Modules 11-12 Assignment

JARED YU

1. Use any math/stat software (e.g., www.numbergenerator.org/randomnumbergenerator) of your choice to find a random number generator to randomly select only 15 rows of Table shown in Problem 13.2. Do Problem 13.2 (a), (b), (c), (d), page 463 of Textbook, using your generated data.
    a. Fit a logistic regression model to the response variable y. Use a simple linear regression model as the structure for the linear predictor.

Ans:
The model was fit in R using the following command, "glm(formula=y~x, family=binomial(), data=df)." It yields a model with the following coefficients, $\hat{\beta}_0 = -6.7973$ and $\hat{\beta}_1 = 0.0002$.

    b. Does the model deviance indicate that the logistic regression model from part a is adequate?

Ans:
From the textbook, we have the deviance defined as follows,

$$D = 2\ln\left[\frac{L(\text{saturated model})}{L(FM)}\right] = 2\sum_{i=1}^{N}\left[y_i \ln\left(\frac{y_i}{n_i\hat{\pi}_i}\right) + (n_i - y_i)\ln\left(\frac{n_i - y_i}{n_i(1-\hat{\pi}_i)}\right)\right],$$

where n_i is the sample size of group i, $n = \sum_{i=1}^{N} n_i$. This statistic follows a χ^2_{n-p} distribution. As a note, for this problem it will be such that $n_i = 1$ for all i.

Using R, this comes out to $D \approx 18.2121$, while the χ^2 critical value is 22.36203. The deviance is less than the critical values and so it indicates that the logistic regression model is adequate. However, looking at the ratio of $\frac{D}{n-p} = 1.4009$, which is a possibly far from unity, which is undesirable.

    c. Provided an interpretation of the parameter β_1 in this model.

Ans: (partially copied from the discussion response)
To understand the parameter β_1, we can look at the input into the sigmoid function,

$$\hat{\eta}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i,$$

which is the fitted value of the linear predictor at a particular value of x (i.e., x_i). We can look also at the fitted value for $x_i + 1$, which is as follows,

$$\hat{\eta}(x_i + 1) = \hat{\beta}_0 + \hat{\beta}_1(x_i + 1).$$

Then taking the difference between these two we get,

$$\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) = \hat{\beta}_1.$$

The $\hat{\eta}(\cdot)$ functions are also known as the *log-odds*, so we can rewrite the difference as follows,

$$\hat{\eta}(x_i + 1) - \hat{\eta}(x_i) = \ln \text{odds}_{x_i+1} - \ln \text{odds}_{x_i} = \ln\left(\frac{\text{odds}_{x_i+1}}{\text{odds}_{x_i}}\right) = \hat{\beta}_1.$$

Taking the antilog of the above difference yields the *odds ratio*,

$$\hat{O}_R = \frac{\text{odds}_{x_i+1}}{\text{odds}_{x_i}} = e^{\hat{\beta}_1}.$$

In the textbook, it is said to be, "the estimated increase in the probability of success associated with a one-unit change in the value of the predictor variable." This can be generalized to a d-unit change as a change of $e^{d\hat{\beta}_1}$. So, for β_1, a \$1 increase in income is associated with a $e^{\hat{\beta}_1}$ change in probability of home ownership.

d. Expand the linear predictor to include a quadratic term in income. Is there any evidence that this quadratic term is required in the model?

Ans:
In this problem, the step was to change the linear predictor to $\beta_0 + \beta_1 x + \beta_2 x^2$. Again, using the "glm()" function, it gave the following estimated coefficients: $\hat{\beta}_0 = -0.6556$, $\hat{\beta}_1 = 0.0028$, and $\hat{\beta}_2 = -2.861 \times 10^{-8}$. Doing the same deviance calculation as before leads to $D \approx 17.1258$, which is still not larger than the corresponding χ^2 critical values of 21.0261. Therefore, we again fail to reject that this model is appropriate. The ratio $D/(n-p)$ is slightly larger this time at 1.4272.

