第 5 章

17、根据题中给出的数据及置信区间的估计式可得以下结论：(并未使用代码，纯计算)

$$\overline{x} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

a、n=25 时：

置信区间：[18.04,21.93]

b、n=100 时：

置信区间：[19.02,20.98]

c、n=400 时：

置信区间：[19.51,20.49]

18、因为均值已经在第 17 题中告诉而且计算出了对于不同 N 的总体均值的范围故直接计算误差范围：

a、21.96-20=1.96,故可得特殊用户子集的误差范围为 1.96。

b、20.98-20=0.98,故可得特殊用户子集的误差范围为 0.98。

c、20.49-20=0.49,故可得特殊用户子集的误差范围为 0.49。

19、

误差范围是对区间估计精确度的度量。数据误差的最大值与最小值之差，亦称为误差范围，它是描述一个数据不确定性的一种度量。误差幅度是一个统计，表示调查结果中随机抽样误差的数量。误差幅度越大，对民意测验结果反映全体人口调查结果的信心就小。当总体样本不完全且结果度量为正方差时，误差幅度将为正，即度量值变化。在近似计算时，必须注意误差范围，以达到所要求的准确程度。误差范围越小表明精确度越大。$\frac{s}{\sqrt{n}}$代表样本均值(x 采样分布的标准偏差)的标准误差，当样本容量很大或者样本可变性较小时，$\frac{s}{\sqrt{n}}$较小。因子 t_(a／2),与样本容量和分析师指定的置信水平(通常为 90-99%)相关，对于更低的置信区间，该值会更小。由于我们不能直接影响样本可变性并且不愿降低置信水平，我们必须致力于增加样本容量，试图提供更加精确的置信区间估计。通过增大样本容量。要想在减小误差范围的同时保持置信水平不变，增大样本容量是唯一的方法。

（1）绝对值相等的正的与负的误差出现机会相同；

（2）绝对值小的误差比绝对值大的误差出现的机会多；

（3）误差不会超出一定的范围。；-

20、

首先已有公式：

$$p \pm Z_{a/2}\sqrt{\frac{p\cdot(1-p)}{n}}$$

根据给出的数据代入公式计算即可：

a、置信水平：90%

比例置信区间计算结果为：[0.0843,0.1156]

b、置信水平：95%

比例置信区间计算结果为：[0.0814,0.1185]

c、置信水平：99%

比例置信区间计算结果为：[0.0756,0.1243]

21、误差范围与第 18 题计算方式相同可得：

$$p \pm Z_{a/2}\sqrt{\frac{p\cdot(1-p)}{n}}$$

a、置信水平：90%

误差范围计算结果为：0.0156

b、置信水平：95%

误差范围计算结果为：0.0185

c、置信水平：99%

误差范围计算结果为：0.0243

22、

答：置信区间是指由样本统计量所构造的总体参数的估计区间。在统计学中，一个概率样本的置信区间（Confidence interval）是对这个样本的某个总体参数的区间估计。置信区间展现的是这个参数的真实值有一定概率落在测量结果的周围的程度，其给出的是被测量参数的测量值的可信程度，即前面所要求的“一个概率”。置信区间越大代表误差范围也就越大，当样本数量一定时，减小误差范围的方法是减小置信区间的大小。

置信区间可以用样本（或重复样本）来表示："如果此过程在多个样本上重复，则包含真实总体参数的计算置信区间（每个样本会有所不同）的分数将趋于 90%。 [4]

置信区间可以用单个样本表示："未来某些实验中计算的置信区间包含总体参数的真实值的概率为 90%。请注意，这是有关置信区间的概率陈述，而不是总体参数。这考虑了从实验前角度与置信区间相关的概率，在相同的上下文中，随机分配处理以研究项目。在这里，实验者列出了他们打算计算置信区间的方式，并在他们进行实际实验之前知道，他们最终计算的间隔具有覆盖真实但未知值的特定机会。 [5]

这与上述"重复抽样"解释非常相似，只不过它避免依赖考虑假设的重

复抽样程序，而该重复可能没有任何有意义的重复。见尼曼建设。

置信区间的解释可以达到类似："置信区间表示总体参数的值，其中参数和观测估计值之间的差值在 10% 级别上不具有统计显著性"。[6]这种解释在使用置信区间来验证其实验的科学文章中很常见，尽管对置信区间的过度也会引起问题。

23、

a、

假设一：捐助者捐赠金额超过 50 美元。

假设二：捐助者捐赠金额不超过 50 美元。

b、

拒绝法则是捐助者捐赠金额是否超过 50 美金。

c、

t_data 的值可理解为在假设的均值 u 之上或之下的标准误差数目，样本均值$\bar{x}$，其中标准误差等于$\frac{s}{\sqrt{n}}$ (粗略地讲，标准误差表示统计量分布的分散程度度量)。

d、

当 t_data 值为极值时，这表明一种零假设(伴随假设值 H)和观测数据之间的冲突。由于数据表示经验证据而零假设仅仅表示一种断言，因此解决这样的冲突有利于数据，因此，当 t_data 为极值时，零假设 H,是拒绝的。

e、

p-值是指:如果我们假定捐助者捐赠金额超过 50 美金为真时，观测样

本统计量至少与真实观测的统计量一样有极端的概率。

f、

```
> tdata=(55-50)/(25/100^0.5)
> 2*(1-pnorm(tdata))
[1] 0.04550026
```

g、

随着样本数量的增加，捐赠金额的均值将不会减小，并且会小于 50 美金。