SKKU Sociology QE Statistic Note

Version 2.0

by Jm Su

https://blog.jmsu.me

Published 2024

Contents

1 基本概念

1.1 样本误差 (표본오차) 与标准误差 (표준오차 SE)

1.1.1 样本误差

由于样本数量一定小于总体数量，因此样本一定无法完全表示总体的实际参数，样本参数与总体参数之间的差异，即样本误差。

样本误差主要分为如下两类:
1. Bias 偏向 2. Chance 偶然

有偏向的误差即样本与总体分布间呈现系统性的，有偏向的误差；而偶然性误差发生在所有类型的样本抽样中，无法避免。

1.1.2 标准误差

为了衡量样本参数与总体参数间的可能差异，可以利用标准误差公式计算

标准误差公式

已知总体方差 σ_Y 时

$$\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}}$$

总体方差未知时，利用样本方差代替

$$s_{\bar{y}} = \frac{s_y}{\sqrt{n}}$$

即可以理解为总体的方差 σ 除以样本总数 n 后对其开平方根

1.2 统计决定错误 (Statistical Decision Error)

通过样本 (표본) 数据对总体 (모집단) 数据进行推断时，一定会存在不可避免的，发生推断错误的概率，为了让统计检验过程更加的可控，通过如下两种错误来表示，或控制错误发生的概率，以试图得到更准确的检验结果，也使统计决策错误可解释。

1.2.1 一类与二类错误

Type 1 Error & Type 2 Error

Type 1 Error: 当 H_0 实际为 $True$ 但是统计检验结果为 $False$ (弃真)
Type 2 Error: 当 H_0 实际为 $False$ 但是统计检验结果为 $True$ (择假)

Type 1 Error 也称为 α $Error$ 因为我们使用 α 代表发生此类弃真事件的概率
Type 2 Error 也称为 β $Error$ 因为我们使用 β 代表发生此类择假事件的概率
* 需要注意的是，此处的' 弃真',' 择假' 都是针对于 H_0 而言的，请不要产生误区

相比于第二类错误 β $Error$, α $Error$ 更常在应用统计学 (~~응용통계학~~) 中被讨论 (样本与总体间无差异，但检验结果显示有差异的弃真概率)，而 β $Error$ 更常在对统计学理论的开发 (~~순수통계학~~) 中讨论，因为其代表着统计检验方法的准确性 (样本与总体间有差异或处理效果，但统计方法没有准确识别而是择假的概率)。

1.2.2 可信区间

由于统计决定错误 (Statistical Decision Error) 的存在，样本对总体参数的估计一定不是毫无偏差的，而为了解释，阐明在某置信水平下 (CI) 这种误差的范围 (폭;width)，采用称之为' **置信区间'** 的区间来表示 **该区间内包含真实总体参数的概率**。

置信区间公式

$$(1-\alpha)CI = \bar{Y} \pm Z_{c.v} \cdot \sigma_{\bar{y}} = \bar{Y} \pm Z_{c.v} \cdot \frac{\sigma_y}{\sqrt{N}}$$

α: 研究者设定的一类错误发生概率，一般为 0.05 即发生一类错误的概率为 5%，对应可信区间即 95%CI，代表的是在重复取样中，有 95% 的置信区间包含了真正的总体参数
$\bar{Y}$: 样本平均
$Z_{c.v}$: 检验统计值，此处为 Z 分布
$\sigma_{\bar{y}}$: 样本对于总体的**标准误差**

在对置信区间的计算中，α 为研究者依据对具体研究的精度要求事先确定，$\bar{Y}$ 为样本的平均值，其是固定的数值，$Z_{c.v}$ 为固定的分布临界值，因此只有标准误差 $\sigma_{\bar{y}}$ 是唯一可变的，影响置信区间幅度范围 (폭;width) 的参数，具体来说，即 N 的可变性，当 N 越大，标准误差相对越小，所对应的与分布统计值的乘积也越小，因此整体置信区间的幅度范围也将变窄。

