### 分析流程

### 分析步骤 1. 对数据进行Shapiro-Wilk（小数据样本，一般样本数5000以下）或者Kolmogorov–Smirnov（大数据样本，一般样本数5000以上）检验，查看其显著性。 2. 若不呈现出显著性(P>0.05)，说明符合正态分布，反之说明不符合正态分布（PS：通常现实研究情况下很难满足检验，若其样本峰度绝对值小于10并且偏度绝对值小于3，结合正态分布直方图、PP图或者QQ图可以描述为基本符合正态分布）。

### 详细结论

**输出结果1：总体描述结果**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 变量名 | 样本量 | 中位数 | 平均值 | 标准差 | 偏度 | 峰度 | S-W检验 | K-S检验 |
| AGE | 15898 | 73.4 | 73.313 | 6.98 | -0.137 | -0.326 | 0.995(0.000\*\*\*) | 0.025(6.3497089141911326e-9) |
| APOE4 | 15898 | 0 | 0.517 | 0.648 | 0.874 | -0.329 | 0.718(0.000\*\*\*) | 0.355(0) |
| CDRSB\_bl | 15898 | 0.5 | 1.219 | 1.516 | 1.783 | 3.719 | 0.786(0.000\*\*\*) | 0.211(0) |
| ADASQ4\_bl | 15898 | 4 | 4.694 | 2.843 | 0.334 | -0.924 | 0.945(0.000\*\*\*) | 0.131(7.991062033103057e-239) |
| ADAS13\_bl | 15898 | 13 | 14.67 | 8.627 | 0.996 | 0.988 | 0.936(0.000\*\*\*) | 0.103(6.970484541777451e-148) |
| ADAS11\_bl | 15898 | 8 | 9.269 | 5.782 | 1.363 | 2.615 | 0.905(0.000\*\*\*) | 0.115(4.69749764359162e-182) |
| 注：\*\*\*、\*\*、\*分别代表1%、5%、10%的显著性水平 | | | | | | | | |

**图表说明：**

上表展示了AGE、APOE4、CDRSB\_bl、ADASQ4\_bl、ADAS13\_bl、ADAS11\_bl描述性统计和正态性检验的结果，包括中位数、平均值等，用于检验数据的正态性。  
1. 通常正态分布的检验方法有两种，一种是Shapiro-Wilk检验，适用于小样本资料（样本量≤5000）；另一种是Kolmogorov–Smirnov检验，适用于大样本资料（样本量>5000）。  
2. 若呈现显著性(P<0.05)，则说明拒绝原假设（数据符合正态分布），该数据不满足正态分布，反之则说明该数据满足正态分布。  
PS：通常现实研究情况下很难满足检验，若其样本峰度绝对值小于10并且偏度绝对值小于3，结合正态分布直方图、PP图或者QQ图可以描述为基本符合正态分布。

**智能分析**

分析项：AGE样本N >= 5000，采用K-S检验，显著性P值为0.000\*\*\*，水平呈现显著性，拒绝原假设，因此数据不满足正态分布。(其峰度（-0.326）绝对值小于10并且偏度（-0.137）绝对值小于3，可以结合正态分布直方图、 PP图或者QQ图进行进一步分析。)

分析项：APOE4样本N >= 5000，采用K-S检验，显著性P值为0.000\*\*\*，水平呈现显著性，拒绝原假设，因此数据不满足正态分布。(其峰度（-0.329）绝对值小于10并且偏度（0.874）绝对值小于3，可以结合正态分布直方图、 PP图或者QQ图进行进一步分析。)

分析项：CDRSB\_bl样本N >= 5000，采用K-S检验，显著性P值为0.000\*\*\*，水平呈现显著性，拒绝原假设，因此数据不满足正态分布。(其峰度（3.719）绝对值小于10并且偏度（1.783）绝对值小于3，可以结合正态分布直方图、 PP图或者QQ图进行进一步分析。)

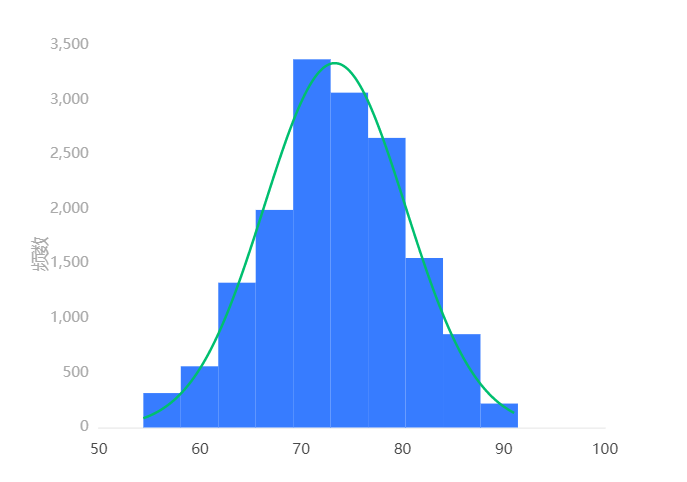
分析项：ADASQ4\_bl样本N >= 5000，采用K-S检验，显著性P值为0.000\*\*\*，水平呈现显著性，拒绝原假设，因此数据不满足正态分布。(其峰度（-0.924）绝对值小于10并且偏度（0.334）绝对值小于3，可以结合正态分布直方图、 PP图或者QQ图进行进一步分析。)

分析项：ADAS13\_bl样本N >= 5000，采用K-S检验，显著性P值为0.000\*\*\*，水平呈现显著性，拒绝原假设，因此数据不满足正态分布。(其峰度（0.988）绝对值小于10并且偏度（0.996）绝对值小于3，可以结合正态分布直方图、 PP图或者QQ图进行进一步分析。)

分析项：ADAS11\_bl样本N >= 5000，采用K-S检验，显著性P值为0.000\*\*\*，水平呈现显著性，拒绝原假设，因此数据不满足正态分布。(其峰度（2.615）绝对值小于10并且偏度（1.363）绝对值小于3，可以结合正态分布直方图、 PP图或者QQ图进行进一步分析。)

