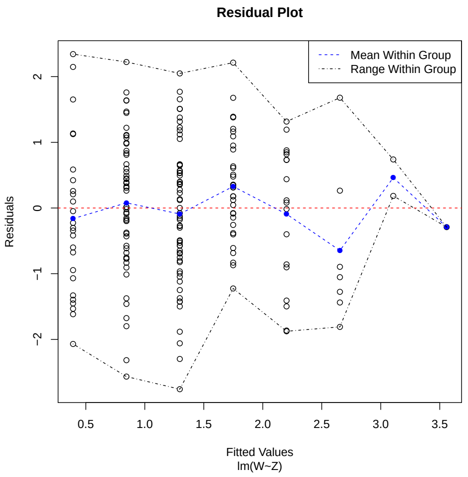
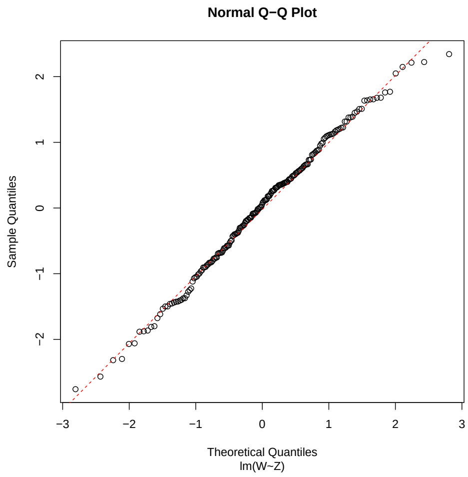
Question1 (10 points)

Consider the following regression model.

$$W = \hat{\beta}_0 + \hat{\beta}_1 Z + \varepsilon$$

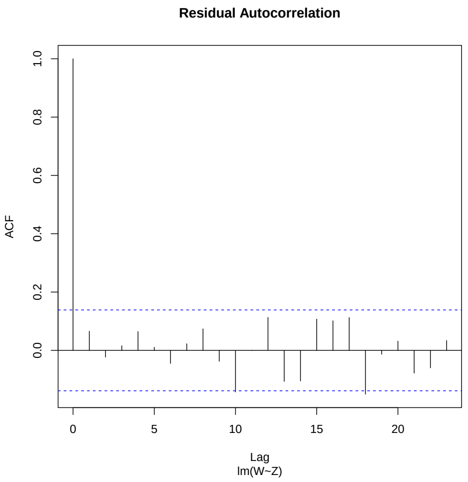
(a) (1 point) Given the following residual plot, is there any evidence against linearity?

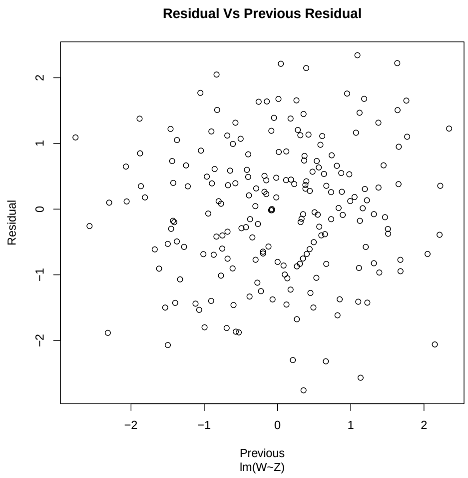




(b) (1 point) How about equal variance?

(c) (1 point) Why do we need both of the following plots for checking independence?





(d) (1 point) According to the QQ-normal plot, is there any evidence against normality?

