# 量刑基准实证研究

白建军*

**内容提要：**作为法制实践离散程度的客观反映，量刑基准是某种犯罪各组权威示范性案例样本之间相互独立的平均刑量。尽管600多个示范性案例的实证研究未能成功找到符合这一概念的量刑基准，但发现宣告刑平均刑量与法定刑中线之间的关系耐人寻味：不论样本的分组较粗还是较细，各组刑量均值之间的轻重顺序基本符合法定刑的轻重顺序。因此，以示范性案例的平均刑量为量刑基准，与以法定刑各种程度等分线或中线为量刑基准两者之间互不排斥而应相互参照。

**关键词：**量刑基准　平均刑量　示范性案例

据权威司法统计，刑事案件的无罪率一般不超过1%，且呈逐年下降趋势。[1]因此，绝大多数案件的当事人以及刑事法官更关心的是，到底根据什么判处多重的刑罚。这就是研究量刑基准的意义所在。

## 一、量刑基准的概念

考察量刑基准的概念有三个角度：规范的、社会的和方法学的。在刑法学上，量刑基准可以被理解为从重或从轻的相对物，或者说，没有任何从重从轻情节的犯罪构成所对应的刑罚量就是引入从重从轻情节时的量刑基准。[2]问题是，既然从规范学角度观察，我们便有理由追问，量刑基准到底是

---

* 北京大学法学院教授。

本文为笔者承担的国家社会科学基金资助项目“刑事司法公正性实证研究”的阶段性成果之一，研究中的资料收集、整理得到北京大学实证法务研究所的陈浩等多位研究人员的极大帮助，在此一并致谢。

〔1〕 全国法院审理刑事案件的无罪率分别为：1997年0.66%，1998年1.03%，1999年0.97%，2000年1.02%，2001年0.88%，2002年0.7%，2003年0.65%，2004年0.44%，2005年0.26%。该数据根据《中国法律年鉴》（1998年版第129页、1999年版第113页、2000年版第121－122页、2001年版第155页、2002年版第144页、2003年版第141页、2004年版第118－119页、2005年版第147－148页、2006年版第110－111页）公布的数字推算而来。

〔2〕 有学者将量刑基准界定为，对已确定适用一定幅度法定刑的抽象个罪，在不考虑任何量刑情节的情况下仅依其构成事实所应当判处的刑罚量。（周光权：《量刑基准研究》，《中国法学》1999年第5期）另一种观点认为，量刑基准是指排出各种法定和酌定情节，对某种仅抽象为一般既遂状态的犯罪构成的基本事实所判处的刑罚。（［日］西原春夫主编：《日本刑事法的形成与特色》，李海东等译，中国法律出版社与日本成文堂1997年联合出版，第150页）还有学者认为，所谓从重（从轻），是指对具有从重（从轻）情节的犯罪人所判处的刑罚比对不具有该从重（从轻）情节时所应判处的刑罚要相对重（轻）些，而不是一律判处法定最高（低）刑或者一律使用较重（轻）刑种、较长（短）的刑期或者一律在法定刑的平均刑期以上（下）考虑判处刑罚。（杨春洗、杨敦先、郭自力主编：《中国刑法论》，北京大学出版社2005年版，第159页）这种说法其实是通过对从重从轻相对物的阐释，表达了何谓量刑基准的理解。

不是法？如果是法，其法律规定何在？如果不是法，我们为什么要遵循它？而且，由于量刑基准到底是个点估计还是个区间估计并无定论，又使其规范学意义变得更加复杂。如果量刑基准是个点，显然没有法律根据。如果量刑基准是个区间值，是否可以将某个罪的法定基本刑度视为该罪的量刑基准？如果这样，量刑基准倒是获得了规范的形式，但这个意义上的量刑基准其实就是从重从轻的活动范围，而从重从轻到底是在基准之内还是之外又变得含混不清。其实，量刑基准问题上的这种纷繁多义恰好可以回溯到法律解释学的两大传统。一方面，按照规则怀疑主义或法官决断论，法律规范的适用过程不可能只导出某种唯一正确的结果，法官必然受制于多种因素的影响，案件的审理和判决实际上具有极大的自由度。〔3〕从这个意义上说，法官最多只能被要求在合法的形式范围内量刑，而不能要求量刑结果都紧紧地靠近某个法律并未明文规定的点或“紧缩”的区间附近，甚至与其相重合。〔4〕因此，人们质疑有些法官匪夷所思的判决的同时，也显出几分无奈。〔5〕这暗示着，在规范学意义上是否获得了正当性，是量刑基准研究的前提。另一方面，按照形式主义或法律决定论，法律意味着某种真理性的完美逻辑体系，按照这个逻辑体系，任何案件中提出的法律问题都必将得到一个唯一正确的答案，因此，对任何行为法律后果的预见都必将是唯一的、确定的，法律必须立足于限制案件审理过程和结果的自由度。〔6〕如此说来，人们对法官量刑统一、精准的渴望，还是有规范学根据的。那么，量刑基准到底是否具有规范的属性呢？笔者认为，量刑基准是一种“法中之法”，是法律的形式理性中隐含着的某种实质理性。尽管没有法律条文明确规定量刑基准，但是，我们还是能够从刑法中关于从轻从重的相关规定中感知它的存在，它是罪刑法定原则的定量展开。退一步说，犯罪构成理论也没有法条的明文规定，但没有哪个法官不去遵循。而理论研究的任务，恰恰是找到规范体系中量刑基准的真正所在。从这个意义上说，所谓量刑基准应当是指量刑情节的规则体系所环绕的轴心，相对这个轴心才有所谓依法从重或依法从轻。

