区间估计和假设检验

最小二乘估计给出的是回归模型参数的点估计（point estimates），表示对回归函数 $E(y)=\beta_1+\beta_2 x$ 的推断（inference）。这里" 推断 ``的意思是，从已知或假定出发，经过推理得出结论。区间估计（interval estimation）和假设检验（hypothesis testing）是统计推断的两个重要工具。IE 和 HT 依赖假设 SR6。如果 SR6 不成立，则当样本容量足够大，可以得到近似结果。

区间估计

当 SR1-SR6 成立时，b_2 服从正态分布，对其进行标准化：

$$Z=\frac{b_2-\beta_2}{\sqrt{\sigma^2/\sum(x_i-\bar{x})^2}}\sim N(0,1)$$

由正态分布表可知，$P(-1.96\leq Z\leq 1.96)=0.95$，从而有：

$$P\left(b_2-1.96\sqrt{\sigma^2\Big/\sum(x_i-\bar{x})^2}\leq\beta_2\leq b_2+1.96\sqrt{\sigma^2\Big/\sum(x_i-\bar{x})^2}\right)=0.95$$

上式定义了一个在 0.95 概率下包含参数 β_2 的区间。其中的两个端点给出了区间估计量（interval estimator）。在重复抽样中，这样构造的区间中有 95% 将会包含真实参数 β_2。这个估计量依赖于未知的误差方差，需要用其估计量代替，但将 σ^2 换成 $\hat{\sigma}^2$ 后，上面的正态分布就变成了具有 2 个自由度的 t 分布：

$$t=\frac{b_2-\beta_2}{\sqrt{\hat{\sigma}^2/\sum(x_i-\bar{x})^2}}=\frac{b_2-\beta_2}{\sqrt{\widehat{\operatorname{var}(b_2)}}}=\frac{b_2-\beta_2}{\operatorname{se}(b_2)}\sim t_{(N-2)}$$

b_1 也是如此。一般意义下，对于满足假定 SR1-SR6 的简单线性回归模型，我们有：

$$t=\frac{b_k-\beta_k}{\operatorname{se}(b_k)}\sim t_{(N-2)},\quad k=1,2$$

上式是简单线性回归模型进行区间估计和假设检验的基础。

t 分布的临界值 t_c 满足：$P(t\geq t_c)=P(t\leq -t_c)=\alpha/2$，其中 α 是概率，通常取 0.01 或 0.05。自由度为 m 的临界值 t_c 即分位数 $t_{(1-\alpha/2,m)}$。容易发现，$P(-t_c\leq t\leq t_c)=1-\alpha$。因此，区间估计可以如下构造：

$$p\left[-t_c\leq\frac{b_k-\beta_k}{\operatorname{se}(b_k)}\leq t_c\right]=1-\alpha$$

从而，

$$P[b_k-t_c\operatorname{se}(b_k)\leq\beta_k\leq b_k+t_c\operatorname{se}(b_k)]=1-\alpha$$

区间端点 $b_k\pm t_c\operatorname{se}(b_k)$ 是随机变量，它们定义了 β_k 的一个区间估计量。当基于某个样本得到 b_k 和 $\operatorname{se}(b_k)$，则得到 β_k 的一个区间估计值，称为 $100(1-\alpha)\%$ 区间估计值（interval estimate），也叫 $100(1-\alpha)\%$ 置信区间（confidence interval）。

假设检验

所谓假设检验，就是将我们对总体特征的猜测与数据样本信息作比较。给定一个经济和统计模型，我们设定关于经济行为的假设，再将这些假设用模型参数来表达。假设检验包括五个要素：

(1) 原假设（null hypothesis）H_0
(2) 备择假设（alternative hypothesis）H_1
(3) 检验统计量（test statistic）
(4) 拒绝域（rejection region）
(5) 结论（conclusion）