**输出结果2：正态性检验直方图**

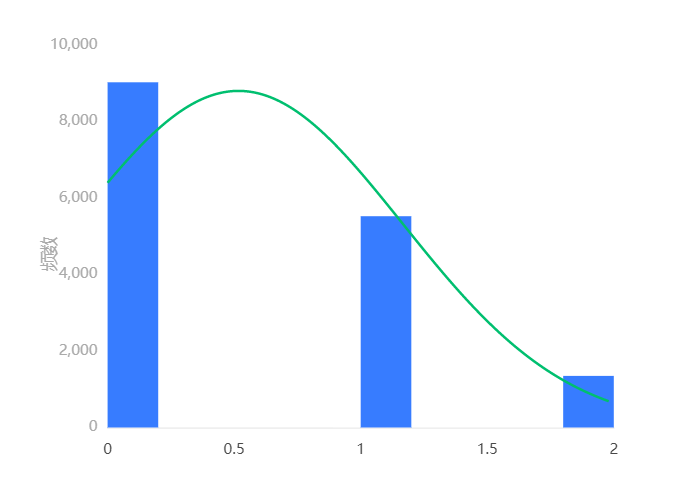
AGE



**图表说明：**

上图展示了AGE数据的正态性检验直方图，若正态图基本上呈现出钟形（中间高，两端低），则说明数据虽然不是绝对正态，但基本可接受为正态分布。

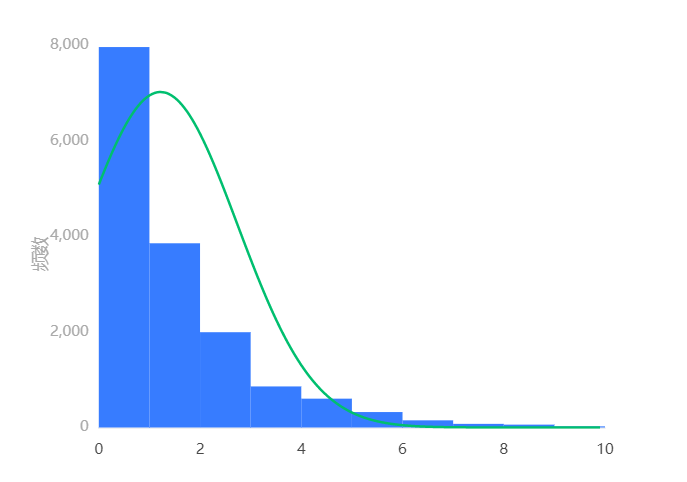
APOE4



**图表说明：**

上图展示了APOE4数据的正态性检验直方图，若正态图基本上呈现出钟形（中间高，两端低），则说明数据虽然不是绝对正态，但基本可接受为正态分布。

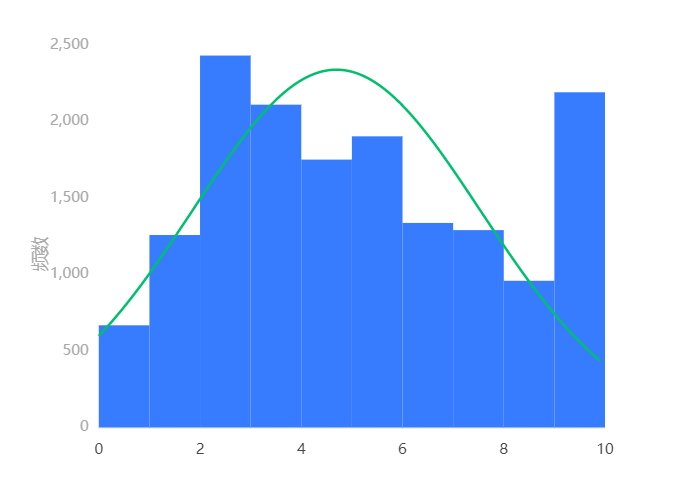
CDRSB\_bl



**图表说明：**

上图展示了CDRSB\_bl数据的正态性检验直方图，若正态图基本上呈现出钟形（中间高，两端低），则说明数据虽然不是绝对正态，但基本可接受为正态分布。

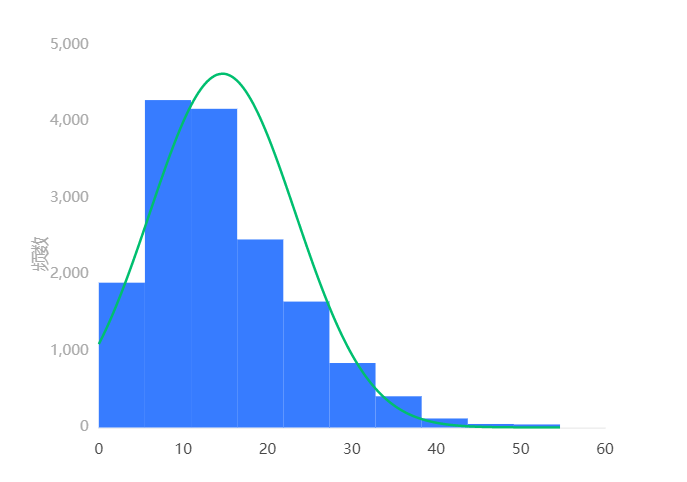
ADASQ4\_bl



**图表说明：**

上图展示了ADASQ4\_bl数据的正态性检验直方图，若正态图基本上呈现出钟形（中间高，两端低），则说明数据虽然不是绝对正态，但基本可接受为正态分布。

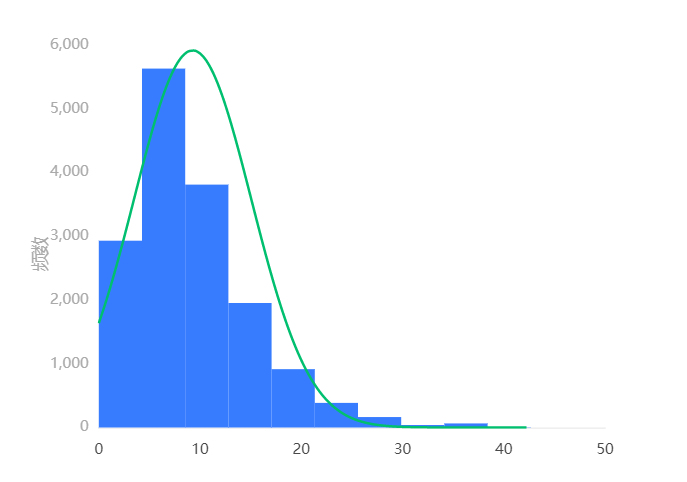
ADAS13\_bl



**图表说明：**

上图展示了ADAS13\_bl数据的正态性检验直方图，若正态图基本上呈现出钟形（中间高，两端低），则说明数据虽然不是绝对正态，但基本可接受为正态分布。

ADAS11\_bl

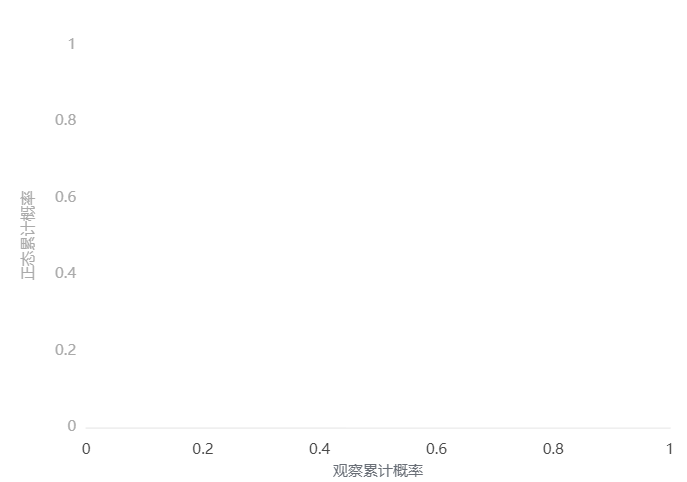


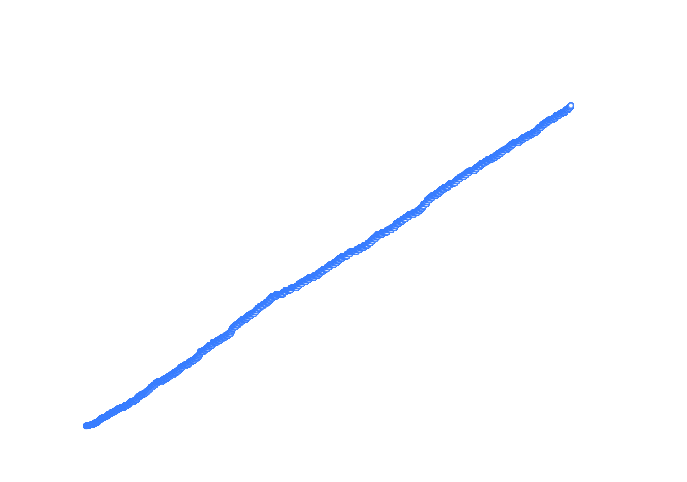
**图表说明：**

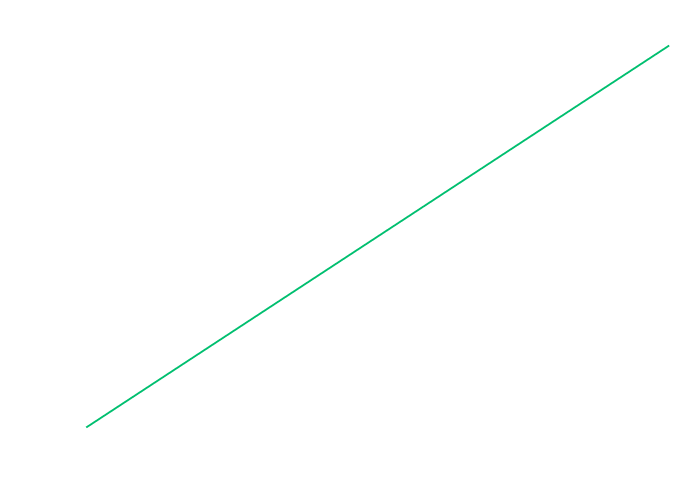
上图展示了ADAS11\_bl数据的正态性检验直方图，若正态图基本上呈现出钟形（中间高，两端低），则说明数据虽然不是绝对正态，但基本可接受为正态分布。

