

终检试验区的关联工具

确保轮胎试验机获取正确的测量值并具有相关性，从而保证轮胎满足顾客需求，对于任何制造商都是一项关键任务

作者：Shaun M. Immel博士，美国Micro-Poise测量系统公司首席技术官

在全球轮胎制造厂中，轮胎终检测试系统一般每天要测试几千条轮胎，以评定多种轮胎的最终质量。

为了完成这种大量的测试工作，大多数工厂都要使用多个轮胎均匀性试验机、动态平衡试验机和几何尺寸测量机械。有时候，这些轮胎制造商需要保证他们的轮胎终检检测设备相互之间具有一致性。

在工业界，这种需求一般指检测设备的相关性。轮胎客户自然想知道他们所购买的轮胎的质量是否在不同的检测设备上进行了正确、一致地评估。

为此，他们希望由不同检测设备检测的一条轮胎的质量检测结果要一致。轮胎制造商要求终检设备供应商验证并担保他们检测设备的相关性。

“轮胎制造商要求终检设备供应商验证并担保检测设备的相关性”

本文将研究并讨论用于评估终检设备相关性（或者是，作者所认为的“测量一致性”）的几种工具。

用于研究这种现象的传统相关方法和工具应用于终检测试和测量设备时有一定的局限性，因为这些工具和方法更适合研究系统的行为，用户设法

找出两个或多个变量之间的内在联系，从而可以对该系统进行控制，或者可预测系统未来的行为。

但对于上述轮胎测试而言，真正所需要的是要确保用于检测同类轮胎的几台检测设备的检测结果相互一致。这种检测一致性实际上是常见测试相关性的一个特例。

在考虑测量一致性问题时，有一点应特别关注。如果一组不同的检测系统（机械或方法）中各个系统都不具备较高水平的测量重复性，则很难评估他们之间的一致性。

由于重复性差，测试结果分布（范围）非常宽，实际上很难评估测试一致性。

这种情况的一个泛例示于图1。在图的左边示出了两个概率分布（一个是黑色，一个是红色），它们的平均值相差一个单位， $\sigma = 1.25$ 单位。

在图1的右边示出了两个相似的概率分布，它们的平均值相差相同的量， $\sigma = 0.25$ 单位。很难判断左方的分布是否不同。但是，对右方的分布进行目测判断就要容易得多。左方存在大范围的分布叠加，很难确定一致性。

为了解决重复性差的问题，推荐在进行相关测试前对检测设备进行调节和维护。这样就可以尽可能地提高重复性。

还建议由每个测试设备/方法获取多个测试值，用各个测试设备/方法所获取结果的平均值进行测量一致性分析。

这样在进行测量一致性分析时有助于消除机械/方法重复性差的影响。

无论采用什么样的标准来评估设备的一致性，都很难回答人们经常会提出的一个问题“这些检测设备结果一致吗？”

在确定这种一致性特征时，每种工具和方法都有其优缺点，但很难对数值计算结果设定限定值来做出相符/不相符判定。

另外，在检测设备的整个测量范围内测量一致性真得就如此重要吗？大多数人都会同意，在质量筛选极限值附近设备测量一致性当然最重要，但对于极高质量和极低质量轮胎而言，由不同设备得到的检测结果不能完全匹配并非很重要，这种情况下设备测量一致性就不太重要了。

