Linear Regression

Mathematical Analysis of Linear Regression

1. To prove: $\mathbb{E}[w] = \beta$
   Given:

$$y = X\beta + \epsilon \tag{1}$$
$$w = (X^T X)^{-1} X^T y \tag{2}$$

   Substituting (1) in (2):

$$\begin{aligned}
w &= (X^T X)^{-1} X^T * (X\beta + \epsilon) \\
w &= ((X^T X)^{-1} * (X^T X)) * \beta + ((X^T X)^{-1} X^T) * \epsilon \\
w &= I * \beta + ((X^T X)^{-1} X^T) * \epsilon \\
\mathbb{E}[w] &= \mathbb{E}[I * \beta + ((X^T X)^{-1} X^T) * \epsilon]
\end{aligned}$$

   Using the linearity of expectation:

$$\begin{aligned}
\mathbb{E}[w] &= \mathbb{E}[I * \beta] + \mathbb{E}[(X^T X)^{-1} X^T) * \epsilon] \\
&= \beta + (X^T X)^{-1} X^T) * \mathbb{E}[\epsilon] \\
&= \beta \text{ (since the mean of } \epsilon = 0)
\end{aligned}$$

2.
   - If the two columns are identical, then the **Cov**[w] will not be well defined since, X would be a singular matrix, i.e. there is n't any inverse of X(det(X=0))
   - The **Cov**[w] will not be well defined again since, X would be a singular matrix, i.e. there is n't any inverse of X(det(X=0)), when $X_i = -X_j$ for $i \neq j$.

Empirical Investigation of Linear Regression

1.
   - Condition number of $X^T X$=22.1549.
   - $\mathbb{E}[w] = \beta = [0\ 1]^T$,
     $\mathbf{Cov}[w] = \sigma^2[(X^T X)^{-1}] = \begin{bmatrix} 1 & -1 \\ -1 & 1.2 \end{bmatrix}$
   - Empirical mean of $w = [0.0322\ 1.0874]^T$
     Empirical Covariance= $\begin{bmatrix} 1.1007 & -1.0476 \\ -1.0476 & 1.1429 \end{bmatrix}$
   - Plots:

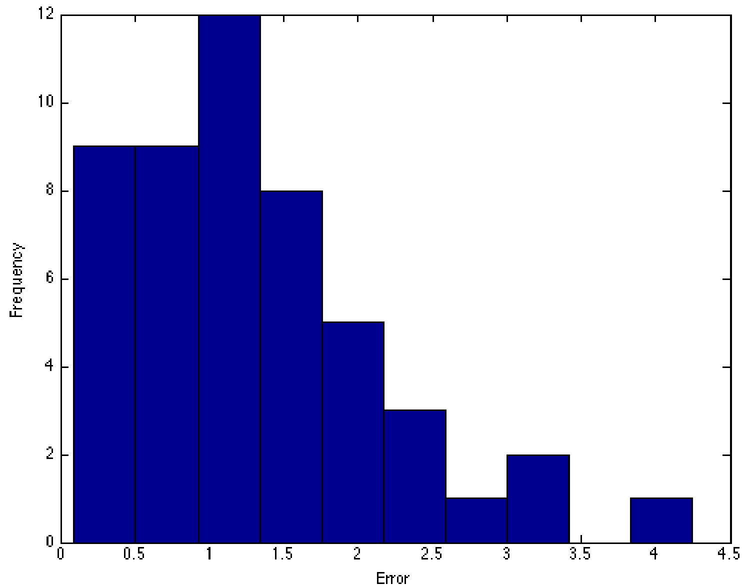

Figure 1: Histogram of the errors

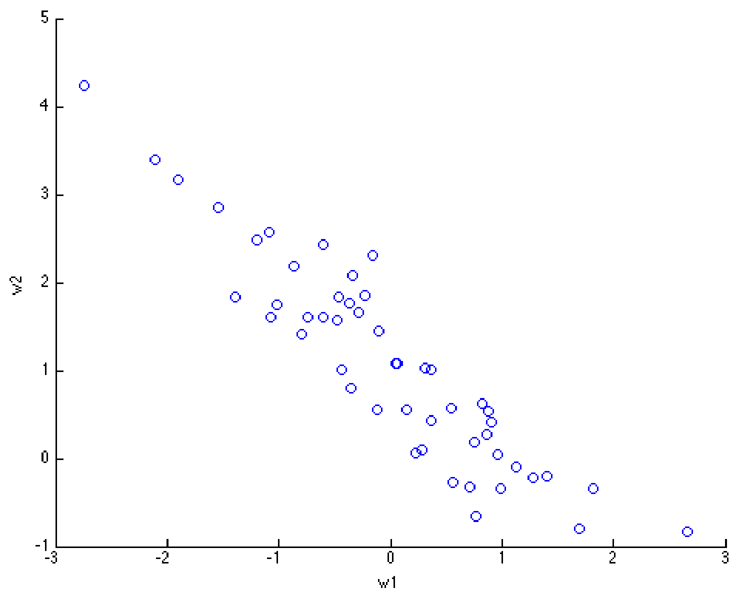

Figure 2: Scatter Plot of w_1 vs w_2

- QQ Plots:

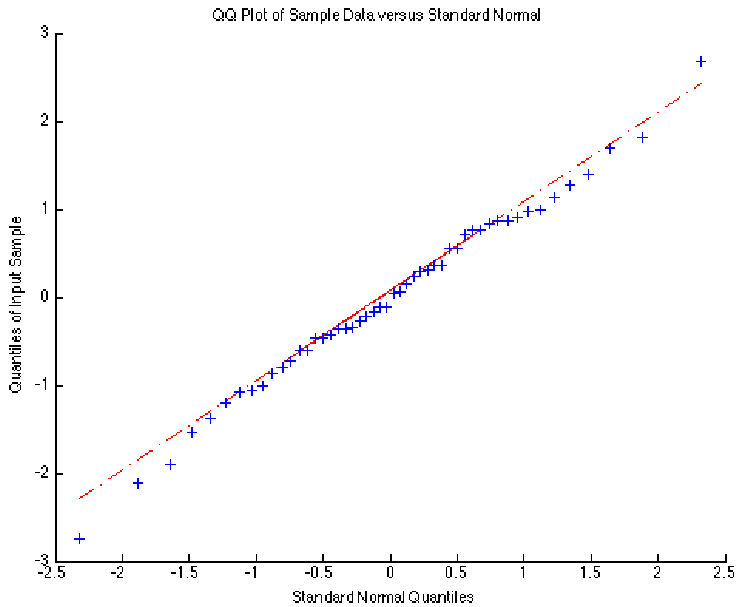

Figure 3: QQ Plot: w_1

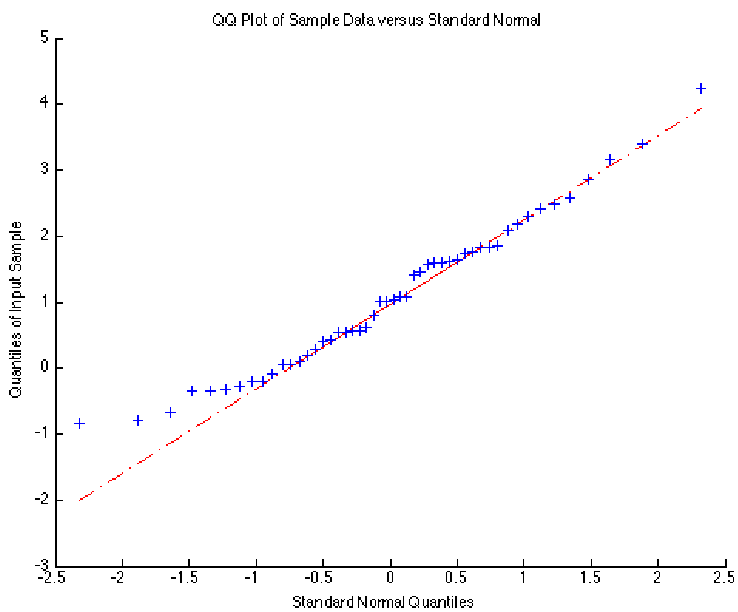

Figure 4: QQ Plot: w_2

2. - Condition number of X^TX=4.2386e+20.
   - $\mathbb{E}[w] = \beta = [0\ 1]^T$

     $$\textbf{Cov}[w] = \sigma^2[(X^TX)^{-1}] = \begin{bmatrix} 320.5 & -310 \\ -310 & 300 \end{bmatrix}$$

