UNIT–5 : Generalized Linear Models (PYQ Solution)

PART–A Answers (≈70 Words Each)

Q1. What is the odds ratio in logistic regression? Write the logit link. (Expected)

Odds of an event $= \dfrac{p}{1-p}$. The logit link makes the log-odds linear in the predictors:

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \quad \Longleftrightarrow \quad p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

The odds ratio for a predictor is e^{β_j}: the factor by which the odds multiply when x_j rises by one unit. $e^{\beta} > 1$ raises the odds, < 1 lowers them, $= 1$ means no effect.

Q2. Differentiate between ridge and lasso regression. (Expected)

Ridge (L2)	Lasso (L1)
Minimise $RSS + \lambda \sum \beta_j^2$	Minimise $RSS + \lambda \sum \lvert\beta_j\rvert$
Shrinks coefficients toward 0, never exactly 0	Can set coefficients exactly to 0 (variable selection)
Good when all predictors matter, handles multicollinearity	Good for sparse models with many useless predictors
Circular constraint region	Diamond-shaped region with corners on axes

Both reduce variance at the cost of a little bias; λ is chosen by cross-validation.

PART–B Answers (≈250 Words Each)

Q1. A logistic model has logit(p) = -4 + 0.8x, where x is hours studied. Find the probability of passing for x = 6 and interpret the coefficient. (Expected)

At $x = 6$: log-odds $= -4 + 0.8(6) = 0.8$.

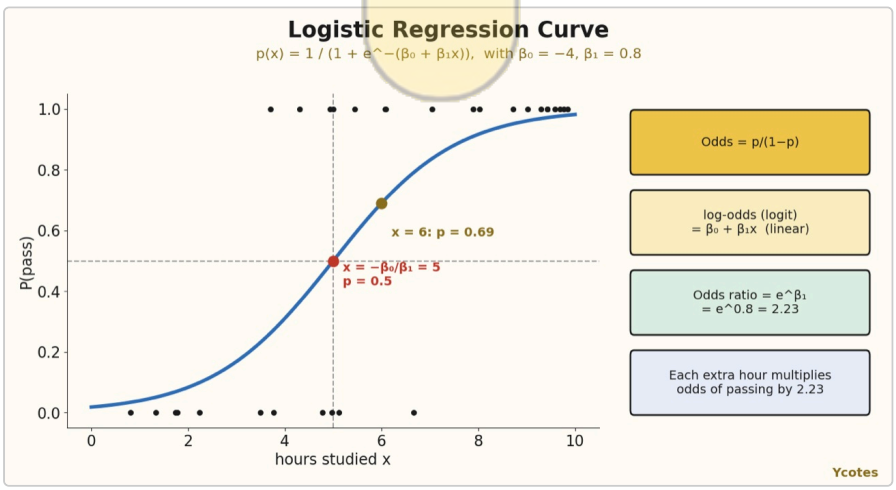
$$\text{odds} = e^{0.8} = 2.226, \qquad p = \frac{e^{0.8}}{1 + e^{0.8}} = \frac{2.226}{3.226} = 0.690$$

Probability of passing = 0.69.

Interpretation: $\beta_1 = 0.8$ means each extra hour studied increases the log-odds of passing by 0.8. The **odds ratio** is $e^{0.8} = 2.23$: every extra hour multiplies the odds of passing by 2.23 (about 123% higher odds).

x (hours)	log-odds	odds	p
3	-1.6	0.202	0.168
5	0.0	1.000	0.500
6	0.8	2.226	0.690
8	2.4	11.02	0.917

The p = 0.5 point is $x = -\beta_0/\beta_1 = 5$ hours.



Q2. Test independence for the 2 x 2 table [[30, 20], [20, 30]] using chi-square. (Expected)

H_0: the two attributes are independent; H_1: they are associated. $\alpha = 0.05$.

Expected frequencies: $E = \dfrac{\text{row total} \times \text{column total}}{N}$. Row totals 50, 50; column totals 50, 50; $N = 100$, so every $E = 25$.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(30-25)^2}{25} \times 4 = 4 \times 1 = 4.0$$

$df = (r-1)(c-1) = 1$, critical $\chi^2_{0.05,1} = 3.841$. Since $4.0 > 3.841$, **reject** H_0 (p = 0.0455): the attributes are associated.

(Condition: every expected count at least 5, which holds.)

Q3. A die-like device gives frequencies 18, 22, 30, 30 over four outcomes in 100 trials. Test whether the outcomes are equally likely (chi-square goodness of fit). (Expected)

H_0: all four outcomes are equally likely, so $E = 100/4 = 25$ each.