**输出结果3：正态性检验P-P图**

AGE



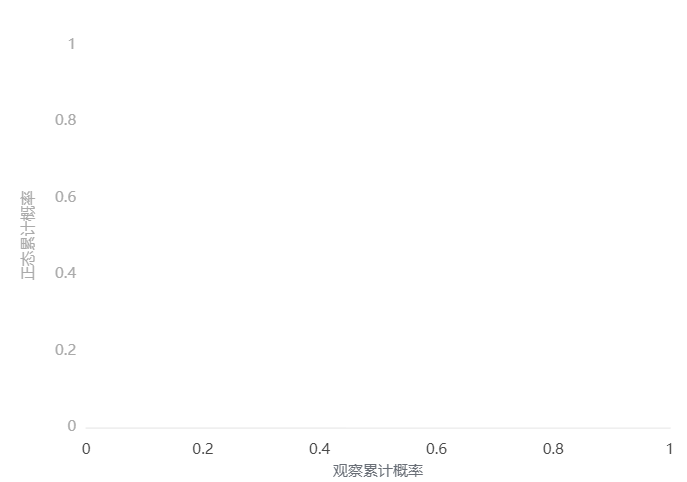


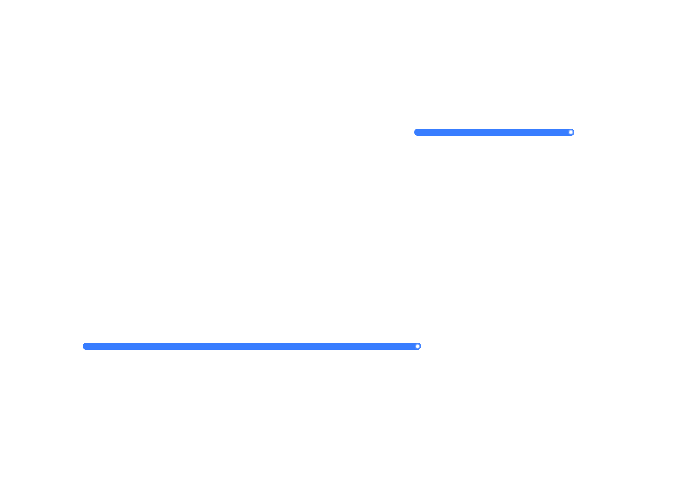


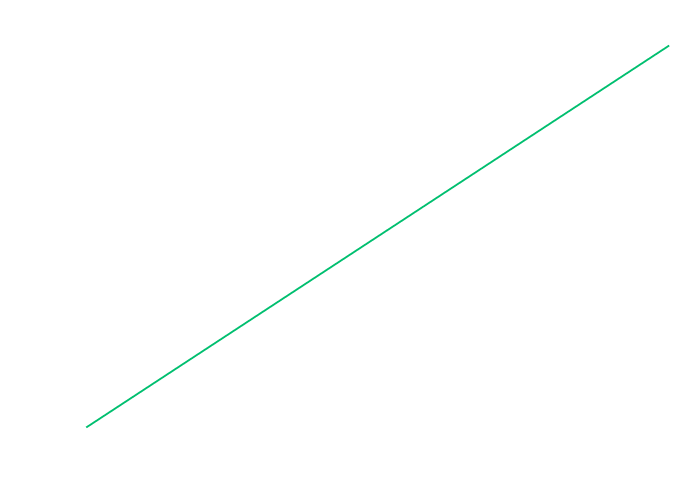
**图表说明：**

上图是AGE计算观测的累计概率（P）与正态累计概率（P）的拟合情况。拟合程度越高越服从正态分布。

APOE4



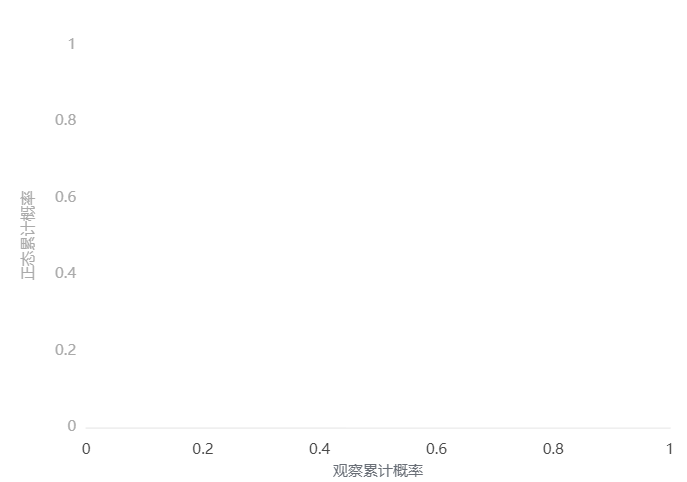


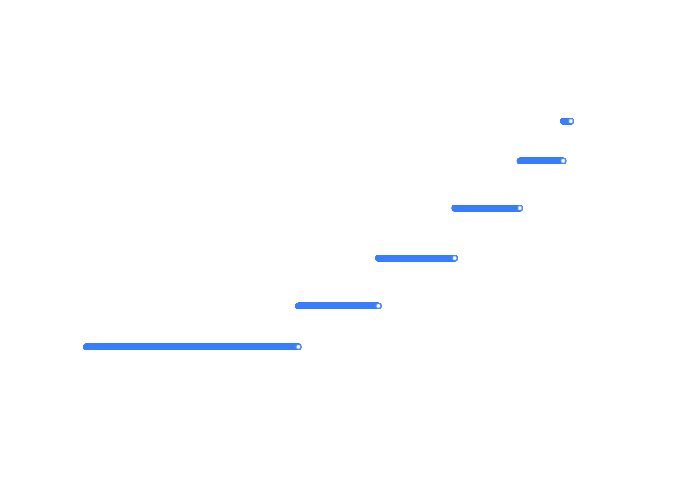


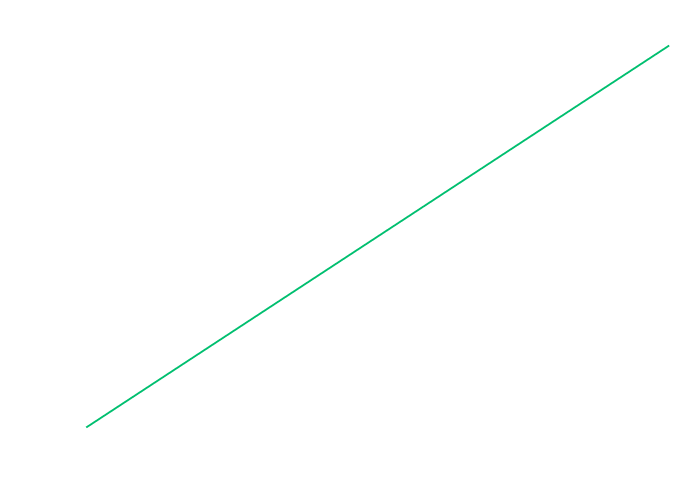
**图表说明：**

上图是APOE4计算观测的累计概率（P）与正态累计概率（P）的拟合情况。拟合程度越高越服从正态分布。

