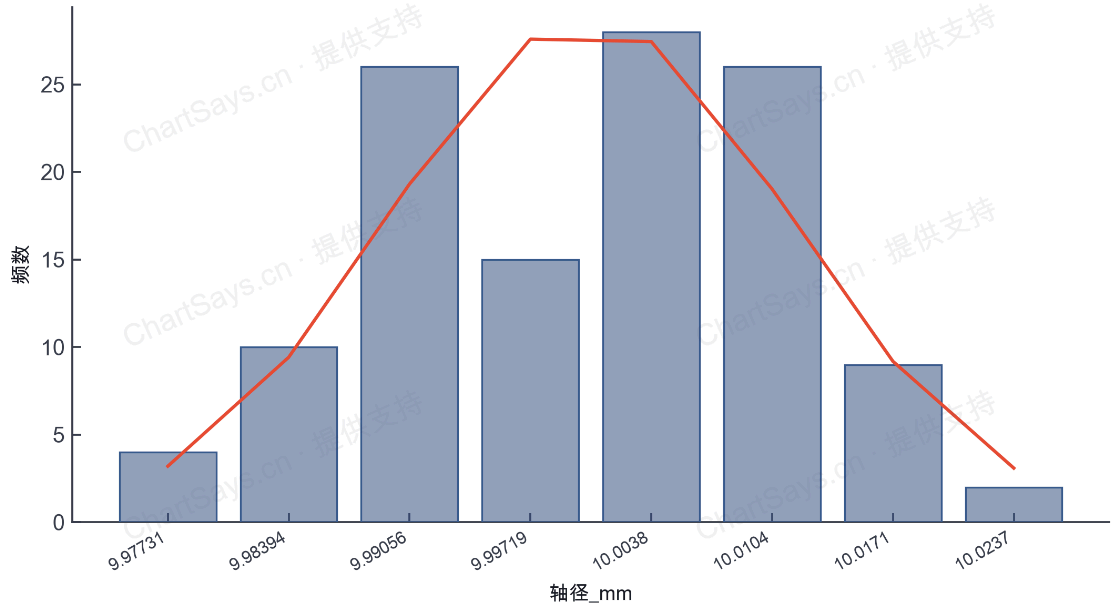
C01

看分布的图

一列数长什么样



2026 年 9 月版

本讲地图：先问一句，再挑图

我想看什么	用哪种图	第几章	平台里在哪儿出
一列数集中在哪、散多开	直方图（四种形式）	第 1 章	基础统计与分布
样本很少，怕分箱骗人	点图	第 2 章	通用图形
几组点想摆在一起	个体值图	第 3 章	通用图形
位置、散布、离群	箱线图	第 4 章	基础统计与分布 / 分组描述统计
低于某值的比例	ECDF 台阶线	第 5 章	通用图形
像不像正态	正态概率图	第 6 章	正态性检验集
该用哪种分布	分布拟合排序	第 7 章	个体分布识别
偏态怎么救	变换前后直方图	第 8 章	数据变换

出处：读图指引（直方图 / 点图 / 个体值图 / 箱线图 / ECDF / 正态概率图 / 分布识别）；八章共用同一批合成数据，末页有清单。

先问「这张图回答什么」，再挑图。

第 1 章　直方图：一列数长什么样

01　第 1 章 · 共 8 章　直方图 Histogram

什么时候用

手上只有一列连续数值（一批轴径、一批等待时间），想先知道它长什么样的时候，直方图就是第一步。

本章要点

- 四种形式：简单、带正态拟合、带组轮廓、带组拟合。
- 分箱数会改变形状：箱少看整体，箱多看细节。
- 四种要警惕的形状：两个峰、右拖尾、一侧削平、随分箱变形。