Previously, the $\hat{\beta}_1$ along was quite small, but this time the $\hat{\beta}_2$ is significantly smaller being roughly equivalent to 0. So, although it seems to pass the goodness-of-fit test, it is difficult to think that this quadratic term is meaningful.

2. Use any math/stat software (e.g., www.numbergenerator.org/randomnumbergenerator) of your choice to find a random number generator to randomly select only 9 rows of Table shown in Problem 13.4. Do Problem 13.4 (a), (b), (c), (d), (e), (f), (g), page 464 of Textbook, using your generated data.
   a. Fit a logistic regression model to the data. Use a simple linear regression model as the structure for the linear predictor.

Ans:
The first step was to create a response variable which is the proportion defined as $y_i = \frac{r_i}{n_i}$. Here, within the dataset $n_i = 500$ for all i. After plugging in the formula "y~x" into the glm() function, it gives the following estimated coefficients, $\hat{\beta}_0 = -2.0852$ and $\hat{\beta}_1 = 0.1358$.

   b. Does the model deviance indicate that the logistic regression model from part a is adequate?

Ans:
Using the same calculation for the deviance from problem 1, the result is that $D \approx 0.0006$, while the corresponding χ^2 critical value is 14.0671. In this case, the deviance is smaller than the critical value and so we say that the logistic regression model is adequate. Looking at the ratio $D/(n-2)$, it comes out to 8.3269×10^{-5}, which is much less than unity.

   c. Draw a graph of the data and the fitted logistic regression model.

Ans:
The below figure shows the fitted and true values on the y-axis and the regressor on the x-axis. The fitted values are shown as "+" signs, while the true values are shown as "Δ" signs.

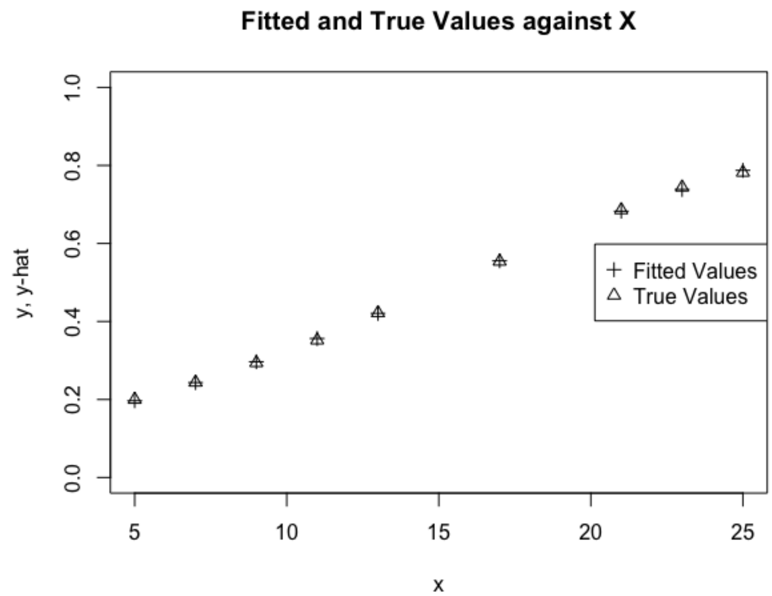

Figure 1 The above figure shows the fitted ($\hat{y}$) and true values (y) against the regressor (x).

d. Expand the linear predictor to include a quadratic term. Is there any evidence that this quadratic term is required in the model?

Ans:

Performing the same deviance calculation leads to $D \approx 0.006$, while the corresponding critical value from the χ^2 is 12.59159. The deviance is less than this so we can again say that the model is appropriate. The corresponding $D/(n-p)$ ratio is also much less than unity at 9.5943×10^{-5}. The same situation applies from before however, where $\hat{\beta}_2 = 5.803 \times 10^{-5}$ is approximately zero. So, it is hard to say that the quadratic term is meaningful or required.

e. Draw a graph of this new model on the same plot that you prepared in part c. does the expanded model visually provide a better fit to the data than the original model from part a?