Outcome	O	E	$(O-E)^2/E$
1	18	25	1.96
2	22	25	0.36
3	30	25	1.00
4	30	25	1.00
Total	100	100	4.32

$df = 4 - 1 = 3$, critical $\chi^2_{0.05,3} = 7.815$. Since $4.32 < 7.815$ (p = 0.229), **accept** H_0: the data are consistent with equally likely outcomes.

PART–C Answers (≈500 Words Each)

Q1. Three groups have scores 6, 8, 7, 9; 10, 12, 11, 9; 14, 13, 15, 16. Perform one-way ANOVA. Explain MANOVA. (Expected)

$H_0 : \mu_1 = \mu_2 = \mu_3$; H_1: at least one mean differs. $k = 3$ groups, $n = 4$ each, $N = 12$.

Group means: 7.5, 10.5, 14.5. **Grand mean** $= 130/12 = 10.833$.

Between-group SS: $SSB = 4[(7.5 - 10.833)^2 + (10.5 - 10.833)^2 + (14.5 - 10.833)^2] = 4(11.11 + 0.111 + 13.44) = 98.67$.

Within-group SS: group 1: $1.5^2 + 0.5^2 + 0.5^2 + 1.5^2 = 5$; group 2: $0.5^2 + 1.5^2 + 0.5^2 + 1.5^2 = 5$; group 3: 5. So $SSW = 15$. Total SS $= 113.67$.

Source	SS	df	MS	F
Between	98.67	2	49.33	
Within	15.00	9	1.667	29.6
Total	113.67	11		

$$F = \frac{MSB}{MSW} = \frac{49.33}{1.667} = 29.6$$

Critical $F_{0.05;2,9} = 4.256$. Since $29.6 > 4.256$ (p = 0.0001), **reject** H_0: at least one group mean differs. A post-hoc test (Tukey) shows which; here group 3 is highest.

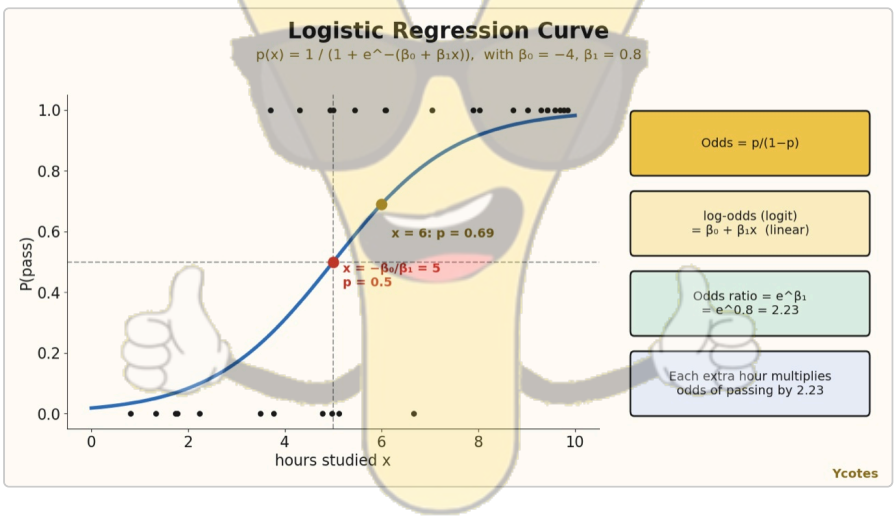
MANOVA (multivariate ANOVA) compares group means on **several dependent variables at once** (for example maths, science and English scores across three teaching methods). It tests whether the vector of means (μ_1, μ_2, μ_3) is equal across groups using the between and within **SSCP matrices** B and W; statistics are Wilks' lambda $\Lambda = \frac{|W|}{|B + W|}$, Pillai's trace, Hotelling-Lawley trace and Roy's largest root. A small Λ means significant group differences. It controls overall Type I error and uses correlations among the outcomes; assumptions are multivariate normality, equal covariance matrices and independence. With one dependent variable it reduces to ANOVA.

Q2. Explain multiple logistic regression and Poisson regression for count outcomes. (Expected)

Both are **generalized linear models (GLMs)**, which have three parts: a distribution for y (binomial, Poisson, normal), a linear predictor $\eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$, and a **link function** $g(\mu) = \eta$ connecting the mean to the linear predictor.