理解量刑基准的第二个角度是社会标准：从为什么要设定量刑基准来论证什么是量刑基准。普遍认同的一种观念是，量刑基准是防止法官腐败、恣意量刑的一种限制。其假定是，只要各种案件事实与一定刑罚量之间的对应是确定的，法官就失去了权力寻租的空间，而这种确定性

〔3〕 规则怀疑主义认为，法院在做出司法裁决的过程中，其实并不真正受到所谓法律规则的制约。法官有高度的自由裁量权，随心所欲地进行判决。法律规则只是达到法官所喜欢的判决的籍口、可供其利用和摆布的手段，并不对法官达致判决结果的思考过程发挥规范作用，因为法律规则具有高度的不确定性，法官可以随意解释有关规则、制造例外情况或在适用规则时做出变通，从而得到他希望做出的结论。参见陈弘毅：《当代西方法律解释学初探》，载梁治平编：《法律解释问题》，法律出版社1998年版，第11页以下。与规则怀疑主义类似，法律现实主义也认为，法官的法律解释是一项完全自由的社会行动。他们不受法律规范的约束，因为法律是不确定的，作为方法论的法律解释理论所确信的那种确定的、稳定的、通过正确的方法可以发现其正确意义的法律，在法律现实主义看来是“基本的法律神话”。参见强世功、赵晓力：《双重结构化下的法律解释——对8名中国法官的调查》，载梁治平编：《法律解释问题》，法律出版社1998年版，第222页以下。

〔4〕 有学者认为，量刑基准由于受事实和法律的不确定性影响，不可能是一个精确的数值或“点”，在很多情况下，他仍然可能是一个幅度（法定刑为绝对确定刑的除外），量刑基准的这种幅度与法定刑幅度颇为相似，只是量刑基准幅度的“域”的范围较为紧缩而已。参见前引〔2〕，周光权文。

〔5〕 参见王恩海：《论量刑基准的确定》，《法学》2006年第11期；朱建华：《量刑的理论与实践研究——量刑基准的确立》，《河北法学》2006年第12期。

〔6〕 法律决定论认为法律给予法官的正确的决定是独一无二的。按照决定论的思维模式，法是全知全能的；法官不能以无法可依为由来拒绝做出判决，而必须通过解释发现包含在法律体系之中的具体的规范。法被理解为一个自我封闭、自我准据、等级森严的体系，一切事实关系都必须而且能够包摄其中。法律决定论的核心在于通过“概念计算”来预测审判结果的理论前提以及相应的制度性设计。这种“可预测性”概念正是M·韦伯关于经济、法律以及社会的宏观理论的基石，并成为描述现代法特征的最基本的指标。参见季卫东：《法律解释的真谛——探索实用法学的第三道路》，载季卫东：《法治秩序的建构》，中国政法大学出版社1999年版，第90页以下。法律是一个全涉的（gapless）的规则体系，它覆盖了社会生活的各个方面，即使在法律条款没有规定的地方，只要运用正确的法律解释方法，就可以发现法律在这方面的态度。参见前引〔3〕，梁治平编书，第222页以下。

的参照点就是量刑基准。不过仔细想来，这个假定的成立其实还有赖于另一个更加前提性的假定——案件审理中，量刑情节本身是否成立不仅不受任何非法律因素的干扰，而且没有任何争议。因为只有认定或不认定某个案件事实，才谈得上应该或可以在此基础上从轻还是从重决定刑罚。不论法官是否腐败，都不会在判决书中使案件事实的认定与刑罚结果之间形成法律上的明显矛盾。换句话说，即使被认定的案件事实与刑罚量之间的对应十分确定，还是不能排除法官为了得出某个结论而能动地认定或不认定某个案件事实的可能性。而且，这种能动性并不全部来自司法腐败，而法官的经验、知识背景、案件本身的非典型程度等多种因素都可能影响这个过程。遗憾的是，有些量刑基准的讨论并未注意到这个前提的前提。实际上，某个证据确凿的案件事实是否成立自首、未遂、从犯等量刑情节，才是法庭真正面对的难题，这当中其实存在着影响最终刑罚量大小的巨大空间。从理论上看，一个案件事实是否存在，不仅在自然意义上与其客观性有关，而且在法律和认识论意义上与该事实的认定者自身的主体性有关。案件事实不仅意味着它客观上是否存在，还可能意味着它的定义者是否需要它存在。由此可见，如果将司法公正、防止腐败的功能赋予量刑基准的话，那么，量刑基准就不应只是个量的概念，而应该是定性与定量相结合的概念。具体到法律世界中，可能有纯粹的定性分析，但从来都不会有纯粹的定量分析，定量从来都离不开准确的定性。从这个意义上说，所谓量刑基准首先应该是指量刑情节（包括法定情节和酌定情节）本身是否成立的事实标准，在其范围以内才谈得上依法从重从轻。

