

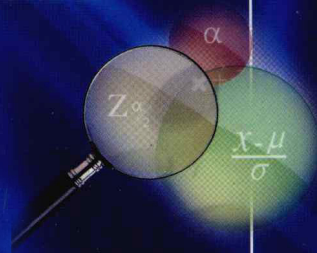
基础培训

新东方AP考试指定辅导教程

AP 统计学

Statistics

- 国内第一本中英文结合教材 ·
- 易于理解的AP统计学知识体系 ·



北京新东方学校AP教研中心 · 主编
徐一丁 · 编著

基础培训

新东方AP考试指定辅导教程

AP 统计学

北京新东方学校AP教研中心·主编

徐一丁·编著

图书在版编目(CIP)数据

AP 统计学 / 徐一丁编著. —北京: 群言出版社, 2010.12

ISBN 978-7-80256-208-0

I. ①A… II. ①徐… III. ①统计学—高等学校—入学考试—自学参考资料 IV. ①C8

中国版本图书馆 CIP 数据核字(2010)第 225089 号

AP 统计学

出 版 人 范 芳

责任编辑 孙春红

封面设计 王 琳 盛 蓝

出版发行 群言出版社(Qunyan Press)

地 址 北京东城区东厂胡同北巷 1 号

邮政编码 100006

网 站 www.qypublish.com

读者信箱 bj62605588@163.com

总 编 办 010—65265404 65138815

编 辑 部 010—65276609 65262436

发 行 部 010—62605588 62605019

经 销 新华书店

读者服务 010—65220236 65265404 65263345

法律顾问 中济律师事务所

印 刷 北京通州皇家印刷厂

版 次 2010 年 12 月第 1 版 2010 年 12 月第 1 次印刷

开 本 880mm×1230mm 1/16

印 张 19.5

字 数 291 千

书 号 ISBN 978-7-80256-208-0

定 价 59.00 元

 [版权所有 侵权必究]

如有缺页、倒页、脱页等印装质量问题,请拨打服务热线:010—62605166。

新东方
NEW ORIENTAL **图书策划委员会**

主任 俞敏洪

委员 (按姓氏笔画为序)

王文山 王 强

包凡一 仲晓红

沙云龙 陈向东

邱政政 汪海涛

周成刚 徐小平

谢 琴 窦中川

序 言

AP(Advanced Placement)课程在中国的成功登陆,无论是对中国高中教育的体系还是对中国高中教育的教学思路都是个不小的冲击。不同于国内高中“一刀切”(不论学习能力强弱,不论不同兴趣爱好,所有高中生都学习统一的标准课程)的教育体系,AP课程作为美国大学理事会(College Board)的明星产品,让学有余力的高中生能提前学习大学课程,不仅体现了分级教学的思想,更是满足了不同学习能力、不同学习兴趣学生学习的需要。

据美国大学理事会的年度报告,申请顶尖名校的学生向大学招生办提供了平均4~5门AP成绩,而AP成绩在所有录取因素中以80.3%的影响力居第一位。因此,在SAT和TOEFL成绩的基础上,AP成绩成为了步入名校竞争的新项目。随着中国学生留学大潮的涌来,加上AP课程在中国本土的开设,AP考试成为了时下最时髦的留学考试之一。

历史的实践告诉我们,无论是科学,还是技术,还是小到一个考试,完全照搬西方肯定是行不通的。因此,AP在中国的教学、考试辅导等工作都已经悄然开展,不管是摸石头过河,还是模拟赶超,AP中国本土化势必进行。基于此,北京新东方学校成立了一支具有博士学位教师组成的AP教研中心团队,大力开展了AP课程的教学教研活动。在近两年的研究过程当中,教材问题成为了当前最棘手的问题。在目前国内市面上,适合中国本土的AP教材几乎空白。为了帮助国内AP学习者更好的学习AP课程和准备AP考试,我们按照AP官方给出的考试大纲,编写了此套中英文结合的AP教材。AP考试不同于托福等语言类考试,它是学科考试。学科知识无论用什么语言来描述都是同样的知识,因此本系列教材适当辅以中文解释,帮助考生更深入的理解。另一方面,为了让考生能够适合未来英文学习和英文考试的需要,本系列教材的定理展开、例题讲解等大部分内容都采用了英文描述。中英结合,易于中国考生对知识点的理解和把握。希望本套教材能给AP学习者助一臂之力。

最后,要感谢关心和支持本套书出版的大愚文化传播公司,北京新东方学校北美部的AP教研团队,是他们的努力才使得本书能够顺利的出版。限于水平,书中难免有不妥之处,望同行、读者不吝赐教。

