UNIT–2 : Probability & Sampling Distributions (PYQ Solution)

PART–A Answers (≈70 Words Each)

Q1. From a lot of 25 items containing 5 defectives, a sample of 4 items is drawn at random (i) without replacement (ii) with replacement. Find the expected number of defectives. (2020, old scheme)

Let X be the number of defectives in the sample.

(i) Without replacement: X is hypergeometric (N = 25, K = 5, n = 4).

$$E(X) = n\frac{K}{N} = 4 \times \frac{5}{25} = 0.8, \quad Var(X) = n\frac{K}{N}\left(1 - \frac{K}{N}\right)\frac{N-n}{N-1} = 0.56$$

(ii) With replacement: X is binomial (n = 4, p = 5/25 = 0.2).

$$E(X) = np = 0.8, \quad Var(X) = npq = 0.64$$

The expected value is the same (0.8) in both cases; only the variance differs.

Q2. Calculate mean and variance of rectangular (uniform) distribution. (2019, old scheme)

For a uniform distribution on $[a, b]$: $f(x) = \dfrac{1}{b-a}$.

$$E(X) = \int_a^b \frac{x}{b-a}dx = \frac{a+b}{2}, \quad E(X^2) = \frac{a^2+ab+b^2}{3}$$

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

Example: for uniform on [0, 6], mean = 3, variance = 3.

PART–B Answers (≈250 Words Each)

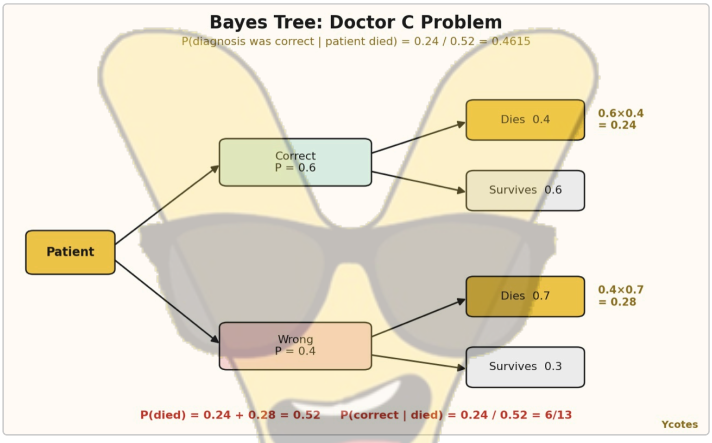
Q1. The chances that doctor C diagnoses disease X correctly is 60%. The chance that a patient dies after correct diagnosis is 40% and after wrong diagnosis 70%. A patient of doctor C, who had disease X, died. What is the chance that his disease was diagnosed correctly? (2019, old scheme)

Let C = correct diagnosis, W = wrong diagnosis, D = patient died.
$P(C) = 0.6,\ P(W) = 0.4,\ P(D|C) = 0.4,\ P(D|W) = 0.7.$

Bayes' theorem:

$$P(C|D) = \frac{P(C)P(D|C)}{P(C)P(D|C) + P(W)P(D|W)} = \frac{0.6 \times 0.4}{0.6 \times 0.4 + 0.4 \times 0.7} = \frac{0.24}{0.52} = \frac{6}{13} \approx 0.4615$$

Answer: the chance the disease was diagnosed correctly is about **46.2%**.



Q2. A firm produces steel pipes in three plants with daily output 500, 1000 and 2000 units and defective fractions 0.005, 0.008 and 0.010. A pipe is selected at random from a day's production and is defective. What is the probability that it comes from the first plant? (2020, old scheme)

Total output = 3500. $P(P_1) = 500/3500 = 1/7$, $P(P_2) = 2/7$, $P(P_3) = 4/7$.
Defective given plant: 0.005, 0.008, 0.010.