Ans:

The plot with the additional estimated probabilities that include the quadratic term can be seen below in Figure 2. This second model is called “Model 2” and is denoted as a red circle. The previous model is “Model 1” and is denoted with the same blue triangle. The fit for these two models seem rather identical visually, and so it is not possible to say that it provides a better fit.

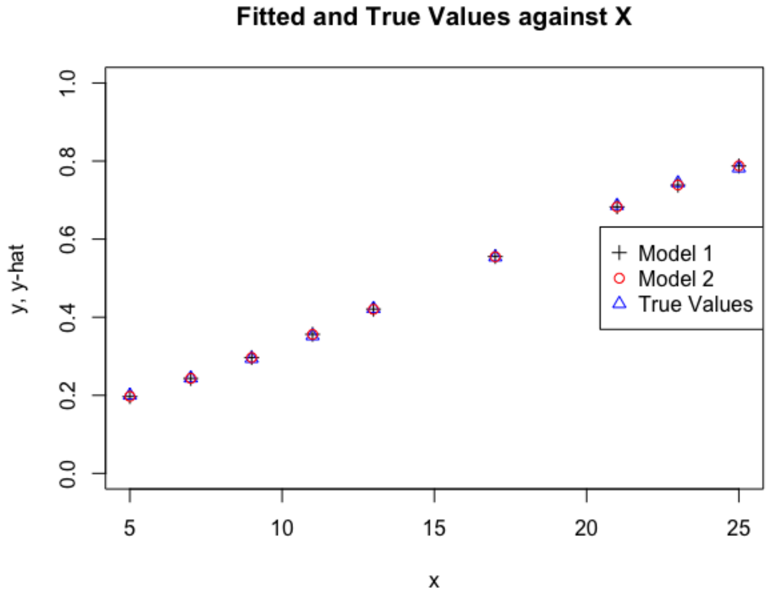

Figure 2 The above figure shows the fitted ($\hat{y}$) and true values (y) against the regressor (x). The difference this time is that a second model is added (model 2) which includes a quadratic term.

f. For the quadratic model in part d, find the Wald statistic for each individual model parameter.

Ans:

The Wald statistic is as follows,

$$Z_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)},$$

which follows a standard normal distribution. The standard error in the denominator can be found by finding the covariance matrix of $\hat{\boldsymbol{\beta}}$,

$$Var(\hat{\boldsymbol{\beta}}) = -\mathbf{G}(\hat{\boldsymbol{\beta}})^{-1} = (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1},$$

where $\mathbf{G}$ is the Hessian matrix found from the log-likelihood function.

The covariance matrix of the coefficients is calculated within the glm() function in R. Taking the diagonal and square rooting it allows for finding the standard errors of the estimated coefficients. Then, dividing each of the estimated coefficients by these standard errors gives the following Wald statistics which can be seen below in Table 1.

Table 1 The below table shows the Wald statistics for each of the parameters in the model.

	β_0	β_1	β_2
Z_0	-0.4651	0.2013	0.0027

From the table, in the second row it shows the Wald statistic, and the columns identifies which of the parameters it corresponds to.

g. Find approximate 95% confidence intervals on the model parameters for the quadratic logistic regression model from part d.

Ans:
Based on the Wald statistic from the previous part, it can be seen that the $(1-\alpha)\%$ confidence interval for β_j would be as follows,

$$\hat{\beta}_j - Z_{\frac{\alpha}{2}}\, se(\hat{\beta}_j) \le \beta_j \le \hat{\beta}_j + Z_{\frac{\alpha}{2}}\, se(\hat{\beta}_j).$$

Plugging in the variables and using 1.96 for the $Z_{\frac{\alpha}{2}}$ leads to the following confidence intervals:

Table 2 The below table shows the 95% confidence intervals for the model parameters.

