Vinayak Nayak
2021fc04135

INTRODUCTION TO STATISTICAL METHODS

Assignment-2

1. 46 women are randomly selected with a mean height of 86" and a standard deviation of 6.2" each. Find the limits in which the heights of the women lie with 95% Confidence Interval.

Assume that the population follows a normal distribution. We can then say that here,

$\bar{x} = 86$, $n = 46$ and $\sigma = 6.2$

Also 95% CI $\Rightarrow z = 1.96$

$\therefore$ The Lower limit of interval $= \bar{x} - \frac{z\sigma}{\sqrt{n}}$

$= 86 - \frac{1.96 \times 6.2}{\sqrt{46}}$

$= 84.208$

Upper limit of interval $= \bar{x} + \frac{z\sigma}{\sqrt{n}}$

$= 86 + \frac{1.96 \times 6.2}{\sqrt{46}}$

$= 87.792$

Ans] The CI for the given claim (95%) is (84.208, 87.792)

Vinayak Nayak
2021fc04135

2. Consider two continuous random variables X & Y with joint pdf

$$f(x,y) = 2/81\, x^2 y \qquad x \in (0,k)\,;\ y \in (0,k)$$
$$= 0 \qquad \text{otherwise.}$$

(a) Find k so that $f(x,y)$ is a valid joint pdf

For a valid joint pdf, $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$

$$\Rightarrow \int_{-\infty}^{0}\int_{-\infty}^{0} f(x,y)\,dx\,dy + \int_{0}^{k}\int_{0}^{k} f(x,y)\,dx\,dy + \int_{k}^{\infty}\int_{k}^{\infty} f(x,y)\,dx\,dy = 1$$

$$\Rightarrow 0 + \int_0^k\int_0^k \frac{2}{81} x^2 y\,dx\,dy + 0 = 1$$

$$\Rightarrow \int_0^k \left[\frac{2x^3 y}{81 \cdot 3}\right]_0^k dy = 1$$

$$\Rightarrow \int_0^k \frac{2k^3}{243}\, y\,dy = 1$$

$$\Rightarrow \frac{2k^3}{243}\left[\frac{y^2}{2}\right]_0^k = 1$$

$$\Rightarrow \frac{\cancel{2}k^3 \cdot k^2}{243 \cdot \cancel{2}} = 1$$

$$\Rightarrow k^5 = 243$$

$$\Rightarrow \boxed{k = \sqrt[5]{243} = 3}$$

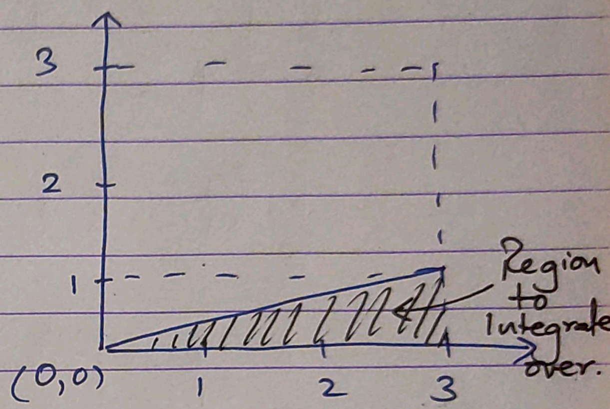
For the given ~~pdf~~ function to be a valid joint pdf, $k = 3$.

Vinayak Nayak
2021fc04135

(b) Find $P(x > 3y)$

Both x & y when in range $(0,3)$, we know we have a non-zero probability density and everywhere else it is zero.