Total probability of defective:

$$P(D) = \frac{1}{7}(0.005) + \frac{2}{7}(0.008) + \frac{4}{7}(0.010) = \frac{0.005 + 0.016 + 0.040}{7} = \frac{0.061}{7}$$

Bayes' theorem:

$$P(P_1|D) = \frac{\frac{1}{7}(0.005)}{\frac{0.061}{7}} = \frac{0.005}{0.061} = 0.0820$$

Simple check with counts: defectives per day = 2.5 + 8 + 20 = 30.5, so share of plant 1 = 2.5/30.5 = **0.082**.

Q3. A random variable X has the probability distribution x: 0 to 7, P(x): 0, k, 2k, 2k, 3k, k^2, 2k^2, 7k^2+k. Find k, P(x<6), P(x>=6), P(0<x<5), the distribution function and the minimum C with P(x<=C) > 1/2. (2020, old scheme)

(i) Find k. $\sum P(x) = 1$:

$$0 + k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 10k^2 + 9k = 1 \Rightarrow 10k^2 + 9k - 1 = 0$$

$(10k - 1)(k + 1) = 0 \Rightarrow k = 0.1$ (reject $k = -1$, probability cannot be negative).

Probabilities for k = 0.1:

x	0	1	2	3	4	5	6	7
P(x)	0	0.1	0.2	0.2	0.3	0.01	0.02	0.17
F(x)	0	0.1	0.3	0.5	0.8	0.81	0.83	1.00

(ii) $P(x < 6) = F(5) = 0.81$; $P(x \geq 6) = 1 - 0.81 = 0.19$; $P(0 < x < 5) = 0.1 + 0.2 + 0.2 + 0.3 = 0.8$.

(iii) Distribution function: $F(x) = P(X \leq x)$ is the cumulative row above (a step function: 0 for x < 1, 0.1 for 1 <= x < 2, 0.3 for 2 <= x < 3, 0.5 for 3 <= x < 4, 0.8 for 4 <= x < 5, 0.81 for 5 <= x < 6, 0.83 for 6 <= x < 7, 1 for x >= 7).

(iv) $F(3) = 0.5$ is not greater than 1/2, while $F(4) = 0.8 > 1/2$. So the **minimum C = 4**.

Q4. Define binomial distribution. Find its mean and variance. (2020, old scheme)

Definition: if an experiment has n independent trials, each with only two outcomes (success p, failure q = 1 - p), the number of successes X has the binomial distribution

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \ldots, n$$

Mean:

$$E(X) = \sum x \binom{n}{x} p^x q^{n-x} = np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} q^{n-x} = np(q + p)^{n-1} = np$$

Variance: $E[X(X-1)] = n(n-1)p^2$, so

$$E(X^2) = n(n-1)p^2 + np, \quad Var(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p) = npq$$

Properties: mean > variance; it is symmetric when p = q = 1/2; for large n and small p it tends to Poisson, and for large n it is approximately normal.

Q5. A Poisson distribution has a double mode at x = 3 and x = 4. What is the probability that x will have one or the other of these two values? (2019, old scheme)

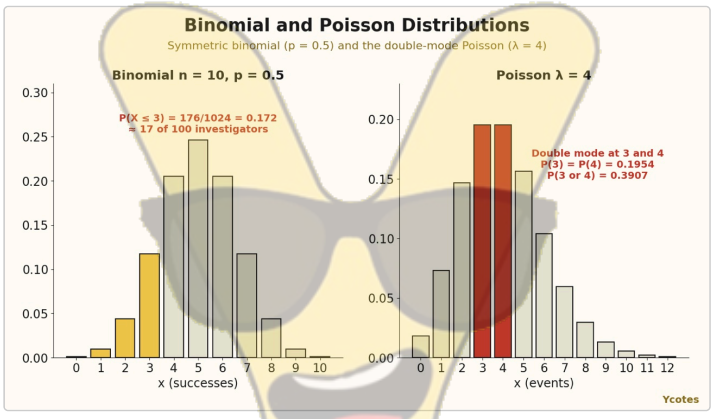
For a Poisson distribution $P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$. A double mode occurs when $P(3) = P(4)$:

$$\frac{e^{-\lambda}\lambda^3}{3!} = \frac{e^{-\lambda}\lambda^4}{4!} \Rightarrow \lambda = 4$$