直方图是一列数的第一张图。

直方图回答什么：峰、对称、红线贴不贴

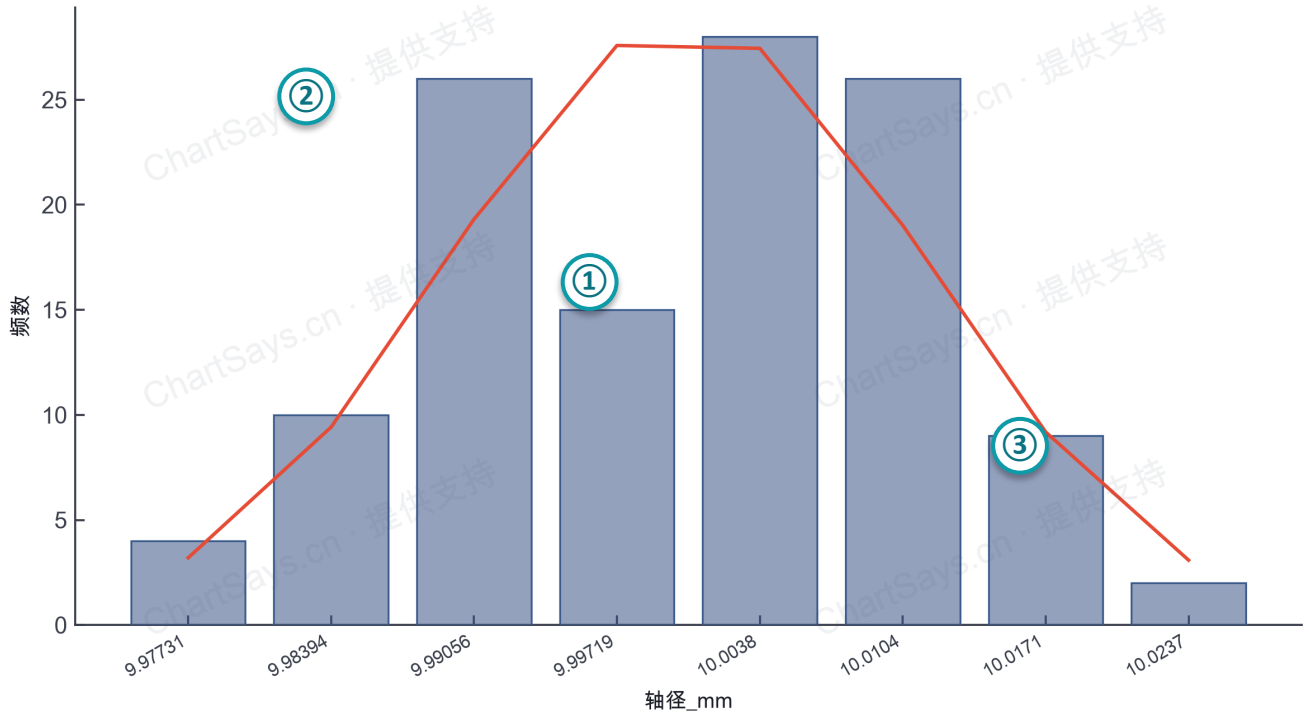


图 1-1　轴径_mm 的直方图，N=120，平台按 auto 分成 8 箱。

chartsays.cn · 图由 ChartSays 生成

这张图回答什么

一列数集中在哪、散多开、几个峰、拖不拖尾——四个问题一张图答完。

图上有什么

柱：落入该区间的观测量，区间由平台自动分箱。

红线：按本列均值与标准差画的正态曲线（纵轴已换算成频数，可直接比柱高）。

横轴：分箱中点；纵轴：频数。

① 峰有几个　② 是否对称　③ 红线贴不贴柱子。

柱看频数，红线看正态，三件事一起看。

直方图的四种形式（一）：简单 / 带正态拟合

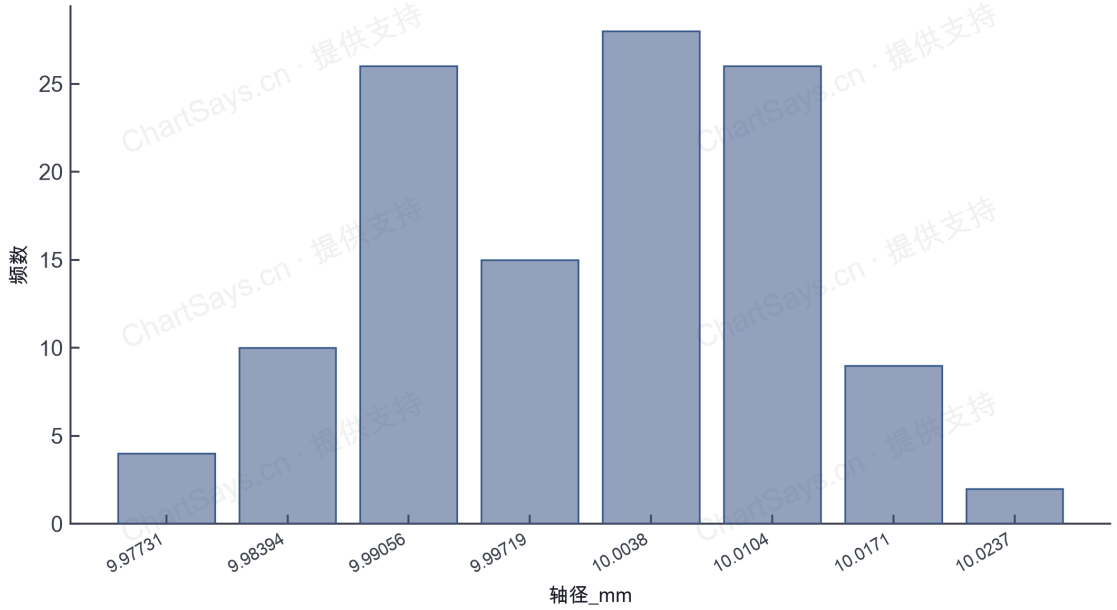


图 1-2　简单直方图：只有柱子加坐标轴。
看形状够用，但它不给你参照物。
出处：读图指引（直方图）。

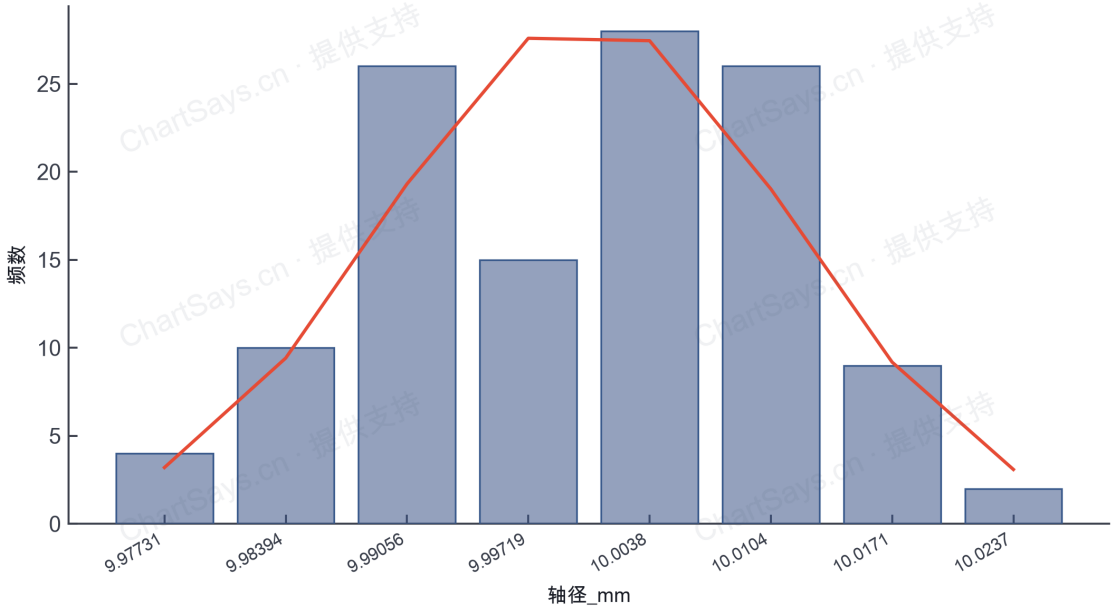


图 1-3　带正态拟合：多一条红色正态曲线。
柱子贴不贴红线，一眼看得出来。
红线 = 按本列均值与标准差画的正态曲线。

chartsays.cn · 图由 ChartSays 生成

简单版看形状，带拟合版看参照。

直方图的四种形式（二）：带组轮廓 / 带组拟合

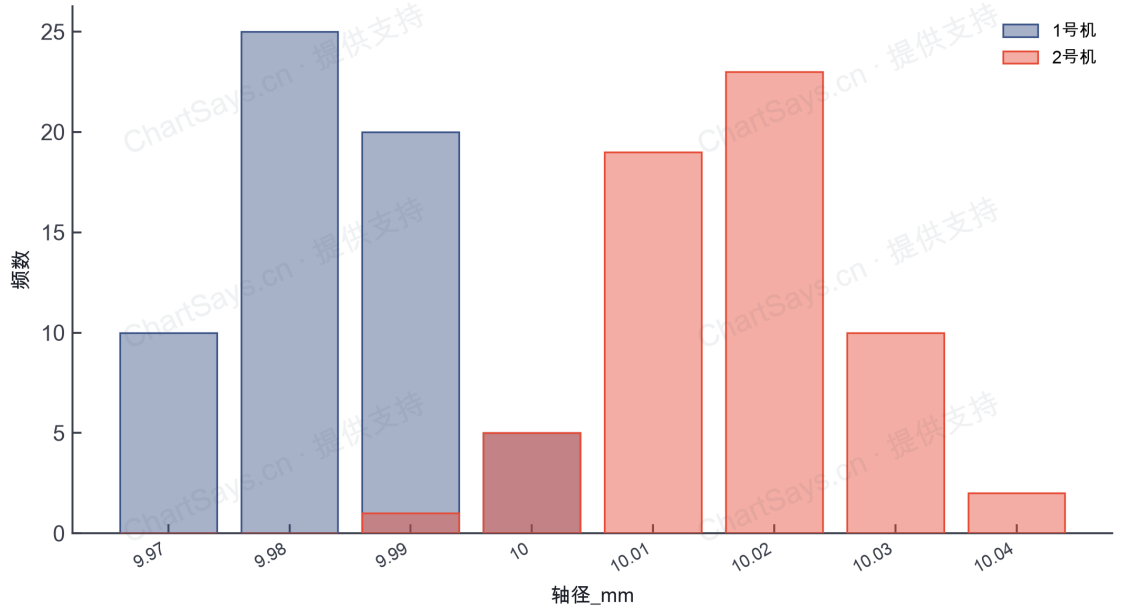


图 1-4　带组轮廓：60 件与 60 件各一色叠在一起。
不分色只看得到两个峰、分不清来源；分色后一眼看出两台机各占一个峰。
（分组着色版式；教材重绘）

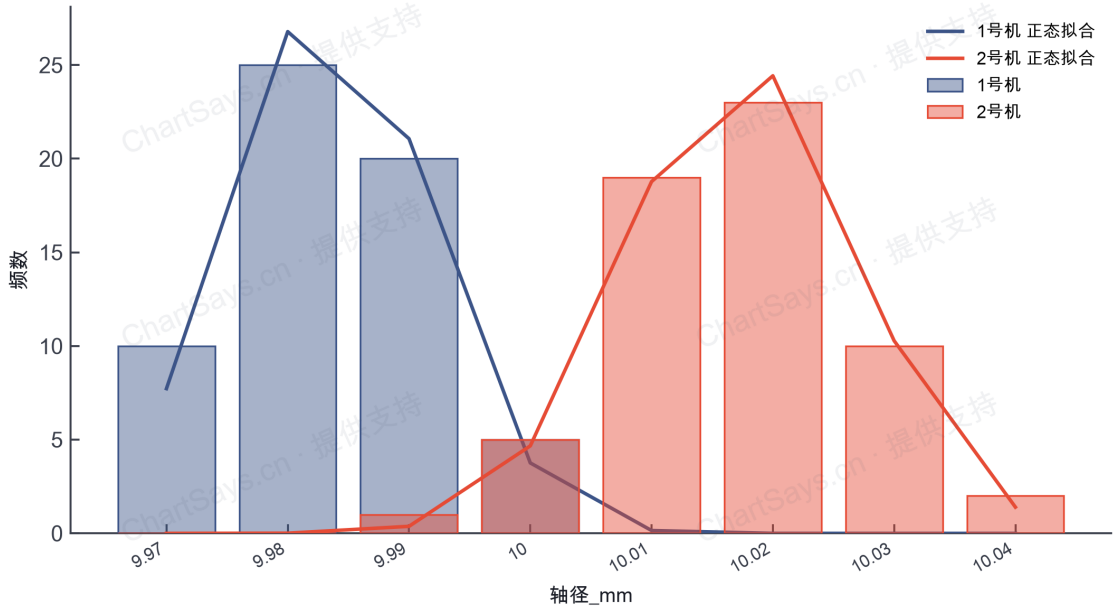


图 1-5　带拟合与组：每组再多一条同色正态线。
轮廓版看两组位置，拟合版看每组像不像正态。
（分组着色版式；教材重绘）

chartsays.cn · 教材据结果重绘

两组画在一张图上，先分色再谈结论。

分箱数：箱少看整体，箱多看细节