在评估检测设备/方法一致性时的另一个重要问题是，没有用于轮胎终检检测设备的真正的NIST可溯源(NIST-traceable)尺度标准。

对于大多数测量装置而言，都存在可靠稳定的NIST可溯源标准，可以对各种测试装置进行比较（NIST是美国

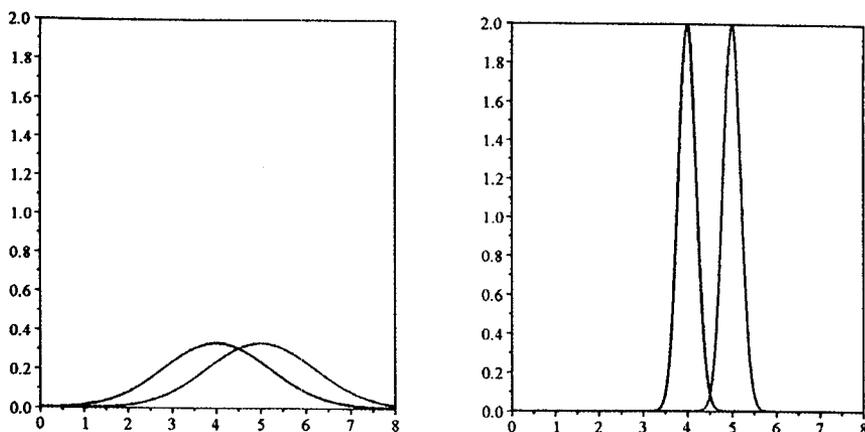


图1：测量一致性对比表明可重复性差（左）和可重复性好（右）

商务部美国国家标准技术研究院)。轮胎终检检测设备没有这种可靠、稳定的标准，没有聚合物轮胎之类的标准装置可用于评估测量一致性或校准终检设备。

在实践中一般使用“标准轮胎”。标准轮胎是从标准生产过程中抽出的轮胎，一般具有固有的测试稳定性，而且在各种测试设备上多次测试，以此进一步稳定其测试值。

但是，由于在测试过程中轮胎会老化、磨损，这些标准轮胎的性能和测试值会随时间而发生变化。

只有选用高质量的机械，正确处理机器设定、调节、维护、校准和其他类似重要因素来确保精确的测量，才可以真正评估测试一致性。

到目前为止，测试一致性的关键是设备的精度。

如果轮胎在完全相同的条件下放置到两个或多个测试设备上，测量值应当总是一致的。在评估测试设备的一致性时，更重要的是在调节或校正设备一致性时，重点永远都要放在设备的精度上，而非追求标准轮胎的测试一致性。

追求标准轮胎测试值的一致性，势必会导致对设备进行人为修正，使其与一种不断变化的目标保持一致，因而偏离了真正精确测试的终极目标。

尽管如此，仍有许多统计工具可用

来描述确定设备相关性的正确过程。

下一节将讨论几种重要工具和方法，这些都是作者所选取的精确评估轮胎试验相关性的相关工具和方法。

工具A：线性回归

用于分析设备一致性（相关性）的最常见的工具（但未必是最佳的工具）是线性回归。

该工具的理论基础是，在比较两个设备的测量值，或将设备的测量值与标准轮胎的值进行比较时，如果设备测量结果一致，就应当得到 $a y = x$ （最小二乘法拟合图线方程）。

即，得到的线性回归方程（ $y = m \cdot x + b$ 的形式）中，应当是 $m = 1$ ， $b = 0$ 。一般通过计算线性回归模型中的决定系数（样本相关系数，用 R^2 表示）来评价该方程的相关水平。

R^2 值越接近1，则两组数据之间存在的相关性就越高。

使用这种基于线性回归的标准存在几个问题。

线性回归模型的相关系数对几个事项很敏感，其中包括（但不限于此）所比较数据的分布。另外，也基本没有 R^2 多大时表示具有一致性的一般原则。

在进行线性回归线性鉴别时，对于b项是否应限定为0（或不是0）存在不同的流派。

如果该项不限制为0，那么其与零有多接近才可以接受呢？在不同的情况下答案是不同的，同时斜率变量 m 接近于1的程度也很难评估。

此处的要点是单个轮胎测试值的分布对分析结果有直接影响，或者，更重要的是，如果仅使用这种方法的话，可能会得出错误的研究结果。

这类分析的结果非常有趣，也很有意义。但是，作者推荐不要只采用这些结果来评估测试是否具有一致性。

工具B：GMUTS

大家所熟知，但在公开出版物中并不多见的汽车和轮胎工业用于评估检测设备一致性的工具是通用汽车公司统一测试规范(GMUTS)。

该工具是通用汽车公司为了保证其组织内部试验的统一而开发的，其中包括对设备重复性和相关性的分析。该方法是为设备评定三个介于1到10之间的数值分级。在这三类分级中，设备必须达到足够高的等级才能被批准用于生产测试。

这三个方面包括：

1. 线性回归方程分级——基于拟合回归线预测每个测量参数标准值的能力，最重要的是接近筛选极限值处的值，在比较测量值或测量值与标准值时，还有在0和2倍筛选极限值附近的值。
2. 单独的回归分级——基于决定系数（即 R^2 ）。
3. 重复性等级——基于测量值的重复性。