第三，更多的量刑基准概念是从怎样确定量刑基准来说明何谓量刑基准。在这方面，我们面临两个二元对立：一是演绎与归纳。演绎法试图自上而下地从立法中寻找量刑基准，归纳法试图采用实证方法自下而上从实际判决中寻找量刑基准。前者突显着量刑基准的规范意义，而后者更看重与实际问题联系在一起的规则以及尊重法官群体的法律实践，包括笔者在内的越来越多的研究者倾向后者。二是中值与均值。中值是一组事件中处于50％位置上的值，不少研究实际上是将法定刑或宣告刑的中值设定为量刑基准。笔者认为，至少从法律适用的结果来看，量刑基准应该是个均值而非中值。理由是：中值只是一组数据的中心点，有一半的数据大于它，一半的数据小于它，而均值是一组数据的平衡点，所有数据到均值的距离的总和为零。所以，均值对数据的变化尤其是奇异值的大小更为敏感，中值则没有这个特性。比如，9个罪犯被判有期徒刑的月数分别为6、12、24、36、48、60、72、84、240。这组数据的中值为48个月，而因为有一个20年有期徒刑（即240个月）其均值则为64.7个月。可见，均值能够灵敏地反映出数据距离平衡点的远近大小，而个别极端数值的有无却不会改变中值的大小。那么，量刑基准应该更关心它两边的数据个数是否相等呢，还是应该更看重个案尤其是极端个案的值距离平衡点的远近大小呢？很显然，量刑基准两边的个案数是否相等只说明较多的个案低于或高于量刑基准。而量刑基准两边个案的刑罚量轻重大小，才说明法官在自由裁量权的行使中到底走了多远。如果量刑基准设置的意义在于限制自由裁量权滥用的话，那么，真正应该关注和调控的，正是能够灵敏反映极端个案刑量值的集中量数，或者说是与离散趋势同在的集中趋势，而这个量数就是均值而非中值。同时从这两个角度观察，我们便有4种可能的量刑基准类型：（1）演绎型中值，如将法定刑中线确定为量刑基准的中线说。[7]（2）演绎型均值，如“法定刑的平均刑期”说。[8]此外，在各种量刑基准的确定方法中，危害行为论也强调从抽象个罪中演绎出量

〔7〕 参见赵廷光：《量刑公正实证研究》，武汉大学出版社2005年版，第27页以下。苏惠渔、张国全、史建三：《量刑与电脑——量刑公正合理应用论》，百家出版社1989年版，第102页以下。

〔8〕 参见杨春洗、杨敦先、郭自力主编：《中国刑法论（第三版）》，北京大学出版社2005年版，第159页。

刑基准，基本上属于演绎型均值说。它在综合了主要因素说〔9〕和重心说〔10〕的基础上认为，实践中存在着大致的、可推演的将危害行为换算成一定刑罚量的方式，所以对抽象个罪中的危害行为在不考虑任何情节时的危害性的考量，以及对抽象个罪在“真空”状态下所对应的刑罚量进行观察与分析，就是确定自然刑罚量即量刑基准的过程。〔11〕(3) 归纳型中值。(4) 归纳型均值。这两种归纳型量刑基准的研究及其实践已经越来越令人瞩目，如美国量刑委员会的《量刑指南》〔12〕、日本学者热衷的量刑计量化实证研究〔13〕、以北京市朝阳区1999年度的刑事案件为样本的实证分析〔14〕、我国某些地方法院所作的积极尝试〔15〕，等等。但是，这些研究基本上没有专门讨论中值与均值的差异。其实，如果归纳重于演绎且均值优于中值的话，作为确定量刑基准的最佳方式，归纳型均值应该是指以权威案例为样本计算出来的平均刑罚量，它可能高于也可能低于法定刑中线，也可能不等于宣告刑中线。最终，归纳型均值力图反映的是犯罪本身的多样性和复杂性。

至此，笔者对量刑基准的理解可以归纳为：第一，量刑基准也是法，是应该得到遵守和适用的“法中之法”。第二，量刑基准也是定性基准，是量刑情节能否成立的事实标准。第三，量刑基准是法制实践离散程度的反映，是权威案例样本的平均刑罚量。为了满足前两个规定，量刑基准的经验来源应该是某类犯罪的全部权威判决。换句话说，量刑基准不仅是个点、圈〔16〕、线、格、域〔17〕、度〔18〕，更是个库，即一组具有指导作用的司法先例。这样，不仅常见案件，而且非典型的疑难案件也有了参照依据。为了满足第三个规定，至少要考虑到以下几种复杂性：首先，基本犯罪构成的从重从轻，与派生构成（加重、减轻）中的从重从轻应该参照不同的量刑基准。比如，入户抢劫未遂在从轻时所参照的基准显然与普通抢劫未遂的从轻不是同一个基准水平。而且，尽管既无法定情节又无酌定情节的“真空”状态的确存在，将这种案件的某个均值确定为量刑基准也不是不可能，但问题是，即使抽空所有量刑情节，一个案件也还是各种主客观方面的不同组合。因此，寄希望于案件中的某一个所谓主要事实（如犯罪数额）所对应的刑量作为量刑基准，也是不科学的。进一步说，基本、派生犯罪构成内部还可能有数个相同或不同方向的量刑情节，此即情节冲突与竞合问题。基于这些考虑，一个能够同时满足各种理论假定又能消化这三类复杂性的量刑基准的操作定义应该具体化为多个均值：即某种犯罪各组权威案例样本之间相互独立的平均刑量，样本的分组首先区分基本犯罪构成与派生犯罪构成（加重、减轻），然后再进一步区分各基本、派生犯罪构成组内的从重、从宽（从轻、减轻、免除）、竞合、基本四类情节。按此界定，以盗掘古文化遗址、古墓葬罪为例，一种犯罪理论上最多可能有三组12个量刑基准：加重犯罪构成组中从重情节组的平均刑量、从宽情节组的平均刑量、情节竞合组的平均刑量、无任何情节的基本情节组的平均刑量；基本犯罪构成组中从重情节组的平均刑量、从宽情节组的平均刑量、情节竞合组的平均刑量、无任