同一列 120 件轴径只改箱数：左 5 箱、右 12 箱。

读图指引（直方图）原话：「直方图的形状会随分箱数改变，小样本请改看点图或正态概率图，那两张不受分箱影响。」

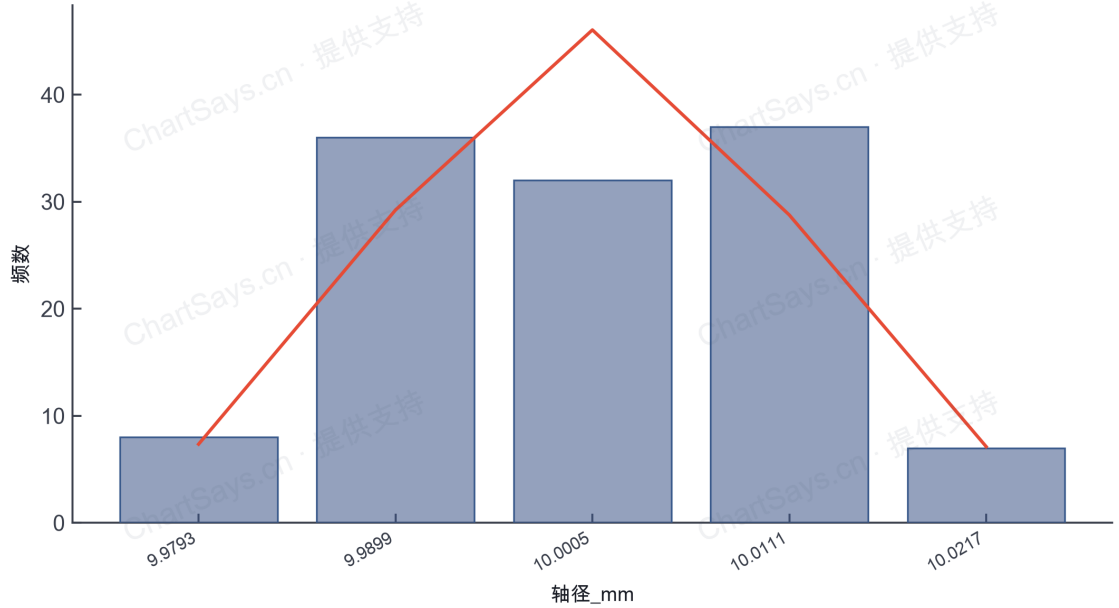


图 1-6　分成 5 箱：只剩一个宽包，细节被抹平。（教材示意，手工改箱数）

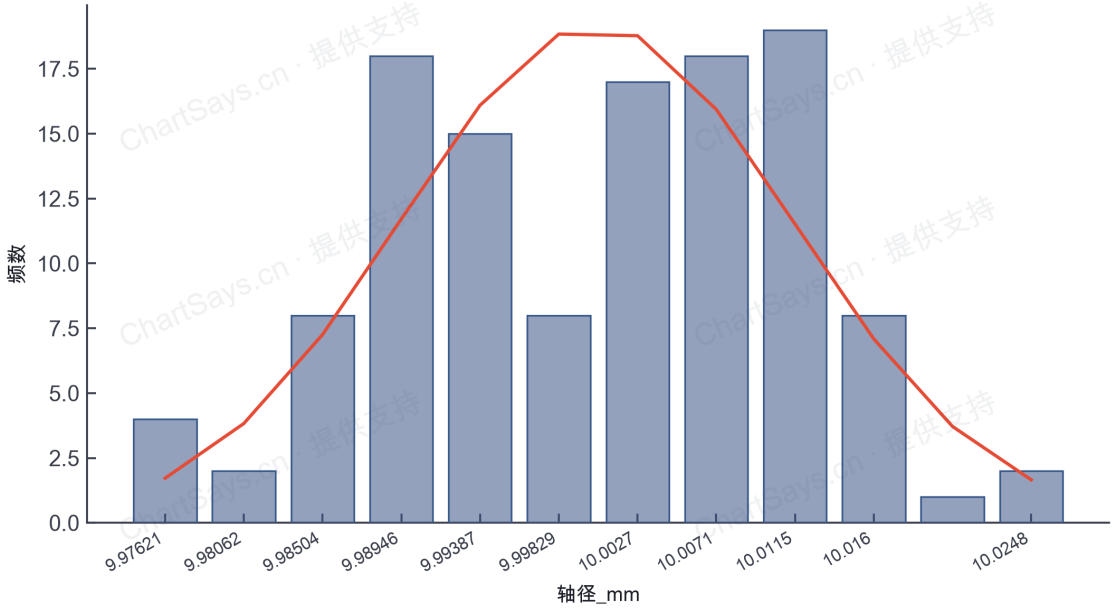


图 1-7　分成 12 箱：峰的细节露出来，也容易带偏。（教材示意，手工改箱数）

箱数少看整体，箱数多看细节。

要警惕的形状（一）：两个峰、右拖尾

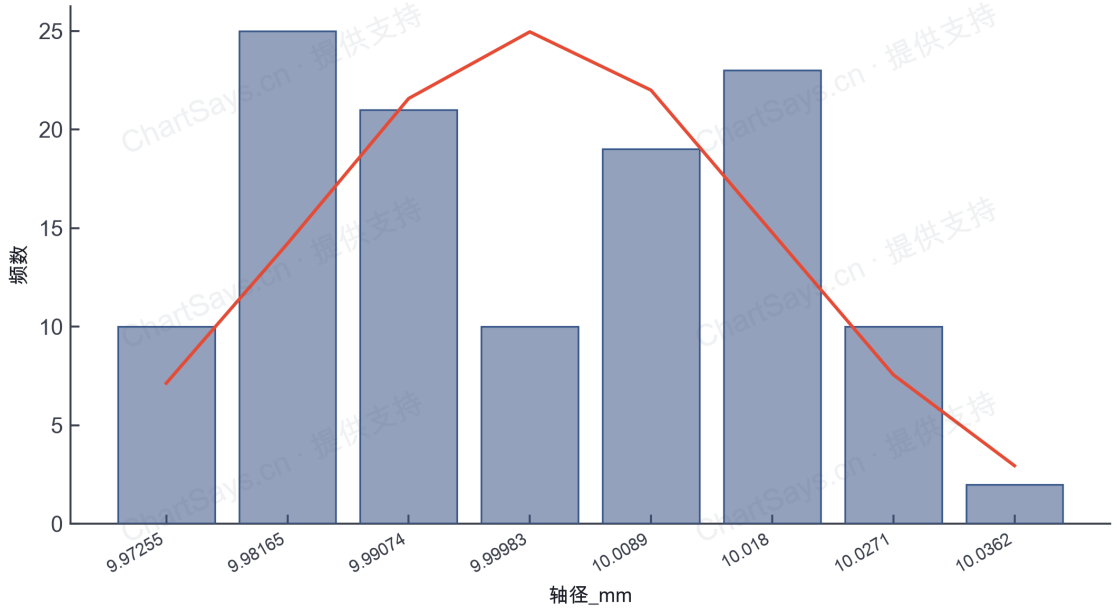


图 1-8　两机台合并成一列，各 60 件。
两个峰离得很开，平台提示「约 2 个明显峰」。
出处：读图指引（直方图）——混了不同来源。

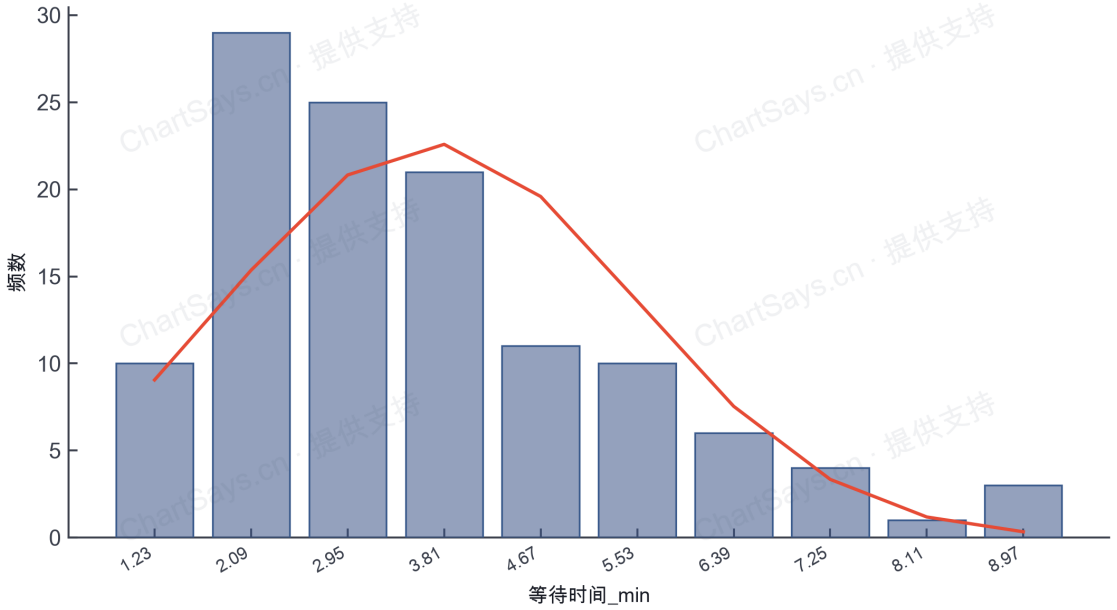


图 1-9　等待时间_min，N=120，偏度 1.0179。
长尾拖向大值一侧，红线压不住右边。
出处：读图指引（直方图）——别硬套正态。

chartsays.cn · 图由 ChartSays 生成

两个峰是混了来源，右拖尾是别硬套正态。

要警惕的形状（二）：一侧被削平

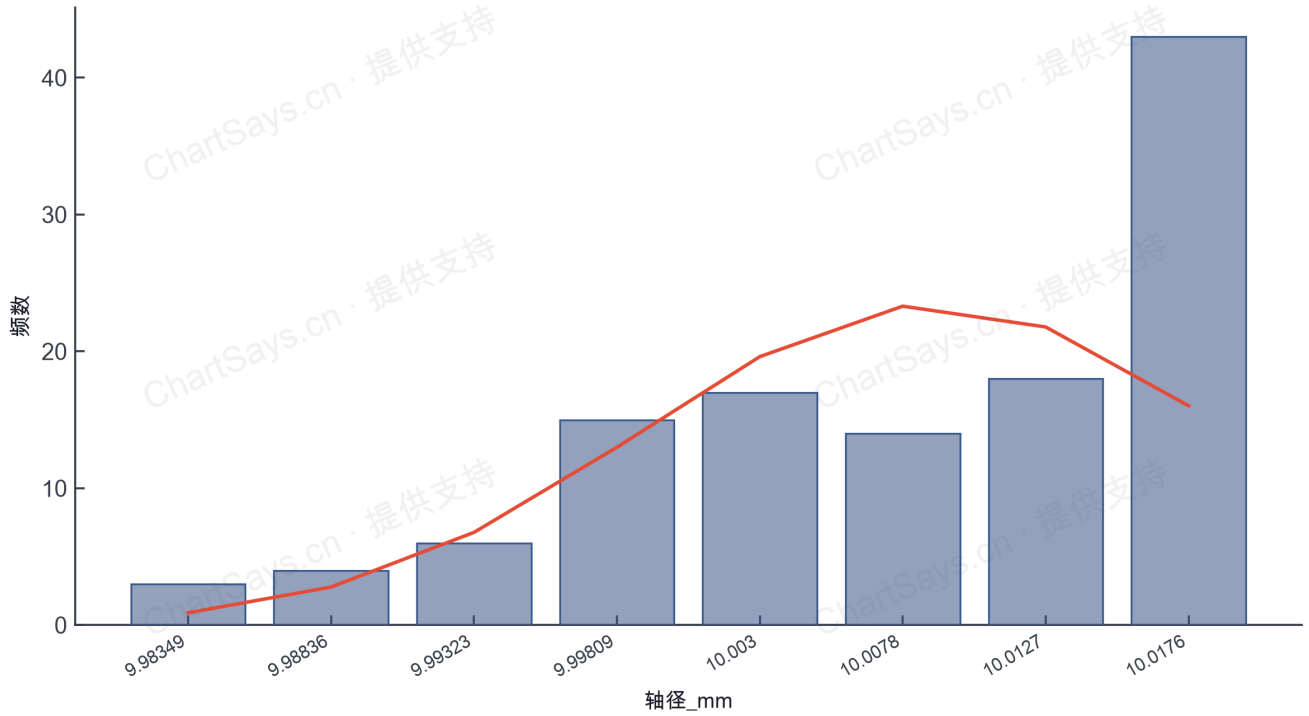


图 1-10　截尾记录 120 件轴径：最右一柱紧贴量程上限 10.02。

chartsays.cn · 图由 ChartSays 生成

要警惕的样子：一侧被削平贴边

最右一根柱子贴在 10.02 上突然拔高——这不是数据长得怪，是超量程被记成了上限。

出处：读图指引（直方图）：「某一侧被削平贴边」→ 数据被截断，先查记录规则。

在哪些方法里出现

120 件里有 26 件正好等于 10.02。截断会把标准差算小、把 PPM 算乐观，单值图、控制图、能力分析都会吃到这个偏差。ECDF 在同一处会「几乎垂直」，可以互相印证，都指向同一处截断。

