

1.3 试验的基本概念与流程

王正明 易泰河

系统工程学院 军事建模与仿真系

2019 年 11 月 8 日

引言

- ① 如何理解试验的意义?
- ② 科学研究的四范式?
- ③ 试验设计发展的四阶段?

教学目的

- ① 掌握试验的基本术语
- ② 树立试验是一个过程的概念;
- ③ 掌握试验的基本流程, 以及每一步骤的具体任务.

1.3 试验的基本概念与流程

1.3.1 试验的基本概念

1.3.2 试验的基本流程

1.3.3 作战实验：恐怖的海峡

- 称衡量试验效果的特征量为**响应**, 也称**指标**, 用字母 y 来表示.
 - 定性响应和定量响应.
 - 单响应试验和多响应试验.
 - 多响应试验通常也转化为单响应试验来处理.
- 称响应的测量值与真值之间的偏差为**误差**
 - 误差是一个随机变量, 一般假定 $\varepsilon \sim N(0, \sigma^2)$.

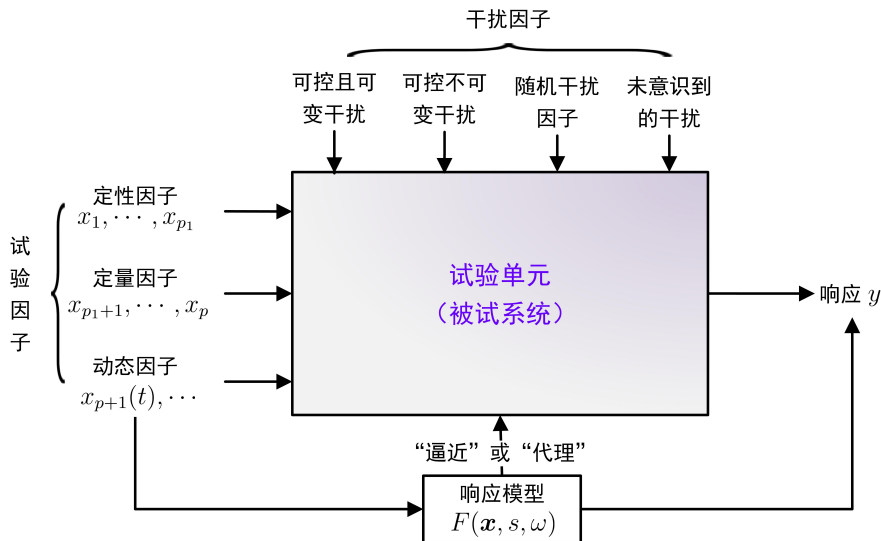
- 称衡量试验效果的特征量为**响应**, 也称**指标**, 用字母 y 来表示.
 - 定性响应和定量响应.
 - 单响应试验和多响应试验.
 - 多响应试验通常也转化为单响应试验来处理.
- 称响应的测量值与真值之间的偏差为**误差**
 - 误差是一个随机变量, 一般假定 $\varepsilon \sim N(0, \sigma^2)$.

- 凡对响应可能产生影响的原因都称为因子, 也称因素.
- 试验因子: 根据试验目的选中的可控的因子
 - 定性因子、定量因子、动态因子
- 干扰因子: 试验因子以外的所有因子
 - 多种分类方式: 随机和非随机、可控和不可控、可测量和不可测量、已知和未知、...
- 称试验因子在试验中所处的状态 (或所取的值) 为水平, 试验因子的水平组合构成处理.

- **试验单元**: 一次试验中试验对象、实验员、工具以及材料等组成的基本单元.
- 试验单元之间的差异可能是造成结果波动的原因之一.
- 试验单元是一个非随机的已知的干扰因子.

- 称因子与响应之间的数量关系为响应函数;
- 响应函数就是因果关系的定量描述;
- 称响应函数的数学模型为响应模型;
- 假定存在一个真实的响应函数, 试验的最终目标是为了探究响应函数.