范 猛

北京新东方学校北美考试部主任

Contents

Chapter 1	Introduction of AP Statistics Exam AP 统计学考试简介	1
Chapter 2	Descriptive Analysis of Categorical Data 分类数据的描述分析	6
2.1	Types of variables 变量类型	8
2.2	Analysis of univariate categorical data 单变量分类数据的分析	10
2.3	Analysis of bivariate categorical data 双变量分类数据的分析	16
Chapter 3	Descriptive Analysis of Continuous Data 连续型数据的描述分析	21
3.1	Analysis of univariate continuous data 单变量连续型数据的分析	23
3.2	Analysis of bivariate continuous data 双变量连续型数据的分析	49
Chapter 4	Sampling, Surveys and Experiments 抽样, 调查和实验设计	79
4.1	Some basic concepts 几个基本概念	82
4.2	Methods of data collection 收集数据的方式	84
4.3	Some simple sampling methods 几种简单的抽样方法	87
4.4	Terms and concept of experiments 实验设计中的基本概念	95
4.5	Single-blind experiment and double-blind experiment 单盲试验和双盲实验	101
4.6	Experiments 实验设计	101
Chapter 5	Probability and Probability Distribution 概率与概率分布	113
5.1	Probability 概率	115
5.2	Basic rules and more probability rules 概率的基本准则及运算法则	118
5.3	Discrete random variable and probability distribution 离散型随机变量及其概率分布	125
5.4	The probability distribution of continuous random variable 连续型随机变量的概率分布	138
5.5	Combining independent random variables 独立随机变量的组合	144
Chapter 6	Statistic and Sampling Distribution 统计量及抽样分布	163
6.1	Basic concepts 基本概念	165
6.2	Sampling distribution of sample mean and central limit theorem 样本均值的抽样分布与中心极限定理	167
6.3	Sampling distribution of a sample proportion 样本比例的抽样分布	171
6.4	Sampling distribution of a difference between two independent sampling means 两个相互独立的样本均值差的抽样分布	172

6.5	Sampling distribution of a difference between two independent sample proportions 两个相互独立的样本比例差的抽样分布	173
Chapter 7	Parameter Estimation 参数估计	183
7.1	General issues 参数估计的一般问题	185
7.2	Interval estimation for a population parameter 一个总体参数的区间估计	194
7.3	Interval estimation for two population parameter 两个总体参数的区间估计	207
7.4	Determine the sample size 样本量的确定	219
7.5	Interval estimation for the slope of the regression line 回归系数的区间估计	223
Chapter 8	Hypothesis Testing 假设检验	240
8.1	Testing a hypothesis 假设检验的基本问题	242
8.2	Inference about one population parameter 一个总体参数的检验	250
8.3	Inference for two population parameters 两个总体参数的检验	260
8.4	Significance test for the slope of the regression line 回归系数(最小二乘回归线斜率)的显著性检验	267
8.5	Test through the confidence interval 用置信区间进行检验	270
8.6	Test for categorical data 分类数据的假设检验	271
Formulas and Tables	公式和图表	297

Introduction of AP Statistics Exam

AP统计学考试简介

AP 是 Advanced Placement 的缩写,即大学先修课程。AP 课程及考试始于 1955 年。美国高中 AP 课程是在美国大学理事会(College Board)的主持下,在高中阶段开设的具有大学水平的课程,涵盖 22 个门类、37 个学科,已在美国 15000 多所高中开设。它可以使高中生提前接触大学课程,避免了高中和大学初级阶段课程的重复。目前,已有 40 多个国家的近 3600 所大学承认 AP 学分作为其入学参考标准,其中包括哈佛大学、耶鲁大学、牛津大学、剑桥大学等世界名牌大学。清华大学、北京大学、北京语言大学等也接受 AP 成绩。

美国每年约有 200 万高中毕业生,他们都要参加美国高考 SAT 和 AP 课程的考试。在美国,初等教育是免费的,但高等教育是收费的。美国高中生会在 11 年级时完成 SAT 的考试,在 12 年级(高中最后一年)则要完成两件大事。其一,根据 SAT 的考试成绩,申请大学和奖学金;其二,选修 AP 课程并准备考试。AP 考试的目的在于,利用高中最后一年免费教育的时间,提前完成一些美国大学的学分课程及考试。否则,在大学阶段完成相同的课程和学分,需要支付高昂的学费。也就是说,AP 课程及考试可以为高中生起到减免大学学分、降低大学教育成本、缩短大学教育时间的目的。对中国学生而言,除可以获得美国大学学分,省时省钱外,还可以在国内提前解决好在美国上大一课程适应难的问题。