	95% Confidence Interval
β_0	$(-10.8161, 6.6677)$
β_1	$(-1.1711, 1.4392)$
β_2	$(-0.0423, 0.0424)$

3. Use any math/stat software (e.g., www.numbergenerator.org/randomnumbergenerator) to randomly select only 30 rows of Table shown in Problem 13.7. Do Problem 13.7 (a), (b), (c), (d), (e), page 466-467 of Textbook, using your generated data.
   a. Fit a Poisson regression model to these data using the log link.

Ans:
The following command was used in R to fit a Poisson regression model, “glm(formula=y~., family=poisson(), data=df).” It yields the following regression coefficients: $\hat{\beta}_0 = -2.5999$, $\hat{\beta}_1 = -0.0023$, $\hat{\beta}_2 = 0.0533$, $\hat{\beta}_3 = -0.0028$, and $\hat{\beta}_4 = -0.0425$.

   b. Does the model deviance indicate that the model from part a is satisfactory?

Ans:
As a note, the previous method of calculating the deviance doesn’t produce the corresponding result as calculated from the glm() function, so in this problem the deviance from the glm() function will be used instead. Here, $D \approx 25.1219$, while the corresponding χ^2 is 37.6525. Therefore, the model seems to be appropriate. The corresponding $D/(n-p)$ ratio is 1.0049, which is highly close to unity. This also indicates that the model is satisfactory.

   c. Perform a type 3 partial deviance analysis of the model parameters. Does this indicate that any regressors could be removed from the model?

Ans:
The partial deviance in this problem is calculated as follows,

$$D(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) = D(\boldsymbol{\beta}_1) - D(\boldsymbol{\beta}),$$

where if $D(\boldsymbol{\beta}_2|\boldsymbol{\beta}_1) \ge \chi^2_{\alpha,r}$, we reject the null hypothesis that $\boldsymbol{\beta}_2 = 0$. In this case, the goal will be to fit the full model including all regressors ($\boldsymbol{\beta}$) and then fit a reduced model excluding one of the regressors ($\boldsymbol{\beta}_1$). The deviance calculation will be the same as before, where it is derived from the glm() function in R. So, the result will be four different tests for each of the four regressors in the model. In all of the tests, the corresponding critical value will be $\chi^2_{0.05,1} = 3.8415$. The table below (Table 3) shows the corresponding partial deviance for each of the regressors. It can be seen that for β_2 and β_4, the decision is to reject the null hypothesis. While for β_1 and β_3, the decision is to fail to reject the null hypothesis. This indicates that possible for β_1 and β_3 that one or both can possibly be removed from the final model.

Table 3 The below table shows the partial deviance for each of the regressors in the model.

	β_1	β_2	β_3	β_4
$D(\boldsymbol{\beta}_2 \mid \boldsymbol{\beta}_1)$	3.5884	16.4329	0.2340	4.2819

d. Compute Wald statistics for testing the contribution of each regressor to the model. Interpret the results of these test statistics.

<u>Ans:</u>
The same steps as in part f of the previous problem for calculating the Wald statistics are repeated here. The result can be seen below in Table 4.

Table 4 The table below shows the Wald statistics for each of the regressors from the model.

	β_0	β_1	β_2	β_3	β_4
Z_0	-2.1562	-1.7655	3.6801	-0.4705	-1.9629

The hypothesis test is as follows,

$$H_0: \beta_j = 0, H_1: \beta_j \neq 0$$

where the critical value at $\alpha = 0.05$ is approximately 1.9600 from the standard normal distribution. Therefore, the decision rule is to reject the null for all the regressors, except β_1 and β_3. This could indicate that all except for x_1 and x_3 don't contribute significantly to the fit, given that the other regressors are held constant. However, it is possibly not practical to try and remove the intercept term.

e. Find approximate 95% Wald confidence intervals on the model parameters.