试验单元、因子和响应共同构成试验的三要素:



例 (Yule-Simpson's Paradox)

研究某一疾病的某种处理方式是否有效, 数据如下

		存活人数	死亡人数	存活率
总体	处理组	200	200	50%
	对照组	160	240	40%
男性	处理组	180	120	60%
	对照组	70	30	70%
女性	处理组	20	80	20%
	对照组	90	210	30%

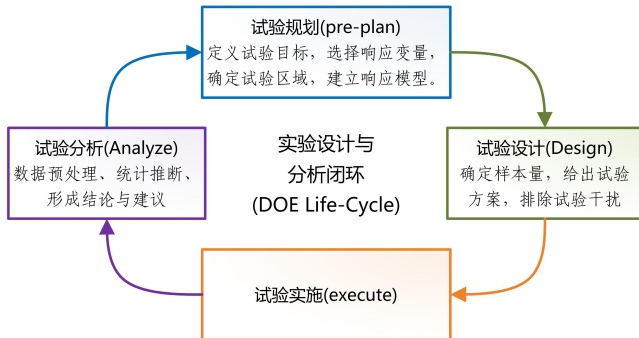
- 响应、因子、水平、处理、试验单元分别是什么?

1.3 试验的基本概念与流程

1.3.1 试验的基本概念

1.3.2 试验的基本流程

1.3.3 作战实验: 恐怖的海峡



- <https://testscience.org>
- R. T. Johnson, et al. Designed experiments for the defense community. Quality Engineering, 24(1):60–79, 2012.
- E. Viles, et al. Planning Experiments, the First Real Task in Reaching a Goal. Quality Engineering, 21(1):44–51, 2008.
- L. J. Freeman, et al. A tutorial on the planning of experiments. Quality Engineering, 25(4):315–332, 2013.

1.3 试验的基本概念与流程

1.3.1 试验的基本概念

1.3.2 试验的基本流程

- (1) 试验规划
- (2) 试验设计
- (3) 试验分析

1.3.3 作战实验: 恐怖的海峡

Step 1 明确试验需求与目的

- **处理比较**: 比较不同处理下系统响应的差异是否统计显著;
- **因子筛选**: 对不同试验因子对响应影响的重要程度排序, 确定引起响应变化的主要因子;
- **系统辨识**: 辨识试验因子与响应之间的行为关系, 获得响应与因子之间的数学模型, 以预测响应;
- **系统优化**: 包括找到使响应达到某特定值 (最大、最小以及达标) 的处理, 和找到使响应达到目标值且受噪声因子影响最小的处理;
- **问题发现**: 哪些处理导致系统失效? 这一类试验目的由于其危险性, 通常需要借助仿真开展.

Step 2 选择响应变量和试验因子

- 响应变量由试验目的决定

- 复杂系统的响应是一套指标体系
- 三级指标体系: 使命 (mission)、任务 (task)、性能 (attribute)
- 多响应试验转化为单响应试验: 评估
- 以 $y \in \mathbb{R}$ 表示响应变量

- 试验因子由试验目的和条件决定

- 由领域专家和统计学家共同从可控因子中挑选
- 注意潜在变量的影响.
- 以 $x = (x_1, \dots, x_p)^T$ 表示试验因子.

Step 2 选择响应变量和试验因子

- 响应变量由试验目的决定
 - 复杂系统的响应是一套指标体系
 - 三级指标体系: 使命 (mission)、任务 (task)、性能 (attribute)
 - 多响应试验转化为单响应试验: 评估
 - 以 $y \in \mathbb{R}$ 表示响应变量
- 试验因子由试验目的和条件决定
 - 由领域专家和统计学家共同从可控因子中挑选
 - 注意潜在变量的影响.
 - 以 $\boldsymbol{x} = (x_1, \dots, x_p)^T$ 表示试验因子.

Step 3 明确试验约束

- 试验资源约束

- 预算、时间、....