一侧削平是记录规则的问题，不是数据长得怪。

第 2 章　点图：样本少的时候先看它

02　第 2 章 · 共 8 章　点图 Dotplot

什么时候用

样本只有几十个，直方图分箱一变形状就变的时候，点图是更老实的那张。

本章要点

- 每个观测画一个点，不丢任何一个数据。
- 多组并排时看几堆点的位置有没有错开。
- 点挤在少数几列上，先查量具分辨力和记录规则。

样本少的时候，点图比直方图可靠。

点图：每个观测一个点，不丢任何数据

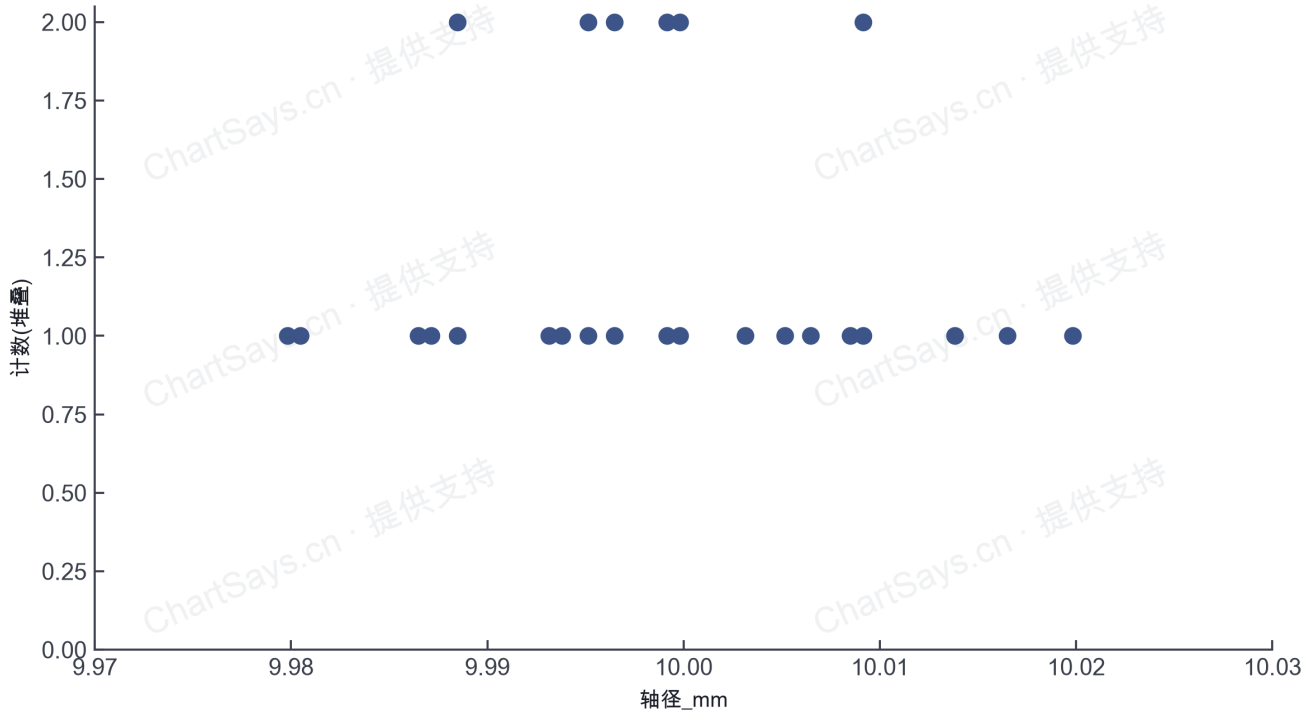


图 2-1　小样本 25 件轴径的点图：点沿值轴堆叠，堆得越高说明这个值附近的件数越多。

chartsays.cn · 图由 ChartSays 生成

这张图回答什么

样本只有 25 件，均值 9.9988、标准差 0.0107；这张图把每个观测都摆出来。

出处：读图指引（点图）——小样本首选它而不是直方图，「直方图的形状会随分箱起点变」。

先看什么、别被什么骗

先看整体形状：单峰还是双峰、左右对不对称；再看有没有远离主体的孤点。

「点只堆在少数几列上」多半是量具分辨力不够，先查最小刻度，别先怀疑工艺。

「分成两簇」多半混了两个总体，先按来源分层。

点图不丢任何一个观测。

点图变体：单列看堆叠，分组看错开

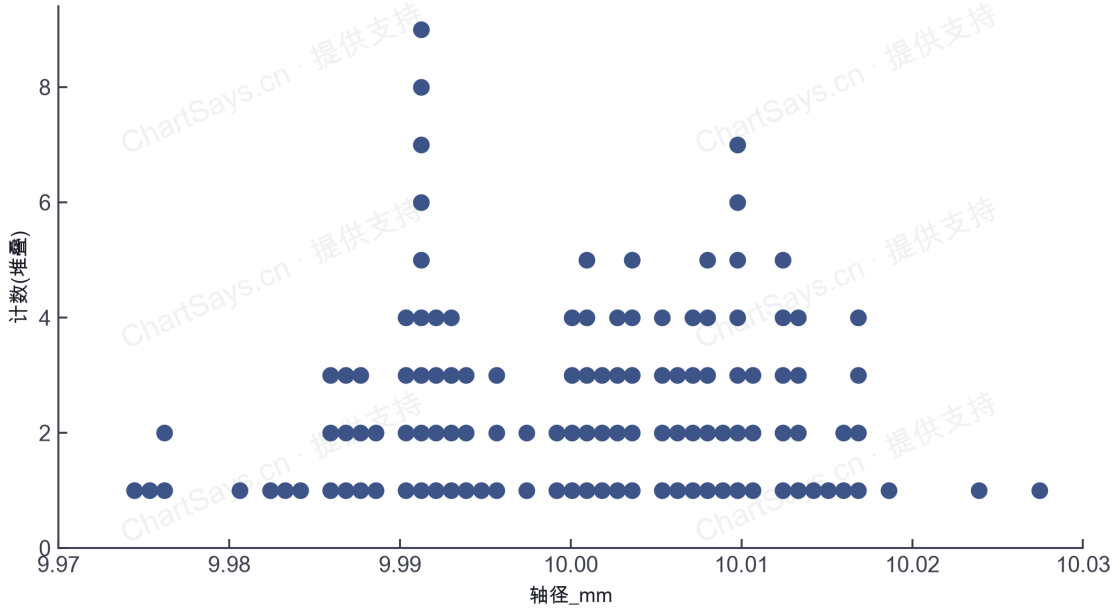


图 2-2　同一张点图换成 120 件全样本。
点多了堆叠最高能到 9 层，孤点还看得见。

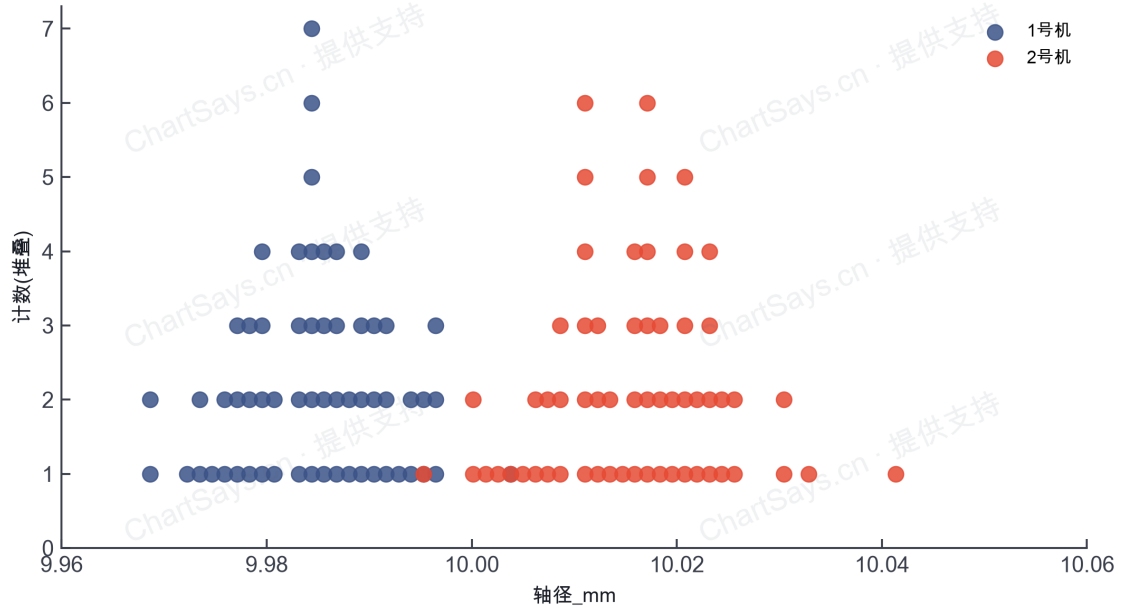


图 2-3　两机台分组点图，各 60 件。
两堆点整段错开：9.9847 对 10.0156。
出处：读图指引（点图）——多组时看几堆点错不错开。

单列看点密不密，分组看点错不开错开。

第 3 章　个体值图：点全画出来

03　第 3 章 · 共 8 章　个体值图 / Individual Value Plot

什么时候用

几组数据要摆在同一张图里比较，又不想让箱线图把点云压成一个盒子的时候。

本章要点

- 每个观测画一个点，按组分列。
- 半透明大圆点是各组均值，点云才是事实。
- 点云中心与大圆点差得远，先查极端值和记录错误。

个体值图：看到多少就是多少。

个体值图：点云是事实，大圆点是均值

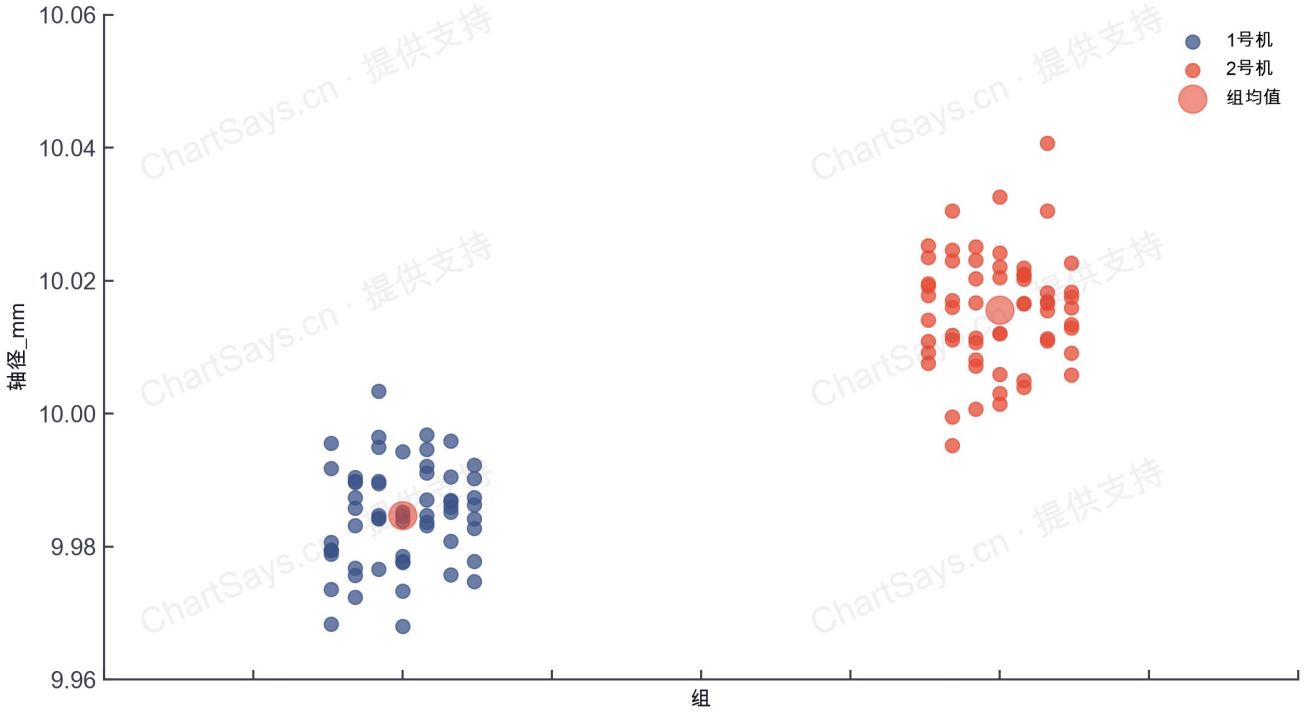


图 3-1　两机台各 60 件的个体值图：点云是原始观测，半透明大圆点是组均值。

chartsays.cn · 图由 ChartSays 生成

这张图回答什么

几组点云的中心有没有错开、散布宽不宽、有没有把某组的均值拽走的点。

出处：读图指引（个体值图）——小样本时用它替代箱线图，「箱线图要先算四分位数，n 很小时那几条线本身就不稳」。

先看什么、别被什么骗

横轴不给组名，靠颜色区分两台机，见图例。

「大圆点落在点云中心之外」→ 有极端值把均值拽走了，先追溯那个点是不是记录错误。

各组点数相差悬殊时，别只看大圆点高低。

大圆点是均值，点云才是事实。

第 4 章　箱线图：五个数交代位置与散布

04　第 4 章 · 共 8 章　箱线图 Boxplot

什么时候用

要多组并排比较位置、散布和离群，又只有很小的版面可用的时候。

本章要点

- 中位线与 Q1、Q3 画成盒子，两条须到 1.5 倍四分位距栅栏以内的最远点。
- 盒子窄＝主体集中；中位线居中、两须相当＝对称。
- 须外圆点不等于坏数据，双峰也会被它完全藏起来。