与标准线性回归相比，该方法的突出优点是其考察几个相关维度，而且实施过程中有严格的原则，按照这些原则可以选取标准轮胎。

这些原则（太多，太详细，本文无法全面叙述）非常具体，可确保决定系数更具有意义，确保轮胎具有足够的稳定性来评估设备的重复性。

该方法中包含了重复性评估，这是极为重要的。如果测量系统的重复性

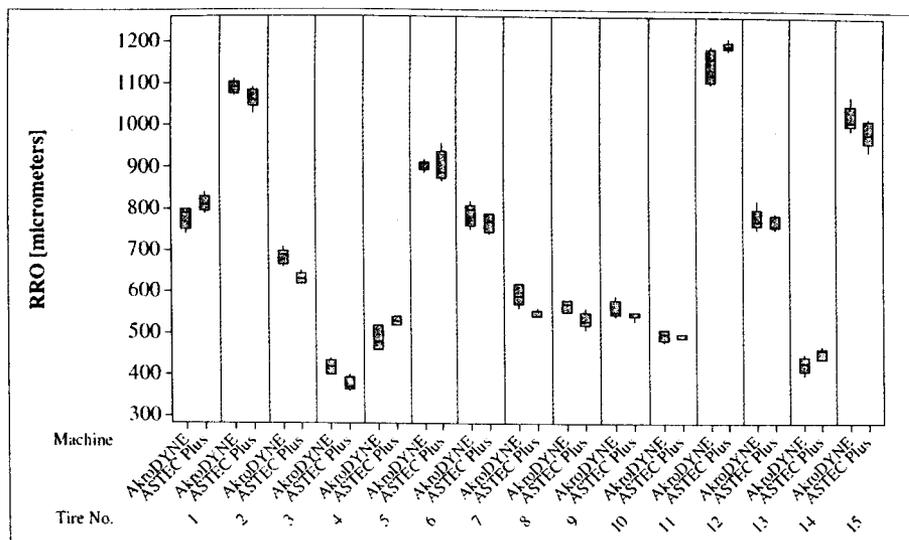


图2：在ASTEC Plus和AkroDYNE系统上测量的15个轮胎组的轮胎径向跳动数据箱线图

和精度不高，就不必谈测试设备的测量一致性。

该方法的另一个重要优点是其强调预测的筛选极限值附近的测量一致性。并不十分注重0和筛选极限值2倍附近测量值的一致性，在这些地方的测量值对于基于产品质量来制定经济的决策并不重要。

执行GMUTS的过程也非常简便。几个标准轮胎或标准轮胎和轮辋总成在待评估设备上多次测试，再与标准值进行比较，如果评估两个设备，则在两设备上多次试验。

通过测量数据拟合修正的线性回归方程。上述(1)项是根据回归方程预测筛选极限值的能力从预定数值表中选出的1到10之间的数值。

这就是说，在计算出回归方程后，用该方程预测在0、所评估特定测量值筛选极限值处，以及在约2倍筛选极限值处的值。

如果回归方程非常理想，就会精确地预测出0、筛选极限值，以及2倍筛选极限值处的值。根据这些值与实际值的接近程度进行分级。

对用于进行精确分级的另一个方程进行了加权处理，与极端情况下相比，使回归方程预测筛选极限值附近的值的能力占有更大的权重。这是一种正确的理念，也是这种测量一致性评估工具的最大优点。

上述评估项(2)和(3)就比较简单

了。根据所计算的 R^2 值（与1.0的接近程度）和标准偏差计算值（与0.0接近程度），从预定的表中选取介于1到10之间的数值等级。

工具C：GR&R

量具再现性和重复性(GR&R)工具一般与测量一致性（相关性）研究没有任何关系；但是，该工具在轮胎自动测量设备评估方面可能有一种新用途。

GR&R中的再现性维度一般用于评估两名检查人员使用特定测量装置所获取的测量值之间的接近程度。

这一方法可用于评估两个自动机械测量值的接近程度（与两名操作人员相比）。

也可用相同的基本GR&R分析结果来解释。该工具的另一个优点来自对测量系统可重复性的评估，及设备测量值之间的一致性。

工具D：相关系数

还有一些其他定义的相关系数可用于评估测量一致性，它们的使用是有限的。

所计算的这些量值包括斯皮尔曼等级相关系数或肯德尔等级相关系数，以及皮尔逊积矩相关系数。斯皮尔曼等级相关系数反映两个变量之间的统计相关性，评估使用一个单调函数，描述两个变量之间关系的好坏程度(Anderson, 2005)。