CDRSB\_bl



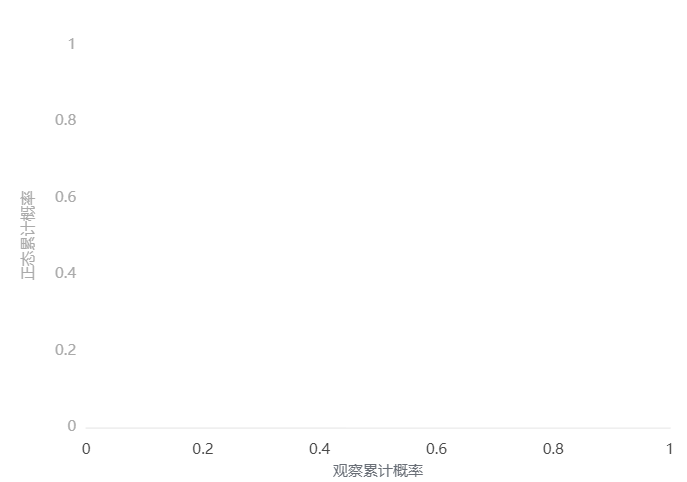


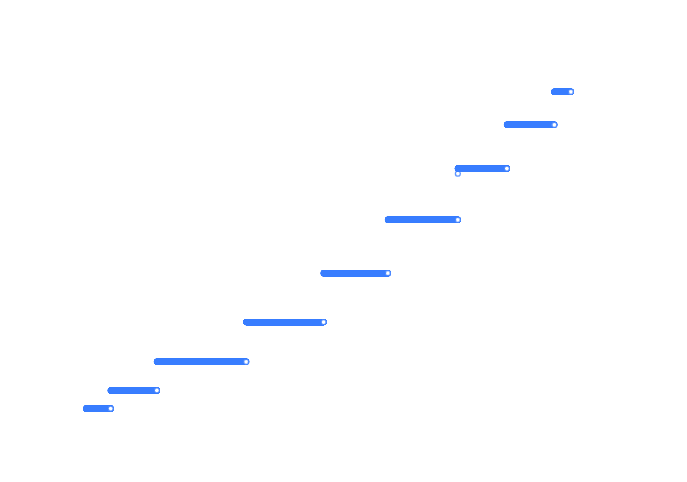


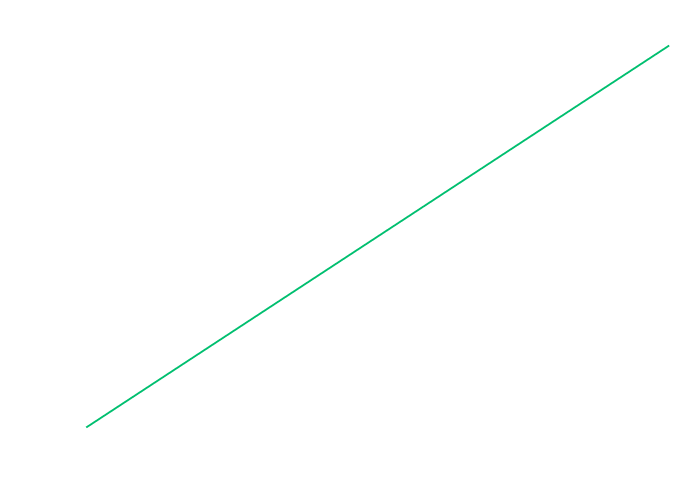
**图表说明：**

上图是CDRSB\_bl计算观测的累计概率（P）与正态累计概率（P）的拟合情况。拟合程度越高越服从正态分布。

ADASQ4\_bl



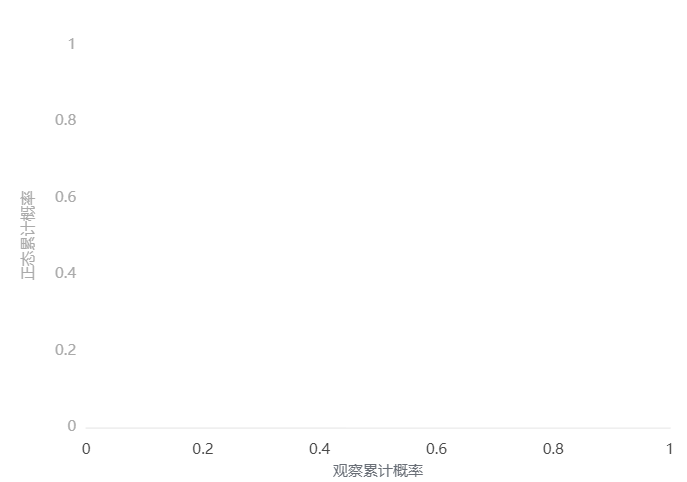


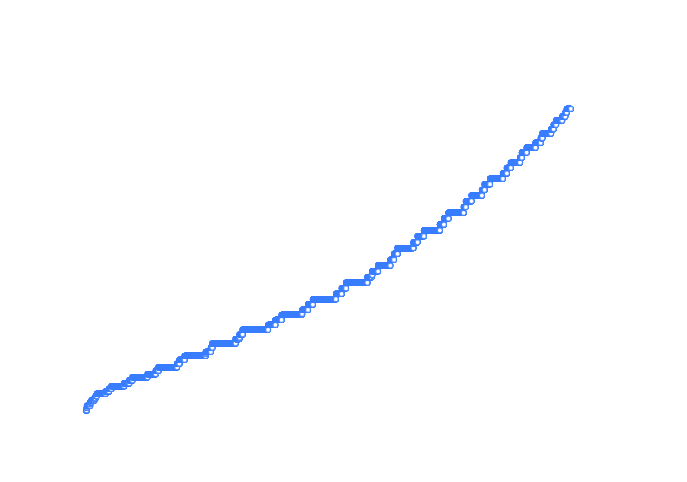


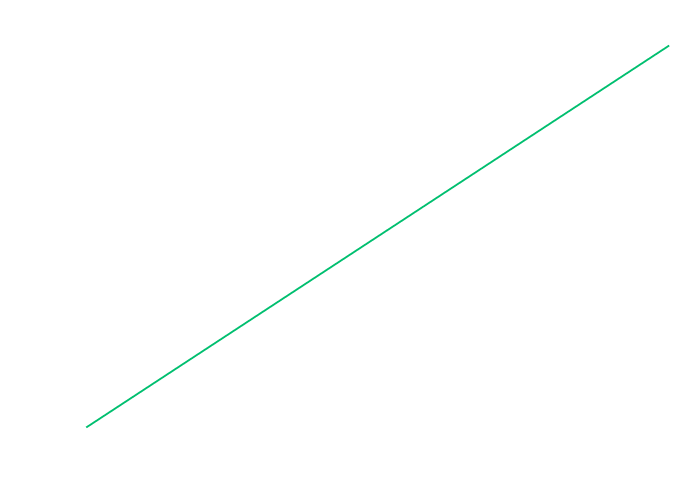
