ST2137 Finals

by Zong Xun

Describing Numerical Data

Summarizing Data

Location is a simple summary of the data. **Variability** is mean squared deviation from the mean.

Skewness

Distribution is positively skewed if the right tail is longer, and negatively skewed if the left tail is longer.

Given a sample size of n, the **sample skewness** is

$$s_k = \frac{\sqrt{n(n-1)}}{n-2} \times \frac{m_3}{m_2^{3/2}}$$

where

$$m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \text{ and } m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3$$

- If s_k = 0, distn is perfectly symmetrical.
- If s_k is greater than $|1|$, distn is highly skewed.
- If s_k is $> |\frac{1}{2}|$ and $< |1|$, distn is moderately skewed.
- If s_k is between $-\frac{1}{2}$ and $\frac{1}{2}$, distn is approximately symmetric.

Kurtosis

Higher values of kurtosis indicate a higher, sharper peak

Given a sample of size n, the sample kurtosis is:

$$\frac{n-1}{(n-2)(n-3)}[\frac{(n+1)m_4}{m_2^2} - 3(n-1)]$$

where

$$m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \text{ and } m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4$$

Association between two variables

An association exists if a particular value for one variable is more likely to occur with certain values of the other variables.

Quantifying with correlation

The **correlation** between two variables, X and Y, is defined as:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}(\frac{X_i - \bar{X}}{s_X})(\frac{Y_i - \bar{Y}}{s_Y})$$

where $\bar{X}, \bar{Y}$ are the sample means. s_X and s_Y are the sample standard deviations of the two variables.

- The two variables have the **same correlation**, regardless of the response or explanatory variable.
- A positive correlation coefficient **does not** indicate a linear relationship, only a positive association.

Graphical Summaries

Histogram

- The overall pattern. Do the data cluster together, or is there a gap in which one or more data deviate from the rest?
- Is the data single modal, bimodal or multimodal?
- Is the distribution skewed? Any suspected outliers?

Boxplots

Able to identify the median, lower, upper quantiles and outlier(s).

- Define max whisker to be $Q_3 + 1.5IQR$, vice versa. Points is out of range from the min to max whisker are considered outliers.
- Define upper whisker to be the maximum point of the data.
- Define extreme outlier to be larger than $Q_3 + 3IQR$ or $Q_1 - 3IQR$

If there are more than 1 variable,

- Is there a trend?
- Compare the range, median, and mean

QQ plots

A QQ-plot plots the **standardized sample quantiles (x-axis)** against the **theoretical quantiles (y-axis)** of a N(0; 1) distribution. If the points on the tail forms a trend deviating from the straight line, there is evidence that the data is not normal.

- Right tail is below the straight line: longer than Normal.
- Right tail is above the straight line: shorter than Normal.
- Left tail is below the straight line: shorter than Normal.
- Left tail is above the straight line: longer than Normal.

Scatterplots

Visualize the association between two quantitative variable.

- Is there any relationship between the two variables?
- Is the association positve or negative? If so, is it linear or non-linear?
- Are there any observations that departs from the overall trend?
- Is the variance of the y-variable stable when the value of x-variable changes?

Robust Estimators

A statistical method is **robust** if it performs adequately even when the assumption is modestly violated.

Location Parameter

Trimmed mean

The $100\alpha\%$ trimmed mean calculated by: discarding the lowest $100\alpha\%$, highest $100\alpha\%$ and take the arithmetic mean of the remaining data.

Winsorized mean

Winsorization replaces extreme data values with less extreme values. The winsorized mean is computed after all the $[n\alpha]$ smallest observations are replaced by x($[n\alpha]$+1), and the $[n\alpha]$ largest observations are replaced by x(n-$[n\alpha]$)).

M-Estimators for Location Parameter

One can obtain more robustness by another function of error than the sum of their squares. We can find the estimator denoted by T - which is a function of $x_1, \ldots, x_n$ and this T is the minimizer of

$$\sum_{i=1}^{n} p(x_i - T)$$

where p is a **non-constant function** that is meaningful.

