1. A student was fitting a simple linear regression model to data

$$\mathbf{Y} = \beta_0 \mathbf{1}_n + \beta_1 \mathbf{X} + \mathbf{e}$$

with $\mathbf{e} \sim N(0, \sigma^2 I_n)$ and $\mathbf{X}$ a $n \times 1$ vector with $\bar{X} = 100$. Since theory suggested that when $X_i = 0$ that $\mathsf{E}(\mathsf{Y_i})$ should be zero, the student set β_0 equal to zero, forcing the fitted regression line to go through the origin.

(a) What is the ordinary least squares estimate of β_1 in this case?

(b) if β_0 is actually **not zero**, is $\hat{\beta}_1$, the OLS estimate of β_1, unbiased?

(c) The student decided to plot the residuals, $\hat{\mathbf{e}} = \mathbf{Y} - \mathbf{X}\hat{\beta}_1$ versus the fitted values $\mathbf{X}\hat{\beta}_1$ as a check and observed the plot (shown on the next page). The student thought that fitted values and residuals should be uncorrelated. Find the $\mathsf{E}(\hat{\mathbf{e}})$ if $\mathsf{E}(\mathbf{Y}) = \beta_0 \mathbf{1}_\mathsf{n} + \beta_1 \mathbf{X}$ with β_0 not equal to zero. Can you explain to the student why the residuals and fitted values appear to be correlated in the residual plot?

(d) The student decided to go back to the computer output. Is there any evidence to suggest that β_0 is not zero? Give appropriate test statistic, distribution and your conclusion.

(e) In the computer output, the F-statistic in the summary for the regression thru the origin is 3232000 with 1 and 99 degrees of freedom. What is the null model in this case? The alternative model? Because the p-value is very small, and this F-statistic is much larger than the F-statistic in the model with the intercept and $\mathbf{X}$, does this mean we should accept the regression thru the origin model? Explain. What can you conclude?

(f) The student noticed that the standard error for $\hat{\beta}_1$ is about 50 times smaller in the regression thru the origin than in the model with the intercept and X, so that confidence intervals for β_1 would be narrower in the regression thru the origin. Isn't a narrower confidence interval always better? Explain.

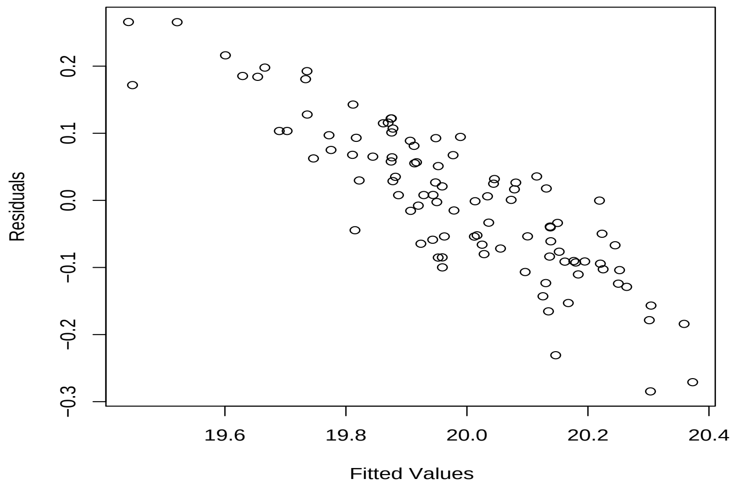

Figure 1: Residual plot from model with regression thru the origin.

```
> summary(lm0) # model without intercept

lm(formula = Y ~ X - 1)

Coefficients:
   Estimate Std. Error t value Pr(>|t|)
X 0.1999504  0.0001112    1798   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1112 on 99 degrees of freedom
Multiple R-Squared:     1,      Adjusted R-squared:     1
F-statistic: 3.232e+06 on 1 and 99 DF,  p-value: < 2.2e-16

> summary(lm1) # Model with intercept

lm(formula = Y ~ X)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 9.821601   0.568709   17.27   <2e-16 ***
X           0.101677   0.005691   17.87   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.05556 on 98 degrees of freedom
Multiple R-Squared: 0.7651,     Adjusted R-squared: 0.7627
F-statistic: 319.2 on 1 and 98 DF,  p-value: < 2.2e-16
```

