lecture2&3

Vincent

2023-02-06

Statistics

Basic components

mean: $E(Y)$

variance: $E(Y-\mu_Y)^2$

skewness: $\frac{E[(Y-\mu_Y)^3]}{\sigma_Y^3}$

kurtosis: $\frac{E[(Y-\mu_Y)^4]}{\sigma_Y^4}$

covariance: $cov(X,Z)=E[(X-\mu_X)(Z-\mu_Z)]=\sigma_{XZ}$ **correlation:** $cor(X,Z)=\frac{cov(X,Z)}{\sqrt{var(X)var(Z)}}=\frac{\sigma_{xz}}{\sigma_x\sigma_z}=r_{xz}$

Two sample testing

Estimator

$\bar{Y}_s-\bar{Y}_l=\frac{1}{n_s}\Sigma_{i=1}^{n_s}Y_i-\frac{1}{n_l}\Sigma_{i=1}^{n_l}Y_i$

Example:

C1 C2

$$\begin{aligned}\bar{Y}_{\text{small}}-\bar{Y}_{\text{large}} &= \frac{1}{n_{\text{small}}}\sum_{i=1}^{n_{\text{small}}}Y_i-\frac{1}{n_{\text{large}}}\sum_{i=1}^{n_{\text{large}}}Y_i\\ &= 657.4-650.0\\ &= 7.4\end{aligned}$$

Hypothesis testing

t-statistics:
Denote: s and l as small and large

$t=\frac{\bar{Y}_s-\bar{Y}_l}{\sqrt{\frac{s_s^2}{n_s}+\frac{s_l^2}{n_l}}}=\frac{\bar{Y}_s-\bar{Y}_l}{SE(\bar{Y}_s-\bar{Y}_l)}$

$s_s^2 = \frac{1}{n_s - 1} \Sigma_{i=1}^{n_s} (Y_i - \bar{Y}_s)^2$

Example:

Size	$\bar{Y}$	s_Y	n
small	657.4	19.4	238
large	650.0	17.9	182

$$t = \frac{\bar{Y}_s - \bar{Y}_l}{\sqrt{\frac{s_s^2}{n_s} + \frac{s_l^2}{n_l}}} = \frac{657.4 - 650.0}{\sqrt{\frac{19.4^2}{238} + \frac{17.9^2}{182}}} = \frac{7.4}{1.83} = 4.05$$

|t| > 1.96, so reject (at the 5% significance level) the null hypothesis that the two means are the same.

Confidence Interval

$\text{CI} = (\bar{Y}_s - \bar{Y}_l) \pm CV \times SE(\bar{Y}_s - \bar{Y}_l)$

Law of Large number

Summary: The Sampling Distribution of $\bar{Y}$

For $Y_1, \ldots, Y_n$ i.i.d. with $0 < \sigma_Y^2 < \infty$,

- The exact (finite sample) sampling distribution of $\bar{Y}$ has mean μ_Y ("$\bar{Y}$ is an unbiased estimator of μ_Y") and variance σ_Y^2/n C2
- Other than its mean and variance, the exact distribution of $\bar{Y}$ is complicated and depends on the distribution of Y (the population distribution)
- When n is large, the sampling distribution simplifies:
  - $\boxed{\bar{Y} \xrightarrow{p} \mu_Y}$ (Law of large numbers)
  - $\boxed{\frac{\bar{Y} - E(\bar{Y})}{\sqrt{\text{var}(\bar{Y})}} \text{ is approximately } N(0,1)}$ (CLT)

P-value

p-value=$Pr_{H_0}[|\bar{Y} - \mu_{Y,0}| > |\bar{Y}^{act} - \mu_{Y,0}|]$

variances of sample Y $= s_Y^2 = \frac{1}{n-1} \Sigma_{i=1}^{n} (Y_i - \bar{Y})^2$

Causality

Selection Bias in ATT with Observational Data

$E[Y_i^1|D_1 = 1] - E[Y_i^0|D_1 = 1] + E[Y_i^0|D_1 = 1] - E[Y_i^0|D_1 = 0]$

ATT:$E[Y_i^1|D_1 = 1] - E[Y_i^0|D_1 = 1]$

Selection Bias: $E[Y_i^0|D_1 = 1] - E[Y_i^0|D_1 = 0]$

Selection Bias in ATE with Observational Data

$E[Y_i^1] - E[Y_i^0] + Pr(D_i = 1)(E[Y_i^0|D_1 = 1] - E[Y_i^0|D_1 = 0]) + Pr(D_i = 0)(E[Y_i^1|D_1 = 1] - E[Y_i^1|D_1 = 0])$

ATE: $E[Y_i^1] - E[Y_i^0]$

Selection Bias: $Pr(D_i = 1)(E[Y_i^0|D_1 = 1] - E[Y_i^0|D_1 = 0]) + Pr(D_i = 0)(E[Y_i^1|D_1 = 1] - E[Y_i^1|D_1 = 0])$

Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals

The Least Squares Assumptions:
1. $E(u|X = x) = 0$ (or we can say that there is no correlation between the focus and other factors, i.e. causality are 0) 2. $(X_i, Y_i), i = 1, ..., n$, are i.i.d. 3. Large outliers are rare $E(X4) < \inf, E(Y4) < \inf$.

$\hat{\beta}$~$N(\beta_1, \frac{\sigma_v^2}{n\sigma_X^4})$, where $v_i = (X_i \breve{} \mu_X)u_i$ In other word, the model is normally distributed

$Test\hat{S}core = \beta_0 + \beta_1 STR$
β_1 is the slope of population regression line or the change in test score for a unit change in STR

t-testing
$\beta_{1,0} = 0$
$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}$

SE($\hat{\beta}_1$)

$var(\hat{\beta}_1) = \frac{var[(X_i - \mu_x)u_i]}{n(\sigma_X^2)^2}$, where $v_i = (X_i - \mu_X)u_i$

$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{1}{n} \times \frac{estimatorof\sigma_v^2}{(estimatorof\sigma_X^2)^2} = \frac{1}{n} \times \frac{\frac{1}{n-2}\Sigma_{i=1}^n \hat{v}_i^2}{\frac{1}{n}\Sigma_{i=1}^n (X_i - \hat{X})^2}$ where $\hat{v}_i = (X_i - \bar{X})u_i$
confidence interval
$\beta_1 =\{\hat{\beta}_1 \pm C \times SE(\hat{\beta}_1)\}$
C is the critical value

The G-M theorem says that among all possible choices of $\{w_i\}$, the OLS weights yield the smallest $var(\hat{\beta}_1)$