Statistics Review

Thomas Lonon

Financial Engineering/Financial Analytics
Stevens Institute of Technology

August 30, 2018

Data

Observations from a **population** is called a **sample.**

- Univariate
- Bivariate
- Multivariate

Descriptive vs. Inferential statistics

Histograms

Constructing a Histogram for Discrete Data:
First, determine the frequency and relative frequency of each *x* value. Then mark possible *x* values on a horizontal scale. Above each value, draw a rectangle whose height is the relative frequency (or alternatively, the frequency) of that value. [1]

Constructing a Histogram for Continuous Data: Equal Class Widths:
Determine the frequency and relative frequency for each class. Mark the class boundaries on a horizontal measurement axis. Above each class interval, draw a rectangle whose height is the corresponding relative frequency (or frequency).[1]

Location

sample mean: for observations $X_1, X_2, \ldots, X_n$ we define

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

sample median: obtained by first ordering the n observations from smallest to largest (with any repeated values included so that every sample observation appears in the ordered list). Then $\tilde{X} =$ the single middle value if n is odd or the average of the two middle values if n is even.[1]

Variability

range: the difference between the largest and the smallest sample values.

sample variance: denoted by s^2 is given by,

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

sample standard deviation: denoted by s, is the (positive) square root of the variance:

$$s = \sqrt{s^2}$$

Law of Large Numbers

Let $X_1, X_2, \ldots, X_i, \ldots$ be a sequence of independent random variables with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}(X_i) = \sigma^2$. Let

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then, for any $\varepsilon > 0$,

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \to 0 \text{ as } n \to \infty$$

[2]

Central Limit Theorem

Let $X_1, X_2, \dots$ be a sequence of independent random variables having mean 0 and variance σ^2 and the common distribution function F and moment-generating function M defined in a neighborhood of zero. Let

$$S_n = \sum_{i=1}^{n} X_i$$

Then

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right) = \Phi(x), -\infty < x < \infty$$

[2]

Confidence Interval

A $\mathbf{100(1-\alpha)\%}$ **confidence interval** for the mean μ of a normal population when the value of σ is known is given by

$$(\bar{X} - z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} * \frac{\sigma}{\sqrt{n}})$$

or, equivalently, by $\bar{X} \pm z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$[1]

note that z_β is the z-score corresponding to β such that for a standard normal random variable Z we have:

$$\mathbb{P}(Z \leq z_\beta) = \Phi(z_\beta) = \beta$$

Def: If Z is a standard normal random variable, the distribution of $U = Z^2$ is called the chi-square distribution with 1 degree of freedom.

Def: If $U_1, U_2, \ldots, U_n$ are independent chi-square random variables with 1 degrees of freedom, the distribution of $V = U_1 + U_2 + \cdots + U_n$ is called the chi-square distribution with n degrees of freedom and is denoted by χ_n^2.[2]

Def: If $Z \sim N(0,1)$ and $U \sim \chi_n^2$ and Z and U are independent, then the distribution of $\frac{Z}{\sqrt{\frac{U}{n}}}$ is called the t-distribution with n degrees of freedom.

Def: Let U and V be independent chi-square random variables with m and n degrees of freedom, respectively. The distribution of

$$W = \frac{\frac{U}{m}}{\frac{V}{n}}$$

is called the F distribution with m and n degrees of freedom and is denoted by $F_{m,n}$.[2]

Theorem: The random variable $\bar{X}$ and the vector of random variables $(X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X})$ are independent.

Corollary: $\bar{X}$ and s^2 are independently distributed.[2]

Theorem: The distribution of $(n-1)s^2/\sigma^2$ is the chi-square distribution with $n-1$ degrees of freedom.

Corollary: Let $\bar{X}$ and s^2 be as given. Then

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim \underline{t_{n-1}}$$

[2]

Properties of t Distributions

student distribution

Let t_ν denote the t distribution with ν df.

1. Each t_ν curve is bell-shaped and centered at 0
2. Each t_ν curve is more spread out than the standard normal curve. fat tail
3. As ν increases, the spread of the corresponding t_ν curve decreases.
4. As $\nu \to \infty$, the sequence of t_ν curves approaches the standard normal curve

[1]

Let $\bar{X}$ and s be the sample mean and sample standard deviation computed from the results of a random sample from a normal population with mean μ. Then a **100$(1-\alpha)$% confidence interval for μ** is

$$(\bar{X} - t_{\alpha/2,n-1} * \frac{s}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1} * \frac{s}{\sqrt{n}})$$

or, equivalently, by $\bar{X} \pm t_{\alpha/2,n-1} * \frac{s}{\sqrt{n}}$
[1]

Population Mean:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Population Total:

$$\tau = \sum_{i=1}^{N} x_i = N\mu$$

Population Variance:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

Sample Mean:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Theorem: With simple random sampling, $\mathbb{E}[\bar{X}] = \mu$

Theorem: With random sampling,

$$\mathbb{V}(\bar{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

[2]

Biased vs Unbiased

Let

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Theorem: With random sampling,

$$\mathbb{E}[\hat{\sigma}^2] = \sigma^2 \left(\frac{n-1}{n}\right)\frac{N}{N-1}$$

Let

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

We have:

$$\mathbb{E}[s^2] = \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \mathbb{E}\left[\frac{n}{n-1}\hat{\sigma}^2\right]$$
$$= \frac{n}{n-1}\mathbb{E}\left[\hat{\sigma}^2\right] = \frac{n}{n-1}\sigma^2\left(\frac{n-1}{n}\right)\frac{N}{N-1}$$

And so,

$$\mathbb{E}\left[\frac{n}{n-1}\hat{\sigma}^2\right] = \sigma^2\frac{N}{N-1}$$

Method of Moments

Let $\mu_k = \mathbb{E}[X^k]$, and then define the **sample moment** as

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

The method of moments estimates parameters by finding expressions for them in terms of the lowest possible order moments and then substituting sample moments into the expressions.[2]

Maximum Likelihood Estimate

Suppose that random variables $X_1, \ldots, X_n$ have a joint density of frequency function $f(x_1, x_2, \ldots, x_n|\theta)$. Given observed values $X_i = x_i$, where $i = 1, \ldots, n$, the likelihood of θ as a function of $x_1, \ldots, x_n$ is defined by

$$\text{lik}(\theta) = f(x_1, x_2, \ldots, x_n|\theta)$$

The **maximum likelihood estimate (mle)** of θ is that value of θ that maximizes the likelihood–that is, makes the observed data "most probable" or "most likely."[2]

Hypothesis Testing

Null Hypothesis: $\mathbf{H_0}$ The default assumption that is believed to be true (e.g. $\mu = 0$)

Alternate Hypothesis: $\mathbf{H_a}$ An alternate interpretation of the results. (e.g. $\mu \neq 0$)

Type I and II Errors

	Actual Positive	Actual Negative
Classified Positive	True Positive	False Positive
Classified Negative	False Negative	True Negative

- Type I error: incorrect rejection of a true null hypothesis (false negative) innocent person go to jail
- Type II error: incorrect failure to reject a false null hypothesis (false positive) guilty person go free

[1] Jay L. Devore. *Probability and Statistics for Engineering and the Sciences*. Brooks/Cole, eighth edition, 2012.

[2] John A. Rice. *Mathematical Statistics and Data Analysis*. Duxbury Press, second edition, 1995.