VATSAV SETHUPATHI

Date
Page

MATH-DATA 363 Exam 1

① (a) Fraction of games won ⇒ $\frac{\text{No. of games with +ve GD}}{\text{Total no. of games}}$

$= \frac{17+6+7+5}{82} = \frac{35}{82} \approx 42.9\%$ games

(b) 1st Quartile ⇒ $1/4 \cdot 82 \Rightarrow 20.5 \Rightarrow \frac{20\text{th game} + 21\text{st game}}{2} = \frac{-2-3}{2} = -2.5$

Median ⇒ $1/2 \cdot 82 = \frac{41\text{st} + 42\text{nd}}{2}$

$= \frac{-1-1}{2} = \boxed{-1}$

3rd Quartile $= 3/4 \cdot 82 = 60.5 = \frac{61\text{th} + 62\text{nd}}{2} = \frac{1+1}{2} = \boxed{1}$

5 number summary

Min: $\boxed{-6}$ 1st Quart: $\boxed{-2.5}$ Median: $\boxed{-1}$ 3rd Quar: $\boxed{1}$ Max: $\boxed{4}$

(c) Mean $= \frac{1}{n}\sum_{i=1}^{n} x_i\, n(x_i) = \frac{(-6\cdot 2)+(-5\cdot 0)+\cdots+(3\cdot 7)+(4\cdot 5)}{82}$

x_i	$n(x_i)$	
-6	2	-12
-5	0	0
-4	4	-16
-3	14	-42
-2	9	-18
-1	18	-18
0	0	0
1	17	17
2	6	12
3	7	21
4	5	20

$= \frac{-21}{82} = \boxed{-0.256}$ is the mean goal difference

(d) (i) Median
Would still remain the same since that is the average of the 41st & 42nd term which were not altered by this change

(ii) The Mean
The value of the mean would increase since we are reducing the sum of total goal difference by improving the margin of loss in 2 of the games

(iii) Standard deviation
This would decrease since by improving the goal difference, we are reducing the spread of the data, meaning the standard deviation will decrease.

Here MGP = Mean Gestation Period
LE = Life Expectancy

② (a)

MGP (x_i)	LE (y_i)	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
214	22	101	10201	8	64	808
56	9	-57	3249	-5	25	285
35	12	-78	6084	-2	4	156
240	30	127	16129	16	256	2032
108	10	-5	25	-4	16	20
25	1	-88	7744	-13	169	1144
			43432			4445

$$\text{Cov}(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \frac{4445}{5} = 889$$ [Working shown above]

(b) For regression line →

$$\hat{\beta} = \frac{\text{cov}(x,y)}{\text{Var}(x)} = \frac{889}{8686.4} = \boxed{0.102}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 14-(0.102)(113) = \boxed{2.474}$$

$$\left[\text{Var}(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{43432}{5} = 8686.4\right]$$

Eq of regression line: $\widehat{LE} = 2.474 + (0.102)\,\widehat{MGP}$

(c) The slope says that for every additional day in gestation, the life expectancy increases by 0.102 years.

(d) $\widehat{LE} = 2.474 + (0.102)(35) = 6.044$ years (predicted)

Our predicted Life expectancy for the koala is approx 6 years. When compared to the actual data, this means that the koala has a life expectancy of 6 years (residual = actual - predicted) more than what we predicted with our regression line.

③ $P\{F\} = 0.52 \quad P\{F^c\} = 0.48$

$P\{B|F\} = 1/200 \qquad P\{B|F^c\} = 1/12$

(a) $P\{B^c|F\} = 1 - P\{B|F\} = 1 - 1/200 = 199/200$

(b) $P\{B\} = P\{B|F\}P\{F\} + P\{B|F^c\}P\{F^c\}$

$= 1/200 \cdot 0.52 + 1/12 \cdot 0.48$

$= \boxed{0.0426}$

(c) $P\{F|B\} = \dfrac{P\{B|F\}P\{F\}}{P\{B\}}$

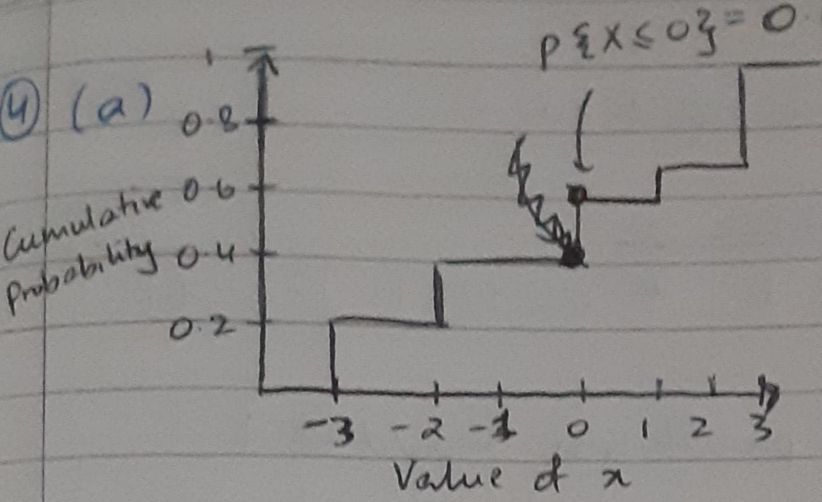
$= \dfrac{P\{B|F\}P\{F\}}{P\{B|F\}P\{F\} + P\{B|F^c\}P\{F^c\}}$

$= \dfrac{1/200 \cdot 0.52}{0.0426} = \dfrac{0.0026}{0.0426}$

$= \boxed{0.061}$

(d) The probability will decrease.