而置信区间的宽幅越窄，其对总体参数的估计也就越精确。

1.3 同方差假定

同方差假定指的是: 误差的发生仅为偶然 (Chance) 情况，误差与独立变量没有系统性的相关关系。

同方差假定

$$Var(Y_i|X_i) = \sigma^2$$

即对于每个自 (独立) 变量 X_i 其对应的因变量 Y_i 的方差是固定值 σ^2，其不随自 (独立) 变量的变化而变化。

1.3.1 关于 t-检验的同方差假定

对于比较两个独立样本的均值差异，t-检验的计算方法如下

独立样本 t-检验

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p/\sqrt{n_1 + n_2}}$$

s_p: 指将两个样本集团的合并方差

可以看出，对两个独立样本进行 t-检验时，其比较的是两个独立样本的均值差异，并除以用两个独立样本的合并方差计算的 标准误差。

合并方差 s_p

$$s_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

即对两个样本集团的方差进行加权平均，以此推断**共同的总体方差**

由此可以看出，独立样本 t-检验是建立在两个样本的方差是**属于同一总体的**，因为其的标准误差计算是基于依据对两个样本各自方差加权平均估计出的**共同总体方差**得出的，换一个角度也就是说，其假设两个样本集团是**同方差**的

独立样本 t-检验的同方差假定

$$Var(Y_1|X_1) = Var(Y_2|X_2) = \sigma^2$$

即同方差假定的拆分后的形式，独立样本 t-检验的自变量为分类变量，对应着两个样本集团 X_1,X_2。即两个样本集团的方差相同，都为 σ^2

1.3.2 关于 OLS 线性回归的同方差假定

对于 OLS 来说，同方差假定意味着模型拟合仅存在偶然误差，**模型不存在未被系统解释的部分**.

OLS 线性回归中同方差的含义

$$Var(Y_i|X_i) = \sigma^2 \rightarrow E(\varepsilon|X) = 0$$

ε 指的是拟合值与实际观测值间的残差

其含义是，同方差假定意味着模型残差的期望为 0，因为模型残差的正负相互抵消。

而当线性模型不符合同方差假定时，意味着残差期望值不为 0，并且残差是随 X 的变化而变化的，这意味着基于现有自变量 X，无法完全解释因变量 Y 的某些特性 (部分)。有可能是自变量与因变量的关系无法仅通过线性关系解释，也有可能是某些关键自变量的遗漏所导致的.

1.3.3 关于 F-检验的同方差假定

F-检验的原理是，求各集团平均组间方差 ($MS_{betweeness}$) 和与平均组内方差 (MS_{within}) 的比率，以此判断处理效果的显著性。因此 F-检验的前提条件是，各组应该满足同方差假定，如果不满足，其导致了在计算方差时，MS_{within} 平均组间方差可能由于某组的方差影响显著增大，导致统计结果的不准确，即 F 值变小，提高了 β 错误的发生概率。

1.4 无偏估计与有偏估计

无偏估计是指样本的统计参数与总体的统计参数相拥有相同期待的情况。

无偏估计

$$E(\hat{\theta}) - \theta = 0$$

$\hat{\theta}$ 为样本统计值

1.4.1 样本平均-无偏估计

样本的平均计算是无偏估计的，因为

样本平均的期待

$$E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{n}x_i) = \frac{1}{n}\sum_{i=1}^{n}E(x_i) = \frac{1}{n}n \cdot \mu = \mu$$

由于 $E(\bar{X}) = \mu$，因此样本平均是无偏估计的

1.4.2 (通过样本平均计算的) 样本方差-有偏估计

当使用总体平均计算的样本方差是无偏的，但是使用样本平均计算的样本方差是对总体的有偏估计.

通过样本平均计算的样本方差

$$E(s_n^2) = E(\frac{\sum_{i=0}^{n}(x_i - \bar{x})^2}{n})$$

由于 $\bar{x}$ 与实际总体 μ 之间有误差，当我们将 $\bar{x}$ 与 μ 之间的误差在方差公式中联系起来，将得到

$$E(s_n^2) = E(\frac{\sum_{i=0}^{n}((x_i - \mu) - (\bar{x} - \mu))^2}{n})$$

依据上式推倒后，得到

$$E(s_n^2) = \frac{n-1}{n}\sigma^2$$

因此我们可以得出，$E(s_n^2) \neq \sigma^2$，因此样本方差对于总体是有偏估计的.