**图表说明：**

上图是ADASQ4\_bl计算观测的累计概率（P）与正态累计概率（P）的拟合情况。拟合程度越高越服从正态分布。

ADAS13\_bl



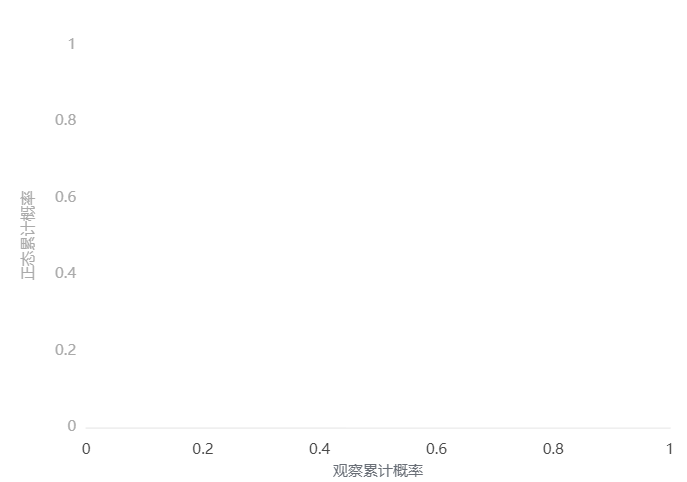


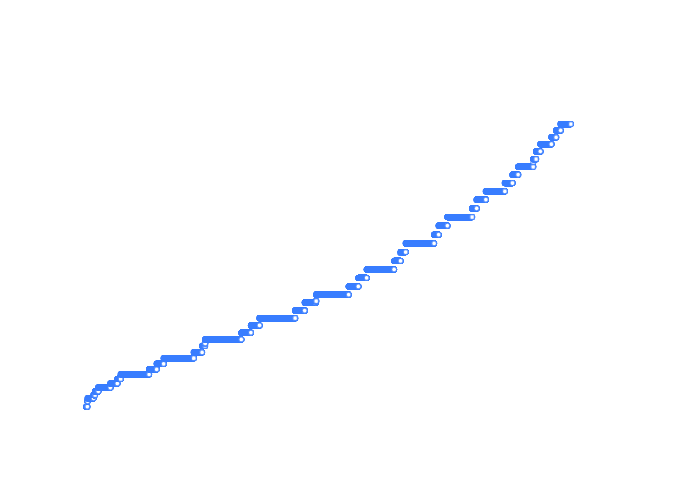


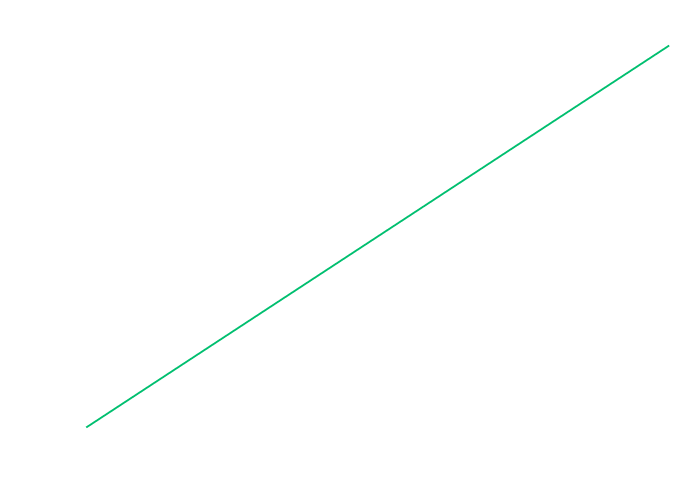
**图表说明：**

上图是ADAS13\_bl计算观测的累计概率（P）与正态累计概率（P）的拟合情况。拟合程度越高越服从正态分布。

ADAS11\_bl





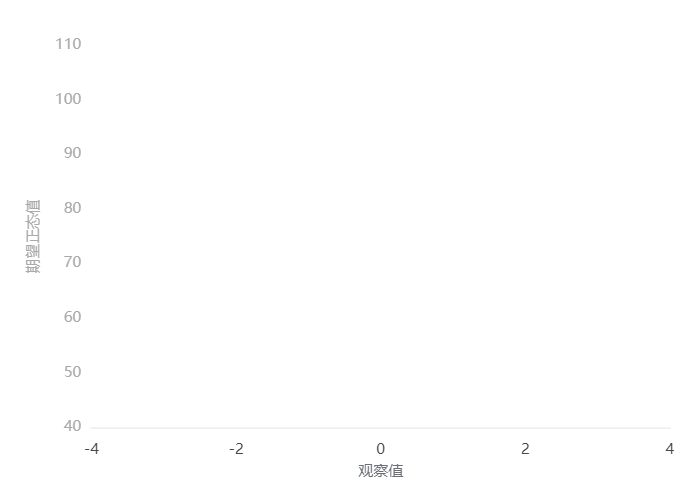


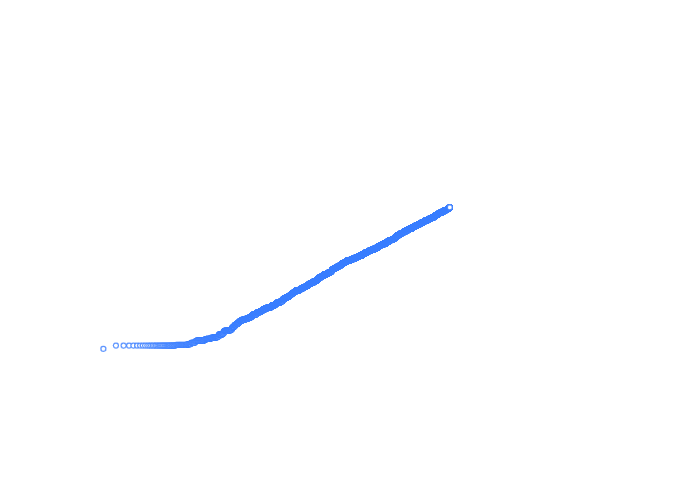
**图表说明：**

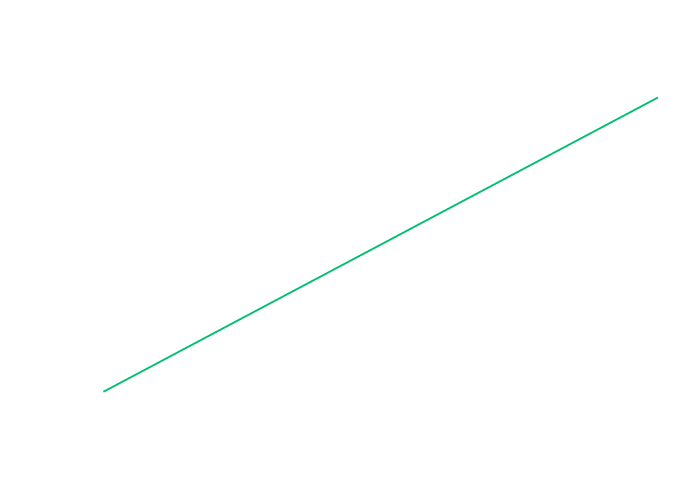
