STATS 2107 Statistical Modelling and Inference II
STATS 7107 Statistical Modelling and Inference

Semester 2, 2020
Sharon Lee

1 Introduction

1.1 Preamble

Suppose $y_1,\ y_2,\ \ldots,\ y_n$ are independent observations drawn from the $N(\mu,\sigma^2)$ distribution. If σ^2 is known, then the following calculations are usually performed.

1. Use the *sample mean* $\bar{y}$ to estimate μ.
2. The *standard error* of $\bar{Y}$ is $\sigma/\sqrt{n}$.
3. To test $H_0:\ \mu = \mu_0$, we calculate the value of the test statistic
$$z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}$$
and *reject* H_0 at the 5% level of significance if and only if $|z| \geq 1.96$, i.e., if and only if
$$\text{P-value} = 2P(Z \geq |z|) < 0.05,$$
where $Z \sim N(0,1)$.
4. A 95% confidence interval for μ is
$$(\bar{y} - 1.96\sigma/\sqrt{n},\ \bar{y} + 1.96\sigma/\sqrt{n}).$$

If σ^2 is not known, the procedures are modified as follows.

1. We use the *sample standard deviation*
$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$
in place of σ.
2. We use $t_{n-1}(0.025)$ in place of 1.96. More precisely, to test $H_0:\ \mu = \mu_0$, we calculate the value of the test statistic
$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$
and *reject* H_0 at the 5% level of significance if and only if $|t| \geq t_{n-1}(0.025)$, where $t_{n-1}(0.025)$ is the number such that $P(T > t_{n-1}(0.025)) = 0.025$, for $T \sim t_{n-1}$.

1.2 Course aims

In this course, we will elaborate on the concepts and techniques discussed above from the following perspective:

- Formulate the underlying statistical concepts.
- Formulate the principles from which hypothesis tests and confidence intervals may be derived as "good procedures".
- Explore the use of probability theory in the derivation of the distributional results needed for statistical inference.
- Develop a framework in which hypothesis tests, regression, analysis of variance and analysis of covariance can be seen as instances of the same theory.
- Apply these methods to data analysis, using the statistical package R.
- Consider the theory of statistical estimation and hypothesis testing in a more general framework.

1.3 Notation and general remarks

- We use uppercase letters, $X,\ Y,\ Z$ to denote *random variables* (RVs).
- We use lowercase letters, $x,\ y,\ z$ to represent real numbers.
- The *probability function* of a *discrete* random variable X is denoted:

  $$p(x) = P(\{X = x\}).$$

  Some important examples are:

  Bernoulli Distribution

  $$p(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0. \end{cases}$$

  Binomial Distribution

  $$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ for } x = 0,\ 1,\ \ldots,\ n.$$

  Geometric Distribution

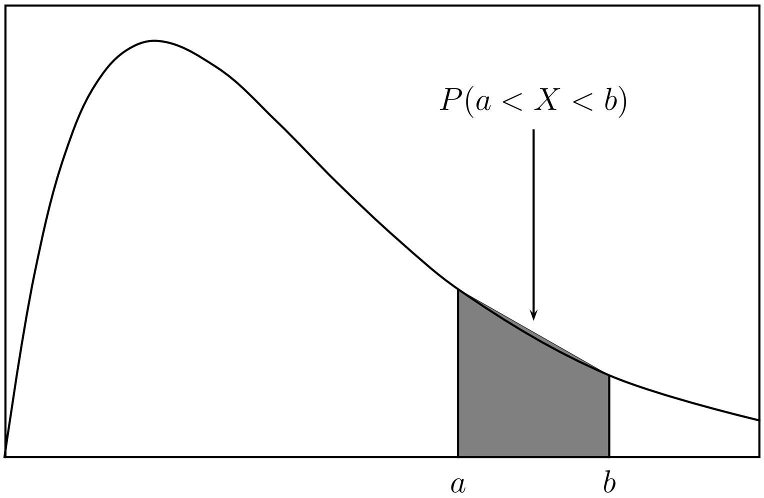
  $$p(x) = p(1-p)^{x-1}, \text{ for } x = 1,\ 2,\ \ldots\ .$$

Poisson Distribution

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \text{ for } x = 0,\ 1,\ 2,\ \ldots\ .$$

- The *probability density function* (PDF) for a continuous random variable X is denoted by $f(x)$. It uses *area* to represent probability.

$$P(a \le X \le b) = \int_a^b f(x)dx$$



Some important examples are:

Uniform Distribution

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise.} \end{cases}$$

Exponential Distribution

$$f(x) = \lambda e^{-\lambda x}, \text{ for } x > 0.$$

Standard Normal $N(0,1)$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Normal $N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- The *cumulative distribution function* (CDF) is defined for both discrete and continuous random variables by

$$F(x) = P(\{X \le x\}).$$

Probability: In probability theory, we use probability distributions to describe the behaviour of random variables. That is, we take the probability distribution as given and make predictions about a (yet to be observed) random variable.

Statistical Inference: Statistical inference is concerned with the inverse problem.

- We begin with data (numbers)

$$y_1,\ y_2,\ \ldots,\ y_n.$$

- We assume the data to be *realizations* of random variables

$$Y_1,\ Y_2,\ \ldots,\ Y_n$$

  with some unknown CDF F.

- We use the data to make conclusions about F.

- Usually, it is assumed that F belongs to a given *family* of distributions, indexed by an unknown *parameter* θ.

  θ may be a *vector* parameter.

  For example, there is a different normal distribution for each pair (μ, σ^2). In this case, we would take θ to be the vector $\boldsymbol{\theta} = (\mu, \sigma^2)$.

Within this framework, the problem of statistical inference can be stated as using the observed data $y_1,\ y_2,\ \ldots,\ y_n$ to draw conclusions about the unknown parameter θ.