箱线图用五个数交代一组数据。

箱线图：位置、箱高、须、圆点

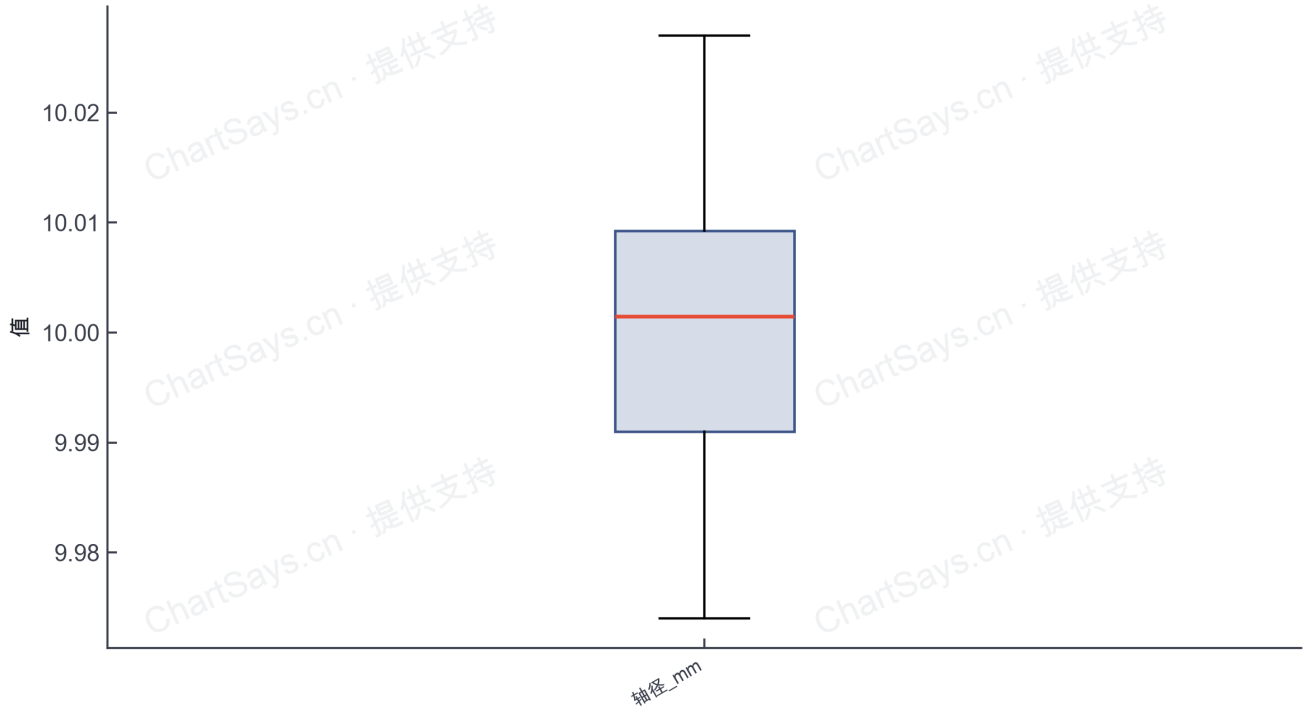


图 4-1　轴径_mm 的箱线图，N=120：盒内是一条中位线，须端是 1.5×IQR 栅栏内的最远点。

chartsays.cn · 图由 ChartSays 生成

这张图回答什么

位置看中位线、散布看盒子高度、对称看两须长度、离群看须外的圆点。

出处：读图指引（箱线图）——「用最小的面积交代位置、散布和对称性，也最适合多组并排比较」。

五个数都是什么

中位线 10.0015，盒子从 Q1 9.991 到 Q3 10.0092，两须到 9.974 和 10.027；须长按 1.5 倍四分位距的约定截断。本例中位线落在盒子正中，形状对称。

位置看中心，散布看箱高，圆点查离群。

箱线图：双峰会被箱子藏起来

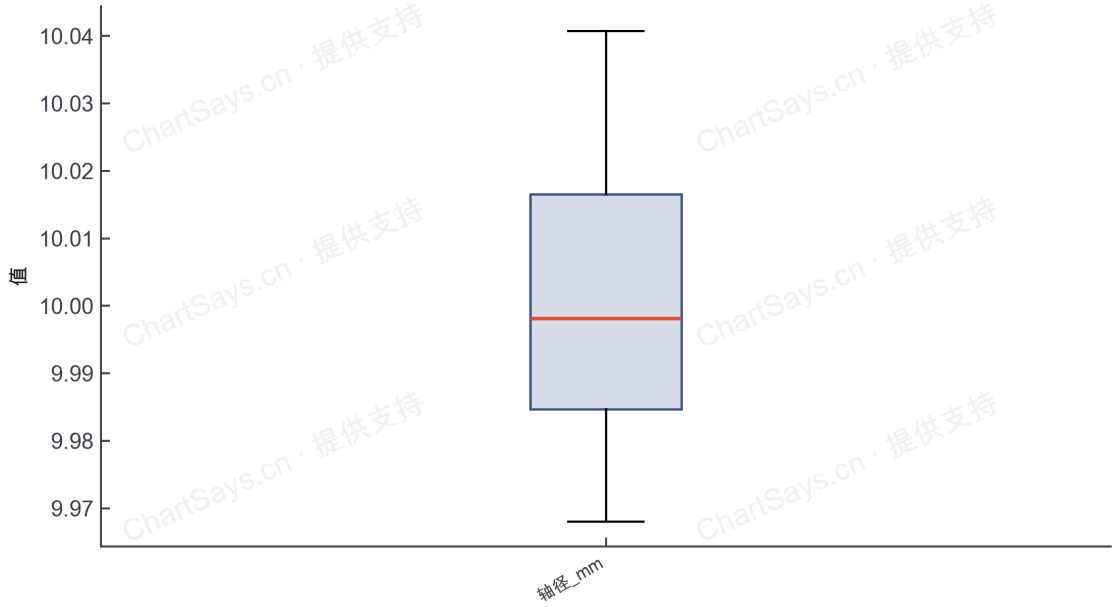


图 4-2　两机台合并的 120 件轴径。
中位线 9.9981，盒子 Q1 9.9847 到 Q3 10.0165，两须到 9.968 和 10.0407。
盒子宽、中位线大致居中，看不出是两个峰拼的——箱线图会把双峰藏起来。