Examples:

- For the function $p(x) = x^2$, the minimizer of $\sum_{i=1}^{n}(x_i - T)^2$ is $\bar{x}$.
- For the function $p(x) = |x|$, the minimizer of $\sum_{i=1}^{n}|(x_i - T)|$ is the sample median.
- For the function

$$p(x) = \begin{cases} 1/2x^2 & \text{for } |x| \le k \\ k|x| - 1/2k^2 & \text{for } |x| > k \end{cases}$$

  then the estimator corresponds to a Winsorized mean
- if we set the function

$$p(x) = \begin{cases} 1/2x^2 & \text{for } |x| \le k \\ 1/2k^2 & \text{for } |x| > k \end{cases}$$

  then the estimator corresponds to a trimmed mean

Scale Parameter

Should remain robust even when a portion of the data points are replaced by arbitrary numbers.

Standard Deviation

Not robust, sensitive to outliers, and may not remain bounded when a single data point is replaced by an arbitrary number.

Interquartile Range

Better than standard deviation, but it is not a robust as well. For a normal distribution, the standard deviation, σ, can be estimated by dividing the interquartile range by 1.35.

$$IQR(X) = \sigma \times IQR(Z)$$

Median Absolute Deviation

The median absolute deviation (MAD) is a robust measure of the variability of a univariate sample of quantitative data.

$$MAD = median_i(|x_i - median_j(x_j)|)$$

For a normal distribution, $1.4826 \times MAD$ can be used to estimate the standard deviation.

$$MAD(X) = \sigma \times MAD(Z)$$

Categorical Data Analysis

A variable is called a categorical variable if each observation belongs to one of a set of categories. To distinguish between quantitative and categorical variables, one can ask if there is a meaningful distance between any two points in the data.

- If the observations can be ordered, the variable is **ordinal**. Otherwise, the variable is **nominal**.

Single Categorical Variable

We can use a frequency table (barplot) to produce the proportion or percentage of categories. The category with the highest frequency is known as the **modal** category.

Two Categorical Variable

- Two categorical variables are **independent** if the population conditional distributions for one of them are identical at each category of the other.
- The variables are **dependent** if the conditional distributions are not identical.

Contingency Table

Important to identify which variable is the **response** variable and which is **explanatory** variable, so that the conditional proportion can be calculated properly. Explanatory variables are often placed along the rows.

Comparing Proportions

Let ϕ_1, ϕ_2 denote the probabilities of success for each row. Let p_1, p_2 denote the proportion of successes for each row. In an ideal scenario, $\phi_1 = p_1$ and $\phi_2 = p_2$.

- The **sample difference**, $p_1 - p_2$ is used to estimate the difference between $\phi_1 - \phi_2$. If this difference is significant, we can infer association between the two variables.
- **Relative risk**: The ratio p_1/p_2 is used to estimate ϕ_1/ϕ_2. If the ratio is significantly different from 1, we can infer association between the two variables.

Odds ratio

$$\theta = \frac{\phi_1/(1-\phi_1)}{\phi_2/(1-\phi_2)}$$

When the two variables are independent, $\phi_1 = \phi_2$, so $\theta = 1$. The further it is away from 1, the stronger the association. Suppose OR = 1.51, then we say that the odds of **success** given **row** is 1.51 times the odds of **success** given **row 2**.

If the order of the rows **or** columns is reversed, the new value of θ is inverse of the original value.

If the table orientation reverses (row become column, and vice versa), the odds ratio **does not** change. This is unlike RR or difference of proportions, whose values depends on **each row**.

Confidence interval for Odds Ratio

The $100\%(1-\alpha)$ confidence interval is formed by

$$exp\{log\hat{\theta} \pm z_{\alpha/2} \times ASE(log\hat{\theta})\}$$

where

$$ASE(log\hat{\theta}) = \sqrt{\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{1}{n_{ij}}}$$

If the CI contains 1, that means that the population OR might be 1, hence the two variables might be independent!

Prospective vs Retrospective Studies

For **prospective studies**, sample subjects are picked randomly from the population. Record the explanatory variable status. Follow the subjects over time to see if they "succeeded" with relevance to the study.
For **retrospective studies**, sample a group of success and failure cases. Further split the success and failure into the explanatory variable status.

- Cheap, quick and fewer subjects are involved.
- Cannot obtain a valid estimate of ϕ_1 and ϕ_2.
- Can only use **odds ratio**.

Chi-squared (χ^2) Test

An indication of the degree of evidence for an association. Do note that χ^2 test **does not depend** on the order in which the rows and columns are listed. Thus, they ignore some information when there is an ordinal variable.