(Generally the double mode is at $\lambda - 1$ and λ when λ is an integer.)

$$P(3) = \frac{e^{-4}4^3}{6} = 0.1954, \quad P(4) = \frac{e^{-4}4^4}{24} = 0.1954$$

$$P(x = 3 \text{ or } 4) = 2 \times 0.1954 = 0.3907$$



Q6. When will a binomial distribution be symmetric? If half the population of a town consumes chocolates and 100 investigators each take 10 individuals, how many investigators would you expect to report that three people or less were consumers? (2019, old scheme)

Symmetric when $p = q = 1/2$, because then $P(x) = \binom{n}{x}(1/2)^n = P(n - x)$.

Here $n = 10,\ p = 0.5$:

$$P(X \le 3) = \left[\binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{3}\right]\left(\frac{1}{2}\right)^{10} = \frac{1 + 10 + 45 + 120}{1024} = \frac{176}{1024} = 0.1719$$

Expected number of investigators $= 100 \times 0.1719 = 17.19 \approx$ **17**.

Q7. Temperature T in May is normally distributed with mean 68 and standard deviation 6. Find the probability that the temperature is (i) between 70 and 80 (ii) less than 60. (2019, old scheme)

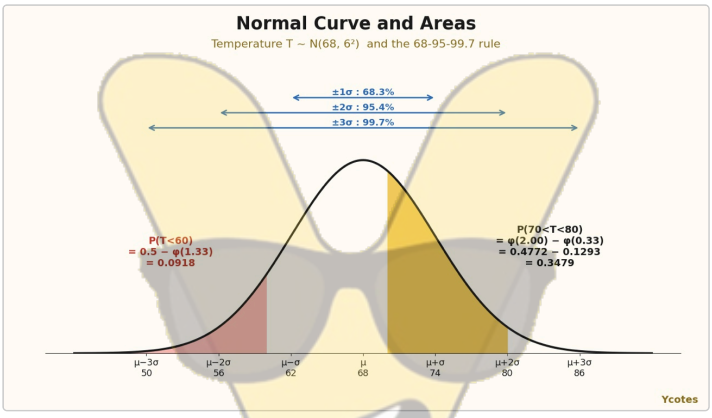
Standardise $z = \dfrac{T - 68}{6}$. Given $\phi(z) = P(0 < Z < z)$: $\phi(2.00) = 0.4772,\ \phi(0.33) = 0.1293,\ \phi(1.33) = 0.4082$.

(i) $z_1 = \dfrac{70 - 68}{6} = 0.33,\ z_2 = \dfrac{80 - 68}{6} = 2.00$

$$P(70 < T < 80) = \phi(2.00) - \phi(0.33) = 0.4772 - 0.1293 = 0.3479$$

(ii) $z = \dfrac{60 - 68}{6} = -1.33$

$$P(T < 60) = P(Z < -1.33) = 0.5 - \phi(1.33) = 0.5 - 0.4082 = 0.0918$$



Q8. Skills are classified A, B, C by breadth index: under 75, between 75 and 80, over 80. Assuming normality, find the mean and standard deviation if A are 58%, B 38% and C 4%. (2020, old scheme)

Let the index be $N(\mu, \sigma^2)$. Given $f(z) = P(0 < Z < z)$: $f(0.20) = 0.08,\ f(1.75) = 0.46$.

- $P(X < 75) = 0.58 = 0.5 + 0.08$, so $z = 0.20$: $\dfrac{75 - \mu}{\sigma} = 0.20$
- $P(X > 80) = 0.04 = 0.5 - 0.46$, so $z = 1.75$: $\dfrac{80 - \mu}{\sigma} = 1.75$

Subtract: $\dfrac{5}{\sigma} = 1.55 \Rightarrow \sigma = 3.23$.