(e) (4 points) Compute the confidence interval for $\mathbb{E}[W \mid Z = 1]$ using the R output.

```
> summary(model)
```

```
Call:
lm(formula = W ~ Z)

Residuals:
     Min       1Q   Median       3Q      Max
-2.75796 -0.70866  0.06424  0.66591  2.34291

Coefficients:
            Estimate Std. Error t value Pr(>t)
```

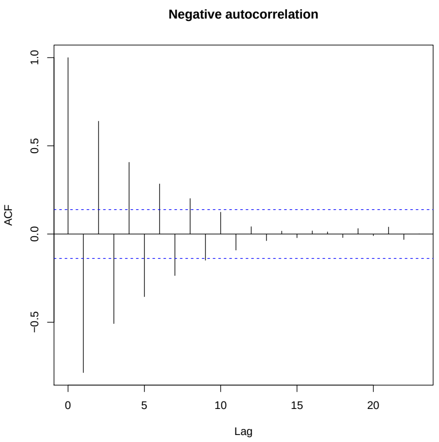
```
(Intercept)  0.39198    0.12941    3.029  0.00278 **
Z            0.45234    0.05339    8.472 5.42e-15 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 1.039 on 198 degrees of freedom
Multiple R-squared:  0.2661,    Adjusted R-squared:  0.2624
F-statistic: 71.78 on 1 and 198 DF,  p-value: 5.422e-15
```

(f) (2 points) Compute the prediction interval for W given $Z = 1$ using the R output.

Question2 (1 points)

Explain why the negative autocorrelation is indicated by the following ACF plot.




```
> n = 200
> y = double(n)
> for (i in 2:n){
+   y[i] = rnorm(1, -0.8*y[i-1], 1)
+ }
>
> acf(y, main = "Negative autocorrelation")
```

Question3 (2 points)

In terms of simple linear regression, write down the formulas for the following two variances using the notation in class and explain the difference between the two variances.

$$\text{Var}\left[Y_{n+1} - \hat{Y}^*_{n+1} \mid X_1, X_2, \cdots, X_n, X_{n+1}\right]$$
$$\text{Var}\left[Y_i - \hat{Y}_i \mid X_1, X_2, \cdots, X_n\right] \qquad \text{for all } i$$

Question4 (3 points)

Consider the model $y_{ij} = \alpha_0 + \alpha_i + \beta_j + \varepsilon_{ij}$, where $i = 1, 2$, $j = 1, 2, 3$, αs are fixed, while β_j and ε_{ij} are random and are all independent. In matrix form, we have $\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, in which $\boldsymbol{\alpha} = \begin{bmatrix}\alpha_0 & \alpha_1 & \alpha_2\end{bmatrix}^{\mathrm{T}}$, $\boldsymbol{\beta} = \begin{bmatrix}\beta_1 & \beta_2 & \beta_3\end{bmatrix}^{\mathrm{T}}$, $\mathbf{y} = \begin{bmatrix}y_{11} & y_{21} & y_{12} & y_{22} & y_{13} & y_{23}\end{bmatrix}^{\mathrm{T}}$, and $\boldsymbol{\varepsilon} = \begin{bmatrix}\varepsilon_{11} & \varepsilon_{21} & \varepsilon_{12} & \varepsilon_{22} & \varepsilon_{13} & \varepsilon_{23}\end{bmatrix}^{\mathrm{T}}$ where $\boldsymbol{\beta} \sim \text{Normal}\left(\mathbf{0}, \sigma^2_\beta \mathbf{I}\right)$ and $\boldsymbol{\varepsilon} \sim \text{Normal}\left(\mathbf{0}, \sigma^2_\varepsilon \mathbf{I}\right)$. State the matrix $\mathbf{X}$ and $\mathbf{Z}$. Find the variance-covariance matrix of $\mathbf{y}$ in terms of σ^2_β and σ^2_ε.

Question5 (4 points)

A statistics professor has been involved in a collaborative research project with two entomologists. The statistics part of the project involves fitting regression models to large data sets. Together they have written and submitted a manuscript to an entomology journal. The manuscript contains a number of scatter plots with each showing an estimated regression line (based on a valid model) and associated individual 95% confidence intervals for the regression function at each x value, as well as the observed data. A referee has asked the following question: "*I don't understand how 95% of the observations fall outside the 95% CI as depicted in the figures.*" Explain how it is entirely possible that 95% of the observations fall outside the 95% CI as depicted in the figures.