- Empirical mean of $w = [0.897\ 0.071]^T$

  Empirical Covariance= $\begin{bmatrix} 378.96 & -366.52 \\ -366.52 & 354.64 \end{bmatrix}$

- Plots:

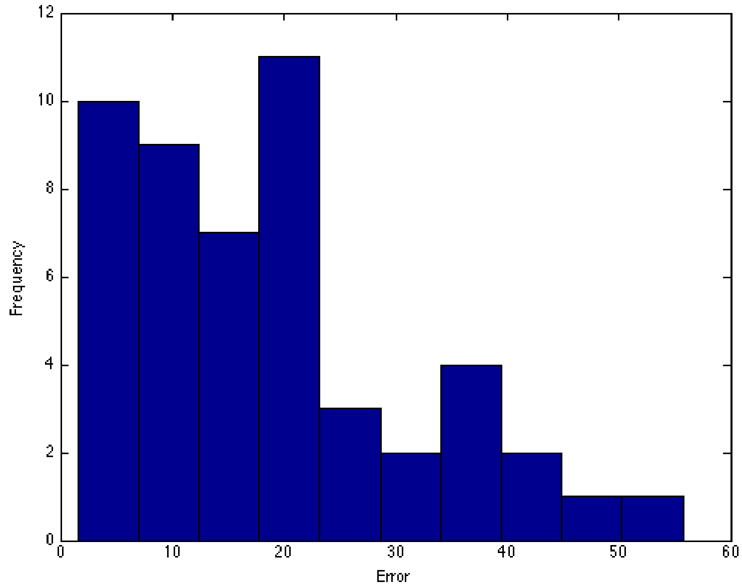

Figure 5: Histogram of the errors

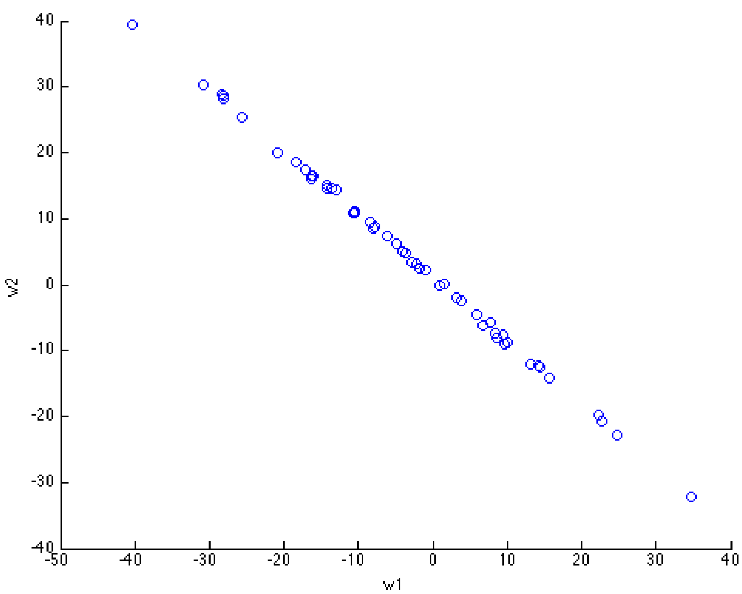

Figure 6: Scatter Plot of w_1 vs w_2

- QQ Plots:

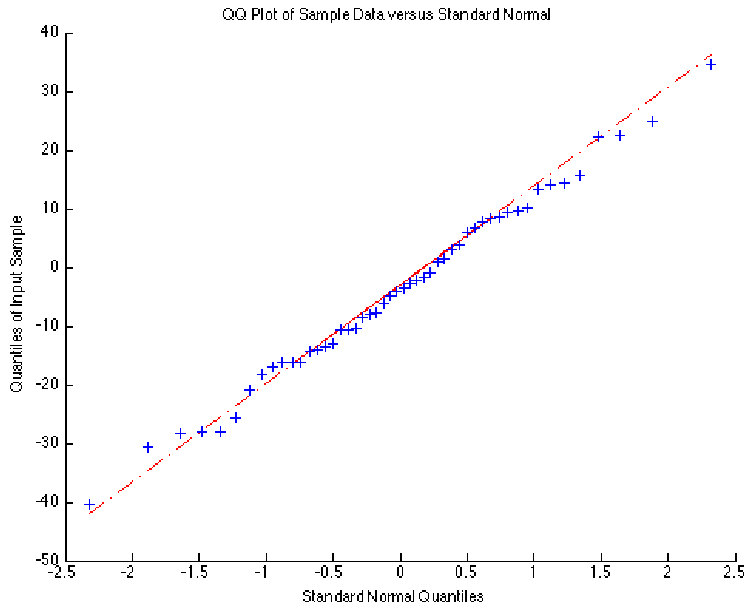

Figure 7: QQ Plot: w_1

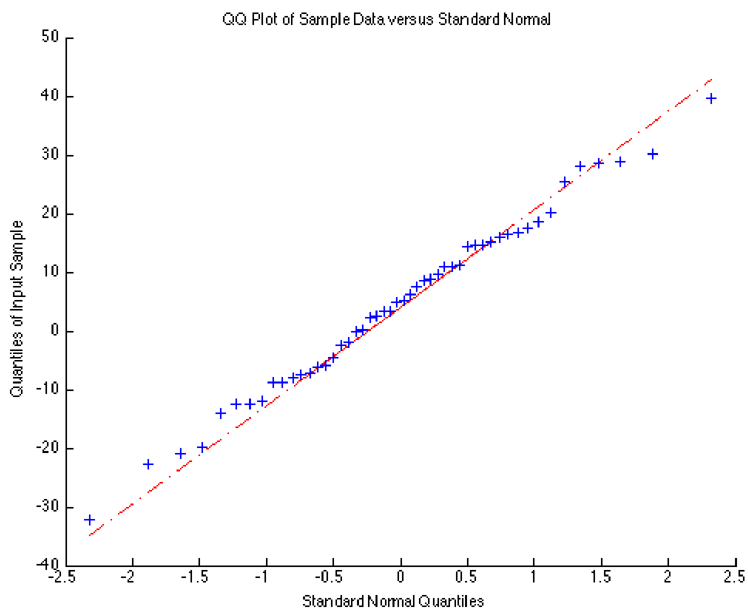

Figure 8: QQ Plot: w_2

Code Snippet:

```
function [w,err] = linReg(X,b)
for i=1:50
```

e=eye(4)*normrnd(0,1,4,1);
y(:,i)=X*b+e;
w(:,i)=inv(X'*X)*X'*y(:,i);
d=w(:,i)-b;%$||w-b||$
err(i)=sqrt(d'*d);
end
end

[w,err]=linReg(X,b);
scatter(w(1,:),w(2,:))
qqplot(w(1,:))
qqplot(w(2,:))
E_{mean}=[mean(w(1,:)) mean(w(2,:))]
E_{cov} =cov(w(1,:),w(2,:))

3. $\mathbf{x}_{test}=[20]$, $\mathbb{E}[w]=0$
Plots:

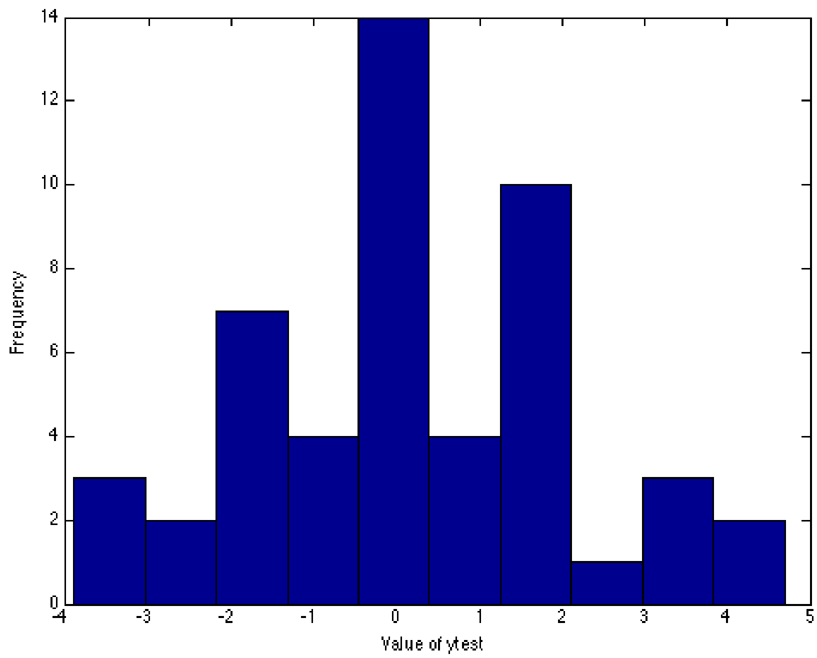

Figure 9: Histogram of the predicted Y given w_1

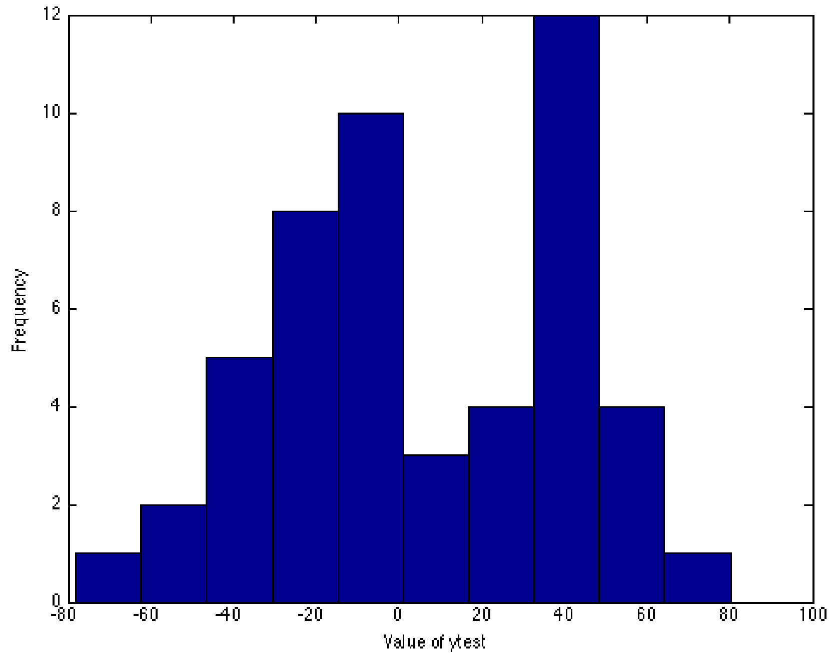

Figure 10: Histogram of the predicted Y given w_1

Code Snippet:

```
for i=1:50
y_test(i)=x_test*w(:,i);
end
```

4.
- For the first experiment both, E[w] and Cov[w] are approximately the same, whereas for Experiment 2, the mean and the Covariance are very different as far as the magnitudes are concerned with respect to the actual mean and Covariance.
- Scatter plots show the relationship between the two weights, whereas QQ plot displays a quantile-quantile plot of the sample quantiles of w versus theoretical quantiles from a normal distribution. If the distribution of w is normal, the plot will be close to linear, which is the case over here. In both the scatter and the QQ-plots w is centered around 0.
- Experiment 1 produced better results for w, as the error rate is much lesser than that obtained for Experiment 2.
- No, If we replace X_2 with -X_2 in the X matrix, w_1 and w_2 are positively correlated with each other which means X_1 and X_2 negatively correlated with each other.