$$\therefore P(x>3y) = \int_0^3 \int_{3y}^3 f(x,y)\,dx\,dy$$



$$= \int_0^1 \int_{3y}^3 \frac{2}{81} \cdot x^2 y\,dx\,dy$$

$$= \int_0^1 \frac{2y}{81} \cdot \left[\frac{x^3}{3}\right]_{3y}^{3} dy$$

$$= \int_0^1 \frac{2y}{81} \left(\frac{27-27y^3}{3}\right) dy$$

$$= \int_0^1 \frac{2y}{\cancel{243}_{9}} \cdot \cancel{27} \cdot (1-y^3)\,dy$$

$$= \int_0^1 \frac{2}{9}(y-y^4)\,dy$$

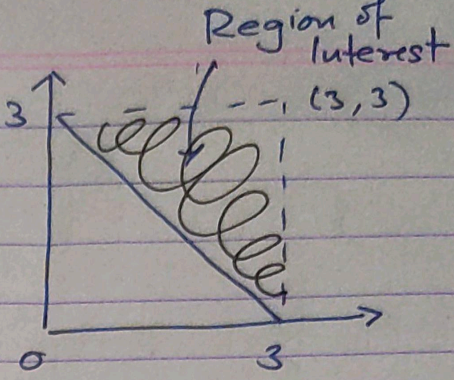
$$= \frac{2}{9}\left[\frac{y^2}{2} - \frac{y^5}{5}\right]_0^1$$

$$= \frac{2}{9}\left[\frac{1}{2} - \frac{1}{5}\right]$$

$$= \frac{\cancel{2}}{\cancel{9}_3} \cdot \frac{\cancel{3}}{\cancel{10}_5} = \frac{1}{3 \cdot 5} = \frac{1}{15}$$

$$\boxed{\therefore P(x>3y) = \frac{1}{15}}$$

Vinayak Nayak
2021fc04135



(b) $P(x+y>3) = ?$

As seen in region of interest,
x limits = $(3-y, 3)$ · y limits $(0,3)$

$$\Rightarrow P(x+y>3) = \int_0^3 \int_{3-y}^3 \frac{2x^2y}{81}\, dx\, dy$$

$$= \int_0^3 \frac{2y}{81} \cdot \left.\frac{x^3}{3}\right|_{3-y}^{3} dy$$

$$= \int_0^3 \frac{2y}{243} \cdot \left(3^3 - (3-y)^3\right) dy$$

$$= \int_0^3 \frac{2y}{243} \left(27 - (27 - 27y + 9y^2 - y^3)\right) dy$$

$$= \int_0^3 \frac{2y}{243} \left(27y - 9y^2 + y^3\right) dy$$

$$= \frac{2}{243} \int_0^3 (27y^2 - 9y^3 + y^4)\, dy$$

$$= \frac{2}{243} \left[27 \cdot \left.\frac{y^3}{3}\right|_0^3 - 9 \cdot \left.\frac{y^4}{4}\right|_0^3 + \left.\frac{y^5}{5}\right|_0^3 \right]$$

$$= \frac{2}{243} \left[\frac{27 \cdot \cancel{27}^{9}}{\cancel{3}} - \frac{9 \cdot 81}{4} + \frac{243}{5} \right]$$

$$= \frac{2}{243} \left[\frac{243 \cdot 1}{4} + \frac{243}{5} \right]$$

$$= \frac{1}{2} + \frac{2}{5} = \frac{9}{10} = 0.9 \Rightarrow \boxed{P(X+Y>3) = 0.9}$$

Vinayak Nayak
2021fc04135

(d) Are X and Y independent? If or if not, mention Cov (X,Y)

$$f(x,y) = \int_0^3 \int_0^3 \frac{2x^2 y}{81}\, dx\, dy$$

Let's find marginal distributions.

$$f(x) = \int_0^3 \frac{2x^2 y}{81}\, dy = \frac{2x^2}{81} \cdot \frac{y^2}{2}\Big|_0^3 = \frac{\cancel{2}x^2}{\cancel{81}9} \cdot \frac{\cancel{9}}{\cancel{2}} = \frac{x^2}{9}$$

$$f(y) = \int_0^3 \frac{2x^2 y}{81}\, dx = \frac{2y}{81} \cdot \frac{x^3}{3}\Big|_0^3 = \frac{2y}{\cancel{81}9} \cdot \frac{\cancel{27}\cancel{9}}{\cancel{3}} = \frac{2y}{9}$$

$$f(x) \cdot f(y) = \frac{x^2}{9} \cdot \frac{2y}{9} = \frac{2x^2 y}{9} = f(x,y)$$

$$\Rightarrow \boxed{f(x,y) = f(x) \cdot f(y)}$$

Joint distribution is the product of marginal distributions.
$\Rightarrow$ X. & Y are independent
$\because$ These are independent $\Rightarrow$ Cov (X, Y) = 0.