For 2x2 Tables

We have the following hypotheses:

$$H_0 : \text{The two variables are independent}$$
$$H_1 : \text{The two variables are dependent}$$

We compare the **expected counts** to the **observed counts**. For a particular cell,

$$\text{Expected Count} = \frac{\text{Row total} \times \text{Column Total}}{\text{Total sample size}}$$

The formula χ^2 test statistic (**with continuity correction**) is:

$$\chi^2 = \sum \frac{(|\text{observed count} - \text{expected count}| - 0.5)^2}{\text{expected count}}$$

p-value is calculated from $\chi^2(1)$ with 1 degree of freedom.

Fisher exact test: Use this when more than 25% of the cell counts have expected values less than 5.
McNemar Test: Test for dependence. Check if the same set of samples is used + before and after comparison is made. The test statistic:

$$\chi_1^2 = \frac{(|b-c|-1)^2}{b+c}$$

only subtract 1 in the numerator if the sample has a small cell count.

For RxC Tables

Generally, χ^2 test can be extended to tables larger than 2 by 2. The only **difference** is the distribution now follows $(r-1)(c-1)$ degree of freedom.

Standardized Residuals

Define **standardized residuals**

$$r_{ij} = \frac{n_{ij} - \mu_{ij}}{\sqrt{\mu_{ij}(1-p_{i+})(1-p_{+j})}}$$

where n_ij is the **observed count**, μ_{ij} is the **expected count**, p_{i+} is the **marginal probability** of row i and p_{+j} is the **marginal probability** of column j. The denominator is the estimated standard error of $(n_{ij} - \mu_{ij})$ under H_0.

- If H_0 is true, then each r_{ij} has a large-sample standard normal distribution.
- If $|r_{ij}|$ exceeds 2, then it indicates lack of fit of H_0 in that cell. Positive means more than expected, vice versa.

Linear-by-Linear test

For ordinal data. To detect a trend, assign scores to categories and measure the degree of linear trend or correlation. The scores should have (1) same ordering (2) reflect the distances between categories.

$$H_0 : \text{The two variables are independent}$$
$$H_1 : \text{The two variables are dependent}$$

The test statistic is calculated by

$$M^2 = (n-1)r^2$$

where r is sample correlation between X and Y,
$\bar{u} = \sum_i u_i p_{i+}$ is the sample mean of row scores,
$\bar{v} = \sum_j v_j p_{+j}$ is the sample mean of the column scores.

$$r = \frac{\sum_{i,j}(u_i - \bar{u})(v_j - \bar{v})p_{ij}}{\sqrt{[\sum_i (u_i-\bar{u})^2 p_{i+}][\sum_j (v_j - \bar{v})^2 p_{+j}]}}$$

where $p_{ij} = n_{ij}/n$; $p_{i+} = n_{i+}/n$; $p_{+j} = n_{+j}/n$.
For large samples, M^2 has approximately a χ^2 distribution with 1 df.

Tests for One and Two Samples

Parametric tests are significance tests that assume some form of distribution that the sample follows.
Non-parametric tests are significance tests that do not assume any form of distribution for the sample.

One sample tests

Parametric Test

One sample t-test to make significance tests about mean.

1. **Assumption:** Sample must come from randomization, and approximately follows normal distribution ($n \geq 30$).
2. **Test Statistics:**

$$T = \frac{\bar{X} - \mu_0}{SE(\bar{X})} \sim t_{n-1}$$

Non-parametric Test

Rough idea of Sign Test: A sample of size n is collected from the population that is *skewed*. As such, we will perform the significance test on the **median**, m_0.

1. Assign a sign to each data point: if $x_i > m_0$ then x_i is assigned positive, negative otherwise. Note that if $x_i = m_0$, the data point **should be removed**.
2. Count the total number of positive vs negative signs. Test statistic: $V = min(V+, V-)$.
3. This test statistic follows a Binomial Distribution, $Bin(n^*, 0.5)$, where n^* is the number of samples after removing points $x_i = m_0$.
4. p-value is calculated for two-sided test.

Idea for Wilcoxon Sign Rank Test: Similarly, the significance test is conducted about the median. Instead of comparing x_i and m_0, we take the difference between the two. Positive sign for positive difference, negative otherwise. **If $V+ \approx V-$, then we have evidence supporting $m_0 = 0$.** Likewise, if $V+ >> V-$, then we have evidence that median is greater than m_0.

Two independent samples test

Parametric Test

Two sample t-test is often used to make significance tests about the difference in mean.