Then $\mu = 75 - 0.20 \times 3.226 = 74.35$.

Answer: mean ≈ 74.35, standard deviation ≈ 3.23.

Q9. For a chip the failure rate is constant and time to failure (hours) has density 4e^(-4x), x >= 0. Calculate the mean time to failure. / If the lifetime of a component has density lambda

e^(-lambda t), find its mean time to failure, variance and failure rate function. (2019, 2020, old scheme)

General case: $f(t) = \lambda e^{-\lambda t},\ t > 0$ (exponential).

$$E(T) = \int_0^\infty t\lambda e^{-\lambda t} dt = \frac{1}{\lambda}, \quad E(T^2) = \frac{2}{\lambda^2}, \quad Var(T) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Failure (hazard) rate: reliability $R(t) = P(T > t) = e^{-\lambda t}$, so

$$h(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

a constant, the memoryless property.

Chip: $\lambda = 4$ per hour, so **mean time to failure = 1/4 hour = 15 minutes** (variance 1/16 h²).

Q10. Ten students got the following marks in Economics and Statistics. Calculate the coefficient of correlation. (2019, old scheme)

Economics (x): 78, 36, 98, 25, 75, 82, 90, 62, 65, 39. Statistics (y): 84, 51, 91, 60, 68, 62, 86, 58, 53, 47. $n = 10$.

Sums: $\sum x = 650,\ \sum y = 660,\ \sum x^2 = 47648,\ \sum y^2 = 45784,\ \sum xy = 45604.$

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\,\sqrt{n\sum y^2 - (\sum y)^2}} = \frac{456040 - 429000}{\sqrt{53980}\,\sqrt{22240}} = \frac{27040}{232.34 \times 149.13} = 0.780$$

r = 0.78: a fairly strong positive correlation between marks in the two subjects.

Q11. Find the coefficient of correlation between X: 34, 40, 27, 17, 36, 19, 38, 23 and Y: 21, 17, 9, 22, 7, 23, 16, 5. (2020, old scheme)

$n = 8.\ \sum x = 234,\ \sum y = 120,\ \sum x^2 = 7404,\ \sum y^2 = 2154,\ \sum xy = 3423.$

$$r = \frac{8(3423) - (234)(120)}{\sqrt{8(7404) - 234^2}\,\sqrt{8(2154) - 120^2}} = \frac{27384 - 28080}{\sqrt{4476}\,\sqrt{2832}} = \frac{-696}{66.90 \times 53.22} = -0.195$$

r = -0.195: a very weak negative correlation, almost no linear relationship.

Q12. Two random variables X and Y have the joint distribution given in a table. Find the marginals, E(X), E(Y), covariance and correlation. (Expected)

Assumed joint pmf $P(X = x, Y = y)$:

	y = 0	y = 1	P(X = x)
x = 0	0.1	0.2	0.3
x = 1	0.2	0.2	0.4
x = 2	0.2	0.1	0.3
P(Y = y)	0.5	0.5	1