④ (a)



(b)

x	$f_X(x)$	$x f_X(x)$	$x^2 f_X(x)$
-3	0.2	-0.6	1.8
-2	0.2	-0.4	0.8
-1	0	0	0
0	0.2	0	0
1	0.1	0.1	0.1
2	0	0	0
3	0.3	0.9	2.7
		0	5.4

$EX = \sum_x x f_X(x) = \boxed{0}$

$EX^2 = \sum_x x^2 f_X(x) = 5.4$

(c)

$Var(x) = EX^2 - (EX)^2 = 5.4 - 0 = \boxed{5.4}$

(d) $E[3-2X] = 3 - 2EX = 3 - 2(0) = \boxed{3}$

$Var(3-2X) = 4\,Var(X) = 4 \cdot 5.4 = \boxed{21.6}$

⑤ (a) For this to be a valid cdf ⇒

~~c(2+(-1)+(-1)^3) = 0~~

$c(2+1+1^3) = 1 \Rightarrow c(4) = 1$ $\boxed{c = 1/4}$

Since the final probability should be 1

(b) $P\{X > 1/4\} = 1 - P\{X \leq 1/4\}$

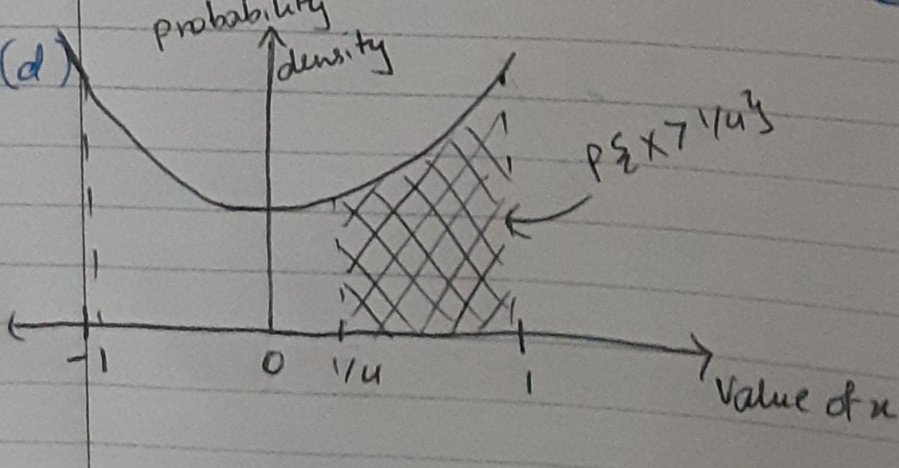
$= 1 - F_X(1/4)$

$\Rightarrow 1 - \frac{1}{4}(2 + 1/4 + 1/64) \Rightarrow 1 - \frac{145}{256} = \boxed{\frac{111}{256}}$

(c) $f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx}\left(\frac{1}{4}(2+x+x^3)\right)$

$= \frac{d}{dx}\left(\frac{1}{2} + \frac{x}{4} + \frac{x^3}{4}\right) = \frac{1}{4} + \frac{3x^2}{4} \Rightarrow \boxed{\frac{1}{4}(1+3x^2)}$

(d)



Mean $= \frac{1}{n}\sum_{i=1}^{n} x_i = \sum_x x p(x)$

$\mathrm{Var}(x) = \boxed{\frac{1}{n-1}} \sum_x (x_i - \bar{x})^2 n(x) = (sd)^2$

$\mathrm{cov}(x,y) = \boxed{\frac{1}{n-1}} \sum (x_i - \bar{x})(y_i - \bar{y})$

EX(Bernoulli) = $0(1-p) + 1(p) = p$

n trials = $EX = p \cdot n$

$\mathrm{Var}(x) = np(1-p)$

r (correl) = $\frac{\mathrm{cov}(x,y)}{S_x S_y}$ | $S^2_{x+y} = S^2_x + S^2_y + 2 r S_x S_y$

Regression: y (Response), x (explanatory)

$\hat{\beta} = \frac{\mathrm{cov}(x,y)}{\cancel{S_x}\ \mathrm{Var}(x)}$ $\quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ $\quad y_i = \alpha + \beta x_i + \epsilon_i$

$S^2_{FIT} = r^2 \cdot S^2_{DATA}$

$\hat{\beta} = r \frac{S_y}{S_x}$ | $S^2_{RES} = (1-r^2) S^2_{DATA}$

If X_1 & X_2 indep $\mathrm{Cov}(X_1, X_2) = 0$

$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$

P(test − | have) = false −ve | true +ve = 1 − FN

P(test + | don't) = false +ve | true −ve = 1 − FP

$P(A|C) = \frac{P(C|A)P(A)}{P(C|A)P(A) + P(C^c|A)P(A)}$ | Dist = $\frac{d}{dx}$ (density)

Binomial ~~Bernoulli~~ MF $\binom{n}{x} p^x (1-p)^{n-x} = P\{S_n = x\}$

Prob Trans = Inv of Dist Funct