对于原假设 $\beta_k = c$，可能有三种备择假设：

- $H_1 : \beta_k > c$。这种形式在经济学中比较常用，因为经济理论往往给出变量之间的关系（符号），比如，食物支出例子中，经济理论认为食物是正常品（normal goods），即 $\beta_2 > 0$，因此，备择假设常设为 $H_1 : \beta_2 > 0$。
- $H_1 : \beta_k < c$。
- $H_1 : \beta_k \neq c$。

与原假设有关的样本信息包含在检验统计量的样本值中。基于统计量的值，我们决定是否拒绝原假设。检验统计量的一个特性是，当原假设成立时，检验统计量的分布是完全清楚的，而当原假设不成立时，它可能具有其他分布。对于我们现在的模型，检验统计量服从 t 分布。当 $H_0 : \beta_k = c$ 成立，我们可以用 c 替代 β_k，从而：

$$t = \frac{b_k - c}{\mathrm{se}(b_k)} \sim t_{(N-2)}$$

拒绝域与备择假设的形式有关。构造拒绝域，除了检验统计量，还必须知道备择假设和显著性水平（level of significance）。拒绝域包含那些不太可能的值，也就是当原假设成立时，这些值出现的概率很小。其中的逻辑链可以表述为："如果检验统计量的值落入拒绝域，则该统计量不太可能服从假定的分布，从而原假设也不太可能为真``。当备择假设成立时，检验统计量的值会不同寻常的大或不同寻常的小（这里的大小取决于显著性水平，一般取 0.01，0.05 或 0.10）。

如果原假设确实为真，而我们拒绝了它，称为第一类错误（Type I error）。犯第一类错误的概率就等于显著性水平，即 $P(\text{Type I error}) = \alpha$。只要拒绝原假设，就会不可避免犯第一类错误，当然如果这种错误成本比较高，可以通过减小 α 来控制。另一方面，如果原假设为假，而我们并未拒绝它，则称为第二类错误（Type II error）。在实际运用中，我们无法控制或计算犯第二类错误的概率，因为它依赖于未知参数 β_k。

假设检验完成后，我们可以给出结论。注意，不要用``接受某假设``的说法，只能说拒绝或者不拒绝原假设。无法拒绝原假设并不意味着原假设为真！

具体地，三种备择假设对应的拒绝域构造如下。

- 单侧检验，备择假设为 ``>": 拒绝域为 $t \geq t_{(1-\alpha,N-2)}$
- 单侧检验，备择假设为 ``<``: 拒绝域为 $t \leq t_{(\alpha,N-2)}$
- 双侧检验，备择假设为 ``≠": 拒绝域为 $t \leq t_{(\alpha/2,N-2)}$ 或 $t \geq t_{(1-\alpha/2,N-2)}$

总结一下，假设检验的标准步骤为：

1. 设定原假设和备择假设
2. 设定统计量及其分布（当原假设成立时）
3. 选择一个显著性水平，决定拒绝域
4. 计算统计量的样本值
5. 给出结论

以食物支出为例。我们想知道到底 β_2 是不是与 0 有显著差异，即显著性检验。步骤如下：

1. $H_0 : \beta_2 = 0, H_1 : \beta > 0$
2. 统计量为前面的 t 统计量，本例中 $c = 0$，因此，$t = \frac{b}{\text{se}(b_2)} \sim t_{(N-2)}$
3. 取 $\alpha = 0.05$。右侧拒绝域的临界值为 95% 分位数：$t_{(0.95,38)} = 1.686$，因此拒绝域为 $t \geq 1.686$。
4. 计算统计量的样本值，由 $b_2 = 10.21, \text{se} = 2.09$，得到 $t = 4.88$
5. 由于 $t = 4.88 > 1.686$，所以我们拒绝原假设，也就是认为收入和食物支出之间存在统计上显著的正相关关系。

实践中，常常用 ***p*-value** 来表达检验结果。规则是，如果 p 值小于或等于显著性水平，则拒绝原假设。