- 试验条件约束

- 导引头挂飞试验速度达不到真实速度;
 - 远程弹道导弹不能开展全程弹道试验;
 - 作战试验受环境构设能力限制;
 -

- 资源约束限制了试验次数, 条件约束限制了试验区域 $x \in \mathcal{X}$.

Step 3 明确试验约束

- 试验资源约束
 - 预算、时间、....
- 试验条件约束
 - 导引头挂飞试验速度达不到真实速度;
 - 远程弹道导弹不能开展全程弹道试验;
 - 作战试验受环境构设能力限制;
 -
- 资源约束限制了试验次数, 条件约束限制了试验区域 $x \in \mathcal{X}$.

Step 4 建立响应模型

- 根据领域知识、历史数据和专家经验, 建立因子与响应之间的响应模型.
- 三类先验信息: 结构信息、机理信息和参数信息.
- 响应模型可抽象地记为 $y = F(\mathbf{x}, s, \omega)$,
 - $s \in \mathcal{S}$ 表示抽象的未知参数, 认知不确定性;
 - $\omega \in (\Omega, \mathcal{A}, P)$ 表示随机因子, 随机不确定性;
 - $F: \mathcal{X} \times \mathcal{S} \times \Omega \mapsto \mathcal{Y}$ 表示已知映射.
- 响应建模十分困难, 决定试验的难易程度!

Step 4 建立响应模型

- 根据领域知识、历史数据和专家经验, 建立因子与响应之间的响应模型.
- 三类先验信息: 结构信息、机理信息和参数信息.
- 响应模型可抽象地记为 $y = F(\mathbf{x}, s, \omega)$,
 - $s \in \mathcal{S}$ 表示抽象的未知参数, 认知不确定性;
 - $\omega \in (\Omega, \mathcal{A}, P)$ 表示随机因子, 随机不确定性;
 - $F: \mathcal{X} \times \mathcal{S} \times \Omega \mapsto \mathcal{Y}$ 表示已知映射.
- 响应建模十分困难, 决定试验的难易程度!

1.3 试验的基本概念与流程

1.3.1 试验的基本概念

1.3.2 试验的基本流程

- (1) 试验规划
- (2) 试验设计
- (3) 试验分析

1.3.3 作战实验: 恐怖的海峡

动词: 试验设计

依据某一准则, 从试验空间 \mathcal{X} 中挑选出部分处理形成试验方案, 本质上是一个优化问题.

Step 1 确定试验次数, 资源消耗与精度构成一对均衡关系:

- 限定试验资源使精度尽可能高
- 限定精度使样本量最小
- 定义损失函数, 如 “1 单位精度 = 1000 单位资源”

Step 2 确定设计的准则, 并依据该准则从 \mathcal{X} 中确定试验方案
 $\xi_n = \{x_1, x_2, \dots, x_n\}$.

Step 3 控制误差, 降低干扰因子的影响.

动词: 试验设计

依据某一准则, 从试验空间 \mathcal{X} 中挑选出部分处理形成试验方案, 本质上是一个优化问题.

Step 1 确定试验次数, 资源消耗与精度构成一对**均衡关系**:

- 限定试验资源使精度尽可能高
- 限定精度使样本量最小
- 定义损失函数, 如 “1 单位精度 = 1000 单位资源”

Step 2 确定设计的准则, 并依据该准则从 \mathcal{X} 中确定试验方案 $\xi_n = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$.

Step 3 控制误差, 降低干扰因子的影响.

名词: 试验设计

① 用矩阵表示一个精确设计 $D_{\xi_n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$

② 离散概率分布表示离散设计 $\xi_k = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_k \\ \omega_1 & \omega_2 & \cdots & \omega_n \end{pmatrix}$

③ \mathcal{X} 上的连续概率分布来表示的设计为逼近设计.

④ \mathcal{X} 上任意分布表示设计, 以该分布的 n 个样本作为试验点.