Multiple logistic regression (binary outcome)

- $y \in \{0, 1\}$, distribution binomial, link logit: $\ln \dfrac{p}{1-p} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$.
- Each e^{β_j} is the **adjusted odds ratio**, controlling for the other predictors.
- Fitted by **maximum likelihood** (not least squares); coefficients tested with the Wald z-test; model compared by deviance or AIC.
- Classify as 1 if $p \geq 0.5$ (threshold can be changed). Example: $\text{logit}(p) = -4 + 0.8\,\text{hours} + 0.5\,\text{attendance}$.



Poisson regression (count outcome)

- y = counts (calls per hour, defects per batch), distribution Poisson, **log link**: $\ln \mu = \beta_0 + \beta_1 x_1 + \ldots$, so $\mu = e^{\beta_0 + \beta_1 x}$ is always positive.
- e^{β_j} is the **rate ratio**: the factor by which the expected count multiplies for a unit rise in x_j. For rates use an **offset** $\ln(\text{exposure})$.
- Example: $\ln \mu = 0.5 + 0.3x$; at x = 2, $\mu = e^{1.1} = 3.0$ events; $e^{0.3} = 1.35$, i.e. 35% more events per unit of x.
- Check **overdispersion** (variance > mean); then use quasi-Poisson or negative binomial.

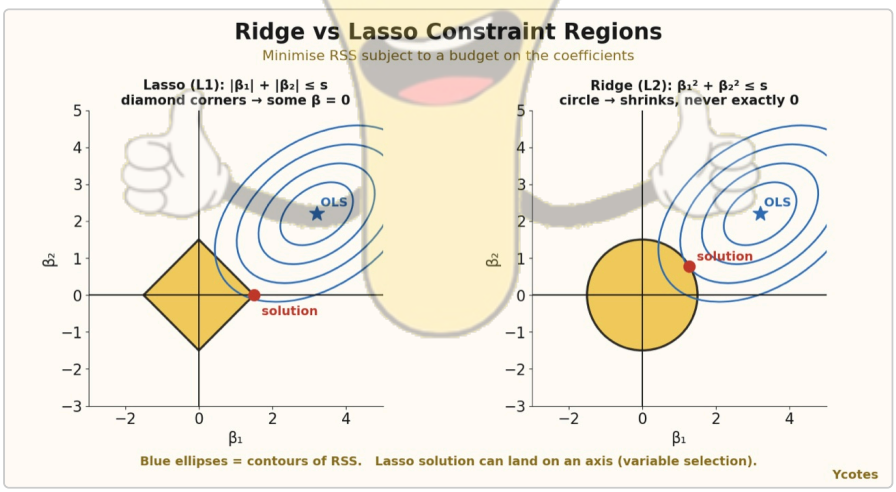
Comparison: linear regression is the GLM with a normal distribution and identity link.

Q3. Explain ridge and lasso regression with the penalty terms, shrinkage and a numerical example. (Expected)

Why: with many or correlated predictors OLS overfits (high variance). Penalised regression adds a cost on coefficient size.

- **Ridge:** $\hat{\beta}^{ridge} = \arg\min \sum(y_i - \hat{y}_i)^2 + \lambda \sum_j \beta_j^2$. Closed form $\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y$. Shrinks all coefficients smoothly; none become exactly zero.
- **Lasso:** $\hat{\beta}^{lasso} = \arg\min \sum(y_i - \hat{y}_i)^2 + \lambda \sum_j |\beta_j|$. No closed form; produces **sparse** models.
- $\lambda = 0$ gives OLS; $\lambda \to \infty$ shrinks coefficients to 0. Choose λ by cross-validation; standardise predictors first (the intercept is not penalised).

Geometry: the solution is where the RSS ellipse first touches the constraint region. Lasso's diamond has corners on the axes, so the touch point is often at a corner (a coefficient = 0). Ridge's circle has no corners.



Numerical 1 (one predictor). x = 1, 2, 3 and y = 2, 3, 7 (no intercept). $X^T X = 14$, $X^T y = 2 + 6 + 21 = 29$.

- OLS: $\beta = 29/14 = 2.071$
- Ridge with $\lambda = 1$: $\beta = 29/(14 + 1) = 1.933$ (shrunk toward 0)

Numerical 2 (orthonormal predictors, OLS $\hat{\beta} = 3.0,\ -1.5,\ 0.4$, $\lambda = 1$).

	β_1	β_2	β_3
OLS	3.0	-1.5	0.4
Ridge: $\hat{\beta}/(1+\lambda)$	1.5	-0.75	0.2
Lasso: $\text{sign}(\hat{\beta})(\lvert\hat{\beta}\rvert - \lambda/2)_+$	2.5	-1.0	**0**

Ridge scales every coefficient by the same factor; lasso subtracts a constant and drops the small one (β_3) to exactly zero, performing variable selection.

