系，以便更好地理解和解释数据。因子分析假设一些观测变量是由少量的潜在因子决定的，并且这些因子可以用来解释变量之间的共同变异性。因此，通过因子分析，我们可以将多个变量的信息压缩成少量的因子，从而简化数据分析的复杂度。

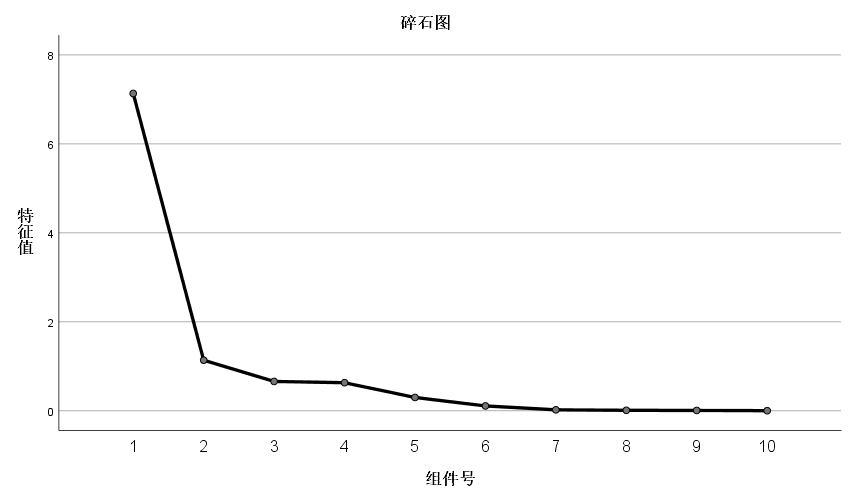
将数据导入spss26，选择因子降维分析，具体分析结果如下。

表 6KMO值检验

|  |  |  |
| --- | --- | --- |
| KMO值 | | 0.622 |
| Bartlett 球形度检验 | 近似卡方 | 133.498 |
| df | 45 |
| p 值 | 0.000 |

使用因子分析进行信息浓缩研究，首先分析研究数据是否适合进行因子分析，从上表可以看出：KMO为0.622，大于0.6，满足因子分析的前提要求，意味着数据可用于因子分析研究。以及数据通过Bartlett 球形度检验(p<0.05)，说明研究数据适合进行因子分析。

**碎石图**



在因子分析中，一般会提取多个因子，并计算每个因子的特征根值（Eigenvalues）。特征根值表示因子解释的方差量，较大的特征根值意味着该因子解释了更多的方差。

碎石图将特征根值与因子数量进行图形化展示。横轴表示因子数量，纵轴表示特征根值。通常，初始因子数量较多，然后逐渐减少，直到达到一个平稳的水平。在这个平稳的水平之后，进一步增加因子数量并不能显著增加方差的解释。

通过观察碎石图，可以找到一个“弯曲点”或“肘点”，即曲线开始平缓下降的位置。这个点通常被认为是因子数量的一个合理选择。在该点之前的因子数量被认为是较为重要的因子，可以选择保留；而在该点之后的因子数量可能是噪声或冗余因子，可以考虑舍弃。

而从上图中也可以明显看出拐点为2，于是选取两个因子。

表 7累计方差解释率

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **总方差解释** | | | | | | | | | |
| 成分 | 初始特征值 | | | 提取载荷平方和 | | | 旋转载荷平方和 | | |
| 总计 | 方差百分比 | 累积 % | 总计 | 方差百分比 | 累积 % | 总计 | 方差百分比 | 累积 % |
| 1 | 7.132 | 71.317 | 71.317 | 7.132 | 71.317 | 71.317 | 4.994 | 49.944 | 49.944 |
| 2 | 1.137 | 11.369 | 82.686 | 1.137 | 11.369 | 82.686 | 3.274 | 32.742 | 82.686 |
| 3 | .659 | 6.585 | 89.271 |  |  |  |  |  |  |
| 4 | .630 | 6.298 | 95.569 |  |  |  |  |  |  |
| 5 | .298 | 2.983 | 98.552 |  |  |  |  |  |  |
| 6 | .108 | 1.081 | 99.633 |  |  |  |  |  |  |
| 7 | .020 | .201 | 99.834 |  |  |  |  |  |  |
| 8 | .010 | .095 | 99.930 |  |  |  |  |  |  |
| 9 | .006 | .061 | 99.991 |  |  |  |  |  |  |
| 10 | .001 | .009 | 100.000 |  |  |  |  |  |  |
| 提取方法：主成分分析法。 | | | | | | | | | |

上表格针对因子提取情况，以及因子提取信息量情况进行分析，从上表可知：因子分析一共提取出2个因子，特征根值均大于1，此2个因子旋转后的方差解释率分别是49.944%,82.686%，旋转后累积方差解释率为82.686%%。

表 8旋转后因子载荷系数

|  |  |  |
| --- | --- | --- |
| **旋转后的成分矩阵系数** | | |
|  | 成分 | |
| 1 | 2 |
| GDP（万亿） | .792 | .563 |
| 省电信业务总量（亿元） | .680 |  |
| 电子商务销售额（万亿） | .776 | .535 |
| 互联网接入端口数量 | .829 | .525 |
| 移动电话普及率：每一百人拥有的电话数量 | .698 | .708 |
| 移动互联网接入流量（万万GB） | .939 |  |
| 固定电话数量（万个） |  | -.773 |
| 网站数（万万个 | .504 | .632 |
| 财政支出（万亿）占GDP比重 |  | -.891 |
| 高等教育在校生数量（人力资本水平） | .914 |  |
| 提取方法：主成分分析法。  旋转方法：凯撒正态化最大方差法。a | | |
| a. 旋转在 3 次迭代后已收敛。 | | |