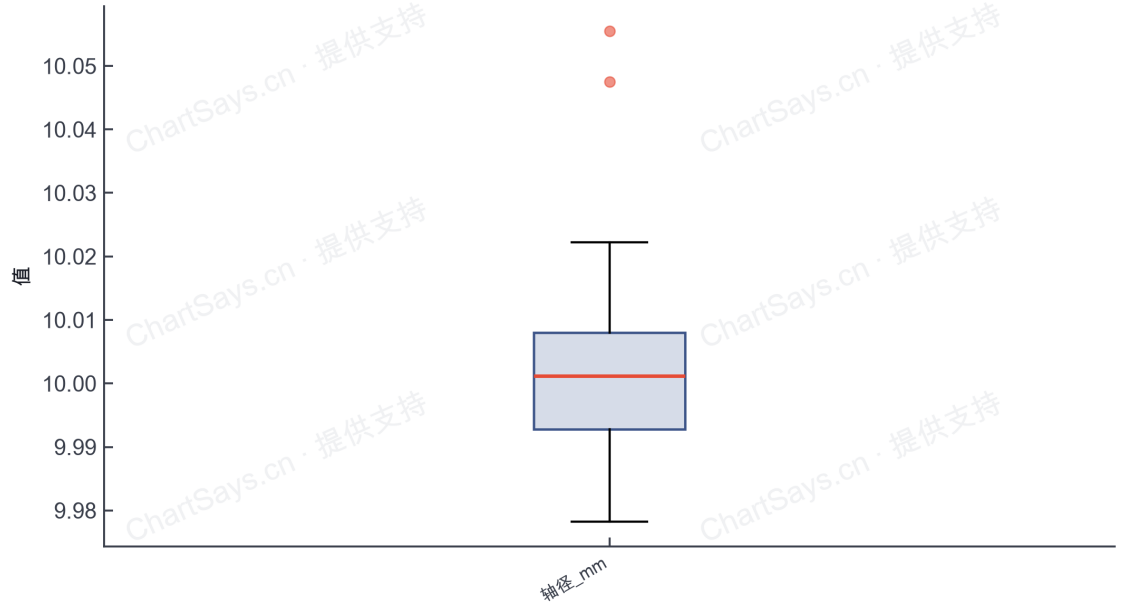


图 4-3　末端翘起的 120 件轴径。
须外 2 个圆点，最长的一条尾巴翘到 10.0555。
出处：读图指引（箱线图）——「须外的点不等于坏数据」。

chartsays.cn · 图由 ChartSays 生成

箱线图会把双峰藏起来，须外圆点别当坏数据删。

第 5 章　ECDF：累计比例怎么读

05　第 5 章 · 共 8 章　经验累积分布 ECDF

什么时候用

要回答「低于某个值的比例是多少」「中位数在哪」这类比例问题的时候。

本章要点

- 横轴是取值，纵轴是累计比例。
- 从 0.5 横切读中位数，从取值竖切读比例。
- 完全不受分箱影响，但尾部只由极少数点决定。

比例类问题，先想到 ECDF。

ECDF：比例问题，一条台阶线

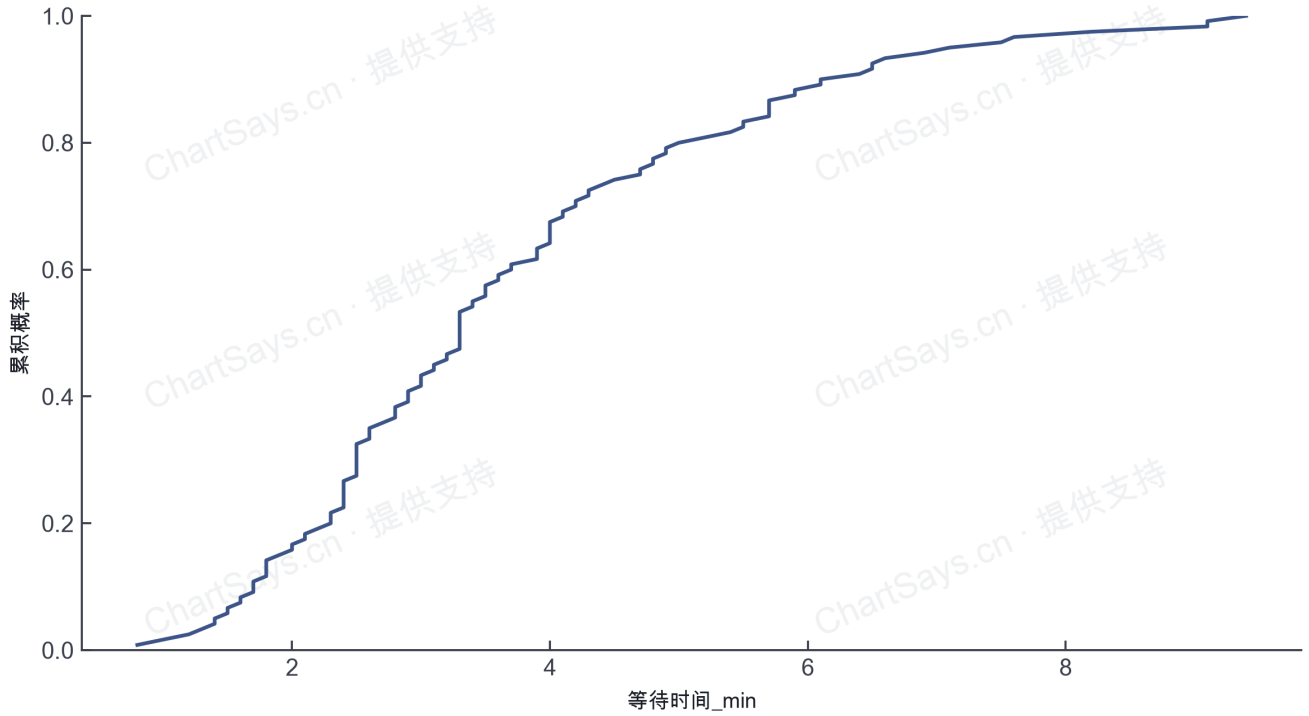


图 5-1　等待时间_min 的 ECDF，N=120：中位数 3.3 落在纵轴 0.5 处。

chartsays.cn · 图由 ChartSays 生成

这张图回答什么

等待时间_min 里「低于 3 分钟的比例是多少」这类问题，直接读一条台阶线。

出处：读图指引（ECDF）——「它比直方图更适合回答比例类问题，而且完全不受分箱方式影响」。

怎么读、别被什么骗

中位数就在纵轴 0.5 的横切点上：3.3。

曲线在某处「几乎垂直」，说明大量观测挤在同一个值上；先查分辨力，再查是否截断。

尾部只由极少数点决定，别拿最末几级台阶外推。

从 0.5 横切读中位数，从取值竖切读比例。

第 6 章　正态概率图：别拿直方图判正态

06　第 6 章 · 共 8 章　正态概率图 / Normal Prob Plot

什么时候用

要判「这列数像不像正态」的时候用它；检验的 p 值卡在边界上、想亲眼看一眼偏离在哪一段，也用它。

本章要点

- 横轴是标准正态下「该在」的位置，纵轴是实际值。
- 点大体贴着红线成一条直线就是正态。
- 不受分箱影响，小样本也能看，是判正态的首选图。
- 四种要认的形状：弓形＝偏斜，S形＝尾巴比正态厚，末端翘起＝离群值，折线＝混了两个总体，先按来源分层。

判正态看概率图，不看直方图像不像。

正态概率图：点贴线就是正态

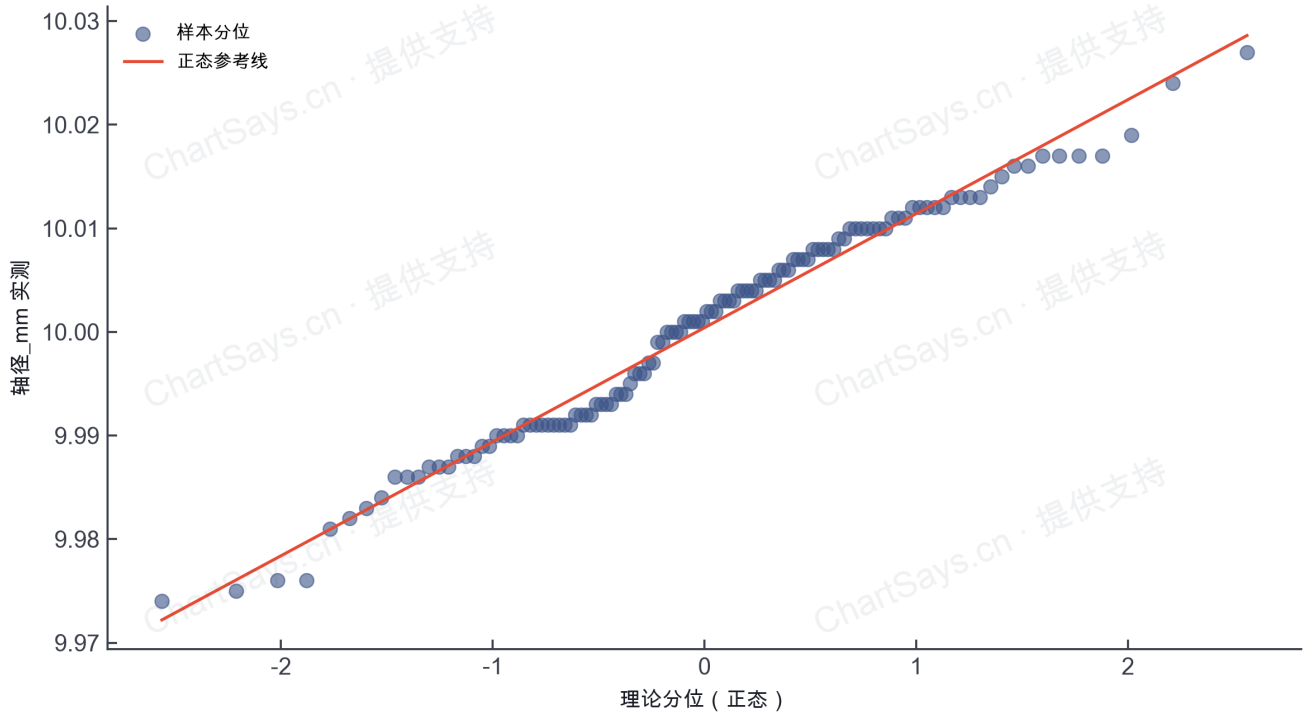


图 6-1　轴径_mm 的正态概率图，N=120：点贴线，末梢只有轻微摆动。

chartsays.cn · 图由 ChartSays 生成

这张图回答什么

把 120 个观测从小到大排好，横轴放它在标准正态下「该在」的位置，纵轴放它「实际」的值；红线是按本列均值与标准差画的参考直线。

出处：读图指引（正态概率图）——「点大体贴着红线成一条直线」就是正态。

为什么不用直方图判正态

直方图的形状会随分箱数改变，概率图不会。

本例 5 项检验里有 1 / 5 判为拒绝正态，平台结论是「多数检验判为正态」；Shapiro-Wilk 的 p 值是 0.0807，大于 0.05。

点贴线是正态，系统性弯才是问题。

四种形状（一）：贴线、弓形

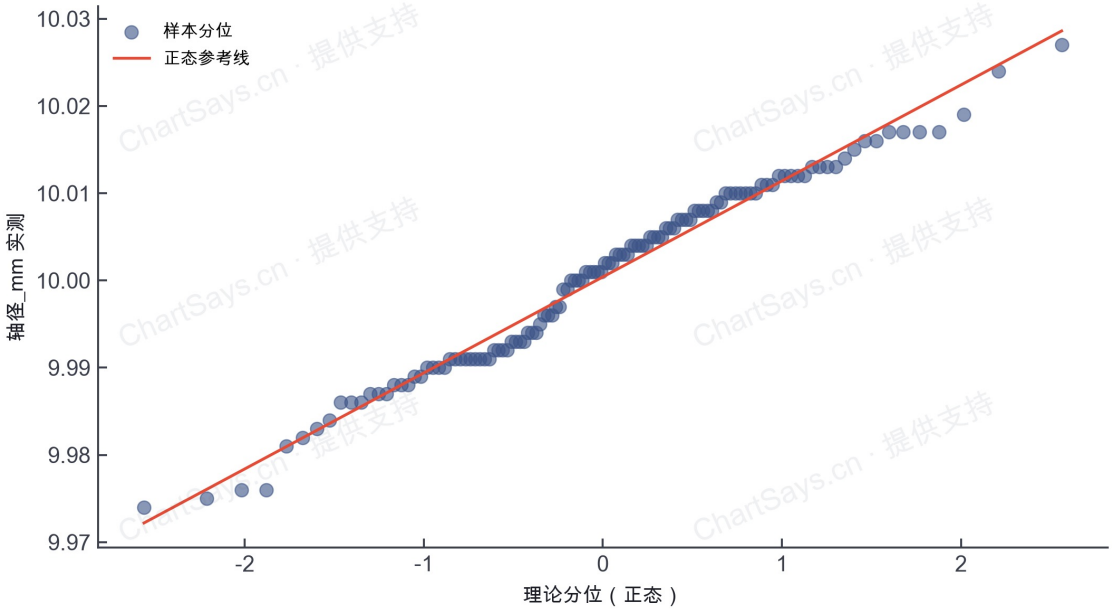


图 6-2　直线：轴径_mm。
点大体贴线，轻微上下摆动是正常抽样波动。
出处：读图指引（正态概率图）——看「有没有系统性的弯」。

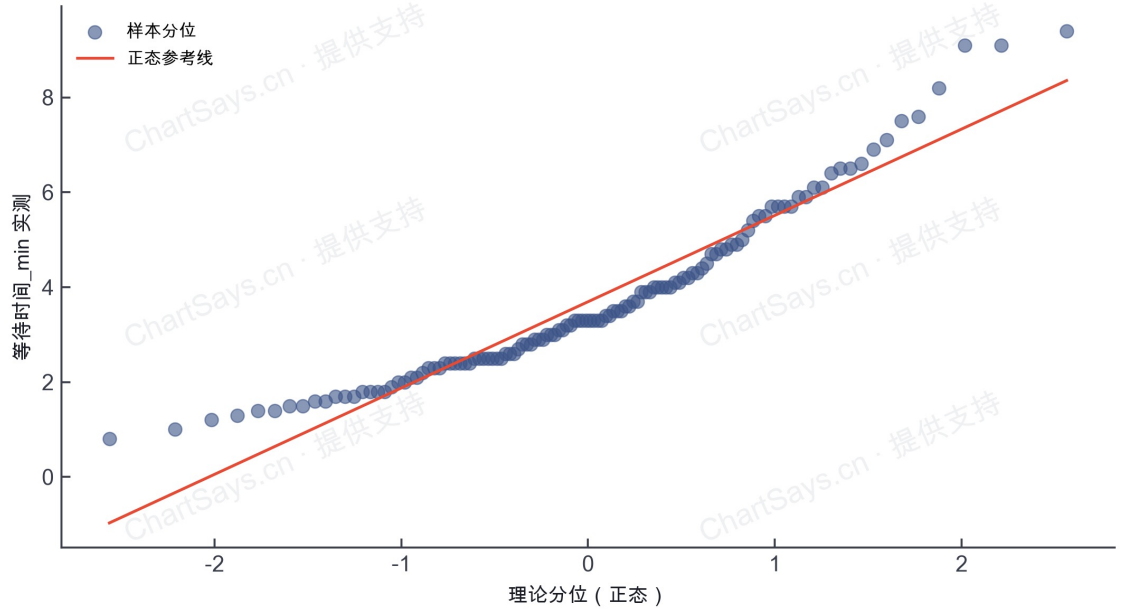


图 6-3　弓形：等待时间_min，偏度 1.0179。
两端都在参考线上方、中段在下方，长尾在翘起的那一边。
出处：读图指引——「先试对数或 Box-Cox 变换」。

