2019

Question 1. (a) Let $\hat{\theta}$ be an estimator of an unknown parameter θ.

i. Define the Bias of $\hat{\theta}$, $\mathrm{Bias}(\hat{\theta})$; [2 marks]

ii. Define the mean squared error of $\hat{\theta}$, $\mathrm{MSE}(\hat{\theta})$; [2 marks]

See lecture notes

(b) Let $X_1, \ldots, X_n$ be an independent random sample from a population X with mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Let

$$\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$$

be an estimator of μ

i. Calculate the Bias of $\bar{X}$, $\mathrm{Bias}(\bar{X})$; [4 marks]

ii. Calculate the mean squared error of $\bar{X}$, $\mathrm{MSE}(\bar{X})$. [6 marks]

Solution

According to the problem statement

$$E(X_i) = \mu \qquad Var(X_i) = \sigma^2$$

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

i) $\mathrm{Bias}(\bar{X}) = E(\bar{X}) - \mu \Rightarrow$

$$\mathrm{Bias}(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu = \frac{1}{n}\sum_{i=1}^{n} E(X_i) - \mu =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mu - \mu = \frac{1}{n}\cdot n\mu - \mu = 0$$

ii) $\mathrm{MSE}(\bar{X}) = Var(\bar{X}) + \mathrm{Bias}^2(\bar{X}) = Var(\bar{X}) =$

$$= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$L(x_1, \ldots, x_n \mid \alpha) = \ldots$

(c) A continuous random variable Y has density function

$$f_Y(y) = \begin{cases} 2\alpha e^{-\alpha y}, & y > 0, \\ 0, & y \leq 0, \end{cases}$$

where $\alpha > 0$ is an unknown parameter to be estimated. Suppose $y_1, \ldots, y_n$ are observations of independent sample random variables $Y_1, \ldots, Y_n$, respectively, all with the same distribution as Y.

i. Find the maximum likelihood ~~estimate~~ estimator for α. [8 marks]

ii. Find the maximum likelihood ~~estimator~~ estimate for α. [3 marks]

Solution:

i) $L(\alpha) = \prod_{i=1}^{n} 2\alpha e^{-\alpha y_i} = 2^n \alpha^n e^{-\alpha \sum_{i=1}^{n} y_i}$

$\ln L(\alpha) = n \ln 2 + n \ln \alpha - \alpha \sum_{i=1}^{n} y_i$

$\dfrac{d \ln L(\alpha)}{d\alpha} = \dfrac{n}{\alpha} - \sum_{i=1}^{n} y_i = 0 \Rightarrow$

$\hat{\alpha} = \dfrac{n}{\sum_{i=1}^{n} y_i}$

ii) $\bar{y} = \dfrac{\sum_{i=1}^{n} y_i}{n} \Rightarrow$ estimate $\hat{\alpha} = \dfrac{1}{\bar{y}}$

observer sample mean

Question 2. (a) Suppose that $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$. Let $X_1, \ldots, X_5$ be an independent random sample from X, with sample variance S_x^2 and $Y_1, \ldots, Y_9$ be an independent random sample from Y, with sample variance S_y^2. Assume that X_i and Y_j are mutually independent for all $i = 1, \ldots, 5$ and $j = 1, \ldots, 9$.

i. Using the results $\frac{4S_x^2}{\sigma_x^2} \sim \chi_4^2$ and $\frac{8S_y^2}{\sigma_y^2} \sim \chi_8^2$ show that

$$\frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} \sim F_{4,8}.$$

[3 marks]

ii. Find the critical value b such that $P(F_{4,8} > b) = 0.05$. [2 marks]

iii. Find the critical value a such that $P(F_{4,8} > a) = 0.95$. [3 marks]

iv. Hence, if $S_x^2 = 10$ and $S_y^2 = 5$, compute a 90% confidence interval for σ_x^2/σ_y^2. [5 marks]

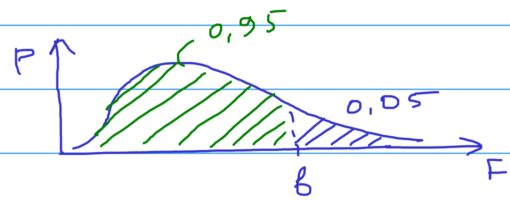
Solution:

i) As X_i and Y_j are mutually independent

$\Rightarrow \frac{4S_x^2}{\sigma_x^2}$ and $\frac{8S_y^2}{\sigma_y^2}$ are independent

$$\Rightarrow F = \frac{\frac{4S_x^2}{\sigma_x^2} \Big/ 4}{\frac{8S_y^2}{\sigma_y^2} \Big/ 8} \sim F_{4,8}$$