〔9〕 认为法定刑的运用基准点的确定应当以对社会危害性大小起主要作用的因素为依据。参见前引〔7〕，苏惠渔等书，第102页以下。

〔10〕 认为在一个以法定刑为纵轴、以发案率为横轴的坐标系中，由随机获取的案例样本组成的散点构成的曲线顶端，即最高发案率，就是该罪的重心，也即该罪的量刑基准。参见郑伟：《重罪轻罪研究》，中国政法大学出版社1998年版，第52页。

〔11〕 前引〔2〕，周光权文。

〔12〕 郑伟：《法定刑的基准点与量刑的精雕细琢——〈美国量刑指南给我们的启示〉》，《人民司法》2003年第7期。

〔13〕 从已经确定的判例中选出作为调查对象的适当标本，再根据达到一定数量的这种事例列出与量刑有密切相关关系的因素并给予评分，从而求解量刑基准。参见前引〔2〕，周光权文。

〔14〕 结果表明，该区法院量刑的平均值一般低于法定刑幅度的中线，而且，将其与1991年以前的上海市及周边地区的调查结果相比，统计结果相当接近。参见阮齐林：《中国刑法上的量刑制度与实务》，法律出版社2003年版，第27页以下。

〔15〕 如江苏省高级人民法院于2004年5月通过的《量刑指导规则（试行）》、江苏省姜堰市人民法院于2003年3月通过、2004年9月修正的《规范量刑指导意见》、山东省淄川区法院研发的量刑软件。

〔16〕 前引〔12〕，郑伟文。

〔17〕 前引〔2〕，周光权文。

〔18〕 前引〔5〕，王恩海文。

何情节的基本情节组的平均刑量；减轻犯罪构成组中从重情节组的平均刑量、从宽情节组的平均刑量、情节竞合组的平均刑量、无任何情节的基本情节组的平均刑量。这样，需要对照的案例只要确定了自己的组别，如是否构成入户抢劫罪未遂，即是否归入抢劫罪加重构成组中的从宽情节组，就可以迅速根据案情及该组的平均刑量决定本案的刑罚量。

## 二、几组示范性案例的平均刑量

完成了量刑基准概念的理论建构之后，便应该进而回答，某某犯罪的量刑基准到底是多少？为此，必须依次满足以下几个条件：

首先，需要测量其量刑基准的那个罪的法定罪刑关系必须在整部刑法罪刑关系体系中符合均衡性要求，不存在系统上偏轻或偏重问题。比如，我国刑法第 195 和 199 条规定，信用证诈骗罪的法定刑上限为死刑，这显然偏重而不符合罪刑均衡的要求。[19]如果权威案例中的法官都照此量刑，其平均刑量很可能超出惩戒预防信用证诈骗犯罪的必要限度。这时，如果将其认定为信用证诈骗罪的量刑基准，实际上是在强化立法缺陷。正是由于存在这种可能性，所以我国现阶段不宜全面照搬美国《量刑指南》的方式，[20]或者逐一采用法定刑中线为量刑基准，而应在现有罪名体系中有选择地设立量刑基准，或者根据罪刑均衡的要求在法定刑幅度内适当调低或调高某罪的量刑基准，实现“依法微调”。

其次，确定了可以设立量刑基准的罪名以后，就需要按照归纳逻辑的要求，建立该罪具有足够数量的权威案例样本库，用来计算各组样本的平均刑量。所以，因其样本的局限性，某个地方法院根据本单位案例所总结归纳出的量刑基准，就无法直接推论到全国各地各级法院。而示范性案例，恰好符合样本的权威性要求。所谓示范性案例，就是指来自最高法院各业务厅、研究机构、出版单位、网站等权威机构公开发布、发表的全部真实审判案例。[21]这个样本库与我国 1997 年以来全部刑事案件总体之间的相似性，就在于其中案件审理结果的示范性。这种“示范性”体现在：第一，由于这几千个案件来自全国各地，时间跨度涵盖了 1997 年至今各个时段的案件，由各地各级法院选送，具有对全国总体时间和空间的代表性；第二，由于是最高法院各权威机构认可并公开的案件，因而具有对司法实践的指导性和模范性；第三，由于其中绝大部分案件属于生效判决，因而具有一定的有效性；第四，由于各地选送案件以及最高法院各单位选取案件时充分考虑到案件类型和性质的多样化，因而对学术研究而言具有一定的标志性；第五，由于是公开发布的案件，因而对公民行为而言具有相当的规范性、模范性和可预测性。示范性案例库不仅满足权威性要求，而且应该具有足够的数量，否则，其平均值的计算就难以稳定在某个水平上。