chartsays.cn · 图由 ChartSays 生成

弓形是偏斜，弯向哪边尾就在哪边。

四种形状（二）：S形（厚尾）、末端翘起

两机台合并的数据（同 P8、P17）概率图上是几段折线——混了两个总体，先分层。

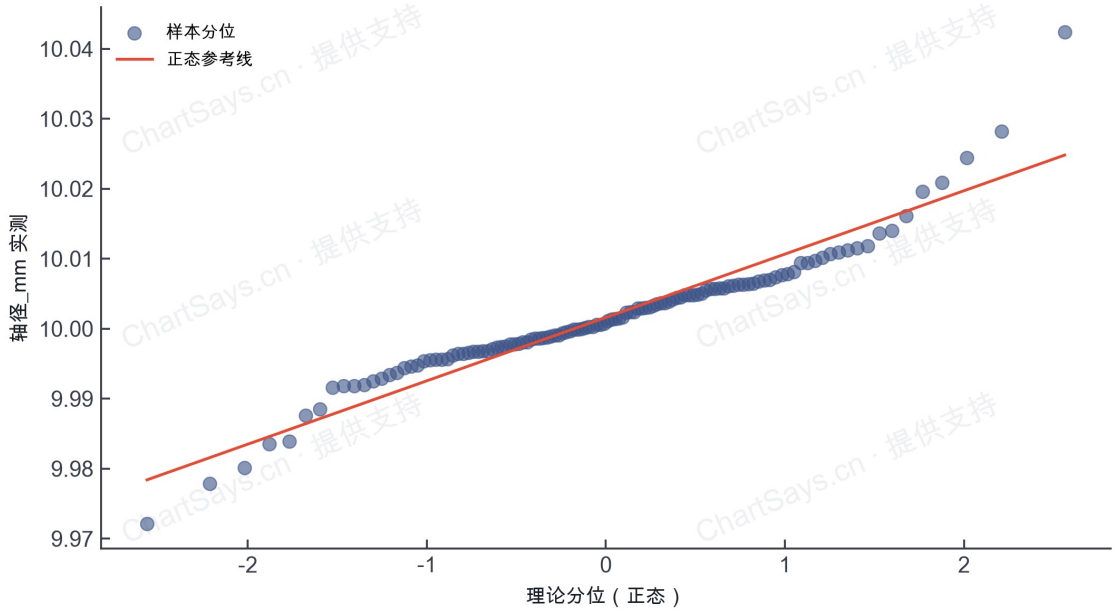


图 6-4　S形（厚尾）：合成的 120 件轴径，偏度 0.5962、峰度 4.0975。左端在线下、右端在上，中段贴线——尾巴比正态厚（出处：读图指引）。

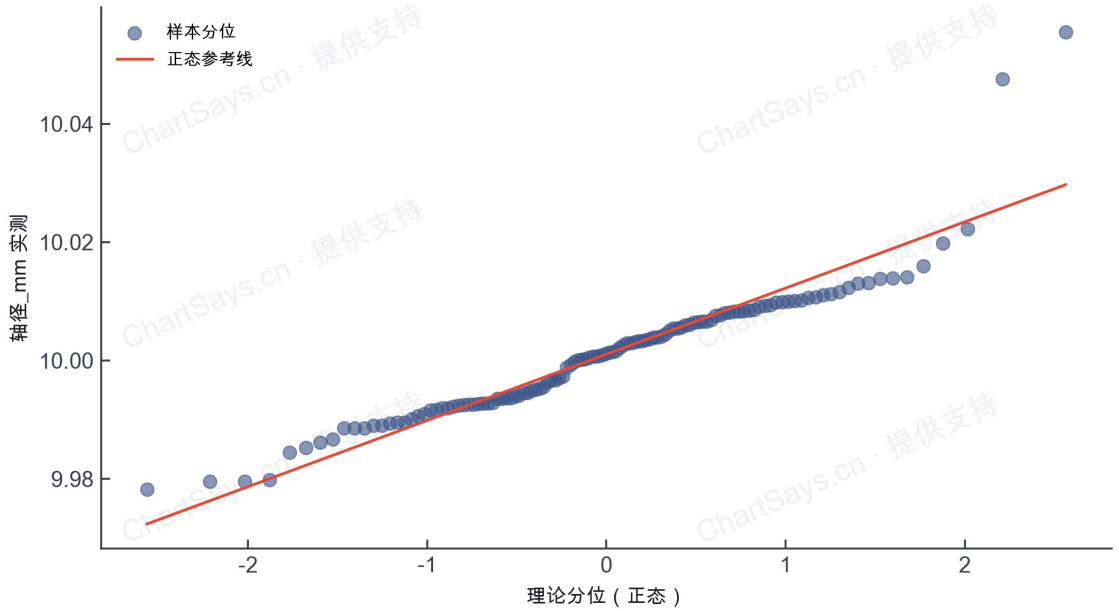


图 6-5　末端翘起：120 件轴径，偏度 1.3314、峰度 5.3642。只有末端一两个点远离直线，其余笔直（出处：读图指引）。

S形是尾巴比正态厚，末端翘起是离群值。

第 7 章　分布拟合：该用哪种分布

07　第 7 章 · 共 8 章　分布识别 / Distribution ID

什么时候用

要用能力分析、控制限或 PPM 的时候：先让候选分布排个序，再挑一个说得过去的。

本章要点

- 13 个候选按 AIC 排序，前 4 名的概率图并排看。
- 点贴对角线的那个就是拟合最好的。
- 头部两名差得少就别急着换，先看尾部和工艺机理。

分布决定方法，先让候选分布排个序。

分布拟合：概率图看尾部，叠加图看主体

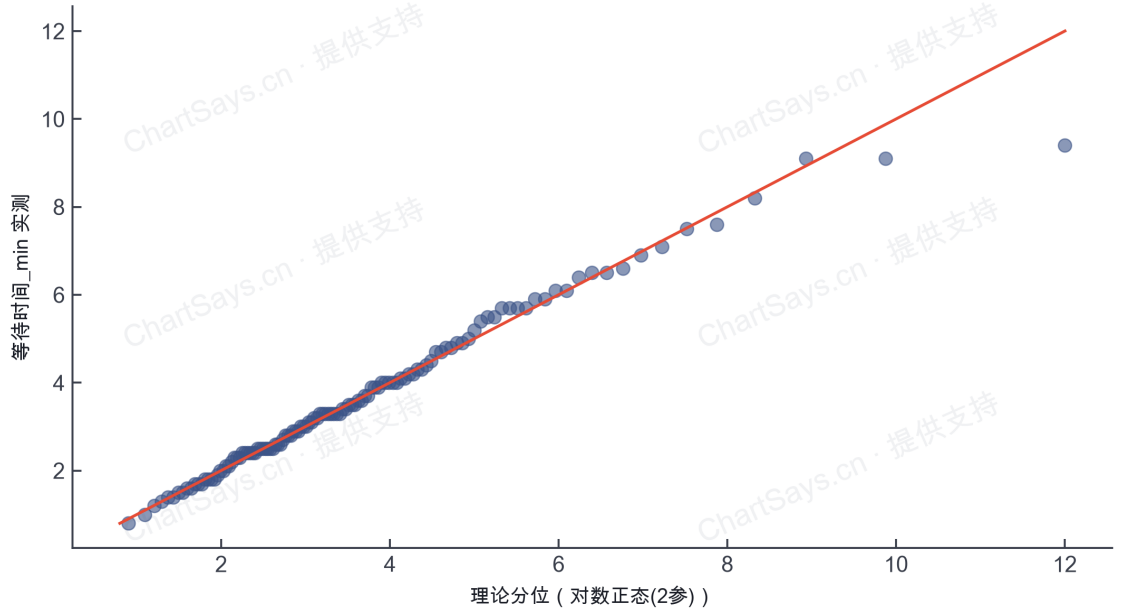


图 7-1　最优候选的概率图：横轴是该分布下的理论分位，纵轴是实测值。对数正态(2参) 的 AD 统计量 0.1895，是 13 个候选里最小的。
出处：读图指引（分布识别）。

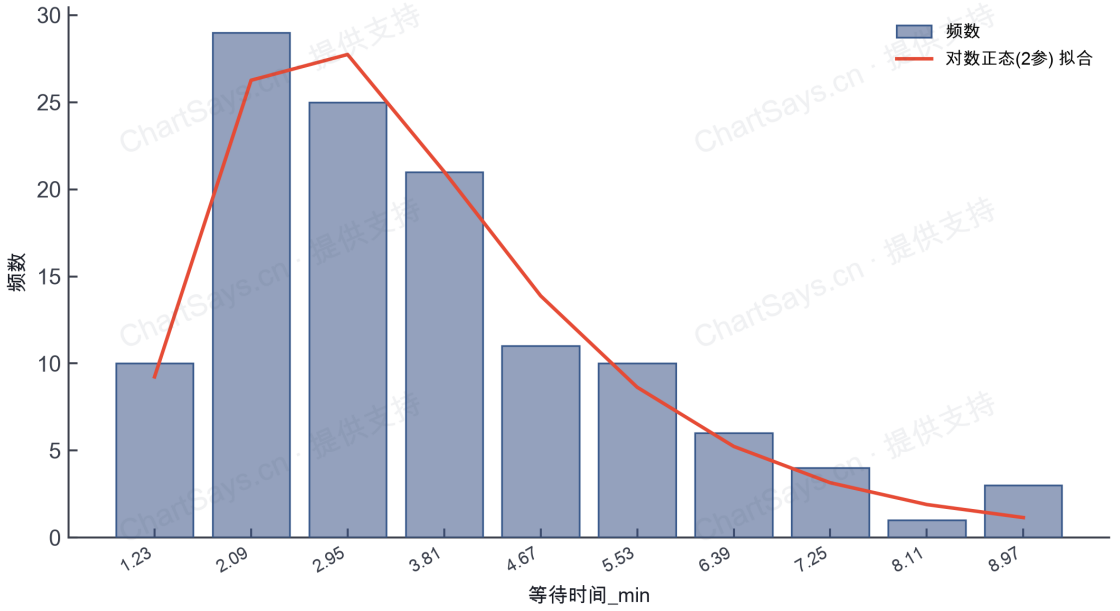
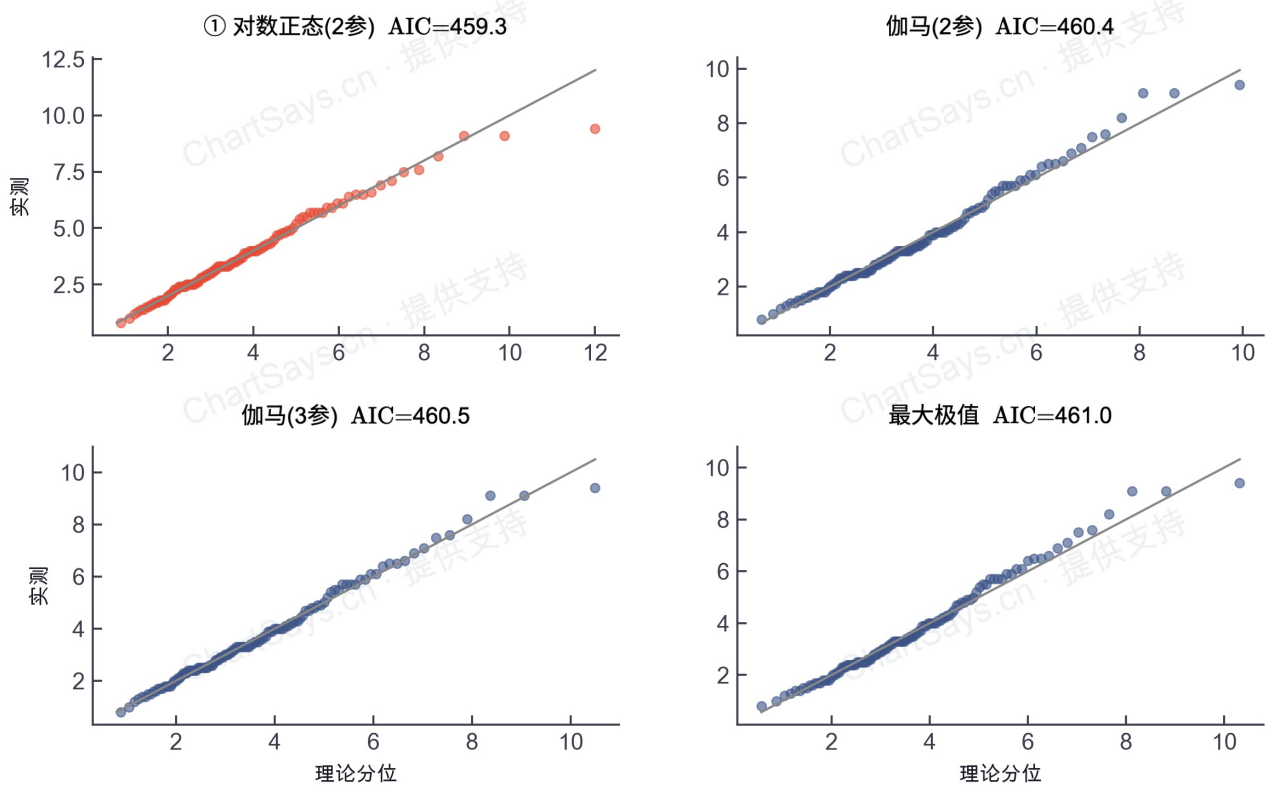