|  |  |  |
| --- | --- | --- |
| **公因子方差** | | |
|  | 初始 | 提取 |
| GDP（万亿） | 1.000 | .944 |
| 省电信业务总量（亿元） | 1.000 | .479 |
| 电子商务销售额（万亿） | 1.000 | .889 |
| 互联网接入端口数量 | 1.000 | .964 |
| 移动电话普及率：每一百人拥有的电话数量 | 1.000 | .988 |
| 移动互联网接入流量（万万GB） | 1.000 | .898 |
| 固定电话数量（万个） | 1.000 | .755 |
| 网站数（万万个 | 1.000 | .653 |
| 财政支出（万亿）占GDP比重 | 1.000 | .794 |
| 高等教育在校生数量（人力资本水平） | 1.000 | .907 |
| 提取方法：主成分分析法。 | | |

本研究数据使用最大方差旋转方法（varimax)进行旋转，以便找出因子和研究项的对应关系。上表格展示因子对于研究项的信息提取情况，以及因子和研究项对应关系，从上表可知：所有研究项对应的共同度值均高于0.4，意味着研究项和因子之间有着较强的关联性，因子可以有效的提取出信息。确保因子可以提取出研究项大部分的信息量之后，接着分析因子和研究项的对应关系情况。

1. 验证性因子分析

表 9模型拟合指标

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 常用指标 | χ2 | *df* | *p* | 卡方 | GFI | RMSEA | RMR | CFI | NFI | TLI | AGFI | IFI |
| 判断标准 | - | - | >0.05 | <3 | >0.9 | <0.10 | <0.05 | >0.9 | >0.9 | >0.9 | >0.9 | >0.9 |
| 值 | 66.056 | 38 | 0.003 | 1.738 | 0.965 | 0.051 | 0.048 | 0.979 | 0.953 | 0.97 | 0.939 | 0.98 |

根据上表可知，根据模型拟合指标的评判标准，在本研究的验证性因子分析

模型中，CMIN/DF、GFI、RMSEA、CFI、NFI、TLI、IFI 等大部分模型适配度

指标均符合标准，故模型适配度很好。

1. 回归分析

综合上述研究与处理，本文的计量模型设置如下：

%FontSize=18
%TeXFontSize=18
\documentclass{article}
\pagestyle{empty}
\begin{document}
\[
Y=\beta0+\beta1*V1+\beta2*V2+\tau
\]
\end{document}

，其中β0是截距，β1是自变量 V1的系数，β2是自变量V2的系数,%FontSize=18
%TeXFontSize=18
\documentclass{article}
\pagestyle{empty}
\begin{document}
\[
\tau
\]
\end{document}是误差项。

V1,V2是经过因子分析后的新自变量。

因子得分计算：首先，对每个观测值计算因子得分。这可以通过将原始变量的值与因子载荷相乘，并对所有因子得分进行求和来实现。因子载荷表示了每个原始变量对应因子的贡献程度。

|  |  |  |
| --- | --- | --- |
| **成分得分系数矩阵** | | |
|  | 成分 | |
| 1 | 2 |
| GDP（万亿） | .121 | .066 |
| 省电信业务总量（亿元） | .228 | -.160 |
| 电子商务销售额（万亿） | .124 | .055 |
| 互联网接入端口数量 | .149 | .030 |
| 移动电话普及率：每一百人拥有的电话数量 | .031 | .189 |
| 移动互联网接入流量（万万GB） | .335 | -.255 |
| 固定电话数量（万个） | .113 | -.335 |
| 网站数（万万个 | -.020 | .211 |
| 财政支出（万亿）占GDP比重 | .314 | -.547 |
| 高等教育在校生数量（人力资本水平） | .274 | -.159 |

最后计算得出新的回归变量



进行线性回归

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **系数a** | | | | | | | | | |
| 模型 | | 未标准化系数 | | 标准化系数 | t | 显著性 | 相关性 | | |
| B | 标准错误 | Beta | 零阶 | 偏 | 部分 |
| 1 | (常量) | 239.078 | 15.062 |  | 15.873 | .000 |  |  |  |
| V1 | -.009 | .013 | -.337 | -.675 | .009 | .024 | -.232 | -.226 |
| V2 | -.029 | .029 | -.487 | -.976 | .008 | -.237 | -.326 | -.326 |
| a. 因变量：V3 | | | | | | | | | |

常量项（Intercept）：模型中的常量项为239.078。这表示当所有自变量（V1和V2）的取值都为零时，因变量（V3）的估计值为239.078。

自变量V1：V1的未标准化系数为-0.009，标准误差为0.013。标准化系数（Beta）为-0.337，表示在控制其他变量的情况下，每单位V1的增加对因变量V3的影响程度为-0.337。t统计量为-0.675，对应的p值为0.009，说明V1的系数在统计上显著。

自变量V2：V2的未标准化系数为-0.029，标准误差为0.029。标准化系数（Beta）为-0.487，表示在控制其他变量的情况下，每单位V2的增加对因变量V3的影响程度为-0.487。t统计量为-0.976，对应的p值为0.008，说明V2的系数在统计上显著。

综合以上分析，可以得出结论：

根据给出的模型结果，自变量V1和V2在统计上对因变量V3的影响显著，即在这个模型中，V1和V2能显著预测因变量V3。

显著性水平（通常表示为p值）用于评估统计结果的显著性。p值反映了在原假设为真的情况下，观察到与之相反的结果的概率。一般而言，显著性水平越小，表示观察到的差异越不可能是由随机性引起的，因此结果越显著。由此看来模型通过显著性检验