肯德尔等级相关系数也是反映两个变量之间的相关性。主要是通过比较一致对和不一致对的个数来进行评估(Correlation (Pearson, Kendall, Spearman), 2009)。

两个变量之间的皮尔逊相关系数定义为两个变量的协方差除以其标准偏差的乘积，介于-1和1之间，具体数值取决于相关性水平，类似于其他已知相关系数(Anderson, 2005)。

在研究测量一致性时这些工具的用途是有限的，因为在考察自变量和因变量之间的一般相关性时，它们更适用于系统行为分析。

“最佳”方法

本文所讨论的几种方法可用于评估测量一致性。但是，仍然存在这样的问题：哪种方法最有效呢？

当然，对于该问题并没有一个可以广为接受的答案，但是，有充分的理由推荐创建GMUTS工具的一些主要原则。使用GMUTS的轮胎测试测量一致性有效工具的要点是：

1. 仔细选取稳定的标准轮胎，减小轮胎和轮胎/设备对测量结果的影响（单个标准轮胎值和感兴趣的筛选极限值附近的平均值足够大的范围）。

2. 以所感兴趣的筛选极限值附近测量值的双权重评估最佳拟合线预测标准轮胎值（或第二个设备值）的能力。

3. 评估设备的可重复性，确保其对分析结果不会产生影响。

4. 评估最佳拟合线的决定系数(R^2)（结果不单独使用，选取标准轮胎有助于确保 R^2 值具有有意义的解释）。

在下一节将使用本文讨论的几种工具进行测量一致性实例分析。

实例分析

深入理解本文所介绍的多种测量一致性工具的最佳方法是进行一个实例分析。

考察一个这样的实例，用Micro-Poise公司制造的TGIS-SL（片状光轮胎几何尺寸检查系统）测量了15个轮胎的径向跳动(RRO)。

这些轮胎由安装在Micro-Poise公司

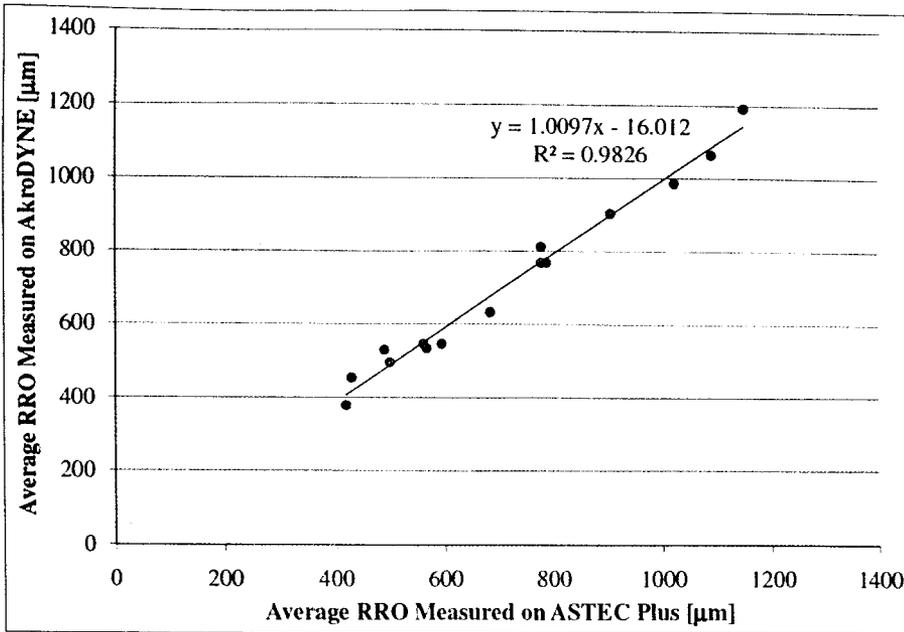


图3：平均轮胎径向跳动数据的线性回归曲线：AkroDYNE上的TGIS与ASTEC Plus上的TGIS的对比

制造的ASTEC PLUS轮胎评估中心上的TGIS-SL系统和另一个安装在AkroDYNE动态平衡试验机上的TGIS-SL系统进行测量。

每条轮胎在每个测量系统上共测量5次。应用前面讨论的三个测量一致性工具对数据进行分析 and 讨论。

评估两组数据之间测量一致性的一种最简单的直观（非解析）方法是绘制箱线图。15个轮胎径向跳动(RRO)测量数据组的箱线图示于图2。

箱线图中垂直条表示每个轮胎5个测量值分布。位于填充矩形中央附近的短水平线表示5个测量值的平均值。每一个轮胎在两个测量平台上获取的5次测量值的平均值似乎“值很相近”。