1.3 试验的基本概念与流程

1.3.1 试验的基本概念

1.3.2 试验的基本流程

- (1) 试验规划
- (2) 试验设计
- (3) 试验分析

1.3.3 作战实验: 恐怖的海峡

从试验数据 $\{(\mathbf{x}_i, y_i) : i = 1, \dots, n\}$ 中提取信息:

- ① 比较 y_1, y_2, \dots, y_n 之间的差异与优劣;
- ② 比较各因子 (x_1, \dots, x_p) 的重要性以挑选出显著因子;
- ③ 获得响应函数的某些“加工”:

$$\int_{\mathcal{X}} F(\mathbf{x}, s, \omega) d\mathbf{x}, \quad \sup_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}, s, \omega), \quad \inf_{\mathbf{x} \in \mathcal{X}} F(\mathbf{x}, s, \omega);$$

- ④ 获得响应函数的水平集 $\{\mathbf{x} \in \mathcal{X} : F(\mathbf{x}, s, \omega) \geq y_0\}$.
- ⑤ 获得响应函数, 即响应模型中的未知参数 s .

主要包括方差分析法和回归分析法.

- 装备试验鉴定工作: 要着力构建先进实用的试验鉴定体系, 摸清武器装备性能底数, 确保武器装备实战适用性.
- 如何理解装备试验鉴定中的“摸边探底”?

1.3 试验的基本概念与流程

1.3.1 试验的基本概念

1.3.2 试验的基本流程

- (1) 试验规划
- (2) 试验设计
- (3) 试验分析

1.3.3 作战实验: 恐怖的海峡

美国兰德 (Rand) 公司以大陆与台湾之间的军事对抗为背景, 于 2000 年发表了一份以"Dire Strait? Military Aspects of the China-Taiwan Confrontation and Options for U.S. Policy" 为题的研究报告.

报告确定了帮助台湾保持足够的防御态势以对抗大陆的关键因素, 提出了美国协助台湾解决这些问题的一系列行动建议.

2005 年, 大陆采用空中进攻和两栖登陆作战的极端作战方案对台发动进攻, 武力统一台湾. 作战过程分为 4 阶段:

- ① 争夺制空权. 两岸空战, 大陆对台空军基地、地空导弹阵地、早期预警雷达和指挥与控制设施实施导弹攻击和空中打击.
- ② 争夺制海权. 包括反航空兵作战、反水面舰艇作战和反潜作战等样式, 可与第一阶段同时进行.
- ③ 大陆夺取制空权后, 通过空中打击集中摧毁沿岸据点、台军炮兵和装甲集群, 全面“软化”岛上防御体系.
- ④ 登陆作战. 包括两栖登陆、伞降突击和直升机攻击, 其目的是在岛上建立坚固的立足点, 打垮台军的抵抗.

兰德公司利用“联合一体化应急作战模型”对争夺制空权空中优势进行了实验.

(1) 试验规划

- 响应变量 y 为推演结果, 取 “胜” 或 “负” ;
- 选择 7 个试验因子, 并分别划分了水平.

(1) 试验规划

因子一 大陆投入攻击的空中部队的规模与编成

机型	基本力量	大规模力量	未来力量
苏-27	72	72	144
强-5	120	216	0
歼轰-7	48	48	72
歼-7	168	288	144
歼-8	144	288	144
歼-10	24	24	48
苏-30	0	0	24
轰-6	48	48	48
预警机载指挥系统	6	6	6
辅助飞机	49	49	49
总计	679	1039	679

(1) 试验规划

因子二 双方拥有超视距中程空空导弹的情况

- 双方均无; 大陆无, 台湾有; 大陆有, 台湾无; 双方均有. 假设只要有, 则数量充足.

因子三 大陆使用的中、近程弹道导弹的数量和质量

- 基本情况 (20 枚 DF-3, 80 枚 DF-21, 50 枚 DF-11, 80 枚 DF-15, 80 枚带 GPS 制导的 DF-15, 共 310 枚)
- 大规模情况 (总数翻倍, 为 620 枚) .