2. Assume that we have a sample of size n, $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ from the model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{1}$$

with $\mathbf{e} \sim N(0, I_n)$ and $\mathbf{X}$ is $n \times p$ of rank p. Suppose we want to find an estimator of $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\mu}}$, which minimizes expected quadratic error loss,

$$\mathsf{E}[(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})^\mathsf{T}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})].$$

where the expectation is taken with respect to the distribution of $\mathbf{Y}$ given $\boldsymbol{\beta}$ in (1).

(a) Ronald knew that out of all the linear unbiased estimates that ordinary least squares $\hat{\boldsymbol{\mu}}_{\text{OLS}} \equiv \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{OLS}}$ has the minimum variance, and suggested that as an estimator. Show that the expected loss for using ordinary least squares is p. *Hint: recall* $\mathsf{E}(\mathsf{Y}^\mathsf{T}\mathsf{A}\mathsf{Y}) = \text{trace}(\mathsf{A}\Sigma) + \boldsymbol{\mu}^\mathsf{T}\Sigma\boldsymbol{\mu}$ *where* A *is a* $n \times n$ *symmetric matrix,* Σ *is the covariance of* $\mathbf{Y}$ *and* $\boldsymbol{\mu}$ *is the* $\mathsf{E}(\mathbf{Y})$.

(b) Thomas was not so worried about being unbiased and decided that his posterior mean of $\boldsymbol{\mu}$, $\hat{\boldsymbol{\mu}}_B \equiv \frac{g}{1+g}\mathbf{X}\boldsymbol{\beta}_{\text{OLS}}$, might be a better choice. Find the expected loss with using the estimator $\hat{\boldsymbol{\mu}}_B$.

(c) If $\boldsymbol{\mu}^T\boldsymbol{\mu} = cn$, for some constant c, can you suggest values of g such the Bayes estimator will do better than ordinary least squares?

3. Consider the simple linear regression model with one predictor,

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \epsilon_i \qquad i = 1, \ldots, n & (2) \\ \epsilon_i &\overset{iid}{\sim} N(0, \sigma_{ee}) & (3) \end{aligned}$$

with independent, identically distributed normal errors with mean 0 and variance σ_{ee}. Suppose that we are unable to observe X_i directly, but instead observe W_i, a noisy version of X_i,

$$W_i = X_i + u_i \tag{4}$$

where $u_i \overset{iid}{\sim} N(0, \sigma_{uu})$ represents measurement error in X_i. For $i = 1, \ldots, n$, assume that the vector

$$\begin{pmatrix} X_i \\ \epsilon_i \\ u_i \end{pmatrix} \overset{iid}{\sim} N\left(\begin{bmatrix} \mu_x \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{xx} & 0 & 0 \\ 0 & \sigma_{ee} & 0 \\ 0 & 0 & \sigma_{uu} \end{bmatrix} \right). \tag{5}$$

(a) Find the joint distribution of (Y_i, W_i) given parameters $(\mu_x, \beta_0, \beta_1, \sigma_{ee}, \sigma_{xx}, \sigma_{uu})'$ using the specifications given by equations $(2-5)$.

(b) Let $\hat{\gamma}_1$ denote the ordinary least squares regression coefficient computed from the regression of Y on W,

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{n}(W_i - \bar{W})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(W_i - \bar{W})^2}.$$

Find $\mathsf{E}[\hat{\gamma}_1]$ and show that as an estimator of β_1, the regression coefficient in (2), that $\hat{\gamma}_1$ is biased towards zero.

Hint: Recall that if $Z = (Z_1, Z_2)'$ has a normal distribution, with partitioned mean and covariance

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

that $Z_1 | Z_2$ is normal with mean $\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Z_2 - \mu_2)$ and covariance $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.