第三，某个罪的案例样本库内，也要实现具体个罪量刑结果的罪刑均衡。否则，组内相对偏轻或相对偏重的案例过多，都可能拉低或抬高刑罚量的平均值，误导量刑基准的适用过程。[22]

第四，准备计算其平均刑量的犯罪，不仅作为抽象个罪和具体个罪都要符合均衡性要求，

---

〔19〕 参见白建军：《罪刑均衡实证研究》，法律出版社 2004 年版，第 284 页。

〔20〕 同上书，第 135 页以下。

〔21〕 这些示范性案例来自：最高人民法院、最高人民检察院：《中国案例指导》，法律出版社；最高人民法院办公厅：《中华人民共和国最高人民法院公报》，人民法院出版社；最高人民法院中国应用法学研究所：《人民法院案例选》，人民法院出版社；国家法官学院、中国人民大学法学院：《中国审判案例要览》，中国人民大学出版社、人民法院出版社；《人民法院裁判文书选》，法律出版社；最高人民法院网站（www.court.gov.cn）；最高人民法院刑事审判第一庭、第二庭：《刑事审判参考》，法律出版社；最高人民法院刑事审判第二庭：《经济犯罪审判指导》，人民法院出版社；最高人民法院审判监督庭：《审判监督指导—审判监督指导与研究》，人民法院出版社。

〔22〕 具体个罪的均衡性计算方法，参见前引〔19〕，白建军书，第 370 页以下。

还应该能够明确区分加重构成、基本构成、减轻构成。这是因为对某个罪的示范性案例首先要划分为加重、基本、减轻三大组，然后才能在各组内进一步分为从重、从宽、基本、竞合各组，最后才能分别计算各小组的平均刑量。其中第二次划分的根据比较明确，即法定情节的各种明文规定，而第一次划分并非易事。例如，抢劫罪、伤害罪的加重、减轻，可以根据是否重伤、死亡、入户等法定情节进行判断。而在纯正情节犯的情况下，刑法的概然性规定使三组犯罪构成之间的区分变得比较模糊。例如，“情节较轻的”故意杀人行为是该罪的减轻构成条件，而何谓情节较轻，却没有明确的指标或者关键词。这样，判断一个案件属于故意杀人罪的基本构成组还是减轻构成组，就只能根据其宣告刑在十年以上还是以下了。而这实际上是由果索因，逻辑上不够严格。

第五，选定可以计算量刑基准的个罪以后，还要使该罪的宣告刑可以被计算平均值。这对有期徒刑而言并不难，但对法定刑中包括死刑、无期徒刑的犯罪而言，就不那么简单。为了使宣告刑实现无量纲化，已有学者作了积极尝试。[23]本研究将死刑、死缓、无期徒刑、有期徒刑、有期徒刑缓刑、拘役、拘役缓刑、管制、没收财产、剥夺政治权利、罚金共11个变量选定为刑量指标，根据相关法律规定、实践经验和反复测试，以SPSS软件的程序格式建立了反映指标间权重关系的宣告刑刑量模型供读者参考：COMPUTE 刑量＝(死刑＊600＋死缓＊400＋无期徒刑＊300 ＋((1－ANY(有期缓月,0)) ＊ (有期月数＊.1 ＋ 有期缓月 ＊ .09 )＋(ANY(有期缓月,0)＊有期月数)) ＋ ((1－ANY(拘役缓月,0))＊(拘役月数＊.04＋拘役缓月＊.08) ＋ (ANY(拘役缓月,0)＊拘役月数))＊.9＋管制月数＊.07)＊.9＋(全部没收＊12＋部分没收＊6＋剥权终身＊10＋剥权月数＊.1＋罚金组＊2)＊.1. EXECUTE .[24]运行这个模型后，不论一个案件的宣告刑中含有何种性质的刑罚，都可以被换算为一个连续变量，以便平均值的计算。

最后，满足以上条件以后，同一犯罪的不同量刑基准之间还应当相互独立，满足统计学上差异显著性要求。如果某罪的示范性案例库中各组样本的12个平均刑量之间未形成显著差异，意味着某组的量刑基准和另一组的量刑基准之间没有显著差别，这等于说有没有从重或从轻甚至是否加重减轻都一样，显然不符合法律本意。

从现有的几千个示范性案例来看，满足或基本满足第一、二个条件的犯罪至少有故意杀人罪、故意伤害罪、抢劫罪、盗窃罪、贪污贿赂罪、走私贩卖运输制造毒品罪等等。至于第三个条件，由于涉及犯罪严重性的去量纲化处理，已经超出了本文容量的要求，因此我们假定上述罪名的示范性案例中罪量与宣告刑的刑量之间都满足均衡性要求。这样，我们可以利用上述刑量模型尝试着计算几种犯罪的量刑基准了。