因此，无需进行任何数值评估，从该图显然就可看出每个轮胎的两组RRO数据之间具有满意的一致性。当然下一步就是就应用本文所述的一些工具对测量一致性进行数值分析。

应用工具A：线性回归

RRO数据的线性回归图线示于图3，图中包含有一条最佳拟合线，其斜率为1.0097，y轴截距为-16.012 μm ，决定系数为 $R^2 = 0.9826$ 。

一般说来，斜率值接近1.0000，y轴截距接近0.0000，决定系数 R^2 接近1.0000就表明测量一致性好。

这些参数的计算值（与实例数据组相关）都接近于理想值，但在确定测量一致性时产生了一些困难。有关斜率、y轴截距和 R^2 的任何指标值在某种程度上都是人为设定的，没有什么物理意义。

该过程中没考虑系统潜在的筛选极限值。相反，而是对于所有测量值都具有相同的权重。需要进一步分析来完成这种评估。

应用工具B：GMUTS

GMUTS不能直接用于RRO数据，因为GMUTS最初是针对均一数据定义的。

如果存在GMUTS RRO指标的话，作者对此也不了解。但是，基于GMUTS过程的一些内在基本原理可以得出一些普遍性结论。一般说来，在GMUTS体系中标准轮胎值的范围介于0到几乎典型筛选极限值的2倍之间。

RRO的典型筛选极限值可能约等于1,000 μm 。所以，按照这些标准，也许可以断定，标准轮胎值的范围不足，其平均值与所预测的筛选极限值不是非常接近。尽管轮胎组存在这些不足之处，但 R^2 值较高，在GMUTS标准中具有较高的分级。

在筛选极限值下评估回归方程得到 $y = 1.0097 \times 1,000\mu\text{m} - 16.012\mu\text{m} = 993.7\mu\text{m}$ 。该结果仅与筛选极限值相

差6.3 μm ，可以认为与理想结果非常接近，表明在筛选极限值附近具有较好的测量一致性。

在GMUTS体系中，还计算了测量数据的标准偏差，并与允许值预定义表进行了比照，确定出从1到10的分级。由于不存在作者了解的数值表，所以该数据的可重复性结果未包括在内。

为了更好的评估可重复性，下一节将讨论在RRO数据组中应用GR&R工具的情况。

应用工具C：GR&R

表1中的结果是将实例数据在Minitab（一种统计工具，用户利用该工具可以分析大量的数据类型，Minitab公司信息，2010）中进行基于方差分析(ANOVA)的量具再现性和可重复性(GR&R)分析获取的，对这些结果进行考察。

在方差分析中确定单个因子的统计显著性时一般要求概率值(P-值)小于0.05(5%)。也就是说，如果P-值小于0.05，就可认为是一个影响数据值变化的统计显著性参数。

表1概率(P)一栏中有两个P-值为0.000。这两个值表明，单个轮胎自身的测量值，以及单个轮胎与测试设备之间的交互作用具有统计显著性。这是预料之中的，因为我们本就希望测试设备具有足够高的可重复性来区分每个轮胎的测量值。

设备因子的P-值为0.295，表明该因子不是显著因子。

换言之，在AkroDYNE动平衡试验机和ASTEC PLUS均匀性试验机上的测试值之间没有统计显著性差别。

进一步考察表2中所示的GR&R结果，可以看到，两个试验机对总变差的贡献仅为0.01%。这一数值是非常小的，说明测量设备之间的差别不是改变测量值的显著因素。

其他可以计算的结果是%研究变差和%公差变差。计算结果示于表3。

汽车工业行动集团(AIAG)公布了%研究变差（评估测量装置的有效性，进行过程改进）和%公差变差（评估测量装置作为筛选工具的有效性）可接受极限值的一般规范。

表1: 量具R&R研究——RRO方差分析方法

来源	DF	SS	MS	F	P
轮胎号	14	8145437	581817	223.057	0.000
设备	1	3038	3083	1.182	0.295
轮胎号 * 设备	14	36517	2608	5.605	0.000
可重复性	120	55840	465		
总数	149	8240877			