为了消除有偏估计产生的偏差，样本方差计算时，可以使用如下调整过的 (adjusted) 方差公式计算，其代表着样本方差对总体方差的**无偏估计**

调整后的样本方差公式

$$E(s_n^2) = \frac{\sum_{i=0}^{n}(x_i - \bar{x})^2}{n-1}$$

$\frac{1}{n-1} \cdot \frac{n-1}{n} = \frac{1}{n}$ 其消除了有偏部分，使样本方差的计算成为无偏

从自由度 (df) 的角度来看，当我们抽取样本时，只有 $n-1$ 个样本是自由选择的，最后一个样本受总体平均期望的限制，其是固定取值，因此需要被排除自样本 n。

2 t-检验 (t-검증)

t 检验的对应变量类型应该是X: 定类变量 对Y: 连续型变量，因为其验证的是两个集团间的平均的差异

2.1 t-检验方法

t-수치

$$t = \frac{\bar{Y} - \mu_y}{\sigma_{\bar{y}}} = \frac{\bar{Y} - \mu_y}{s_y/\sqrt{N}}$$

$\bar{Y}$: 표본평균（样本平均）
μ_y: 모집단평균（总体平均）(期望)
$\sigma_{\bar{y}}$: **표준오차 (标准误差)** 指每次抽样均值与总体均值 (期望) 间的差异 (的标准化)

s_y: **표본표준차 (标本标准差)** 指每个抽样观测值与样本均值 间的差异 (的标准化)
N: 표본수 (标本数量)

t 检验用于检测单一平均或两集团平均差异
检验原理是，通过 t 数值算样本与总体间的（用标准误差衡量）偏差是多少，将 t 值带入 t 分布与 $t_{c.v.}$(拒绝值) 进行比较，$t_{c.v.}$ 依据三个要素决定：

$t_{c.v.}$ 의영향요소

1. 자유도 (df)(自由度)
2. 단측 or 양측검증 (单，双侧检验)
3. α 수준

- **自由度 (df)**: 指得是样本的自由程度，在 t 检验中 df 为 N-1，因为最后一个决定的样本被之前的 N-1 个样本确定了他可能的取值范围，所以自由度为 N-1.
- **单双侧检验**: 根据假设的类型,$t_{c.v.}$ 在分布中的位置
- α **水准**: 即人为设定的弃真概率 (Type 1 Error)

最后将 t-수치与 $t_{c.v.}$ 比较后，得出是否拒绝 H_0 假设.

2.2 t 与 z 的差异

功能上，z 只可用于单一平均检验，t 可用于单一平均检验和两集团平均检验.

z-수치

$$z = \frac{\bar{Y} - \mu_y}{\sigma_{\bar{y}}} = \frac{\bar{Y} - \mu_y}{\sigma_y / \sqrt{N}}$$

计算方法上, z 和 t 唯一的区别是对于 $\sigma_{\bar{y}}$ 标准误差计算上的差异，z 利用总体 (모집단) 的 (σ_y) 标准差计算标准误差，但是总体标准差一般情况下很难得到，因此 t 采用了 (s_y) 样本标准差替代总体标准差.

与直觉相反的是, t 相比于 z 的优势在对于**小样本**的检验上依旧存在, 因为根据中心极限定理，小样本的情况，样本抽样很可能不符合正态分布，因此 z 无法使用, 并且 t 因为使用了样本标准差作为计算值, 所以反而受抽样偏差影响更小.

在大样本的情况下，t 也具有相对 z 的优势，即可以不考虑总体是否符合正态分布，因为抽样自动实现了正态分布 (中心极限定理).

2.3 关于单双侧检验

单侧检验指拒绝值在概率分布的一侧, 即假设为 $H_0 <?$ 或 $H_0 >?$ 的情况
双侧检验指拒绝值在概率分布的两侧，即假设为 $H_0 =?$ 或 $H_0 \neq ?$

3 卡方检验 (x^2-검증)

x^2 检验 (Chi-square) 的对应变量类型应该是X: 定类变量 对Y: 定类变量，其通过判断集团观测频数与其期望频数间的差异而验证两个或多个集团间的**关联性**.