1. 抢劫罪的量刑基准。按照刑法第263条的规定，抢劫罪只有基本构成和加重构成而不存在减轻构成。我们只需计算该罪基本构成中的基本、从重、从宽、竞合，以及加重构成中的基本、从

〔23〕 参见赵廷光：《量刑公正实证研究》，武汉大学出版社2005年版，第15页以下。

〔24〕 其中，有期徒刑的月数是整个模型的基数，因此，模型中的一个死刑立即执行相当于600个月的有期徒刑，一个死缓相当于400个月的有期徒刑，一个无期徒刑相当于300个月的有期徒刑。另外，主刑与附加刑的数量关系被确定为9∶1的关系，即主刑的权重系数为0.9，而附加刑的权重系数为0.1。还有，宣告刑量化处理的难点在于缓刑的存在。按照刑法第72条、第73条规定，同等有期徒刑或拘役刑期情况下，有缓刑的判决应该轻于没有缓刑的情况。但是，如果我们用有期徒刑或拘役的月数减去缓刑月数，势必形成有缓刑的判决之间，缓刑越长刑量越小的结果，显然与实际不符。因此，我们只能分别对有期徒刑（拘役）和缓刑进行加权后再相加，以其和反映刑量大小。这样，即使在判处有期徒刑6个月缓刑5年这种最极端的情况下，其刑量值也仅等于单处有期徒刑6个月的情况。而只要缓刑小于5年，其刑量值都将小于单处有期徒刑6个月的情况。结果，将有期徒刑和其他刑种进行同量纲化处理中，利用该模型就能在犯罪人被判处有期徒刑时有缓刑和无缓刑两种情况下都能够实现具体的一个结果值。同理，拘役和拘役缓刑的权重也是根据极端值获得的。

重、从宽、竞合这 8 组样本的平均刑量，并对它们之间的差异显著性进行统计检验。过程是，在抢劫罪示范性案例库中运行刑量模型〔25〕，SPSS 便立即给出该案例库中 222 个样本的刑量值，然后便可计算各组刑量的均值。第一个结果是，基本构成组的平均刑量为 57.19，大约相当于 60 到 66 个月之间的有期徒刑，而加重构成组的平均刑量为 223.72，大约相当于 20 年有期徒刑与无期徒刑之间的刑罚。而且，两者都略低于该罪法定刑中线。独立样本 T 检验的结果表明，两者差异显著，说明加重构成组刑量的集中趋势的确明显重于基本构成组，符合理论预期。〔26〕但是，如果以这两个均值为抢劫罪的量刑基准，显然过粗。第二个结果是，对 8 个组别的平均刑量进行方差分析的结果表明，除了未出现足够数量的基本构成中的从重案例以外，其余 7 组均值由轻到重依次为：(1) 基本构成从宽组的刑量均值为 38.44，大约相当于 42 个月有期徒刑，与加重构成各组的刑量均值之间均差异显著，但与基本构成其余各组之间都无显著差异。(2) 基本构成基本组的刑量均值为 75.19，大约相当于 84 个月有期徒刑，与加重组中基本、从重、竞合各组的刑量均值之间均差异显著，但与基本构成其余各组及加重从宽组之间无显著差异。(3) 基本构成竞合组的刑量均值为 85.16，大约相当于 96 个月有期徒刑，只与加重基本、加重从重组的刑量均值之间差异显著，但与其余各组均无显著差异。(4) 加重构成从宽组的刑量均值为 124.18，大约相当于 138 个月有期徒刑，与基本构成从宽组及加重各组的刑量均值之间差异显著，但与其余各组均无显著差异。(5) 加重构成竞合组的刑量均值为 271.30，大约相当于无期徒刑剥夺政治权利终身，与基本构成组中基本、从宽、加重构成从宽等组的刑量均值之间差异显著，但与其余各组均无显著差异。(6) 加重构成基本组的刑量均值为 299.93，大约相当于无期徒刑剥夺政治权利终身，并没收全部个人财产，与基本构成各组及加重构成从宽组之间差异显著，但与加重组中其余各组之间无显著差异。(7) 加重构成从重组的刑量均值为 377.23，大约相当于死缓，剥夺政治权利终身，没收全部个人财产，与基本构成各组及加重构成从宽组之间差异显著，但与加重组中其余各组之间无显著差异。可见，尽管 7 组的均值之间看上去大小不等，且顺序上符合理论预期，但由于基础数据不够丰富，有的组别仅有三、四个案例，所以，7 个均值之间的方差分析结果除加重构成组中竞合与从宽两组以外，没有一对相邻均值间差异显著。因此，这 7 个均值因不符合独立性要求而无法成为抢劫罪的量刑基准。