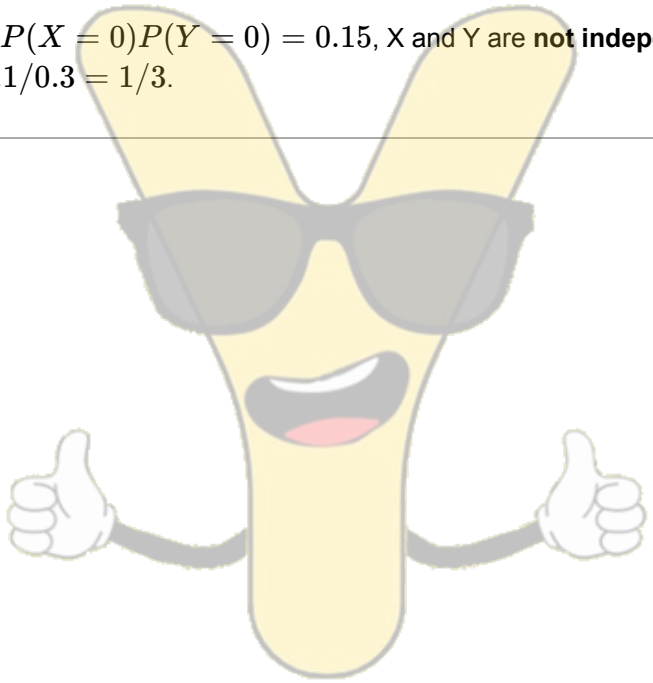
Marginals are the row and column totals. $E(X) = 0(0.3) + 1(0.4) + 2(0.3) = 1$; $E(Y) = 0.5$. $E(XY) = 1 \cdot 1(0.2) + 2 \cdot 1(0.1) = 0.4$.

$$Cov(X, Y) = E(XY) - E(X)E(Y) = 0.4 - 0.5 = -0.1$$

$Var(X) = E(X^2) - 1 = 1.6 - 1 = 0.6$; $Var(Y) = 0.25$.

$$\rho = \frac{-0.1}{\sqrt{0.6 \times 0.25}} = -0.258$$

Since $P(0, 0) = 0.1 \neq P(X = 0)P(Y = 0) = 0.15$, X and Y are **not independent**. Conditional: $P(Y = 1|X = 2) = 0.1/0.3 = 1/3$.



PART–C Answers (≈500 Words Each)

Q1. A random variable X has density f(x) = 4x(9 - x^2)/81 for 0 <= x <= 3. Calculate (i) P(1 < X < 2) (ii) P(X > 2). (2019, old scheme)

Check it is a PDF: $\int_0^3 \frac{4x(9-x^2)}{81}dx = \frac{4}{81}\left[\frac{9x^2}{2} - \frac{x^4}{4}\right]_0^3 = \frac{4}{81}\left[\frac{81}{2} - \frac{81}{4}\right] = 1.$

CDF: for $0 \le x \le 3$, $F(x) = \frac{4}{81}\left(\frac{9x^2}{2} - \frac{x^4}{4}\right) = \frac{2x^2}{9} - \frac{x^4}{81}.$

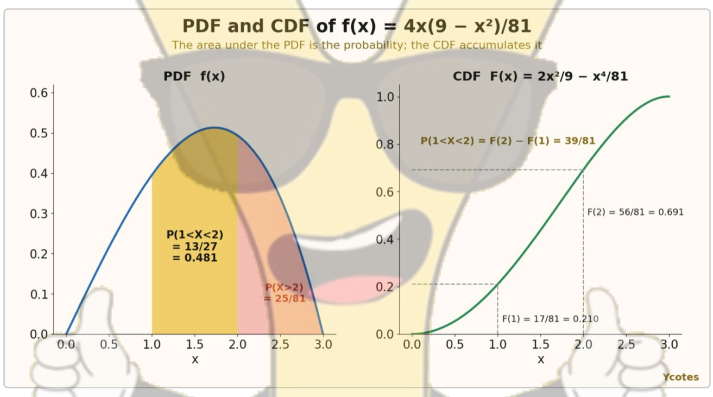
(i) $F(1) = \frac{2}{9} - \frac{1}{81} = \frac{17}{81}$, $F(2) = \frac{8}{9} - \frac{16}{81} = \frac{56}{81}.$

$$P(1 < X < 2) = F(2) - F(1) = \frac{39}{81} = \frac{13}{27} = 0.4815$$

(ii)

$$P(X > 2) = 1 - F(2) = 1 - \frac{56}{81} = \frac{25}{81} = 0.3086$$

Also $E(X) = 8/5 = 1.6$ and $Var(X) = 11/25 = 0.44$.



