1. Consider the simple linear regression model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \; i = 1, \ldots, n$. The $\epsilon_i$ have zero means, common variance $\sigma^2$, and zero correlations. The coefficients are estimated using least squares.

(a) Show that $\text{Var}[\hat{\beta}_0]$ is minimized if $\sum x_i = 0$.

(b) Suppose that all $x_i$ lie in the bounded interval $[a, b]$ and $n$ is even. Then $\text{Var}[\hat{\beta}_1]$ is minimized if half of the $x_i$ are equal to $a$ and the others are equal to $b$. 
2. Consider the one way ANOVA model $Y_{ij} = \mu + \alpha_i + e_{ij}$, $i = 1, 2, 3$, $j = 1, \ldots, J$, where the $e_{ij}$ are i.i.d. $N(0, \sigma^2)$. Show how to test $H_0: \mu + \alpha_2 = (\mu + \alpha_1 + \mu + \alpha_3)/2$. Give the formula for your test statistic, its distribution under $H_0$, and an expression for its power function.

3. A simple random sample of size $n$ is selected from a population $U$ of size $N$. The population mean $\bar{y}_U$ of a variable $y$ is estimated by the sample average $\bar{y}$. The loss due to sampling error is $\lambda|\bar{y} - \bar{y}_U|$ and the cost of sampling is $C = c_0 + c_1 n$. Ignoring any finite population corrections, show that the most economical value of $n$ is $\left[\frac{\lambda S}{2(n-1)^{1/3}}\right]^2$, where $S^2 = (N-1)\sum_{i\in U}(y_i - \bar{y}_U)^2$ is the population variance.

4. The general mixed effects model is

$$Y = X\beta + Zu + e = X\beta + \sum_{j=1}^{m} Z_j u_j + e$$

Assume $X$ has $n$ rows and full column rank $p$, that $Z_j$ has full rank $q_j$, that $q = \sum_j q_j$, that $u_j \sim N(0, \sigma_j^2 I_{q_j})$, that $e \sim N((0, \sigma_0^2 I_n)$, and that the random vectors $u_1, \ldots, u_m, e$ are mutually independent.

(a) Find $V = \text{Var-Cov}[Y]$.

(b) If $V$ is known, show that the maximum likelihood estimator of $\beta$ is $\hat{\beta}_{GLS} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$ and that its variance-covariance matrix is $(X^T V^{-1} X)^{-1}$.

(c) If $V$ is unknown but a consistent estimator $\hat{V}$ is available, show that

$$\hat{\beta} = (X^T \hat{V}^{-1} X)^{-1} X^T \hat{V}^{-1} Y$$

is approximately $N(\beta, (X^T V^{-1} X)^{-1})$ as $n \to \infty$. You may assume regularity conditions as needed.
5. In the regression model \( Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i, \ i = 1, \ldots, n, \) the predictors have identical sample correlations \( r_{jk} = s_{jk}/(s_j s_k) \equiv r, \) where
\[
\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}, \quad s_{ij} = \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{n-1}, \quad \text{and} \quad s_j = \sqrt{s_{jj}}.
\]
Define new predictors
\[
\begin{align*}
z_{i1} &= (x_{i1} + x_{i2} + x_{i3})/\sqrt{3}, \\
z_{i2} &= (x_{i1} - x_{i2})/\sqrt{2}, \\
z_{i3} &= (x_{i1} + x_{i2} - 2x_{i3})/\sqrt{6}.
\end{align*}
\]
(a) Determine the matrix \( C \) so that the new design matrix is written as \( Z = XC. \)

(b) Determine the least squares fit of the model \( Y = Z\gamma + e \) in terms of the correlations \( r_i = \text{corr}(Y, x_j). \) Use this result to determine the estimate \( \beta \) in the original model \( Y = X\beta + e. \)

(c) Comment on the situation when \( r = -1/2. \)

6. A rough measurement \( x, \) made on each unit in a population \( U, \) is related to the true value \( y \) of the unit by the equation \( x = y + d + e, \) where \( d \) is a constant bias and \( e \) is a measurement error uncorrelated with \( y. \) The variable \( e \) has population mean 0 and population variance \( S^2_{eU}. \) The population is assumed to be infinite. Assume a simple random sample of size \( n \) is selected and that the goal is to estimate the population mean \( \bar{x}_U. \) Compare the variances of

(a) the difference estimate \( \bar{y} + (\bar{x}_U - \bar{x}) \)

(b) the regression estimate \( \bar{y} + b(\bar{x}_U - \bar{x}), \) using the value of \( b \) that gives minimum variance.

The notation \( \bar{y} \) represents the sample mean. The variances may depend upon \( S^2_{yU}, \) the population variance of \( y. \)
7. Let $Y_i, i = 1, \ldots, I$, be independent binomial($n_i, \pi_i$) variables with $\pi_i > 0$ and
\[ \log \left( \frac{\pi_i}{1 - \pi_i} \right) = \alpha + \beta_i. \]
(a) Observe that the model is overparameterized. Assume $\beta_I = 0$ and derive the maximum likelihood estimators of the remaining parameters.
(b) Under the null hypothesis $H_0$: $\beta_1 = \cdots = \beta_I$, find the MLE of $\alpha$.
(c) How would you test $H_0$? State your test statistic and its null distribution. You may assume that all $n_i$ are large.

8. Let $Y_{ij} = \mu + \alpha_i + b_j + e_{ij}, i = 1, \ldots, I$ and $j = 1, \ldots, J$. The $b_j$ are i.i.d. $N(0, \sigma_b^2)$, the $e_{ij}$ are i.i.d. $N(0, \sigma_e^2)$ and the $b_j$ and $e_{ij}$ are mutually independent.
(a) Write out the ANOVA table for these data, giving sums of squares, degrees of freedom and expected mean squares. Show that the sums of squares in your ANOVA table are independent.
(b) Provide a 95% confidence interval for $\alpha_1 - \alpha_2$.
(c) How would your answer to (b) change if the $b_j$ were fixed but unknown parameters?
(d) Find unbiased point estimates of $\sigma_b^2$ and $\sigma_e^2$. Would you recommend that a practitioner should use them?