(1) 试验规划

因子四 大陆拥有的先进空射精确制导弹药的数量

- 划分为基本情况 (300 枚)
- 大规模情况 (3000 枚)

因子五 台湾空军出动战斗架次的能力

- 三水平: 基准架次的 100%、75% 和 50%.

因子六 台湾空军机组人员的素质

- 二水平: 达到美空军飞行员水平的 80% 和 60%.

(1) 试验规划

因子七 美空中力量参与台湾防御的程度.

- 没有美国部队参战
- 1 个舰载机联队在台湾以东海域作战
- 驻扎在日本冲绳嘉手纳空军基地的 1 个美国空军联队参战, 有 72 架 F-15C 战斗机
- 1 个舰载机联队和 1 个战斗机联队参战
- 2 个舰载机联队参战
- 2 个舰载机联队和 1 个战斗机联队参战

(1) 试验规划

- 兰德公司把其它干扰因子固定在设定的水平上.

表: 台湾空军编成

机型	数量
F-16A/B	162
"幻影" 2000	54
"经国" 号 (IDF)	126
E-2T 机载预警与控制系统	8
其它各型飞机	44
总计	394

表: 台湾地对空导弹战斗序列

导弹类型	导弹连数量
"爱国者" PA C-2 型导弹	9 (每连 6 部 4 管发射架)
改进型 "霍 克" 型导弹	36 (每连 18 部 3 管发射架)
"天弓" 型导弹	6 (每连 6 部 4 管发射架)

(2) 试验设计

- 采用全面实施的实验方案, 共得到

$$3 \times 4 \times 2 \times 2 \times 3 \times 2 \times 6 = 1728$$

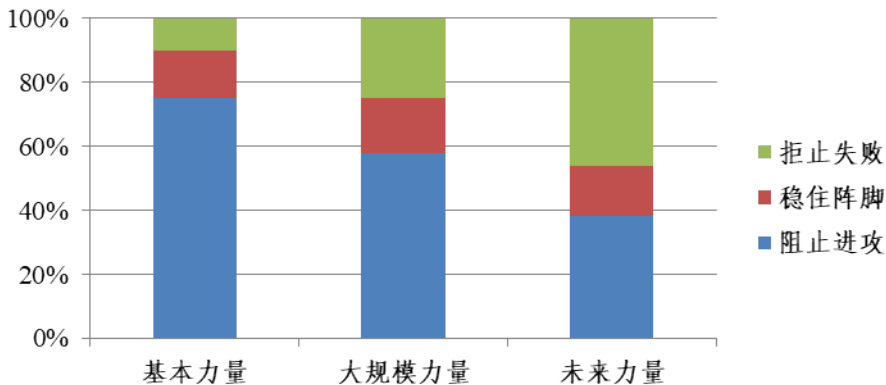
个试验点.

- 在作战仿真中, 称一个实验点为一个作战想定样本.
- 如何建立响应模型?

(3) 试验分析

- 兰德公司对诸因子的各种水平下推演结果的占比做了统计;
- 总体结论: 大陆近期内攻占台湾的任何企图都会是一场失败概率相当高的血腥战争.

(3) 试验分析



(3) 试验分析

其它 6 个因子在绿色的结果中的取值情况:

因子	“基本力量”	“大规模力量”	“未来力量”
美国的有限介入	100%	81%	62%
大陆拥有超视距武器优势	80%	71%	51%
台湾出动飞机架次的能力受到压制	80%	75%	64%
台湾训练质量低劣	67%	65%	54%
大陆拥有更多战术导弹	56%	54%	51%
大陆拥有更多精确制导弹药	56%	54%	48%

大陆空军在出动“基本力量”同时拥有超视距武器优势的样本数量 ($1728 \times 1/3 \times 1/4 = 144$) 与大陆空军在出动“基本力量”且获胜的样本数量比值.