3.1 卡方检验方法

如上所述，卡方检验通过判断频数 (빈도, Frequencies) 差异来确定集团间的关联性, 所以 x^2-수치的计算需要先输出变量的交叉表 (교차분석표, Cross-tabulation), 交叉表会显示出各变量的**观测频数** (f_{ij}), 并在表行 or 列的边界显示出行计和列计 (Marginals). 根据行计和列计, 我们可以计算出**期望频数** ($\hat{f_{ij}}$), 通过观测频数和期望频数完成对 x^2 的计算, 如下:

기대빈도

$$\hat{f_{ij}} = \frac{(f_{i\cdot})(f_{\cdot j})}{N}$$

f_{ij}: 관찰빈도 (Observed Cell Frequencies)
$f_{i\cdot}$: 观测频度的**行计**
$f_{\cdot j}$: 观测频度的**列计**
$\hat{f_{ij}}$: 기대빈도 (Expected Cell Frequencies)

x^2-수치

$$x^2 = \sum_{i=1}^{R}\sum_{j=1}^{C}\frac{(\hat{f_{ij}} - f_{ij})^2}{\hat{f_{ij}}}$$

R: Row 指行
C: Col 指列
对于公式，其加和了各个格子 (房)(Cell) 的观测频数与期待频数间的差，并通过自身平方并除以期待频数去掉了 $\pm$ 号并标准化，这意味着 x^2 统计量越大，观测频数与期待频数间的差异也就越大.

换一种方式来说，在自由度 (df),α 水准相同的情况下，x^2 越大，越有可能拒绝 H_0 假设.

자유도 (df)

$$df = (R-1)(C-1)$$

注. 对于 x^2 分布，N >=50 展现其分布的稳定性，因此 x^2 检验最好适用于大样本

4 ANOVA 분석 (F-statistic)

ANOVA 区别与 t-检验和 z-检验，它可以实现两个及其以上集团间的平均差异比较, 变量类型为 X: 定类变量 Y: 连续型变量，假设如下

4.1 假设陈述

对于 ANOVA 分析，其包含了不同的解释方法和检验思路

ANOVA 가설진술방식 1

$H_0 : \mu_j = \mu \ for \ all \ j$
$H_1 : \mu_j \neq \mu for \ some \ j$

μ_j: 集团 j 的平均

以对比平均差异的方式建立假设，即零假设为：所有集团的平均相同

ANOVA 가설진술방식 2

$H_0 : a_j = 0 \ for \ all \ j$
$H_1 : a_j \neq 0 \ for \ some \ j$

a_j: 集团 j 的处理效果 (처리효과)(Population treatment effect)

以对比处理效果的方式建立假设

4.1.1 什么是处理效果 (Treatment effect)？

(集团) 处理效果：集团平均与总平均的差异

$$a_j = \bar{Y_j} - \bar{Y}$$

一元变量的一般分析模型为

$$Y_{ij} = \mu + a_j + \varepsilon_{ij}$$

即，集团内的个体值由 总平均 +(集团) 处理效果 + 随机误差决定

将以上公式合并，得出的一般模型为

$$Y_{ij} = \bar{Y} + (\bar{Y_j} - \bar{Y}) + \varepsilon_{ij}$$

4.1.2 组内，组间均方

ANOVA 实际是通过 F-检验确定组内，组间均方的比值确定处理效果的显著与否。三种假设虽然各有不同，但是只是解释视角的差异，本假设是最贴近 ANOVA 的实际计算方法的假设.

ANOVA 가설진술방식 3

$H_0 : MS_B <= MS_W$
$H_1 : MS_B > MS_W$

MS_B: 组间均方 (집단간평균자승)
MS_W: 组内均方 (집단내평균자승)

以对比组间, 组内均方的方式建立假设

4.2 F-检验 (F-검증)

4.2.1 通过组内，组间均方度量处理效果和随机误差

所谓均方，指平方和 (자승합)(Sum of squares)(SS) 的平均，所以均方的计算基于平方和的计算，如下

자승합 SS

$$SS_{Total} = \sum_{j=1}^{J}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y})$$

$$SS_W = \sum_{j=1}^{J}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_j)$$

$$SS_B = \sum_{j=1}^{J}\sum_{i=1}^{n_j}(\bar{Y}_j - \bar{Y})$$

$$SS_{Total} = SS_B + SS_W$$

SS_T: 总平方和
SS_W: 组内平方和 SS_B: 组间平方和

为何说平方和是处理效果和随机误差的度量呢，公式中能看到组间平方和是通过计算各组平均 和总体平均间 的差异，并将其平方加和. 同处理效果一样, 表示的是组间的差异性.
而如果对上节的一元一般变量模型公式进行左右互换，消除可得到：随机误差(ε_{ij}) 的计算方法为 $Y_{ij} - \bar{Y}_j$ 同组内平方和一样.