Vinayak Nayak
2021fc04135

3. A public health official claims that mean home water use = 300 gallons per day. To verify this claim, a study involving 12 homes was instigated and average daily water uses of these household were as follows

275, 280, 277, 301, 258, 264, 273, 306, 295, 281, 284, 312

Do the data contradict official statement with 1% level of significance?

Observations:-

275, 280, 277, 301, 258, 264, 273, 306, 295, 281, 284, 312

Sample mean $= \frac{\sum x_i}{n} = \frac{3406}{12} = 283.33$

Sample standard deviation (s) $= \sqrt{\frac{\sum(x - x_i)^2}{n-1}} = 16.59$

Claim $\mu_0 = 300$ no of observations $n = 12$

Null hypothesis: $\mu = \mu_0$

Alternate hypothesis: $\mu > \mu_0$ (Consider mean home water use is more than 300 gallons/day)

Test stat $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

$= \frac{283.33 - 300}{16.59/\sqrt{12}} = \frac{-16.67}{16.59/\sqrt{12}} = \frac{-16.67}{4.789}$

$= -3.48$

Degrees of freedom $= n - 1 = 12 - 1 = 11$

Critical regim $= t^{11}_{0.01} = 2.718$

We can see that $t^{11}_{0.01} > t \Rightarrow$ We will accept the null hypoth

$\therefore$ The data does not contradict the official statement at 1% level of significance.

Vinayak Nayak
2021fc04135

4. Data was collected to determine if there's a difference between mean IQ scores of children in two institutions A & B. A random sample of 100 children from institute A and 60 children from institute B are taken and following data is collected.

	Institute A	Institute B
n	100	60
Mean	102.2	105.3
Stand. Dev.	11.8	10.6

Is the data significant enough at 5% level for us to reject the hypothesis that mean scores of children from A and B are the same?

Solution:

Null Hyp: There is no significant difference between mean scores of childeen ($\mu_A = \mu_B$)

Alt. Hyp: The two means are different $\mu_A \neq \mu_B$.

$\because$ Both institute A & institute B has a sufficiently large no. of examples ($n \geq 30$), we can select a z-statitstic and this is a two tailed test.

For 5% level of significance, for z we know limits are -1.96 and 1.96 respectively.

$$\text{Stat } z = \frac{\overline{X_A} - \overline{X_B}}{S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Now, $$S_p = \sqrt{\frac{(n_1-1)\, s_1^2 + (n_2-1)\, s_2^2}{(n_1-1) + (n_2-1)}}$$

$$= \sqrt{\frac{(100-1) \cdot 11.8^2 + (60-1) \cdot 10.6^2}{(100-1) + (60-1)}}$$

Vinayak Nayak
2021fc04135

$$\therefore S_p = \sqrt{\frac{99 \times 11.8^2 + 59 \times 10.6^2}{99 + 59}}$$

$$= \sqrt{\frac{20414}{158}} = 11.367$$

$$z \text{ stat} = \frac{102.2 - 105.3}{11.367 \times \sqrt{\frac{1}{100} + \frac{1}{60}}}$$

$$= \frac{-3.1}{11.367 \times 0.1633} = \frac{-3.1}{1.856}$$

$$z \text{ stat} = -1.670$$

Our statistic lies within [-1.96, 1.96]. This means we cannot reject the null hypothesis.

⇒ Given the data, we cannot say that the mean scores of children from A & B differ much or we can say mean scores of children from A & B are same.