申请美国名牌大学时需要向学校招生办提供以下材料:申请文(个人陈述)、年级排名、在校总平均成绩,入学考试(SAT)成绩、AP/A-LEVEL 成绩、推荐信、工作及实践经验等信息。通过美国大学录取顾问委员会公布的影响大学录取因素比较分析可以知道:AP 成绩以 80.3% 的影响力居第一,对大学录取过程起决定性作用,因为它向学校充分展示了学生的才智、专长及学生能力。因此学习 AP 课程,参加 AP 考试不仅可以抵大学学分,它更可以让学生在名校录取的竞争中脱颖而出,占尽先机。

AP 统计学考试于 1997 年由美国大学理事会(College Board)开始提供。在 1997 年,约有 7667 名学生参加了 AP 统计学考试(人数最多)。自此以后,参加 AP 统计学考试的学生数量增长迅速。到 2010 年,参加 AP 统计学考试的学生达到约 13 万。AP 统计学是 AP 考试中参考人数最多的考试之一。2010 年,AP 统计学考试的平均分在 2.84 分(满分 5 分)。多数大学接受 4 分或 5 分的成绩,另有少数大学接受 3 分的成绩。在 2010 年,约有 12.8% 的学生的分数为 5 分,约有 22.4% 的学生的分数为 4 分,约有 23.5% 的学生的分数为 3 分,共有约 58.7% 的学生的 AP 统计学成绩在 3 分以上。

开设 AP 统计学课程的学校或者自学的学生,应该在高一高二进行合理的时间安排,确定课程计划,以保证把学习统计学应具备的知识先行学习完毕。由于 AP 统计学是一门

大学水平的课程，所以具有一定的挑战性。

一、考试报名

AP 官方报名网址：<http://www.collegeboard.com>

香港考务局报名网址：<https://www.hkeaa.edu.hk>

AP 统计学课程考试费用：约 87 美元

AP 统计学课程每年 3 月上旬截止报名，5 月初考试。

二、考试内容

AP 统计学考试内容包括以下几部分：

主要内容	考试占比
Exploring data(描述统计)	20% – 30%
Sampling and experiments (抽样和实验设计)	10% – 15%
Anticipating patterns(probability and random variables)(概率和随机变量)	20% – 30%
Statistical inference(统计推断)	30% – 40%

具体见考试说明《AP STATISTICS Course Description-Effective Fall 2010》

下载地址：

<http://apcentral.collegeboard.com/apc/public/repository/ap-statistics-course-description.pdf>

三、考试概况

1. AP 统计学考试题型及时间安排

Multiple choice (选择题) 无计算器	40 题/90 分钟	每个选择题答对得 1 分，不答得 0 分，答错不扣分。卷面得分乘以 1.25 为最后得分。如果所有题都对，则得 50 分。
Free response (自由问答题)	总计 6 题/90 分钟	
A 部分：有计算器	共 5 题/60 – 65 分钟	每题答对得 4 分，卷面得分乘以 1.875 得到最后得分。即每题 7.5 分，总计 37.5 分。
B 部分：有计算器	1 题/25 – 30 分钟	答对得 4 分，卷面得分乘以 3.125 得到最后得分。即 12.5 分。

2. AP 统计学成绩的分数对应表及得分比例

下表是 2007 年 AP 统计学的分数对应表，每年该表都会随着问题的难度而发生微小的变化。

AP statistics exam	
考试实际分数	五分制
60 – 100	5
45 – 59	4
32 – 44	3
23 – 31	2
0 – 22	1

AP 统计学得分比例(2010 年)

考试得分	Statistics	
	人数	百分比
5	16681	12.8
4	29108	22.4
3	30472	23.5
2	23599	18.2
1	30039	23.1
学生总数	129899	
3 分及以上	76261	58.7
平均分	2.84	
标准差	1.35	

3. 考试成绩评定

AP 考试的成绩评定实行 5 分制。

5 分——具有非常好的资格；

4 分——具有好的资格；

3 分——具有资格；

2 分——可能有资格；

1 分——不予推荐。

一般而言, 3 分及以上的成绩为大多数大学所接受, 即可以在进入大学后折抵大学的学分。少数顶尖大学要求 4 分或 5 分才能折抵大学学分。

4. 统计学考试允许使用的图形计算器(参见 AP 官网公布的图形计算器)