(3) 试验分析

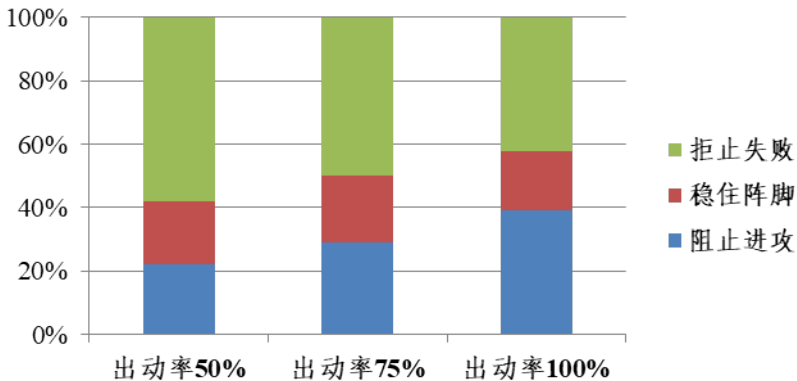


图: 无美军介入时台湾基地运转能力对作战结果的影响

兰德公司建议台湾应对机场的被动防御和快速修复措施给予更多的关注.

(3) 试验分析

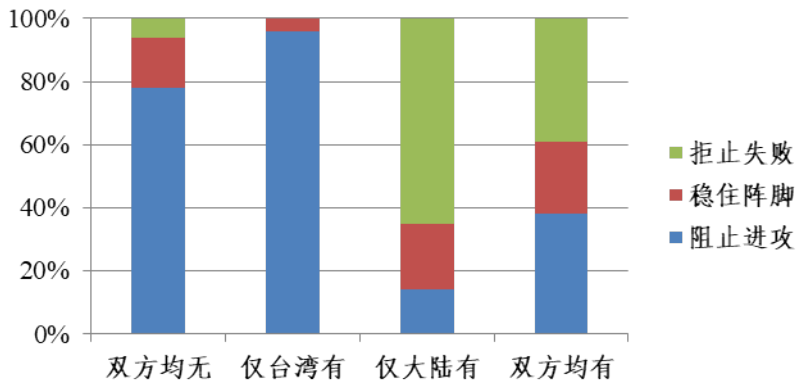


图: 先进的空对空武器 (超视距空空导弹) 对作战结果的影响

兰德公司建议一旦大陆从俄罗斯获得超视距空空导弹, 美国应立即向台湾移交 AIM-120 超视距空空导弹。

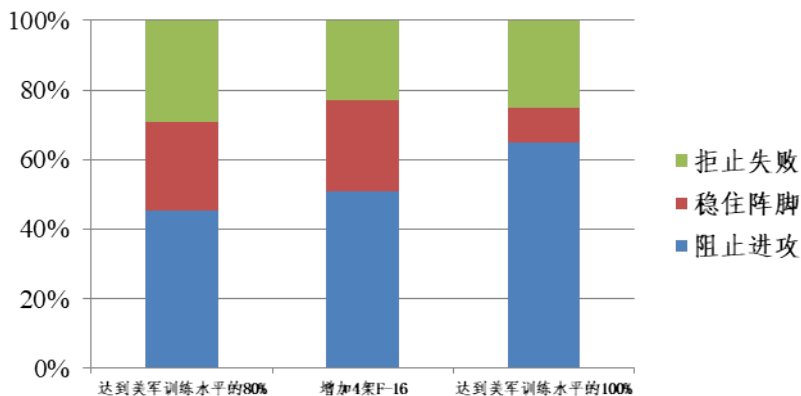


图: 台湾空军训练质量对作战结果的影响

中间一列表示台湾空军训练质量达到美军训练水平的 80% 的条件下, 给台湾空军增加 24 架 F-16 时的作战结果. 兰德公司给出指出, 在短时间内, 提高台湾空军的训练质量比单纯提供更多的 F-16 更有利于台湾夺取空中优势.