Q2. A random variable X has density f(x) = 2e^(-2x), x >= 0. Find (i) the moment generating function (ii) the first four moments about the origin. (2019, old scheme)

(i) MGF:

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx}\, 2e^{-2x}dx = 2\int_0^\infty e^{-(2-t)x}dx = \frac{2}{2-t}, \quad t < 2$$

which can be written $M_X(t) = \frac{1}{1 - t/2} = 1 + \frac{t}{2} + \frac{t^2}{4} + \frac{t^3}{8} + \dots$

(ii) Moments about the origin: $\mu'_r = \frac{d^r}{dt^r} M_X(t)\Big|_{t=0}$ = coefficient of $t^r/r!$.

r	Derivative at 0	μ'_r
1	$2/(2-t)^2$	$1/2$
2	$2\cdot 2/(2-t)^3$	$1/2$
3	$2\cdot 6/(2-t)^4$	$3/4$
4	$2\cdot 24/(2-t)^5$	$3/2$

In general $\mu'_r = r!/2^r$. Hence mean $= 1/2$, variance $= \mu'_2 - \mu'^2_1 = 1/2 - 1/4 = 1/4$.

Q3. Let lambda and mu_r denote the mean and rth central moment of a Poisson distribution. Obtain the recurrence formula mu_(r+1) = r lambda mu_(r-1) + lambda d(mu_r)/d(lambda). (2020, old scheme)

For a Poisson variate with mean λ, $p(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$ and the mean is λ.

$$\mu_r = \sum_{x=0}^{\infty}(x-\lambda)^r \frac{e^{-\lambda}\lambda^x}{x!}$$

Differentiate with respect to λ (the mean also depends on λ):

$$\frac{d\mu_r}{d\lambda} = -r\sum(x-\lambda)^{r-1}\frac{e^{-\lambda}\lambda^x}{x!} + \sum(x-\lambda)^r \frac{d}{d\lambda}\left[\frac{e^{-\lambda}\lambda^x}{x!}\right]$$

Now $\dfrac{d}{d\lambda}\left[\dfrac{e^{-\lambda}\lambda^x}{x!}\right] = \dfrac{e^{-\lambda}\lambda^x}{x!}\left(\dfrac{x}{\lambda} - 1\right) = \dfrac{e^{-\lambda}\lambda^x}{x!}\cdot\dfrac{x-\lambda}{\lambda}$.

So

$$\frac{d\mu_r}{d\lambda} = -r\mu_{r-1} + \frac{1}{\lambda}\mu_{r+1}$$

Multiply by λ and rearrange:

$$\boxed{\mu_{r+1} = r\lambda\mu_{r-1} + \lambda\frac{d\mu_r}{d\lambda}}$$

Use: with $\mu_0 = 1,\ \mu_1 = 0$: $\mu_2 = \lambda$, $\mu_3 = 2\lambda\mu_1 + \lambda\dfrac{d\mu_2}{d\lambda} = \lambda$, $\mu_4 = 3\lambda\mu_2 + \lambda\dfrac{d\mu_3}{d\lambda} = 3\lambda^2 + \lambda$. So skewness $= \mu_3^2/\mu_2^3 = 1/\lambda$ and $\beta_2 = 3 + 1/\lambda$.

Q4. Find the mean and variance of the exponential distribution f(x) = lambda e^(-lambda x). Show that for dp = a e^(-x/c) dx, 0 <= x < infinity, the mean and standard deviation are each equal to c. (2020, old scheme)

Part 1: $f(x) = \lambda e^{-\lambda x},\ \lambda > 0,\ x > 0.$

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}\ \text{(by parts)}, \qquad E(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

$$Var(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Part 2: $dp = ae^{-x/c}dx$. Total probability is 1:

$$\int_0^\infty ae^{-x/c}dx = ac = 1 \Rightarrow a = \frac{1}{c}$$

So $f(x) = \frac{1}{c}e^{-x/c}$, an exponential with $\lambda = 1/c$. Using Part 1:

- Mean $= 1/\lambda = c$
- Variance $= 1/\lambda^2 = c^2$, so standard deviation $= c$

Direct check: $E(X) = \frac{1}{c}\int_0^\infty xe^{-x/c}dx = \frac{1}{c}\cdot c^2 = c$ and $E(X^2) = \frac{1}{c}\cdot 2c^3 = 2c^2$, so $Var = 2c^2 - c^2 = c^2$. Hence **mean = standard deviation = c**.

Q5. Ten competitors in a beauty contest were ranked by three judges. Use rank correlation to determine which pair of judges has the nearest approach to common tests in beauty. (2019, old scheme)

Spearman's rank correlation: $\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$, with $n = 10$, so $n(n^2-1) = 990$.

Ranks: J1 = 1, 6, 5, 10, 3, 2, 4, 9, 7, 8; J2 = 3, 5, 8, 4, 7, 10, 2, 1, 6, 9; J3 = 6, 4, 9, 8, 1, 2, 3, 10, 5, 7.

Pair	Differences d	$\sum d^2$	ρ
J1, J2	-2, 1, -3, 6, -4, -8, 2, 8, 1, -1	200	$1 - 1200/990 = -0.212$
J1, J3	-5, 2, -4, 2, 2, 0, 1, -1, 2, 1	60	$1 - 360/990 = 0.636$
J2, J3	-3, 1, -1, -4, 6, 8, -1, -9, 1, 2	214	$1 - 1284/990 = -0.297$

Conclusion: the first and third judges have the highest positive rank correlation (0.636), so they have the **nearest approach to common tests in beauty**. The second judge's taste differs from both (negative correlation).

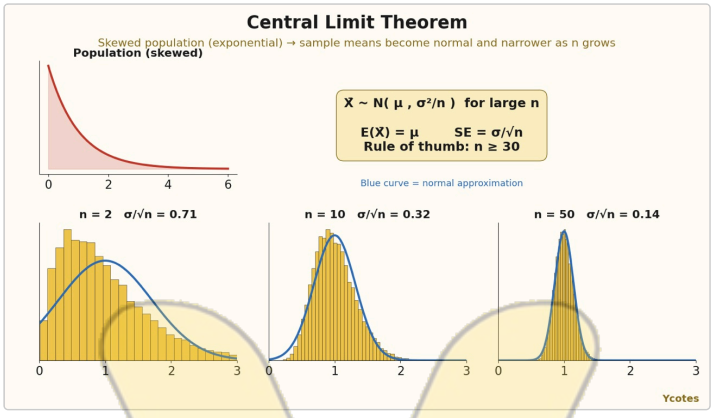
Q6. State the Central Limit Theorem. Use it, and the normal approximation to the binomial,

to solve a numerical. (Expected)

CLT: if $X_1, \ldots, X_n$ are independent observations from any population with mean μ and finite variance σ^2, then for large n (usually $n \geq 30$) the sample mean is approximately normal:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \qquad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

The population need not be normal; the standard error is $\sigma/\sqrt{n}$.



Numerical 1. A population has mean 50 and SD 10. For a sample of n = 100, find $P(\bar{X} > 52)$.
$SE = 10/\sqrt{100} = 1$, $z = (52 - 50)/1 = 2$, so $P = 0.5 - 0.4772 = 0.0228$.

Numerical 2 (normal approximation to binomial). A fair coin is tossed 100 times. Find $P(X \geq 60)$.
$\mu = np = 50$, $\sigma = \sqrt{npq} = 5$. With continuity correction, $P(X \geq 59.5)$: $z = \dfrac{59.5 - 50}{5} = 1.9$,
$P = 0.0287$ (exact binomial value 0.0284).

The approximation works when $np \geq 5$ and $nq \geq 5$.