这些规范中规定, 值小于10%为最佳, 10%和30%之间为可以接受, 大于30%为不可接受。

重点看一下表3中的设备项, %研究变差是1.04%, %公差变差是1.51%。这些值与AIAG公布的极佳值相比较小。

利用该方法的一个优点是同时还计算了设备的可重复性。

在该例中, 总研究变差的重复性为8.89%, 公差变差的为12.94%。按照AIAG规范, 这两种数值都是可以接受的。

该实例的整个过程, 包括可重复性和测量一致性元素, 都得到了可以接受的研究变差和公差变差。

底线

如前例所述, 有几种工具可用于评估两个或多个终检测试设备之间的测量一致性。

对每一个工具都必须检验和监控, 确保正确使用这些工具, 精确地评估设备的一致性。有助于保证该过程正

表2: GR&R %贡献结果

合计量具 R&R	900.3	1.53
-可重复性	465.3	0.79
-重现性	434.9	0.74
-设备	6.3	0.01
-设备*轮胎	428.6	0.73
样件到样件	57920.9	98.47
全变差	58821.1	100.00

确性的关键因素可分为三大类。

首先, 在不考虑设备精度的情况下进行设备一致性评估在确定两个或多个设备是否一致时是一项重大缺陷。轮胎制造商选取他们的测量一致性标准时, 非常重要的一点是, 不要按以前选取的标准轮胎值调节新设备, 而是要把重点放在提高各个设备的精度上。

终检测试设备要精确才能保证未来的测量一致性。标准轮胎值只能在不完全清楚设备精度的情况下对设备进

行检修时作为制造商的参考值。

第二, 在选用评估设备一致性的最佳工具时需要考虑几个因素:

- 首先选取一组标准轮胎, 使其与制造场合的要求完全匹配。对这些轮胎进行重复测试, 非常重要的一点是要更新标准轮胎值, 或者经过一段时间后选取新的标准轮胎来替换磨损的轮胎, 长期保持精度和一致性。
- 设备一致性评估工具应当特别注重预测的筛选极限值附近的值, 因为在极高或极低测量值处的设备相关性并不重要。
- 不要在制造商已知不具可重复性的设备上浪费时间和精力。一些设备本质上就不具备较高的可重复性, 如果是这样的话, 对这些设备进行测量一致性分析是极为困难的。

第三, 制造商若想正确地评估设备的一致性, 与了解这些过程和方法的设备供应商合作是非常关键的。似乎这是很显然的, 但制造商遵守这一点非常重要。

这可能意味着, 测试一致性(相关性)好, 可正确测试的设备之间存在差异, 意味着测试一致性差的机械在浪费时间和资源。

显然, 这些决定不只是影响制造设备本身。做出这些决定时, 需要进行精密的规划以及与最了解这些工具的供应商合作。

每家轮胎制造商的目标都是保证投放市场的轮胎与他们的描述一致。终检工作的功能就是对此进行验证。

确保轮胎试验机获取正确的测量值并具有相关性是达到用户要求的关键步骤, 用户们都希望轮胎拥有良好的行驶性能。●

参考文献

- 1) Anderson, S. A. (2005). Statistics for Business and Economics. Mason: Thomson South Western.
- 2) Correlation (Pearson, Kendall, Spearman). (2009). July 27, 2010, from Statistics Solutions: <http://www.statisticssolutions.com/methods-chapter/statistical-tests/correlation-pearson-kendall-spearman/>
- 3) Minitab Company Information. (2010). July 27, 2010, from Minitab - Software for Quality Improvement: <http://www.minitab.com/en-US/company/default.aspx>

表3: 1,000μm过程公差的GR&R %研究偏差和%公差偏差

来源	StdDev (SD)	(6 * SD)	(%SV)	(SV/Toler)
合计量具R&R	30.004	180.03	12.37	18.00
-可重复性	21.572	129.43	8.89	12.94
-重现性	20.885	125.13	8.60	12.51
-设备	2.515	15.09	1.04	1.51
-设备*轮胎号	20.703	124.22	8.54	12.42
样件到样件	240.668	1444.01	99.23	144.40
全变差	242.531	1455.18	100.00	145.52

可区分类别数 = 11