2. 盗窃罪的量刑基准。按照刑法第 264 条的规定，盗窃罪有基本构成、加重构成、特别加重构成和最加重构成，而不存在减轻构成。考虑到数据有限，将三个层次的加重构成合并处理，认为是一个加重构成。因此，我们只需计算盗窃罪基本构成中的基本、从重、从宽、竞合，以及加重构成中的基本、从重、从宽、竞合这 8 组样本的平均刑量，并对它们之间的差异显著性进行统计检验。然而，盗窃罪示范性案例库中只有 148 个案例，而且，按照有关司法解释，该库中没有从重的情形。于是，我们只能得到盗窃罪基本构成基本组、从宽组、加重构成基本组、从宽组共 4 组样本的刑量均值。在该罪示范性案例库中运行刑量模型得到每个样本的刑量值后，均值分析的第一个结果是，盗窃罪基本构成组的平均刑量为 36.66，大约相当于 42 个月的有期徒刑，高于该罪基本构成法定刑上限；而加重构成组的平均刑量为 101.14，大约相当于 114 个月有期徒刑，明显低于该罪加重构成法定刑中线。独立样本 T 检验的结果表明，两者差异显著，说明加重构成组刑量的集中趋势的确明显重于基本构成组，符合理论预期。当然，如果以这两个均值为盗窃罪的量刑基准也不够精准。第二个结果是，对 4 个组别的平均刑量进行方差分析的结果表明，各组均值由轻到重依次为：(1) 基本构成从宽组的刑量均值为 14.31，大约相当于 16 个月有期徒刑，与加重构成各组的刑量均值之间差异显著，但与基本构成基本组之间无显著差异。(2) 基本构成基本组的刑量均值

〔25〕 为了便于分析，此样本库不包括数罪的情况，如遇数罪情形，只需将抢劫罪的量刑基准应用于分别量刑过程即可。下同。

〔26〕 差异显著与否的标志是 p 值的大小，当 p 值≤0.05 时，才认为均值间差异显著，如果 p 值＞0.05，则认为均值间差异不符合统计学要求。下同。

为49.53，大约相当于54个月有期徒刑，仅与加重构成基本组的刑量均值之间差异显著，但与其余各组之间无显著差异。(3) 加重构成从宽组的刑量均值为71.37，大约相当于80个月有期徒刑，与基本构成从宽组及加重构成基本组的刑量均值之间差异显著，但与基本构成基本组无显著差异。(4) 加重构成基本组的刑量均值为126.33，大约相当于140个月有期徒刑，与其余各组之间均差异显著。可见，尽管4组的均值之间看上去大小不等，且顺序上符合理论预期，但4个均值之间的方差分析结果仅有加重构成基本组与从宽两组间差异显著，其余相邻均值间差异均不显著。再加上从重及竞合的情形未出现，因此，这4个均值无法成为盗窃罪的量刑基准。

3. 故意伤害罪的量刑基准。根据刑法第234条的规定，故意伤害罪有基本构成也有加重构成。由于在故意伤害罪示范性案例库中138个样本中没有出现逆向竞合的情形，所以，除了基本构成与加重构成两大组之外，我们只需计算该罪基本构成中的基本、从重、从宽，以及加重构成中的基本、从重、从宽这6小组样本的平均刑量，并对它们之间的差异显著性进行统计检验。这个过程的第一个结果是，故意伤害罪基本构成组的平均刑量为21.98，大约相当于24个月有期徒刑，而加重构成组的平均刑量为112.03，大约相当于126个月有期徒刑。前者略高于该罪基本构成法定刑中线，后者明显低于该罪加重构成法定刑中线。独立样本T检验的结果表明，两者差异显著，说明加重构成组刑量的集中趋势的确明显重于基本构成组，符合理论预期。但是，如果以这两个均值为抢劫罪的量刑基准，同样过粗。第二个结果是，对6个组别的平均刑量进行方差分析的结果表明，各组均值由轻到重依次为：(1) 基本构成从宽组的刑量均值为18.31，大约相当于20个月有期徒刑，与加重构成中基本组和从重组的刑量均值之间均差异显著，但与其余各组之间无显著差异。(2) 基本构成基本组的刑量均值为23.12，大约相当于24个月有期徒刑，与加重构成各组的刑量均值之间均差异显著，但与基本构成其余各组之间无显著差异。(3) 基本构成从重组的刑量均值为27.00，大约相当于30个月有期徒刑，只与加重构成从重组的刑量均值之间差异显著，而与其余各组均无显著差异。(4) 加重构成从宽组的刑量均值为88.81，大约相当于96个月有期徒刑，与基本构成基本组及加重构成从重组的刑量均值之间差异显著，但与其余各组均无显著差异。(5) 加重构成基本组的刑量均值为123.27，大约相当于136个月有期徒刑，与基本构成组中基本、从宽、加重构成从重等组的刑量均值之间差异显著，但与其余各组均无显著差异。(6) 加重构成从重组的刑量均值为259.65，大约接近于无期徒刑剥夺政治权利终身，与其他所有各组之间均差异显著。可见，尽管6组的均值之间看上去大小不等，且顺序上符合理论预期，但相邻数据间只有加重构成基本组与加重构成从重组之间差异显著，其余各对相邻均值间差异都不显著。因此，这6个均值因不符合独立性要求而无法成为故意伤害罪的量刑基准。