图 7-2　最优分布的直方叠加：拟合线同时压住峰位与两侧尾形。
出处：读图指引（直方叠加）——「比只贴住峰顶更可信」。

chartsays.cn · 图由 ChartSays 生成

概率图看尾部，直方叠加看主体。

分布拟合：候选分布一次排完



chartsays.cn · 图由 ChartSays 生成

这张图回答什么

13 个候选分布先按 AIC 排完，图上是排在前 4 名的，每个格子里点贴对角线的程度就是它像不像。

怎么读、别急着换

13 个候选里最优的是对数正态(2参)，AIC 459.35；紧跟其后的是伽马(2参)，AIC 460.36，ΔAIC 只有 1.01（小于 2），两者难分伯仲，先看尾部和工艺机理。

正态排在第 10 名，AIC 487.18，比最优大 27.83。

「所有格子都偏离对角线」→ 候选里没有合适的，先做变换或分层。

两个候选差得少，就别急着换分布。

第 8 章　变换：偏态怎么救回来

08　第 8 章 · 共 8 章　Box–Cox 变换 Transform

什么时候用

数据明显右拖尾、正态检验把 p 值压到很低，但后续方法又要求近似正态的时候。

本章要点

- 先看变换前后的两张直方图，别看 λ 的数字。
- Box–Cox 只在数据为正时可用，λ 是估出来的。
- 变换改变了数据本身的含义，结论要换回原量纲说。

偏态先别硬套正态，先试变换。

变换前后：柱子从歪到正

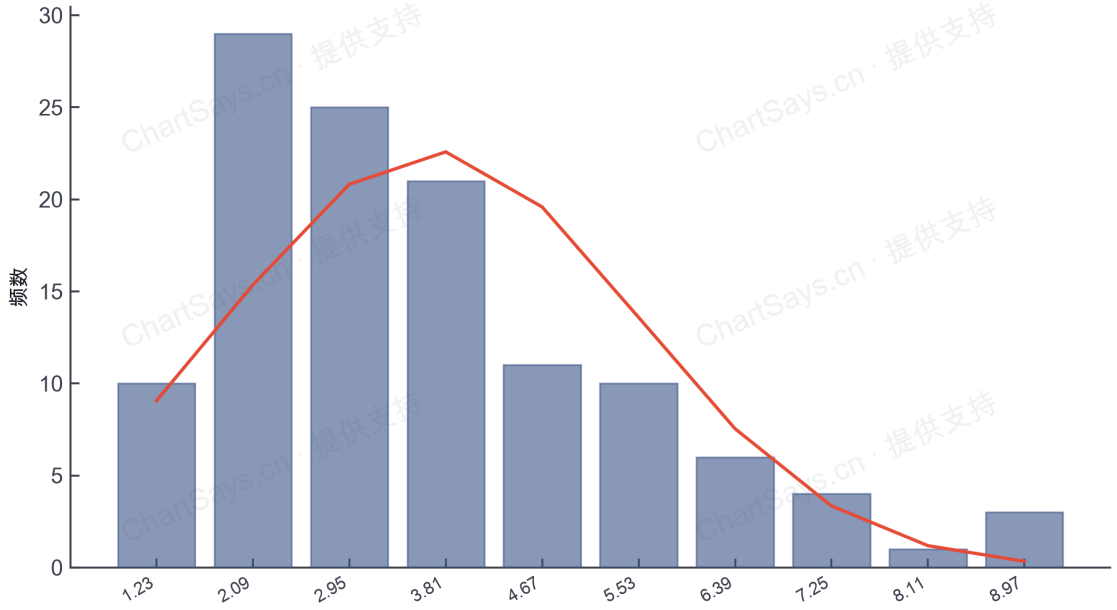


图 8-1　变换前：等待时间_min，N=120。
偏度 1.0179，Shapiro p 值 <0.001。
出处：读图指引（直方图）——「明显右拖尾」别硬套正态。

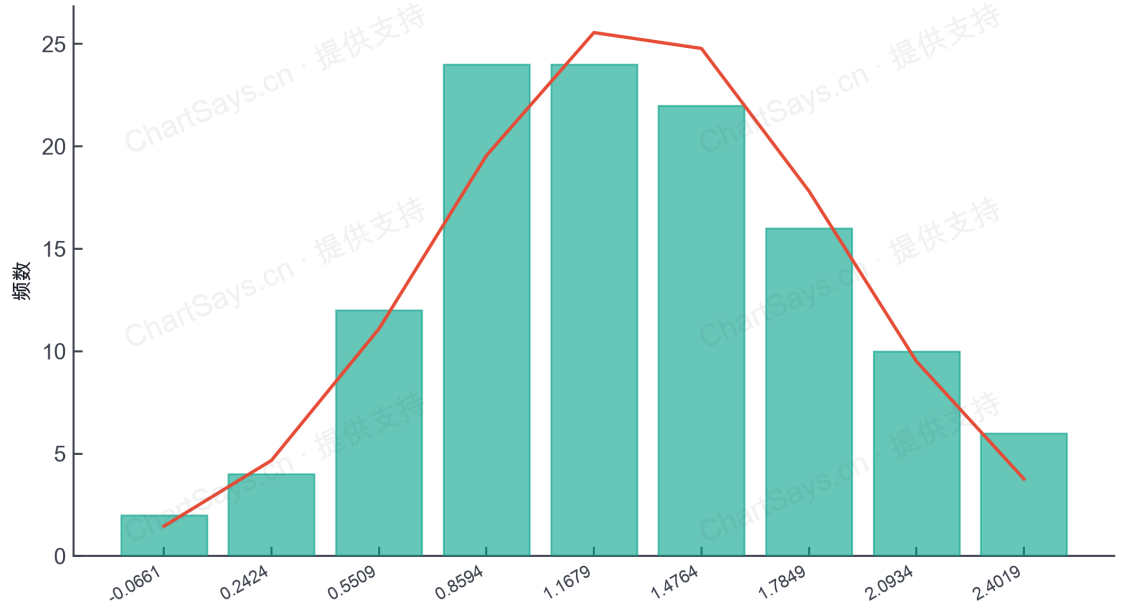


图 8-2　变换后：Box-Cox 取 λ=0.1151。
偏度 -0.0052，Shapiro p 值 0.9017。
出处：读图指引（正态概率图）——目检拿不准时以检验结论为准。

chartsays.cn · 图由 ChartSays 生成

λ 是估出来的，换一批数据就会变。

图之间怎么选：换图的几种情形

你现在的情形	先看这张图
样本只有几十个，还想看形状	点图（直方图会随分箱变）
几组要摆在一起比	个体值图 / 箱线图
只关心比例、分位数	ECDF 台阶线
像不像正态	正态概率图 + 正态性检验
该用哪种分布	分布拟合排序 + 概率图
偏态想救回来	变换前后直方图

出处：读图指引（直方图 / 点图 / 个体值图 / 箱线图 / ECDF / 正态概率图 / 分布识别）。

挑图三步：先问要回答什么（形状、比较、比例还是分布），再看样本量够不够（几十个以下先点图），最后问分箱会不会骗你（形状敏感就换 ECDF 或概率图）。

样本量决定用哪张图，要回答什么决定看哪里。

常见问题（一）：直方图最容易踩的六个坑

直方图能判正态吗?

不能。形状随分箱变，判正态要看概率图和正态性检验。

看到两个峰怎么办?

先别算均值：多半混了两个来源，按机台或批次分层各画一张，比一比就看得见。

ECDF 和直方图差在哪?

ECDF
不受分箱影响，直接读比例与分位数；直方图看形状更直观，也更容易被分箱带偏。

柱子越多越好吗?

不是越多越好。箱少看整体，箱多看细节；样本少的时候，箱子多了全是噪声。

须外的点能删吗?

不能。1.5
倍四分位距是画图约定不是检验，先追溯工况再决定要不要处理。

样本多少算少?

几十个以下：分箱一变形状就变，先看点图和正态概率图。

直方图管形状，判正态要看概率图。

常见问题（二）：箱线图、ECDF 与选图

01

箱线图藏起双峰

盒子只给五个数，两个峰在箱线图上和单峰长得一样。形状问题请配合直方图或点图一起看。

02

须外圆点别乱删

须外的点是按 1.5×IQR 这条画图约定判出来的，不是检验，别一律当坏数据删掉。要正式判定单个极端值，用基础统计量里的 Grubbs 检验，先追溯工况，再看怎么处理。

03

尾部台阶别外推

ECDF 的尾部只由极少数点决定，最末几级台阶只能看个大概，不能拿来外推极端分位数。

出处：读图指引（箱线图 / ECDF / 正态概率图）；Grubbs 检验出自 ChartSays 报告的基础统计量。

须外圆点别删，尾部台阶别外推。

用这份数据自己跑一遍

1. 打开内置示例「螺栓扭矩能力分析」（80 个拧紧扭矩）
2. 选「基本统计量」，度量列选「扭矩_Nm」，加正态拟合线
3. 再点「正态性检验」和「分布识别」，对照本讲的判断顺序

扫码查看方法说明

其中轴径、两机台、等待时间、截尾、小样本取自教材配套合成数据包；末端翘起、厚尾按随机种子 20260917 生成。

chartsays.cn/methods/basic-stats

下一讲

02 做比较与看构成的图
箱线图、帕累托图、条形图

判读口径出自 ChartSays 读图指引（直方图 / 点图 / 个体值图 / 箱线图 / ECDF / 正态概率图 / 分布识别）。

数据：轴径、两机台、等待时间、截尾、小样本、末端翘起、厚尾，七份均为合成数据。

© 2026 ChartSays · 2026 年 9 月版