AP 统计学最常用的计算器是 TI - 83、TI - 84。

Casio	Hewlett-Packard	Texas Instruments
FX - 6000 series	HP - 9G	TI - 73
FX - 6200 series	HP - 28 series *	TI - 80
FX - 6300 series	HP - 38G *	TI - 81
FX - 6500 series	HP - 39 series *	TI - 82 *
FX - 7000 series	HP - 40 series *	TI - 83/TI - 83 Plus *
FX - 7300 series	HP - 48 series *	TI - 83 Plus Silver *
FX - 7400 series	HP - 49 series *	TI - 84 Plus *
FX - 7500 series	HP - 50 series *	TI - 84 Plus Silver *
FX - 7700 series		TI - 85 *
FX - 7800 series	Radio Shack	TI - 86 *
FX - 8000 series	EC - 4033	TI - 89 *
FX - 8500 series	EC - 4034	TI - 89 Titanium *
FX - 8700 series	EC - 4037	TI - Nspire *
FX - 8800 series		TI - Nspire CAS *
FX - 9700 series *	Sharp	
FX - 9750 series *	EL - 5200	Other
FX - 9860 series *	EL - 9200 series *	Datexx DS - 883
CFX - 9800 series *	EL - 9300 series *	Micronta
CFX - 9850 series *	EL - 9600 series *†	Smart ²
CFX - 9950 series *	EL - 9900 series *	
CFX - 9970 series *		
FX 1.0 series *		
Algebra FX 2.0 series *		

四、主要备考资料

1. 考试说明《AP STATISTICS Course Description-Effective Fall 2010》。考试说明为考生必读书, 不仅列有考试大纲, 并且还有真题及评分示范。在备考时要重点研究考试说明。

2. AP 官网上的所有 free-response 都是可以免费下载的。

网址: <http://www.collegeboard.com/student/testing/ap/statistics/samp.html?stats>

五、备考时间安排

AP 统计学一般是一学期的课程, 考试安排在 5 月份。要想在一个有限的集中时间内完成学习和考试的任务, 必须做好时间安排。下面给出一个建议的时间表。

时间	周数	安排
12.1 - 12.5	1	了解 AP 统计考试。下载考试说明以及考试报名的情况, 准备计算器(建议使用 TI-83 或 TI-84), 学习一些基本的统计术语及其英文表述
12.6 - 12.31	4	将教材从头到尾看一遍, 要求基本看懂、看通(建议参加一个辅导班, 疏导一遍考试内容的框架)
1.1 - 2.28	8	再次研读教材, 将所有例题和练习题至少做两遍, 并学习计算器的应用。
3.1 - 4.15	6	研究真题, 做模拟题(建议至少做两遍), 按照考试要求的时间进行练习, 总结知识点及解题方法
4.16 - 4.30	2	回顾知识点, 继续做真题(建议至少做两遍), 查漏补缺
5.1 - 考试	1	考前心态调节, 重点温习, 准备考试用品

描述统计主要包括两部分，分类数据的描述分析以及连续型数据的描述分析。本章是其中一部分——分类数据的描述分析。描述统计的这两部分内容共占考试内容的 20% ~ 30%，其中包括选择题 8 ~ 12 个，问答题 1 ~ 2 个。需要注意的是描述统计部分的考试重点主要集中在连续型数据的描述分析，即第三章的内容。

主要内容

- Analysis of univariate categorical data 单变量分类数据分析
- Analysis of bivariate categorical data 双变量分类数据分析
- Frequency distribution table 频数分布表
- Bar chart 条形图
- Pie chart 饼图
- Contingency table 列联表

【 词汇梳理 】

Descriptive statistics 描述统计

Category 类别

Categorical variable 分类变量

Qualitative variable 定性变量

Metric variable 数值型变量

Discrete variable 离散型变量

Continuous variable 连续型变量

Quantitative variable 定量变量

Univariate data 单变量数据

Bivariate data 双变量数据

Frequency 频数

Relative frequency 相对频数

Cumulative frequency 累积频数

Percentage 百分比

Frequency distribution table 频数分布表

Cumulative frequency charts 累积频数图

Bar chart 条形图

Pie chart 饼图

Contingency table 列联表

Independent 独立

【知识精讲】

2.1 Types of variables 变量类型

个体(individual)就是一组数据描述的对象。个体可以指人,也可以是动物或者其他东西。

变量(variable)是说明现象某种特征的概念,其特点是从一次观察到下一次观察,其观察结果会呈现出差别或变化。同一个变量对于不同个体,可能有不同的取值。如“受教育程度”,“身高”等都是变量。变量的具体取值称为变量值。比如身高可以是 160cm, 165cm, 170cm 等等,这些数字就是变量值。

变量可以分为以下两种类型。

2.1.1 Categorical variable(分类变量)

分类变量(categorical variable)是说明事物类别的一个名称。也称定性变量(qualitative variable),其取值是分类数据。如“性别”就是分类变量,其变量值为“男”或“女”。

- A variable is categorical if it can only take on a fixed number of values.