4. 贪污受贿罪的量刑基准。根据刑法第383、386条的规定，贪污罪和受贿罪既无减轻构成又无加重构成，但按其刑度及数额可将133个贪污受贿示范性案例分为3大组。[27]由于3组中都没发现从重的情形，因此，我们可以按从宽情节的有无，将其分为6小组。在得到每个样本的刑量值以后，3大组均值分析的结果是：(1) 最重组（犯罪数额在10万元以上）的刑量均值为182.62，大约相当于18年有期徒刑，明显低于10年到死刑的法定刑中线，与其余各组的刑量均值之间均差异显著。(2) 特重组（数额在5万元以上不满10万元）的刑量均值51.38，大约相当于60个月有期徒刑，恰好位于5年到无期徒刑的法定刑底线，与最重组的刑量均值之间差异显著，但与较重组之间无显著差异。(3) 较重组（数额在5千元以上不满5万元）的刑量均值为22.10，大约相当于24个月有期徒刑，大大低于1年到10年有期徒刑的法定刑中线，与最重组的刑量均值之间差异显著，但与特重组无显著差异。6小组均值由重到轻依次为：(1) 最重基本组的刑量均值为224.22，大约相当于20年有期徒刑与无期徒刑之间的刑罚，与其余各组的刑量均值之间均差异显著。(2)

〔27〕 由于数额不满5千元的案例只有两个，因而不进入分析。

最重从宽组的刑量均值为 136.18，大约相当于 154 个月有期徒刑，除特重从宽组以外，与其余各组的刑量均值之间均差异显著。（3）特重基本组的刑量均值为 53.44，大约相当于 60 个月有期徒刑，只与最重组的刑量均值之间差异显著，但与其余各组均无显著差异。（4）特重从宽组的刑量均值为 47.56，大约相当于 54 个月有期徒刑，仅与最重基本组的刑量均值之间差异显著，但与其余各组均无显著差异。（5）较重基本组的刑量均值为 16.44，大约相当于 20 个月有期徒刑，仅与最重组的刑量均值之间差异显著，但与其余各组均无显著差异。（6）较重从宽组的刑量均值为30.60，大约相当于 34 个月有期徒刑，反而高于较重基本组，仅与最重组的刑量均值之间差异显著，但与其余各组均无显著差异。可见，尽管 6 组的均值之间看上去大小不等，且顺序上基本（较重组中两组除外）符合理论预期，但 6 个均值之间的方差分析结果仅有前三组之间相邻的均值差异显著，其余相邻均值间差异均不显著。因此，这 6 个均值也无法成为贪污受贿罪的量刑基准。

## 三、平均刑量的双重性

几种案件的实证分析并未找到符合上述概念和条件的量刑基准。然而，从失败中我们却看到了新的意义：首先，以上分析可见，样本分组越细，越难以满足不同刑量均值之间的独立性要求。这说明研究失败的原因之一很可能与样本规模数量偏小、在内容上不典型案例的比重较大有关。其结果，很可能因样本离散性较大造成刑量均值间方差检验不合格。因此，作为提取量刑基准的事实基础，某罪案例库的一个基本分组单位，如基本构成中的从重情节组，至少应有 30 个样本。这样，对既无减轻构成又无加重构成的罪名而言，至少要有从重、从宽、竞合、基本 4 个单位共 120 个案例构成的示范案例库。对既有基本构成又有加重构成的罪名而言，就要有 8 个单位共 240 个案例构成的示范案例库。对贪污罪这种有 4 个刑度的罪名而言，原则上就要有 16 个单位共 480 个案例构成的示范案例库。而且，在内容构成上只要疑难、非典型案件的比例适当大于实际分布即可。建立这样一个大型案例库并从中提取量刑基准，不仅抽样过程要符合随机性要求以保证案例的代表性，而且需要最高司法机关科学导向、依法遴选并定期维护以体现宽严相济的刑事政策。这无疑是一种立足成文法体制吸收判例法合理性的积极尝试。

更重要的是，上述几组实证分析发现，宣告刑的平均值与法定刑之间的关系耐人寻味。首先，不论样本的分组较粗还是较细，各组刑量均值之间的轻重顺序基本符合法定刑的轻重顺序。这说明法官们的操作与立法预期两者之间存在高度的等级相关，不存在系统性紊乱或失衡，这显然与法官群体不约而同依法量刑的努力有关。因此，以示范性案例的平均刑量为量刑基准，与以法定刑各种程度等分线或中线为量刑基准两者之间互不排斥而应相互参照。进一步说，上述几组观察毕竟证明，法定刑的刑度越重，宣告刑平均刑量低于法定刑中线的趋势就越明显。这个现象的意义是，作为量刑基准，平均刑量可以是一种应然的裁判准则——只要当下的案情与相应示范性案例之间没有显著差异，就应参照该组案例的平均刑量决定刑罚；另一方面，作为司法实践的观察结果，平均刑量又是刑罚资源实际投入的一种实然的客观反映——宣告刑与法定刑的上述分离也许说明法官群体在自觉体现刑法的谦抑性，也许说明立法上刑度的分割点需要调整，[28]还可能是犯罪现象本身以及犯罪与犯罪控制之间多种可能关联性的反映。集应然与实然于一身的这种双重性，使平均刑量既是本研究的一个认识结果，又是进一步探索的新起点。

〔28〕 例如刑法第 383 条的数额分组显然已经不太适应贪污受贿犯罪的现实。