1. **Assumptions:**
   - Independent, and random samples.
   - The population variance of each group is the same.
   - Each group is approximately normal.
2. **Null hypothesis:** $H_0 : \mu_1 - \mu_2 = 0$.
3. If the population variance are equal, we denote the pooled estimate of the common variance to be

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

4. **Test Statistic:**

$$T = \frac{(\bar{X} - \bar{Y}) - 0}{se}$$

where the standard error is computed as

$$SE = S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

follows a T-distribution with $n_1 + n_2 - 2$ df.

5. **Confidence interval:**

$$(\bar{X} - \bar{Y}) \pm t_{n_1+n_2-2, 1-\alpha/2} \times SE$$

Non-parametric test

Man-Whitney U-Test is also called as **Wilcoxon Rank Sum Test**. Let $X_1, \ldots, X_n$ be IID with cdf F, and $Y_1, \ldots, Y_m$ be IID with cdf G. The null hypothesis $H_0 : F = G$. There many different ways of representing the **test statistic**.

1. All $n+m$ observations are pooled together and a **rank** is assigned to them based on their magnitude. The smallest point is assigned the rank 1.
2. Define the rank sum scores

$$R_x = \sum_{i=1}^{n} Rank(X_i), \quad R_y = \sum_{i=1}^{m} Rank(Y_i)$$

   The idea is that under H_0, the ranks are uniformly distributed from $1, \ldots, m+n$.
3. Note that $R_x + R_y = (m+n)(m+n+1)/2$ is fixed.
4. Let n be $min(n,m)$, and compute the sum ranks R from that sample.
5. The test statistic is $W = min(R, R')$ where $R' = n(m+n+1) - R$.
6. We reject H_0 if W is too small.

Note: If two points share the same score, they share the same rank (i.e. 3.5 = (3 + 4) / 2)

Two dependent samples tests

Parametric

We will perform **paired T-test**.

1. Calculate the difference, D between each pair of data points. **We assume that the different pairs are independently distributed.**
2. Calculate the mean and SD of D.
3. The null hypothesis is $D_0 = 0$.
4. The test statistic is $t = \frac{\bar{D}-0}{SE(D)} \sim t_{n-1}$

Non-parametric

Same as one sample non-parametric test.

Analysis of Variance (ANOVA)

Generalization for comparing 2 independent samples.

Parametric Test

The F-test compares the variance in each group mean from the overall group variance. If the variance within groups is smaller than the variance between groups, the F-test will find a higher F-value, and therefore a higher likelihood that the difference observed is real and not due to chance.

We derive a test for the **null hypothesis** that all means are **equal**. The test statistic is given by

$$F = \frac{SS_B/(I-1)}{SS_W/[I(J-1)]}$$

Under the assumption that the errors are normally distributed, the null distribution of F is the F-distribution with (I-1) and I(J-1) degrees of freedom.

Assumptions

1. Random samples
2. Equal variance across groups
3. Independence of errors
4. Normal distribution of errors
5. Additivity of treatment effects

Non-parametric Test

Kruskal-Wallis test is a generalization of the Mann-Whitney U-Test. Observations are assumed to be independent, but no particular form of distribution is assumed. The observations are pooled together and ranked.

The null hypothesis is that the groups belong to the **same distribution**.

Multiple Comparison Tests

We will focus on **Bonferroni's Correction** and **Tukey**.

Instead of controlling "locally" the probability of type 1 error for each individual test, we control "globally" the probability of at least a type 1 error among all these k test, by a value that is called **family error rate** a. The role of this family significance level a to the k tests is the same as the role of individual significance level α to each test, $\alpha = a/k$.

Regression Analysis

Given two continuous variables X and Y, the **Pearson correlation coefficient** is defined as:

$$p = \frac{cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

1. A positive value indicates that X and Y moves in the same direction, vice versa.
2. Indicates the strength of the relationship
3. Does not tell us what type of relationship (i.e. linear)

Assumptions

With only 2 variables X and Y with sample $(x_1, y_1), \ldots, (x_n, y_n)$, the form of a simple model is:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n$$

where β_0 is the intercept and β_1 is called the slope/gradient.

1. *Relationship between X and Y are linear* (scatterplot).
2. The error term, $e_i \sim N(0, \sigma^2)$. We refer to this as the **normality assumption** and **constant variance assumption**.
3. The error term is assumed to be uncorrelated for all $i \neq j$.
4. When the model has more than one regressor, we further assume that the regressors are uncorrelated.

With the above assumptions, we derive that y_i's are independent and normally distributed $\sim N(\beta_0 + \beta_1 x_i, \sigma^2)$.

