1/I/3H Statistics

Derive the least squares estimators $\hat{\alpha}$ and $\hat{\beta}$ for the coefficients of the simple linear regression model

$$Y_i = \alpha + \beta(x_i - \bar{x}) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where $x_1, \ldots, x_n$ are given constants, $\bar{x} = n^{-1}\sum_{i=1}^n x_i$, and ε_i are independent with $\mathrm{E}\,\varepsilon_i = 0$, $\mathrm{Var}\,\varepsilon_i = \sigma^2$, $i = 1, \ldots, n$.

A manufacturer of optical equipment has the following data on the unit cost (in pounds) of certain custom-made lenses and the number of units made in each order:

No. of units, x_i	1	3	5	10	12
Cost per unit, y_i	58	55	40	37	22

Assuming that the conditions underlying simple linear regression analysis are met, estimate the regression coefficients and use the estimated regression equation to predict the unit cost in an order for 8 of these lenses.

[*Hint: for the data above,* $S_{xy} = \sum_{i=1}^n (x_i - \bar{x})y_i = -257.4$.]

1/II/12H Statistics

Suppose that six observations $X_1, \ldots, X_6$ are selected at random from a normal distribution for which both the mean μ_X and the variance σ_X^2 are unknown, and it is found that $S_{XX} = \sum_{i=1}^6 (x_i - \bar{x})^2 = 30$, where $\bar{x} = \frac{1}{6}\sum_{i=1}^6 x_i$. Suppose also that 21 observations $Y_1, \ldots, Y_{21}$ are selected at random from another normal distribution for which both the mean μ_Y and the variance σ_Y^2 are unknown, and it is found that $S_{YY} = 40$. Derive carefully the likelihood ratio test of the hypothesis H_0: $\sigma_X^2 = \sigma_Y^2$ against H_1: $\sigma_X^2 > \sigma_Y^2$ and apply it to the data above at the 0.05 level.

[*Hint:*

Distribution	χ_5^2	χ_6^2	χ_{20}^2	χ_{21}^2	$F_{5,20}$	$F_{6,21}$
95% percentile	11.07	12.59	31.41	32.68	2.71	2.57

]

2/I/3H Statistics

Let $X_1, \ldots, X_n$ be a random sample from the $N(\theta, \sigma^2)$ distribution, and suppose that the prior distribution for θ is $N(\mu, \tau^2)$, where σ^2, μ, τ^2 are known. Determine the posterior distribution for θ, given $X_1, \ldots, X_n$, and the best point estimate of θ under both quadratic and absolute error loss.

2/II/12H Statistics

An examination was given to 500 high-school students in each of two large cities, and their grades were recorded as low, medium, or high. The results are given in the table below.

	Low	*Medium*	*High*
City A	103	145	252
City B	140	136	224

Derive carefully the test of homogeneity and test the hypothesis that the distributions of scores among students in the two cities are the same.

[*Hint:*

Distribution	χ_1^2	χ_2^2	χ_3^2	χ_5^2	χ_6^2
99% percentile	6.63	9.21	11.34	15.09	16.81
95% percentile	3.84	5.99	7.81	11.07	12.59

]

4/I/3H Statistics

The following table contains a distribution obtained in 320 tosses of 6 coins and the corresponding expected frequencies calculated with the formula for the binomial distribution for $p = 0.5$ and $n = 6$.

No. heads	0	1	2	3	4	5	6
Observed frequencies	3	21	85	110	62	32	7
Expected frequencies	5	30	75	100	75	30	5

Conduct a goodness-of-fit test at the 0.05 level for the null hypothesis that the coins are all fair.

[*Hint:*

Distribution	χ_5^2	χ_6^2	χ_7^2
95% percentile	11.07	12.59	14.07

]

4/II/12H Statistics

State and prove the Rao–Blackwell theorem.

Suppose that $X_1, \ldots, X_n$ are independent random variables uniformly distributed over $(\theta, 3\theta)$. Find a two-dimensional sufficient statistic $T(X)$ for θ. Show that an unbiased estimator of θ is $\hat{\theta} = X_1/2$.

Find an unbiased estimator of θ which is a function of $T(X)$ and whose mean square error is no more than that of $\hat{\theta}$.