上图是ADAS11\_bl计算观测的累计概率（P）与正态累计概率（P）的拟合情况。拟合程度越高越服从正态分布。

**输出结果4：正态性检验Q-Q图**

AGE



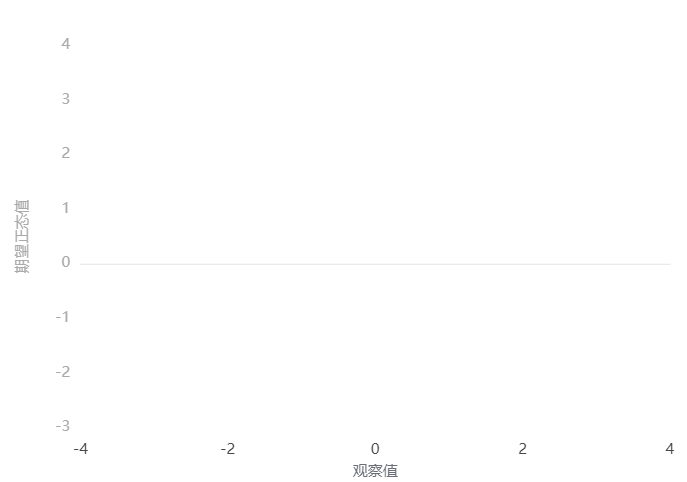


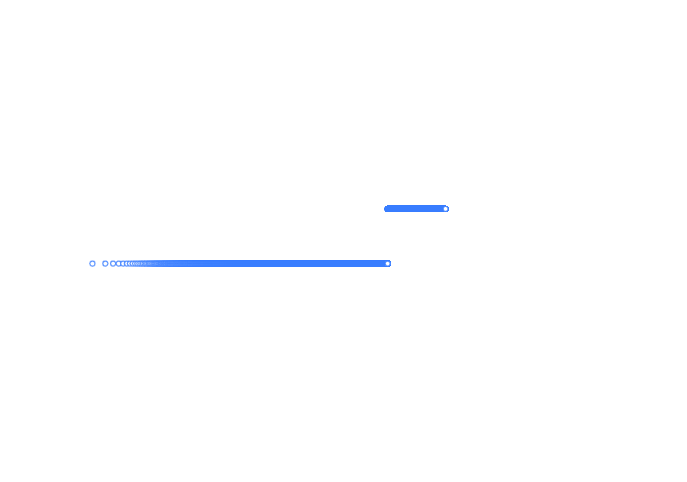


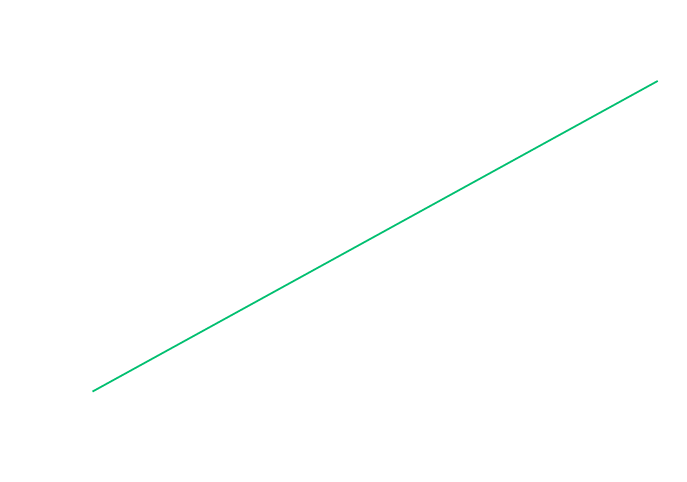
**图表说明：**

Q-Q图，全称“Quantile Quantile Plot”。用图形的方式比较观测值与预测值（假定正态下的分布）不同分位数的概率分布，从而检验是否吻合正态分布规律。并且将实际数据作为X轴，将假定正态时的数据分位数作为Y轴，作散点图，散点与直线重合度越高越服从正态分布，散点差异愈大越不服从正态分布，请视实际情况而定。

