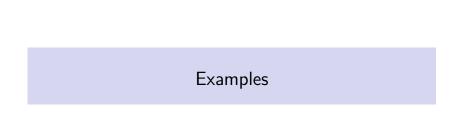
# Lecture 2: Preliminaries and One-sample problem

Pratheepa Jeganathan

04/05/2019



# Example 1.7 (Spatial Ability Scores of Students)

- Data on a student's spatial ability using four tests of visualization.
- ► For each student, a single score representing their overall measure of spatial ability.
- ► The spatial ability scores for 68 female and 82 male high school students enrolled in advanced placement calculus classes in Florida.
  - What is the distribution of spatial ability scores for the population represented by this sample of data?
  - ► Does the distribution for the male students appear to possess different characteristics than that of the female students?
- ► These questions are problems in density estimation

## Example 1.8 (Sunspots)

- Data on mean monthly sunspot observations collected at the Swiss Federal Observatory in Zurich and the Tokyo Astronomical Observatory from the years 1749 to 1983.
- Excessive variability over time, obscuring any underlying trend in the cycle of sunspot appearances.
- ▶ No apparent analytical form or simple parametric model.
- Powerful method for obtaining the trend from a noise in this case is wavelet estimation and thresholding.



#### **Notations**

- X: random variable
- x: realizations (observed random variables)
- f(x): probability density function (pdf)
- ▶  $F_X(x) = P(X \le x)$ : cumulative distribution function (cdf)
- $X_1, \dots X_n$ : random sample (independent and identically distributed)

#### Distribution-free test statistic

- ▶ Test statistic:  $T(\cdot) = T(X_1, \dots, X_n)$ , function of the data.
  - Example:  $T = \frac{\bar{X} \mu}{s/\sqrt{n}}$ , where  $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$  and  $s^2 = \frac{\sum_{i=1}^{n} \left(X_i \bar{X}\right)^2}{n-1}$ ,  $\mu$  is known under  $H_0$ .
- Distribution-free test statistic
  - ightharpoonup Example:  $\mathcal{U} = \mathsf{MVN}\left(oldsymbol{\mu} = (\mu, \cdots, \mu), oldsymbol{\Sigma} = \sigma^2 \mathbf{I}\right)$ 
    - $\qquad \qquad T_1 = \frac{\bar{X} \mu}{\sigma / \sqrt{n}} \sim \mathsf{N} \, (0, 1).$
    - $T_2 = \frac{\bar{X} \mu}{s/\sqrt{n}} \sim \mathsf{t}_{n-1}.$
- Nonparametric distribution-free test statistic
  - ▶ The class  $\mathcal{U}$ ,  $T(\cdot)$  is distribution free over contains more than one distributional forms.

- ► Distribution-free confidence interval, distribution-free multiple comparison procedure, distribution-free confidence band,
  - comparison procedure, distribution-free confidence band, asymptotically distribution-free test statistic, asymptotically distribution-free multiple comparison procedure, and asymptotically distribution-free confidence band.

#### Rank statistic

- ▶ Absolute rank: For any random variable  $Z_1, \dots, Z_n$ , the absolute rank of  $Z_i$ , denoted by  $R_i$  is the rank of  $|Z_i|$  among  $|Z_1|, \dots, |Z_n|$ .
- Rank statistic: A statistic T (R) based only on the ranks of a sample is rank statistic.
  - ightharpoonup T(R) is distribution-free over iid joint continuous distribution.
- ▶ Signed rank: The signed rank of  $Z_i$  is  $R_i\psi_i$ , where

$$\psi_i = \begin{cases} 1, & \text{if} & Z_i > 0, \\ 0, & \text{if} & Z_i < 0. \end{cases}$$
 (1)

- Signed rank statistic: A statistic  $T(\psi, \mathbf{R}) = T(R_1\psi_1, \cdots, R_n\psi_n)$  that is a function of  $Z_1, \cdots Z_n$  only through the signed ranks is the signed rank statistic.
  - $T(\psi, R)$  is ditribution-free over iid joint continuous distribution symmetric about 0.

Sign test (Fisher) - paired replicates data/one-sample data

## Sign test

- ▶  $Z_1, \dots Z_n$  random sample from a continuous population that has a common median  $\theta$ .
  - ▶ If  $Z_i \sim F_i$ ,  $F_i(\theta) = F_i(Z_i \le \theta) = F_i(Z_i > \theta) = 1 F_i(\theta)$ .
- Hypothesis testing:
  - $H_0: \theta = 0$  versus  $H_A: \theta \neq 0$ .

# Sign test (Cont.)

- ▶ Sign test statistic:  $B = \sum_{i=1}^{n} \psi_i$ .
- Motivation:
  - When  $\theta$  is larger than 0, there will be larger number of positive  $Z_i s big B$  value big big B value big big B value big b
- ▶ Under  $H_0$ ,  $B \sim (n, 1/2)$
- ▶ Significance level  $\alpha$ : probability of rejecting  $H_0$  when it is true.
- Note
  - choices of  $\alpha$  are limited to possible values of the  $B\sim (n,1/2)$  cdf.
  - compare the distribution of B under H<sub>0</sub> and the observed test statistic value.

# Sign test (Cont.)

- Rejection regions
  - ▶  $H_A$ :  $\theta > 0$ , Reject  $H_0$  if  $B \ge b_{\alpha;n,1/2}$ .
  - ▶  $H_A$ :  $\theta$  < 0, Reject  $H_0$  if  $B \le n b_{\alpha:n,1/2}$ .
  - ▶  $H_A: \theta \neq 0$ , Reject  $H_0$  if  $B \geq b_{\alpha/2;n,1/2}$  or  $B \leq n b_{\alpha/2;n,1/2}$ .

# Large-Sample Approximation (Sign test)

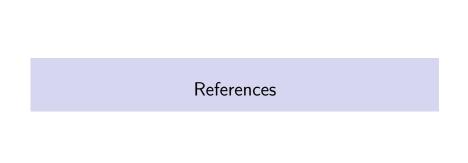
$$ho B^* = rac{B - \mathbb{E}_0\left(B\right)}{\mathbb{V}_0\left(B\right)^{1/2}} \sim \mathsf{N}\left(0,1\right) \quad \text{as} \quad n o \infty \quad , \text{where}$$

$$ightharpoonup \mathbb{E}_0\left(B
ight) = rac{n}{2} ext{ and } \mathbb{V}_0\left(B
ight) = rac{n}{4}$$

- Rejection regions
  - ▶  $H_A$ :  $\theta > 0$ , Reject  $H_0$  if  $B^* \ge z_\alpha$ .
  - $H_A: \theta < 0$ , Reject  $H_0$  if  $B^* \leq -z_{\alpha}$ .
  - ▶  $H_A$ :  $\theta \neq 0$ , Reject  $H_0$  if  $B^* \geq z_{\alpha/2}$  or  $B \leq -z_{\alpha/2}$ .

## Ties (Sign test)

- Discard zero Z values and redefine n.
- ▶ If too many zeros, choose alternative statistical procedure (Chapter 10)



### References for this lecture

HWC: Chapter 1.2

HWC: Chapter 1.3

HWC: Chapter 3.4-3.6