(3) 试验分析

美国只需介入少许力量就能大幅提高台湾抵御大陆进攻的成功率。

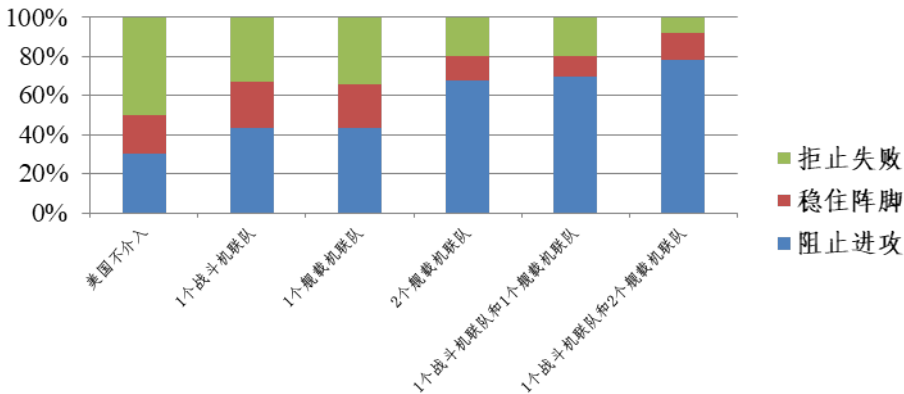


图: 美国介入的价值

(3) 试验分析

美国干预的作用在当前就至关重要, 而且会越来越关键. 且在近期内支持台湾所需要的兵力大大低于五角大楼通常认为实施一场大规模战区战争所需要的兵力规模 (在“沙漠风暴”期间, 美军出动了 10 个战斗机联队和 6 个航母作战大队) .

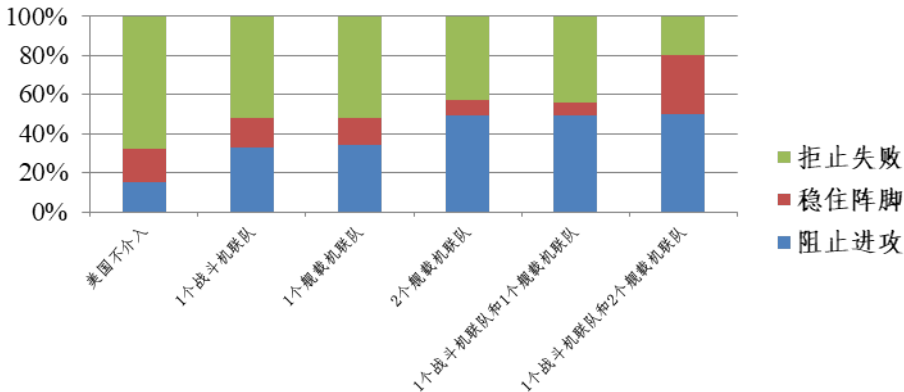


图: 大陆“未来力量”条件下美国介入的价值

总结

- 试验是一个过程：试验规划、试验设计、试验实施、试验分析
- 试验规划：明确实验需求、选择响应与因子、刻画试验约束、建立响应模型
- 试验设计：动词、名词
- 试验分析：试验目标的形式化描述
- 作战实验：恐怖的海峡

习题

- 1 查阅文献, 了解抽样与试验设计的联系和区别.
- 2 安装 RStudio, 并利用 “swirl” 包自学 R 语言.
- 3 请查阅 Fisher, Yates, Box, Taguchi 中某一位的生平并梳理其在试验设计领域的主要贡献.
- 4 从恐怖的海峡的实验结果, 你能分析出 7 个因子中, 哪个因子对空战结果影响最大吗?
- 5 访问宾夕法尼亚大学艾伯利理学院 (PennState Eberly College of Science) 试验设计课程网站:
<https://newonlinecourses.science.psu.edu/stat503/node/1/>