APOE4



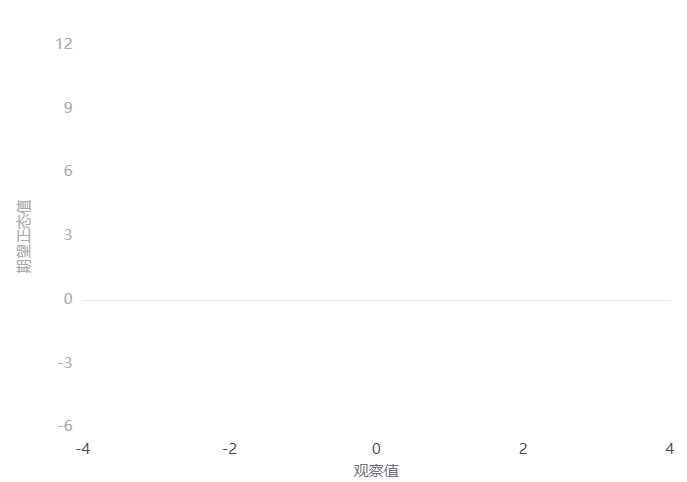


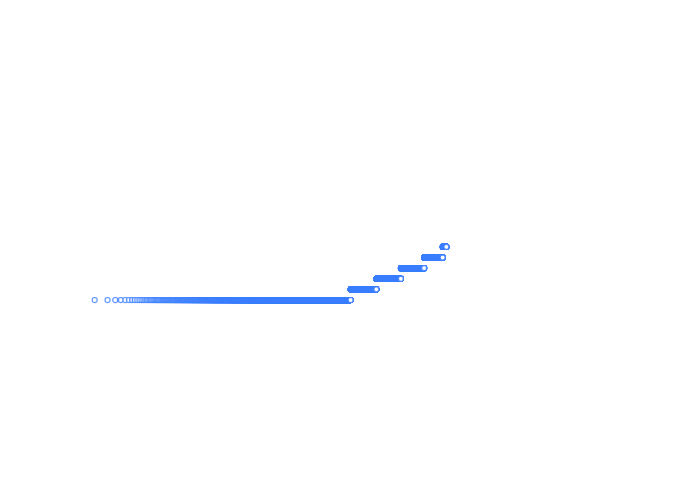


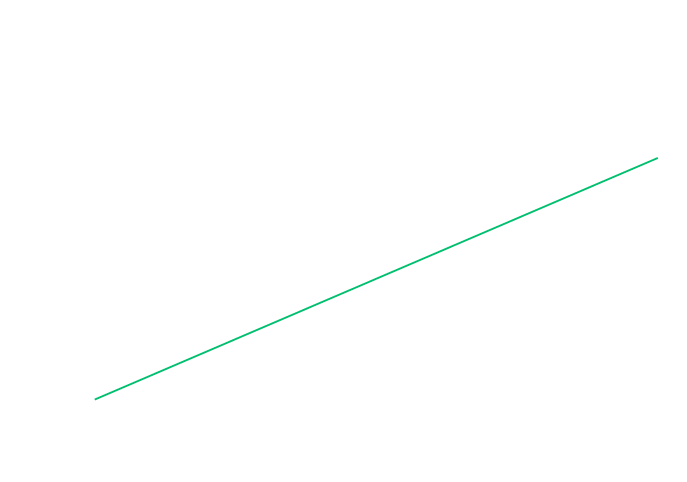
**图表说明：**

Q-Q图，全称“Quantile Quantile Plot”。用图形的方式比较观测值与预测值（假定正态下的分布）不同分位数的概率分布，从而检验是否吻合正态分布规律。并且将实际数据作为X轴，将假定正态时的数据分位数作为Y轴，作散点图，散点与直线重合度越高越服从正态分布，散点差异愈大越不服从正态分布，请视实际情况而定。

CDRSB\_bl



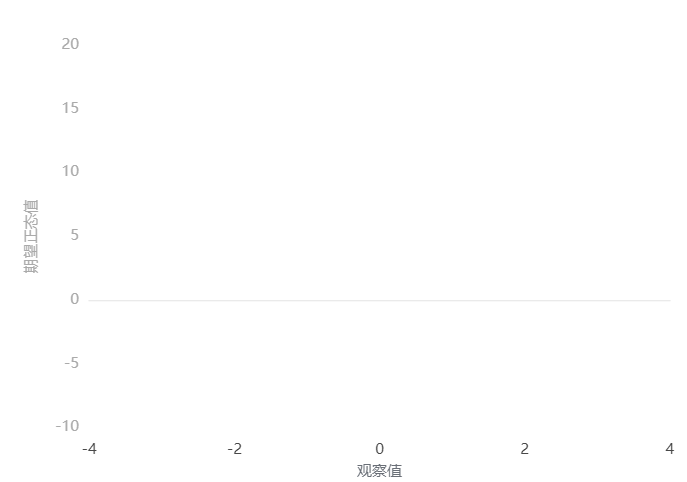


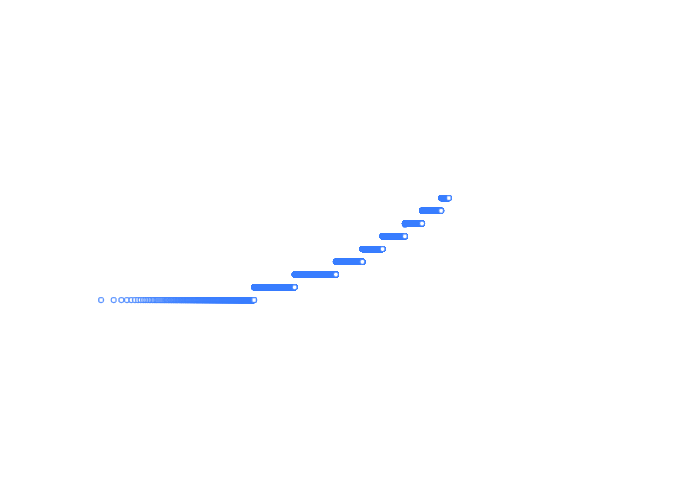


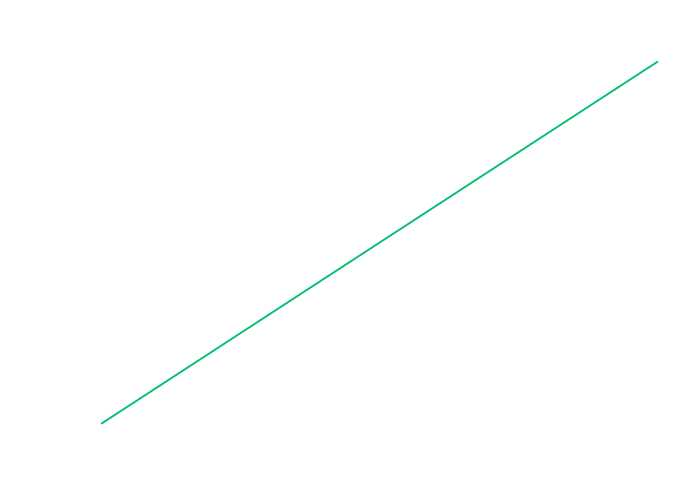
**图表说明：**

Q-Q图，全称“Quantile Quantile Plot”。用图形的方式比较观测值与预测值（假定正态下的分布）不同分位数的概率分布，从而检验是否吻合正态分布规律。并且将实际数据作为X轴，将假定正态时的数据分位数作为Y轴，作散点图，散点与直线重合度越高越服从正态分布，散点差异愈大越不服从正态分布，请视实际情况而定。

