STA 106 Notes - M. Pouokam
Dylan M Ang
March 31, 2022

Contents

1 Terminology

Subject : A person, place or thing from which we measure data.
Population : The collection of all subjects of interest.
Sample : A subset of the population, from which we collect data.
Response/Dependent Variable : The variable which is of primary interest.
Explanatory/Independent Variable : The variable which we believe helps explain some of the variance in the response variable.
For this class, the response variable is numerical , and the explanatory variable/s are categorical . This is often phrased as "how does this numerical variable differ by group?"
Random Variable : A variable whose outcome is random. These are typically denoted by capital letters.

2 Notation

Let Y = the random variable denoting all possible values of the response variable.
Let y = an observed value of Y (in other words, measured observations).
Let Y_{ij} = the rv denoting all possible values of the jth observation in group i.
Let y_{ij} = the ith observed value of the jth group in Y.
As an example, let $i = 1$ denote sex M, and $i = 2$ denote sex F.

- Y_{13} = All possible values of height for the 3rd male. The outcome is random and unknown.

- y_{13} = 72 inches. The specific observed value of the 3rd male once measured.

μ_i = The population mean for group i (a single value).
$\bar{Y}_i$ = All possible values of the sample mean for group i.
$\bar{y}_i$ = A specific, observed value of $\bar{Y}_i$. In other words, the mean of one given sample.
σ_i = The population standard deviation for group i.
S_i = All possible values of the sample standard deviation.
s_i = A specific, observed value of S_i. In other words, the standard deviation of one given sample.
The book does not make this distinction $\implies Y = y$ and $\bar{Y} = \bar{y}$
Parameter : The unknown population value of some statistic. For example μ_i, σ_i. These values are constant (non-random) if we could measure the population we would know the true value.
The goal of statistics 106 is to estimate parameters with sample values, and use the assumed distribution of those sample values to form Hypothesie Tests (HTs) and Confidence Intervals (CIs) .

3 Mean and Variance of RVs

Let Y_i be drawn from a distribution with population mean μ_Y and population standard deviation σ_Y.
Let the mean of $Y_i = \mu_{Y_i} = E\{Y_i\} = \mu_Y$.
Let the standard deviation of $Y_i = \sigma_{Y_i} = \sigma\{Y_i\} = \sigma_Y$.

3.1 Linear Combinations of RVs

Let Y^* be a linear combination of Y where $a, b \in \mathbb{R}$, then

Combination	Mean	Variance
$Y^* = a + Y$	$\mu_{Y^*} = a + \mu_Y$	$\sigma^2_{Y^*} = \sigma^2_Y$
$Y^* = bY$	$\mu_{Y^*} = b\mu_Y$	$\sigma^2_{Y^*} = b^2\sigma^2_Y$
$Y^* = a + bY$	$\mu_{Y^*} = a + b\mu_Y$	$\sigma^2_{Y^*} = b^2\sigma^2_Y$

3.2 Summation Identities

Let $Y_1, Y_2, \ldots, Y_n$ be RVs

$$\begin{aligned} E\left\{\sum_{i=1}^{n} Y_i\right\} &= E\{Y_1 + Y_2 + \cdots + Y_n\} \\ &= E\{Y_1\} + E\{Y_2\} + \ldots E\{Y_n\} \\ &= \sum_{i=1}^{n} E\{Y_i\} \end{aligned}$$

If $Y_1, Y_2, \ldots, Y_n$ are independent RVs,

$$\sigma^2\left\{\sum_{i=1}^{n} Y_i\right\} = \sum_{i=1}^{n} \sigma^2\{Y_i\}$$

Let $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, Y_i is independent with mean μ_Y and std.dev. σ_Y.

$$\begin{aligned} E\{\bar{Y}\} &= E\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i\right\} = E\left\{\frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n)\right\} \\ &= \frac{1}{n}E\{Y_1 + Y_2 + \cdots + Y_n\} \\ &= \frac{1}{n}\Big(E\{Y_1\} + E\{Y_2\} + \cdots + E\{Y_n\}\Big) \\ &= \frac{1}{n}\sum_{i=1}^{n} E\{Y_i\} = \frac{1}{n}\sum_{i=1}^{n} \mu_Y \\ &= \frac{1}{n}(\mu_Y + \mu_Y + \cdots + \mu_Y) = \frac{1}{n}(n * \mu_Y) \\ E\{\bar{Y}\} &= \mu_Y \end{aligned}$$

$$\begin{aligned} \sigma^2\{\bar{Y}\} &= \sigma^2\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i\right\} = \left(\frac{1}{n}\right)^2 \sigma^2\left\{\sum_{i=1}^{n} Y_i\right\} \\ &= \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \sigma^2\{Y_i\} \end{aligned}$$

4 Normal RVs and χ^2 RV

A mormal RV follows a bell curve created by a probability density function (pdf).

If Y is normally distributed with mean μ_Y and std dev σ_Y, we say that $Y \sim N(\mu_Y, \sigma_Y)$

$Y \sim N(\mu_Y, \sigma_Y) \implies Y^* = a + bY \sim N(a + b\mu_Y, b\sigma_Y)$

From this we can get two more results,

If $Y_1, \ldots, Y_n$ independent and $Y_i \sim N(\mu_Y, \sigma_Y)$, then

1. $\bar{Y} \sim N(\mu_Y, \sigma_Y/\sqrt{n})$
2. $\sum Y_i \sim N(n\mu_Y, \sqrt{n\sigma_Y})$

The standard normal distribution is a specific linear combination of a general normal distribution, denoted Z.

Let $Y \sim N(\mu_Y, \sigma_Y)$

$$Z = \frac{Y - \mu_Y}{\sigma_Y} = \frac{Y}{\sigma_Y} - \frac{\mu_Y}{\sigma_Y}$$

$$E\{Z\} = \frac{-\mu_Y}{\sigma_Y} + \mu_Y(\frac{1}{\sigma_Y}) = 0$$

$$\sigma_Z^2 = \left(\frac{1}{\sigma_Y}\right)^2 \sigma_Y^2 = 1$$

Therefore $Z \sim N(0, 1)$