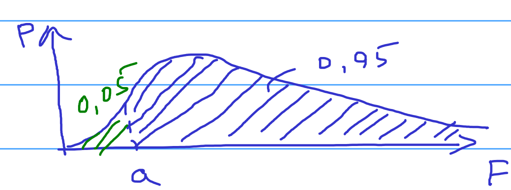
ii) From a table



$b = F_{0{,}95,4,8} = 3.838$

$P(F > 3{,}838) = 0{,}05$

iii)



$a = F_{0{,}05,4,8} = \frac{1}{F_{0{,}95,8,4}}$

$a = \frac{1}{6{,}041} = 0{,}166 \Rightarrow P(F > 0{,}166) = 0{,}95$

iv)

$$P(0{,}166 < F < 3.838) = P\left(0{,}166 < \frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} < 3.838\right) = 0{,}95$$

$S_x^2 = 10$, $S_y^2 = 5$

$$P\left(0{,}166 \cdot \frac{5}{10} < \frac{\sigma_y^2}{\sigma_x^2} < 3.838 \cdot \frac{5}{10}\right) = P\left(\frac{2}{3.838} < \frac{\sigma_x^2}{\sigma_y^2} < \frac{2}{0{,}166}\right) = 0{,}95$$

$$\Rightarrow \quad P\left(0{,}521 < \frac{\sigma_x^2}{\sigma_y^2} < 12{,}048\right) = 0{,}95$$

$\Rightarrow$ the 95% CI for $\frac{\sigma_x^2}{\sigma_y^2}$ is $(0.521, 12.048)$.

(b) If there is no seasonal effect on human births, we would expect equal numbers of children to be born in each of the four seasons (winter, spring, summer and fall). A student took a survey from his 1st year class and found that, of the 120 students in the class, 25 were born in winter, 35 in spring, 32 in summer, and 28 in fall. He wondered if the excess in the spring was an indication that births were not uniform throughout the year.

i. What is the expected number of births in each season if there is no seasonal effect on birth? [2 marks]

ii. Compute the χ^2 statistic for the χ^2 goodness of fit test. [4 marks]

iii. How many degrees of freedom does the χ^2 statistic have? [1 mark]

iv. Perform a χ^2 goodness of fit test of the hypothesis that there is no seasonal effect on human births at the $\alpha = 0.05$ significance level. What do you conclude? [5 marks]

Solution:

If we assume that there is no seasonal effect, then $\theta = \frac{1}{4}$ is a part of births in each season

For the sample $n = 120$ we can fill in the table

Season	Winter	Spring	Summer	Autumn
Number of stud., O_i	25	35	32	28

i) The expected number of births for each season $= \frac{1}{4} \cdot 120 = 30 \Rightarrow$

Season	Winter	Spring	Summer	Autumn
nE_i	30	30	30	30

ii) χ^2 statistic $\quad G = \sum_{i=1}^{n} \frac{(O_i - nE_i)^2}{nE_i}$

$$g = \frac{(25-30)^2}{30} + \frac{(35-30)^2}{30} + \frac{(32-30)^2}{30} + \frac{(28-30)^2}{30}$$

$g = 1,933$

iii) There $k-1$ degrees of freedom

$\Rightarrow \quad G \sim \chi^2_3$

iv) The critical value $\chi^2_{0,05,3} = 7,815$

$g < 7,815 \Rightarrow$

What do you conclude?