Linear Regression

The **Ordinary Least Squares** (OLS) method picks the line that minimizes the term sum of the squared residuals.

In a **simple linear model**, we derive the estimation for parameters B_0 and B_1 by solving the partial derivatives of the loss function, MSE.

Properties of OLS estimators

- The estimators follows a normal distribution.
- The sum of residuals and residuals weighted by corresponding regressor's values or the fitted values is zero.

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} x_i e_i = \sum_{i=1}^{n} \hat{y}_i e_i = 0$$

- Sum of predictions vs actual, $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$

Frequently used notations

- Denote

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad S_{xy} = \sum_{i=1}^{n} y_i (x_i - \bar{x})$$

- The sum of squared residuals,

$$SS_{res} = \sum_{i=1}^{n} [e_i^2 = (y_i - \hat{y_i})^2]$$

- Denote

$$SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2 \implies SS_T = SS_R + SS_{res}$$

where SS_R is the regression sum of squares and has $df = k$, and k refers to the number of regressors.

Estimating variance

We denote $MS_{res} = SS_{res}/(n-k)$ as an unbiased estimator of σ^2, where $(n-k)$ is the number of degrees of freedom and k is the number of regressors. $\sqrt{MS_{res}}$ is called the **residual standard error.**

Standardized Residuals

Standardized residuals is defined as

$$\frac{e_i}{\sigma\sqrt{1 - h_i i}}$$

Since σ is unknown, we can estimate it by $\sqrt{MS_{res}}$. If all assumptions made for the model are met, SR $\sim N(0, 1)$.

Testing Hypothesis

There are two kinds of tests that can be conducted.

- T test: testing significance of one regressor **given the other regressors in the model** $\sim t_{n-k-1}$.

$$H_0 : B_i = 0 \quad H_1 : B_i \neq 0$$

- F test: testing significance of whole model.

$$H_0 : \forall_i \in \{1..n\},\ B_i = 0 \quad H_1 : \exists_i \in \{1..n\},\ B_i \neq 0$$

- Test Statistic: $\frac{SS_R/k}{SS_{res}/(n-k-1)} \sim F_{k,n-k-1}$.

In a simple model, T test is **equal** to F test, $t = \sqrt{F}$

What plots to make?

- Plot the r_i's on the y-axis against Y_i/X_i on the x-axis.
- Create a histogram of the r_i's.
- Create a QQ-plot of the r_i's.

What to look out for?

- Residual plots against Y/X should be scatter randomly about 0, within the interval (-3, 3).
- No funnel shape (Constant variance violated).
- Non-normality in QQ plot (Residuals are not normal)
- A curved band when plotting Y against X. (take $logy$, $1/y$, adding higher order terms etc)

Standard answer: The histogram and QQ plot indicate the normality of the SR. The plot of SR versus the fitted response shows the randomness of the SR within the range of -3 to 3 with no pattern nor trend.

Coefficient of Determination, R^2

The proportion of total variation of the response (about the sample mean $\bar{Y}$) that is explained by the model. It does not indicate the appropriateness of a linear model.

In a **simple model**, $R^2 = cor(X, Y)^2$.

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{res}}{SS_T}$$

In a **non-simple model**, look at the adjusted R^2, which penalizes insignificant terms.

$$R^2_{adj} = 1 - \frac{SS_{res}/(n-p)}{SS_T/(n-1)}$$

Adding more terms generally increases R^2. If there are X values with different Y values, then R^2 can never be 1.

Influential points

To detect the influential points, one may consider the hat diagonal values h_{ii} in conjunction with the standardized residuals. Observations with large hat diagonals $(h_{ii} > 2(k+1)/n)$ and large residuals are likely to be influential. Hat matrix is defined as :

$$H = X(X'X)^{-1}X'$$

Any data point that has Cook's distance larger than 1 can be considered as an influential point.

Multiple Linear Regression

We derive the estimation of $\hat{B}$ by solving

$$\hat{B} = (X'X)^{-1}X'y$$

The fitted model is then

$$\hat{y} = X\hat{B}$$

Indicator Variables

An **indicator variable** takes on the value 1 if a category is observed, and 0 otherwise. A categorical variable with **k categories** requires **k - 1 indicators**. The category without an indicator variable is the **reference group**. Variance is assumed to be equal for all levels of the categorical variable.