分类数据(categorical data)是只能归于某一类别(category)的非数字型数据。它是对事物进行分类的结果,数据表现为类别,是用文字来表述的。例如,人口按照性别分为男、女两类;为了便于统计处理,对于分类数据可以用数字代码来表示各个类别,比如用 1 表示“男性”,0 表示“女性”。

Note: Categorical data is also referred to as qualitative data.

2.1.2 Metric variable(数值型变量)

数值型变量(metric variable)是说明事物数字特征的一个名称。其取值是数值型数据,如“年龄”、“时间”、“产品质量”等都是数值型变量,这些变量可以取不同的数值。也称定量变量(quantitative variable)。

数值型变量根据其取值的不同又可以分为 discrete variable(离散型变量)与 continuous variable(连续型变量)这两种。显然,离散型变量是指其取值可以一一列举的变量。比如一天发几封邮件,这就是离散型变量。连续型变量是可以在一个或多个区间中取任何值的变量,它的取值是连续不断的,不能一一列举,比如“年龄”、“身高”、“温度”、“体重”等都是连续型变量。

我们更多的会研究 continuous variable(连续型变量)。

- A variable is continuous if it can take on any value.

有一种变量，其取值是顺序数据，这是说明事物有序类别的一个名称，称之为“顺序变量”。如“产品等级”就是一个顺序变量，其变量值可以为“一等品”，“二等品”，“三等品”，“次品”等；“受教育程度”也是一个顺序变量，其变量值可以为“小学”、“初中”、“高中”、“大学”等。

Example 1 :

In application to serve a local airport, an airline supplies the following projections. Which of the following is categorical?

- A. projected number of flights per week
- B. projected number of passengers per flight
- C. projected cities to be connected
- D. projected gross weight of aircraft
- E. projected advertising volume in local media

答案：C。

解析：本题考点为分类变量的性质。只有 C: projected city 是一个分类变量，因此选 C。

Example 2 :

The height of John is what type of data?

- A. categorical
- B. quantitative and continuous
- C. quantitative and discrete
- D. categorical and quantitative
- E. categorical and continuous

答案：B。

解析：身高可以取某个区间内的任何值，属于连续型变量，因此选 B。

2.1.3 Univariate data and bivariate data(单变量数据和双变量数据)

对每个个体只进行一项测量(一个变量)，就得到单变量数据。例如，只对学生登记身高，这样得到的数据即是单变量数据。

对每个个体进行两项测量(两个变量)，就得到双变量数据。例如，对学生登记其身高和体重，这样得到的数据即是双变量数据。

- If we take only one measurement on each object, we get **univariate data**.

- With two measurements on each object, we get **bivariate data**.

对数据的展示有不同的方法，数据类型不同，往往其展示方法也会有所差别，数据的展示方法主要有 tabular methods(表格法)，graphical method(图示法)，以及 numerical method(数值法)。下面，我们从单变量分类数据和双变量分类数据分别讨论其展示方法。

Example 3:

Which of the following are examples of continuous data?

- I. the speed of a car
 - II. gender of persons
 - III. the temperature of New York
 - IV. the weight of a person
 - V. outcomes of a football game
- A. I, III and IV
 - B. II and V
 - C. I, III, and V
 - D. II, III, and IV
 - E. I, II, and IV

答案: A。

解析: I 汽车速度, III 纽约的气温, IV 人的体重都是 continuous variable。而 II 性别(男、女), V 足球比赛的结果(输, 平, 赢), 均是分类变量。因此本题选 A。

2.2 Analysis of univariate categorical data 单变量分类数据的分析

2.2.1 Frequency distribution table(频数分布表)

分类数据是对事物的一种分类。因此首先列出所分的类别，然后数出每一类别的频数，计算其相对频数、百分比等，即可形成一张频数分布表。

频数 (frequency) 是落在某一特定类别中的数据个数。

相对频数 (relative frequency) 是落在某一特定类别的数据个数除以样本总个数的值。

百分比 (percentage) 是相对频数乘以 100 得到的结果，用%表示。

- The **frequency** of a value is the number of times that observation occurs. Frequency is usually denoted by f .
- The **relative frequency** of a value is the ratio of the frequency to the total number of