Probability for ML Assignment

Ibrahim
Sobh

① I) a) $\mu_x = 68{,}4$; $\mu_y = 60{,}2$

- $Var(X) = \dfrac{\sum_{i=1}^{i=5} (x_i - \mu_x)^2}{m}$ $\longrightarrow$ (number of specimen)

$$= \frac{\left((44-68{,}4)^2 + (65-68{,}4)^2 + (71-68{,}4)^2 + (75-68.4)^2 + (87-68.4)^2\right)}{5}$$

$\boxed{Var(X) = 200.64}$

- $Var(Y) = \dfrac{\sum_{i=1}^{i=5} (y_i - \mu_y)}{m}$

$$= \frac{\left((40-60{,}2)^2 + (60-60{,}2)^2 + (59-60{,}2)^2 + (65-60{,}2)^2 + (77-60{,}2)^2\right)}{5}$$

$\boxed{Var(Y) = 142.96}$

- $Cov(X,Y) = \frac{1}{m}\left(\sum_{i=1}^{i=5} (x_i - \mu_x)(y_i - \mu_y)\right)$

$$= \frac{\left((44-68{,}4)(40-60{,}2) + (65-68{,}4)(60-60{,}2) \ldots\right)}{5}$$

$\boxed{cov(X,Y) = 166.92}$ $\longrightarrow$ (Equal behavior)

I) b) $\text{cov}(X,Y) = 166.92 > 0$ means

if $x \nearrow \Rightarrow y \nearrow$ (Extra information)
and if $x \searrow \Rightarrow y \searrow$

$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{200.64} = 14.16$

$\sigma_Y = \sqrt{\text{Var}(Y)} = \sqrt{142.96} = 11.96$

$$\text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{166.92}{169,35}$$

$$\boxed{\text{Corr}(X,Y) \simeq 0,9856}$$

c) in b) we got $\text{Corr}(X,Y) \simeq 0,9856$

which is very close to $\underline{\underline{1}}$

Which means that X and Y behave the same or have the same behavior (strong positive association)

d) if the $\text{Corr}(X,Y) = 0$ it means that X and Y dont have a lineaire relationship (Uncorrelated)

(on a side note $\text{Corr}(X,Y) = 0$ does not imply that X, Y are independent)

2 II) a) I will probably try to take samples of this data (1000 students grades) Let us say, 10 as sample size and I take at least 30 samples or more ($n \geq 30$) (n = number of samples) I do the average of each sample and I apply the Central limit theorem that clearly state, I can deal with my data as a normal distribution thanks I can see where most of the grades will sit between using the (68.5%, 95%, 99%) rule

when $\sigma_{s_n} = \sqrt{n}\,\sigma$ and $\sigma = \sqrt{\frac{(x-\bar{x})^2}{n-1}}$

- the more I take samples the more accuracy I will get on my curve.

b) I will deal with this as a Uniform random Variable distribution where

X: the grade of a student

and ($0 \leq X \leq 20$) (Also Uniform)

where $P(0 \leq x \leq 20) = 1$

$P(0) = P(1) = P(20) = \frac{1}{20}$

c) $P(X \geq 12) = P(X = 12) + P(X > 12)$

$= \frac{1}{20} + \frac{8}{20} = \frac{9}{20}$

(because in the example P is equally likely.

$P(X \geq 12) = \frac{9}{20} \simeq 0{,}45$

d) $E(x) = \dfrac{\overset{\text{lowest}}{a} + \overset{\text{height}}{b}}{2} = \dfrac{0+20}{2} = 10$

When we have equal probability for each event.

$f(x) = \dfrac{1}{b-a} = \dfrac{1}{20}$

III a)

$\{ S$: Person is sick $\{ H$: Person healthy

$P(S) = \dfrac{100{,}000}{10{,}000{,}000} = 0{,}01 \simeq 1\%$ of population is sick

$+/H$: False positive

$-/S$: False negative

$P(H) = 0{,}99 \simeq 99\%$

$\begin{cases} + : \text{positive test} \\ - : \text{negative test} \end{cases}$

$P(+/S) = 0{,}99$

$P(+/H) = 0{,}05$

bayes

$P(H/-) = \dfrac{P(-/H) \cdot P(H)}{P(-)}$

$\begin{cases} P(-/S) = 1 - P(+/H) = 0{,}01 \\ P(-/H) = 1 - P(+/H) = 0{,}95 \end{cases}$

$P(-) = P(-/S)\,P(S) + P(-/H) \cdot P(H)$

$= 0.01 \cdot 0{,}01 + 0{,}95 \cdot 0{,}99$

$= 0{,}0001 + 0{,}9405 = 0{,}9406$

③

III) a) $P(S) = 0{,}01$
$P(H) = 0{,}99$

Combinees

$P(+/S) = 0{,}99$
$P(+/H) = 0{,}05$

$P(-/S) = 0{,}01$
$P(-/H) = 0{,}95$

$P(-) = 0{,}9406$

$$P(H/-) = \frac{P(-/H) \cdot P(H)}{P(-)}$$

$$= \frac{0{,}95 \cdot 0{,}99}{0{,}9406} = 0{,}999$$

$$\simeq 99.9\%$$

→ it is actually really good to have tested negative and actually the test being correct (you are healthy 99.9 times out of 100) good test results

b) $$P(S/+) = \frac{P(+/S) \cdot P(S)}{P(+)}$$

$$P(+) = P(+/S)\,P(S) + P(+/H) \cdot P(H)$$
$$= 0{,}99 \cdot 0{,}01 + 0{,}05 \cdot 0{,}99$$
$$= 0{,}0099 + 0{,}0495$$
$$P(+) \simeq 0{,}0594$$

- $P(S/+) = \frac{0,99 \cdot 0,01}{0,0594} = 0,166$

$\simeq 16,6\%$

- Pretty low percentage of
White means → false negatives results
are really heigh ↗↗
only 16,6% of whom tested
positive are actually sick.
test accuracy of detecting sick
people is bad (83,9% False negative)

c)

$$P(H_2/+) = \frac{P(+/H_2) \cdot P(H_1)}{P(+)}$$

When $(P(H_1) = P(H/+))$

First test

$$P(H_2/+) = \frac{P(+/H_2) \cdot P(H_1/+)}{P(+)}$$

- $P(H/+) = 1 - P(H/-) = 0,001$

$$P(H_2/+) = \frac{0.05 \times 0.001}{0,0594} = 0,0008$$

$\simeq 0,08\%$

0,08% the ones tested (+) are false positives

the test accuracy of detecting False positives is great as only 0,08% of