Interaction with Categorical Variables

It is possible that one of the variable may become a linear form of another variable $\implies$ modify/drop the term

Simulation

Monte Carlo Simulation for Approximation

The estimator has a true sampling distribution under certain assumptions/conditions, but often, this is hard to achieve in reality. We can approximate via monte carlo simulation.

1. Generate M independent datasets.
2. Compute the estimator T for each dataset.
3. If M is large enough, the set of simulated estimators should be good approximations to the true properties of the estimator.

Comparing Estimators of Mean

The 3 common estimators for the mean μ are: (1) sample mean, (2) sample median, (3) trimmed mean.

We can based the performance of the estimator on:

- Mean = $1/M \sum_{m=1}^{M} T_m^k$
- Bias = $E(T^K) - \mu$
- SD = $\sqrt{Var(T^K)}$
- MSE = $Bias^2 + SD^2$

Coverage probability of Confidence Intervals

Form M confidence intervals, and determine the proportion of intervals with mean μ inside.

Properties of Hypothesis Test

To evaluate whether sample can achieve the advertised α:

- Generate data under H_0 and calculate proportion of rejections when H_0 is in fact true.

To estimate the **power** of the test:

- Generate data under some **alternative** H_1.
- Approximate the true probability of reject H_0 when H_0 is false.

Resampling

Bootstrap

The distribution of the finite population represented by the sample can be regarded as a pseudo-population with similar characteristics as the true population.

Differences between Simulation and Bootstrap

1. Simulation generates samples from completed specified distribution.
2. (Parametric) Bootstrap: fits a distribution for the given sample, and then generates random samples from this fitted distribution.
3. (Non-parametric) Bootstrap: does not fit any distribution to the given sample, just generates random samples from the empirical distribution of the sample. Empirical distribution:

$$f_n(x) = \begin{cases} 1/n, & x = x_1, x_2, \ldots, x_n \\ 0, & \text{otherwise} \end{cases}$$

Steps of the Bootstrap Estimation

1. For each bootstrap replicate, indexed $b = 1, 2, \ldots, B$
   - Generate bootstrap sample $x^{*(b)} = X_1^*, \ldots, X_n^*$ by sampling with replacement from the observed sample $x_1, \ldots, x_n$. This is the **nonparametric**. For **parametric**, sample from a known distribution $f_X(x, \alpha)$.
   - Compute the value of the estimator from the bth bootstrap sample, denoted as $\hat{\theta}^{*(b)}$.
2. The bootstrap estimate of $F_{\hat{\theta}}$ is the empirical distribution of these replicates.
3. The bootstrap estimate is used to estimate the standard error, bias and confidence interval of an estimator.

Standard Error of an Estimator

The bootstap estimate of the SE is the sample standard deviation of the bootstrap replicates, which is

$$\hat{se} = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}^{*(b)} - \bar{\hat{\theta}}^*)^2}$$

where $\bar{\hat{\theta}}^* = \frac{1}{B} \sum_{i=1}^{B} \hat{\theta}^{*(b)}$.

Bias of an Estimator

The bootstap estimate of the bias of an estimator is the difference between the **mean** of the bootstrap replicates and $\hat{\theta}$.

$$\hat{bias}(\hat{\theta}) = \bar{\hat{\theta}}^* - \hat{\theta}$$

where $\bar{\hat{\theta}}^* = \frac{1}{B} \sum_{i=1}^{B} \hat{\theta}^{*(b)}$.

Confidence Interval of an Estimator

1. **Basic Bootstrap Confidence Interval**
   The $100(1-\alpha)\%$ confidence limits is given by

$$(2\hat{\theta} - \hat{\theta}^*_{1-\alpha/2}, 2\hat{\theta} - \hat{\theta}^*_{\alpha/2})$$

   where $\hat{\theta}^*_{1-\alpha/2}$ is the α sample quantile from the empirical distribution function of the replicates.
2. **Percentile Bootstrap Confidence Interval**
   The $100(1-\alpha)\%$ confidence limits is given by

$$(\hat{\theta}^*_{\alpha/2}, \hat{\theta}^*_{1-\alpha/2})$$

3. **Normal Bootstrap Confidence Interval**
   The normal bootstrap CI constructs the CI based on the assumption that the distribution of the estimator is normally distributed, $N(\theta + bias, variance)$.
   The $100(1-\alpha)\%$ confidence limits is given by

$$(\hat{\theta} - bias \pm z_{1-\alpha/2} \times \sqrt{variance})$$