

生物统计学-期中测试 Mid-term Exam

1. 假设零假设成立，以下哪个陈述是正确的？ [5 分]

- (1) 具有更大样本的研究相较于具有更小样本的研究更有可能获得 $P < 0.05$ 的结果。
- (2) 具有更大样本的研究相较于具有更小样本的研究更不太可能获得 $P < 0.05$ 的结果。
- (3) 具有更大样本的研究相较于具有更小样本的研究同样可能获得 $P < 0.05$ 的结果。

2. 对于以下每个研究，请说明哪个是解释变量，哪个是响应变量，以及说明这个研究是观测性的(observational)还是实验性的(experimental)。 [10 分]

- (1) 林业研究人员想比较高海拔地区生长的树木与低海拔地区生长的树木生长速度。他们使用从天然森林中砍伐的一组树木的树环之间的间距来测量生长速度。
- (2) 一种自然界中的蜘蛛的雌性经常吃掉试图与它们交配的雄性。研究人员进行了一项研究，将一组雄性蜘蛛进行切除一条腿的处理（使其更脆弱并容易被捕食），而另一组蜘蛛雄性则保持完整（未受伤），进而研究这两组雄性蜘蛛在交配期间的生存情况是否有差异。

3. 我们从人类基因组中得到一个随机抽样的 100 个基因的样本（如 Lecture04-uncertainty_and_probability 第 10 页），基因片段长度的中位数（median）为 2640.5 个核苷酸请指出以下各种说法正确还是错误。 [9 分]

- (1) 所有人类基因的中位基因长度为 2640.5 核苷酸。
- (2) 所有人类基因的中位基因长度估计为 2640.5 核苷酸。
- (3) 样本中位数具有具有标准误的抽样分布。

4. HIV 病毒与细菌和多细胞生命相比，具有较高的突变率。其中许多突变对病毒不利，导致其复制速度较慢。研究人员据此构建了一个庞大的已知 HIV 突变体及其复制速率的数据库。我们使用计算机从该数据库中随机抽取了 100 个突变体，计算了样本的复制速率的均值和中位数。我们多次重复了该过程，每次都计算了样本的中位数和均值。样本均值分布（均值的抽样分布）的标准差为 0.00073；样本中位数分布（中位数的抽样分布）的标准差为 0.0224。 [10 分]

- (1) 复制速率均值的标准误是多少？
- (2) 复制速率中位数的标准误是多少？

(3) 如果从数据库中随机抽取 100 个突变体的样本数据, 哪个中心趋势测量值可能更精确, 均值还是中位数? 请解释你的答案。

5. 一项在上海市闵行区进行的研究中, 研究者通过户籍登记确定了 500 个家庭的随机样本。他们共发出了 500 份问卷, 针对每个家庭的一名成年人调查了其对共享单车的态度。其中 80 份问卷被填写并寄回给研究人员。请问: [10 分]

(1) 那 80 个寄回问卷的家庭是否可以被视为家庭的随机样本? 请解释。

(2) 会影响调查结果的偏差类型是什么?

6. 弗朗西斯·高尔顿 (1894 年) 提供了以下关于至少飞行 90 英里(miles)的 3207 只信鸽的飞行速度的数据 (见下图; 横坐标(x)为速度, 单位是码/分钟 (yards/min), 纵坐标(y)为信鸽的频数, 单位通常为“个”, 故省略)。请回答: [15 分]

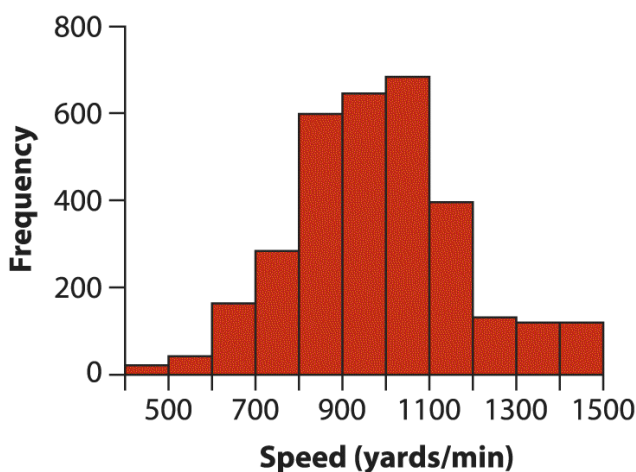
(1) 这是什么类型的图表?

(2) 检查图表并用肉眼直观地估计均值 (mean, 取到最接近的 100 码/分钟, 如 1000 码/分钟, 而不用精确到 1030 码/分钟)。解释如何获得这个估值。

(3) 检查图表并用肉眼直观地估计中位数的估值 (median, 取到最接近的 100 码/分钟)。解释如何获得这个估值。

(4) 检查图表并用肉眼直观地估计众数 (mode, 出现频数最多的值) 的大致值或区间 (取到最接近的 100 码/分钟)。解释如何获得这个估计。

(5) 检查图表并用肉眼直观地估计标准差的大致值 (standard deviation, SD, 取到最接近的 50 码/分钟)。解释如何获得这个估计 (提示: 95% 的观测值范围大致等于 $\text{mean} \pm 2\text{SD}$)。



7. 假设孟德尔的一个实验中有 1600 株豌豆植株，其中有 900 株是高植株并且具有绿色的豆荚，300 株是高植株并且具有黄色的豆荚，300 株是矮植株并且具有绿色的豆荚，还有 100 株是矮植株并且具有黄色的豆荚。[17 分]

(1) 对于这批植株，"高"和"绿色豆荚"是互斥的特征吗？

(2) 对于这批植株，"高"和"绿色豆荚"是独立的特征吗？（提示：可先分别计算出"高"和"绿"各自发生的概率）

8. 一项临床试验的目的是测试一种新疗法是否会影响患有严重疾病的患者的康复率。结果为零假设 H_0 : 治疗没有效果"被拒绝，P 值为 0.04。研究人员使用了显著性水平 $\alpha=0.05$ 。请说明以下各项结论是否正确。如果不正确，请解释原因。[24 分]

(1) 治疗只有很小的效应。

(2) 治疗有一些效应。

(3) 发生第一类错误的概率为 0.04。

(4) 发生第二类错误的概率为 0.04。

(5) 如果显著性水平设置为 $\alpha=0.01$ 而不是 0.05，零假设就不会被拒绝。