ADASQ4\_bl



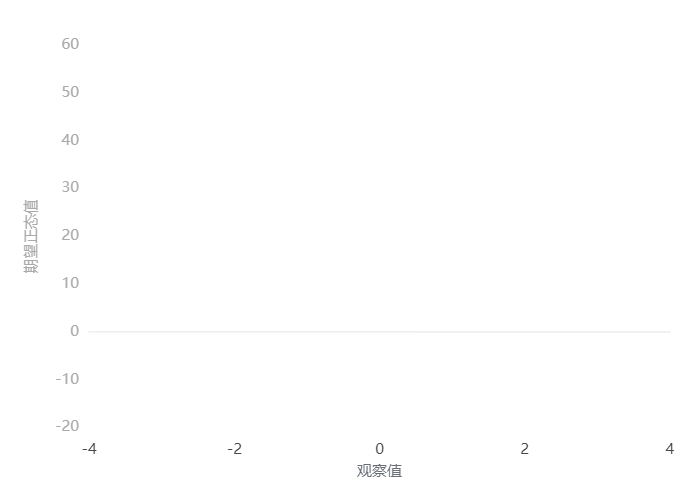


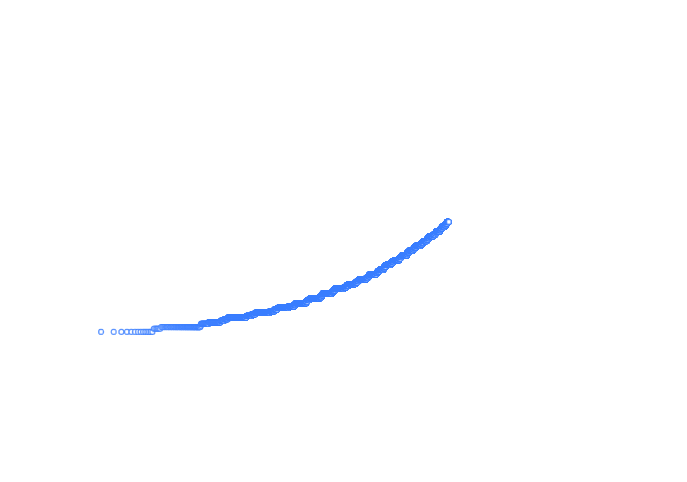


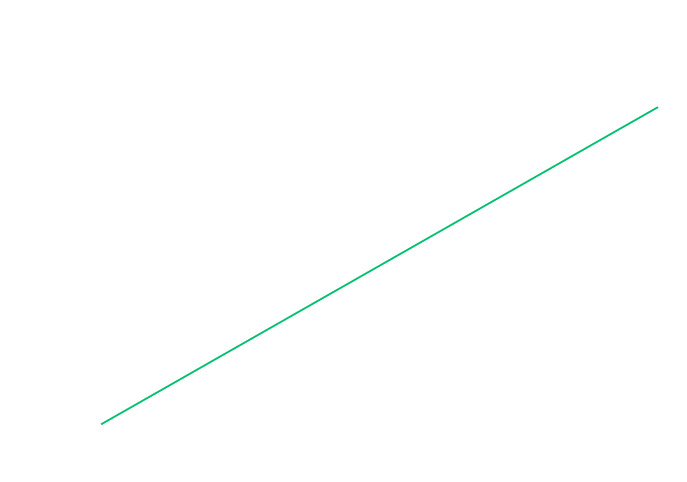
**图表说明：**

Q-Q图，全称“Quantile Quantile Plot”。用图形的方式比较观测值与预测值（假定正态下的分布）不同分位数的概率分布，从而检验是否吻合正态分布规律。并且将实际数据作为X轴，将假定正态时的数据分位数作为Y轴，作散点图，散点与直线重合度越高越服从正态分布，散点差异愈大越不服从正态分布，请视实际情况而定。

ADAS13\_bl



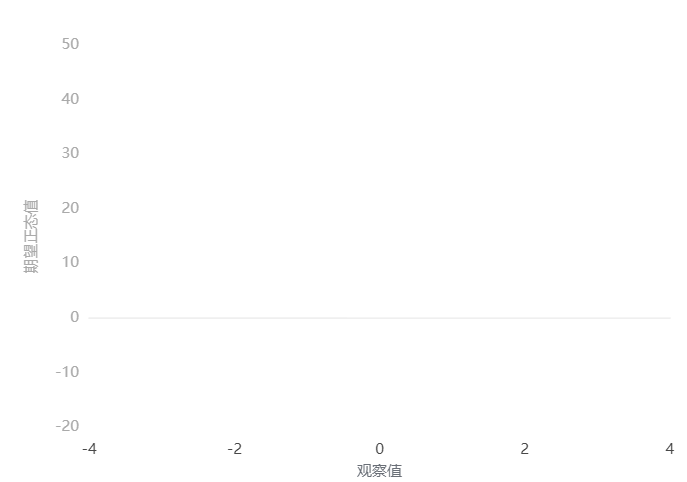


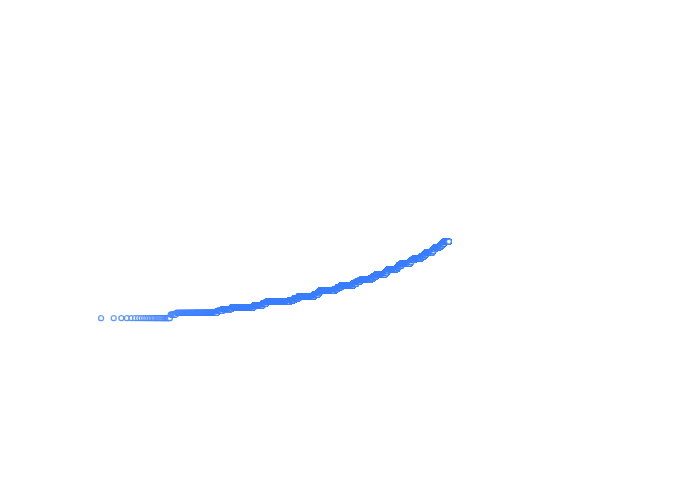


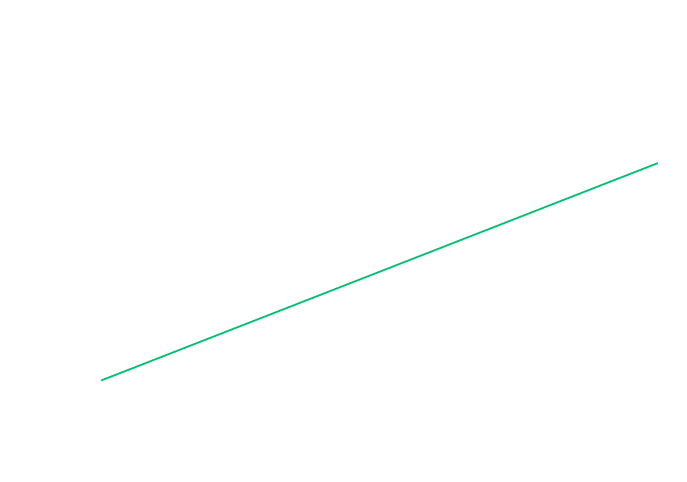
**图表说明：**

Q-Q图，全称“Quantile Quantile Plot”。用图形的方式比较观测值与预测值（假定正态下的分布）不同分位数的概率分布，从而检验是否吻合正态分布规律。并且将实际数据作为X轴，将假定正态时的数据分位数作为Y轴，作散点图，散点与直线重合度越高越服从正态分布，散点差异愈大越不服从正态分布，请视实际情况而定。

ADAS11\_bl







**图表说明：**

Q-Q图，全称“Quantile Quantile Plot”。用图形的方式比较观测值与预测值（假定正态下的分布）不同分位数的概率分布，从而检验是否吻合正态分布规律。并且将实际数据作为X轴，将假定正态时的数据分位数作为Y轴，作散点图，散点与直线重合度越高越服从正态分布，散点差异愈大越不服从正态分布，请视实际情况而定。

### 参考文献 [1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from https://www.spsspro.com. [2] 宗序平, 姚玉兰. 利用Q-Q图与P-P图快速检验数据的统计分布[J]. 统计与决策, 2010(20):2.