集团间的差异显著性即,处理效果大于随机误差 (同假设.2),而具体实现是通过比较组间平均和组内平均实现的,如果组间平方和 (均方) 显著大于组内平方和 (均方), 那么可以判定：集团间是存在显著变异性的，即平均存在显著差异 (同假设.1).

4.2.2 F 值的计算

具体通过 **F-分布检验**显著性，如下

F-검증

$$MS_B = SS_B/df = SS_B/J-1$$
$$MS_W = SS_W/df = SS_W/N-J$$
$$F = MS_B/MS_W$$

MS 为均方，在 3.12 节已经定义，F 值即组间均方和组内均方的比值, 也就是说组间均方越大，组内均方越小的情况下，F 值越大，集团间的显著性也越大.

4.3 ANOVA 的不足

如果仔细观察 3.1 节的三种假设，可以发现 ANOVA 并不能测定

- 哪些集团间存在差异
- 处理效应大小的评估

因此需要其他统计值和事后比较来补足

4.3.1 评估处理效应的大小: η^2

η^2

$$\eta^2 = \frac{SS_B}{SS_T} = 1 - \frac{SS_W}{SS_T}$$

取值范围为 0～1，指自变量对因变量处理效应的百分比，0 即自变量对因变量没有影响 (完全独立).
一元 ANOVA 的情况下，取值为 0.2(20%) 已经是不错的结果.

4.3.2 事后比较 (사후비교)(post hoc comparisons)

通过事后比较确定哪些集团间存在差异.
(사회통계학 -김상옥 p.142)

5 Linear Regression

线性回归假设自变量 X 到因变量 Y 的因果线性关系,变量类型为X：连续型变量 Y：连续型变量 . 一元线性回归适用于单一 X，多元线性回归适用于多 X（多自变量）.

5.1 线性回归模型

관찰값 (Y_i) 모델

$$Y_i = a + b_{yx}X_i + e_i$$

Y_i: 관찰값 (观测值)
a: 절편 (截距)
b_yx: 회귀계수 (回归系数)
e_i: 무작위오차 (随机误差)(Random errors)**(residual[残差])**
X_i: 自变量取值

观测值指实际观测到的 Y 的取值; **截距**是线性回归方程中的常数之一，指方程距离原点 0 的偏离;

回归系数是线性回归方程中的常数, 指自变量 X 对因变量 Y 的影响权重, 线性回归的核心就是估计出此权重，完成线性方程; **随机误差**在线性回归中也叫残差，指预测值与观测值间的距离 (差异).

예측값 ($\hat{Y}_i$) 모델

$$\hat{Y}_i = a + b_{yx}X_i$$

$\hat{Y}_i$: 线性模型预测值
与上面的观测模型比较, 可以看出预测模型去掉了了随机误差，因此也可得出随机误差等于

$$e_i = Y_i - \hat{Y}_i$$

这也很好解释，因为真实观测值-模型预测值，即误差，也叫残差.

5.2 计算线性回归常数

회귀계수 b_{yx}

$$b_{yx} = \frac{s_{xy}}{s_x^2}$$

s_{xy}: xy 의 공변량 (协方差)
s_x^2: x 의 변량 (方差)

절편 a

$$a = \bar{Y} - b_{yx} \cdot \bar{X}$$

通过以上公式计算出的回归直线叫做最小二乘回归线，其代表着这条直线拥有最小的随机误差 (e)(残差)，即公式 1最小化.

$$\sum_{i=1}^{n} e_i^2 \tag{1}$$

5.3 验证总体

모집단모델

$$Y_i = a + \beta_{yx} X_i + e_i$$

$b_y x$: 회귀계수 (样本回归系数)
$\beta_y x$: 모집단 회귀계수 (总体的回归系数)

上述计算的所有回归统计量都是基于样本变量 计算得出的，而其对于总体的显著性如何？这个问题需要通过检验过程得出.

5.3.1 线性回归和 ANOVA

线性回归和 ANOVA 有很大的相似，ANOVA 中的总平方和 (SS_T) 等于组内平方和 (SS_W) 加上组间平方和 (SS_B)，其表示了因变量 (Y_i)，因处理效果 (a) 加上随机误差 (ε) 对于其平均的总变异性.