Question 3

(a) Outline the steps in carrying out a statistical hypothesis test. [5 marks]

See lecture notes

(b) Let X_1 and X_2 be two random variables with distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Assuming that X_1 and X_2 are independent then $X_1 + X_2 \sim N(\mu, \sigma^2)$.

i. What is μ in terms of μ_1 and μ_2? [1 mark]

ii. What is σ^2 in terms of σ_1^2 and σ_2^2? [1 mark]

Solution:

From probability module, following properties of expectation and variance:

$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$\Rightarrow \mu = \mu_1 + \mu_2$

$\sigma^2 = \sigma_1^2 + \sigma_2^2$

(c) An animal park has two types of giraffe: the reticulated giraffe and the Masai giraffe. The height of adult female reticulated giraffes $X_r \sim N(\mu_r, \sigma_r^2)$ with mean $\mu_r = 438$cm, but unknown variance σ_r^2.

i. Suppose that there are n female reticulated giraffes in the park, with sample mean $\bar{X}_r$ and sample variance S_r^2, what distribution does the following statistic have?

$$T = \frac{\bar{X}_r - \mu_r}{S_r/\sqrt{n}}.$$

[2 marks]

ii. The heights of 9 adult female Masai giraffes in the park are measured and recorded below:

435cm, 440cm, 450cm, 425cm, 460cm, 465cm, 455cm, 425cm, 450cm

Calculate the sample mean $\bar{x}_m$ and the sample variance s_m^2 of this dataset.

[4 marks]

Solution:

i) As there is no information about we need to consider 2 cases:

if $n > 30 \Rightarrow T$ can be approximated by the standard normal distribution

if $n < 30 \Rightarrow T$ has t-distribution with $n-1$ degrees of freedom

ii) $\bar{x}_m = \frac{\sum_{i=1}^{n} x_i}{n} = 445$ cm $\qquad S_m^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = 212{,}5$

iii. Assuming the height of adult female Masai giraffes is also normally distributed, with mean μ_m and variance $\sigma_m^2 = \sigma_r^2$, perform a hypothesis test at the 5% significance level with the null hypothesis $\mu_m = \mu_r$ and the alternative hypothesis $\mu_m > \mu_r$. What is the conclusion?

[6 marks]

$H_0: \quad \mu_m = \mu_r = 438\,\text{cm}$

$H_a: \quad \mu_m > \mu_r$

Test statistic : $\quad T = \dfrac{\bar{X}_m - \mu_m}{S_m/\sqrt{n}} \sim t_{n-1}$

under H_0

$$T_{obs} = \frac{\bar{x} - \mu_r}{S_m/\sqrt{n}} = \frac{445 - 438}{\sqrt{\frac{212{,}5}{9}}} \approx 1{,}44$$

The critical value $\quad t_{crit} = t_{0{,}95{,}8} = 1{,}860$

$\Rightarrow T_{obs} < t_{crit} \Rightarrow$

What do you conclude?

iv. A new collection of 6 female Cape giraffes is brought to the wildlife park. The heights of these giraffe are measured and the mean height and sample variance are found to be $\bar{x}_c = 455$cm and $s_c^2 = 200\cancel{\text{cm}}$, respectively. Using the same sample data as in part (ii), perform a suitable hypothesis test at the 5% significance level to test the null hypothesis $\sigma_c^2 = \sigma_m^2$ against the alternative hypothesis $\sigma_c^2 < \sigma_m^2$, where σ_c^2 is the variance of the height of Cape giraffes. What is the conclusion?

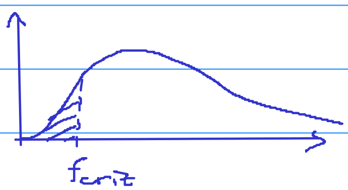
[6 marks]

$H_0: \sigma_c^2 = \sigma_m^2$

$H_a: \sigma_c^2 < \sigma_m^2$ left tailed

Test statistic $T = \dfrac{S_c^2/\sigma_c^2}{S_m^2/\sigma_m^2} \sim F_{6-1,9-1}$

Under H_0 $\Rightarrow$ $T_{obs} = \dfrac{S_c^2}{S_m^2} = \dfrac{200}{212{,}5} = 0{,}941$



$f_{crit} = F_{0{,}05,5,8} = \dfrac{1}{F_{0{,}95,8,5}} = \dfrac{1}{4{,}818} = 0{,}208$

$T_{obs} > f_{crit} \Rightarrow$

What do you conclude?

Alternatively :

$T = \dfrac{S_m^2/\sigma_m^2}{S_c^2/\sigma_c^2} \sim F_{8,5} \Rightarrow T_{obs} = \dfrac{212{,}5}{200} = 1.0625$

$f_{crit} = F_{0{,}95,8,5} = 4.818 \Rightarrow T_{obs} < f_{crit} \Rightarrow$

What do you conclude?