对于线性回归模型，其因变量类型为连续型，因此其统计名称与 ANOVA 的分组平方和不同，主要使用如下统计名称

자승합 (SS) in Linear Regression

$$SS_T = SS_{Error} + SS_{Regression}$$

$$SS_{Error} = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)$$

$$SS_{Regression} = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y}_i)$$

SS_{Error}: 指预测值与观测值间的差异平方和，即残差平方和，简称 SS_E
$SS_{Regression}$: 指预测值与观测平均值间的差异平方和, 简称 SS_R

可以看出，SS_E 和 SS_R等同于 ANOVA 的SS_W 和 SS_B，即随机误差和处理效果.

5.3.2 R^2 决定系数

在 ANOVA 中，η^2 用于测量自变量 X 对因变量 Y 的处理效应百分比，其使用 SS_B/SS_T 计算得出，即组间平方和占全体平方和的比例.

线性回归中，计算处理效果的方法与 ANOVA 相同，只要求出回归平方和占全体平方和 的**比例**就可以了，即:

R^2 결정계수 (Coefficient of Determination)

$$R_{yx}^2 = \frac{SS_R}{SS_T}$$

R^2 的取值范围为 0～1，表示因变量的总变异中可以被回归模型解释的比例.

附. 对 R^2 开根号，可以得出 X 与 Y 的相关系数 ($r_x y$)，其取值为-1～1, 可表示负相关或正相关关系.

5.3.3 t-检验 (检验回归系数 b_{yx} 的显著程度)

b_{yx} 에 대한 t-검증

$$t = \frac{b_{yx} - \beta_{yx}}{SE(b)}$$

零假设为总体回归参数为 0，即 XY 无线性关系，因此 $\beta_{yx} = 0$.
$SE(b)$: b 的标准误差 (计算公式参照 p.158 공식 11-6)

6 Logistic Regression

逻辑回归作为线性回归的变种，适用于X: 定类或连续型变量 Y: 定类变量，利用线性模型拟合因变量发生的概率.

6.1 逻辑回归模型

逻辑回归通过对线性模型进行 Sigmoid 变换得到可预测二分类概率的曲线模型, sigmoid 模型如下

Sigmoid model

$$S(x) = \frac{1}{1+e^{-x}}$$

e: 自然常数

将线性模型传入 Sigmoid 函数进行转换，即得到了 logistic 模型

Logistic model

$$\hat{Y} = \frac{1}{1+e^{-(a+b_{yx}X_i)}}$$

$\hat{Y}$: 事件发生的概率，取值范围 0～1

可以看出 logistic 模型实际上就是将 linear regression 的输出值通过 sigmoid 函数转换为 0～1 区间的" 概率"，因此其优势是可以依据线性模型进行解释.

logistic 模型的线性表示

$$log(\frac{\hat{Y}}{1-\hat{Y}}) = a + b_{yx}X_i$$

$\hat{Y}$ 代表事件发生的概率，那么与之相对 $1-\hat{Y}$ 代表事件不发生的概率

6.2 Odd ratio

Odd ratio 指赔率比，即事件发生概率和事件不发生概率间的比值, 其取值范围为 $0 \sim \infty$.
如上 logistics 模型线性表示公式能看出，线性回归模型对应的结果正是对数后的赔率比，即 log(Odd ratio), 因此在解释自变量 X_i 对因变量 Y_i 的影响时，其实是赔率比的变化.

6.3 对回归系数 b 的解释

在解释时，一般对 $exp(b)$ 进行解释，因为其消除了 log 函数，因此 $exp(b)$ 的值即赔率比 (Odd ratio). 可以将其解释为：当自变量增大 1 个单位, 事件发生的概率是事件不发生概率的? 倍.

即当 $exp(b) < 1$ 时，自变量的增加与因变量发生概率间成负关系；$exp(b) > 1$, 自变量的增加与因变量发生概率间成正关系；当 $exp(b) = 1$，即自变量和因变量之间无关联.

7 附表

7.1 T.1 各统计方法的变量适用类型

各统计方法适用类型表

X \ Y	定类型	连续型
定类型	X^2-검증 로직회귀	z,t-검증 ANOVA(F-검증)
连续型	로직회귀	선형회귀

7.2 版本内容

7.2.1 version 2.0 (2024)

- 新增了关于논자시关联的相关概念内容
- 修正了翻译不妥当的部分

7.2.2 version 1.1 (2023)

- 更新了对应名词的韩中翻译