<u>Ans:</u>
In this problem, the same steps from part g of the previous problem will be repeated here. The 95% Wald confidence intervals can be seen below in Table 5.

Table 5 The below table shows the 95% confidence intervals for the model parameters.

	95% Confidence Interval
β_0	$(-4.9633, -0.2366)$
β_1	$(-0.0049, 0.0003)$
β_2	$(0.0249, 0.0817)$
β_3	$(-0.0147, 0.0090)$
β_4	$(-0.0849, -0.0001)$

4. Do Problem 13.14, page 468 of Textbook

The exponential probability density function is

$$f(y, \lambda) = \lambda e^{-\lambda y} \text{ for } y, \lambda \geq 0$$

Show that the exponential distribution is a member of the exponential family.

<u>Ans:</u>
From the textbook, it states that distributions which are members of the exponential family have the general form

$$f(y_i, \theta_i, \phi) = \exp\left\{\frac{[y_i\theta_i - b(\theta_i)]}{a(\phi)} + h(y_i, \phi)\right\},$$

where ϕ is a scale parameter and θ_i is called the natural location parameter. For members within this family,

$$\mu = E(y) = \frac{db(\theta_i)}{d\theta_i}$$
$$Var(y) = \frac{d^2b(\theta_i)}{d\theta_i^2}a(\phi) = \frac{d\mu}{d\theta_i}a(\phi)$$

We know from the exponential distribution that $\mu = \frac{1}{\lambda}$ and $Var(y) = \frac{1}{\lambda^2}$. Therefore, we need it to be that $\frac{db(\theta_i)}{d\theta_i} = \frac{1}{\lambda}$ and $\frac{d^2b(\theta_i)}{d\theta_i^2}a(\phi) = \frac{1}{\lambda^2}$. Starting with μ, we can do the following. Let $\underline{\theta_i = \lambda}$ which leads to $\underline{b(\theta_i) = b(\lambda) = \ln\lambda}$. Therefore,

$$E(y) = \frac{db(\theta_i)}{d\theta_i} = \frac{db(\lambda)}{d\lambda} = \frac{d}{d\lambda}\ln\lambda = \frac{1}{\lambda}$$

Then, let $\underline{a(\phi) = -1}$, so it follows that,

$$Var(y) = a(\phi)\frac{d\mu}{d\theta_i} = a(\phi)\frac{d\mu}{d\lambda} = (-1) - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

If we set $\underline{h(y_i, \phi) = 0}$, then we also get that,

$$f(y_i, \theta_i, \phi) = \exp\left\{\frac{[y_i\theta_i - b(\theta_i)]}{a(\phi)} + h(y_i, \phi)\right\}$$
$$= \exp\left\{\frac{[y_i\lambda - \ln\lambda]}{-1} + 0\right\} = \exp\{-y_i\lambda + \ln\lambda\} = \lambda e^{-\lambda y_i} \blacksquare$$

5. Use any math/stat software (e.g., www.numbergenerator.org/randomnumbergenerator) to randomly select only 10 rows of Table shown in Problem 13.26 on page 471. Do Problem 13.26, using your generated data.

<u>Ans:</u>

The following Poisson regression model was fit using the following command, "glm(formula=y~., family=poisson(), data=df)." The table below (Table 6) shows the estimated coefficients for the model. The following are the regressors: $x_1 = \text{Temperature}$, $x_2 = \text{Oil}$, and $x_3 = \text{Time}$.

Table 6 The table below shows the estimated coefficients for the model.

$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
8.4197	−0.2944	−0.1419	−0.0283

The deviance for the model is $D \approx 86.4248$, while the corresponding χ^2 critical value is approximately 12.5916. The deviance is larger than the critical value and so we say that the Poisson regression model is not appropriate. Furthermore, the $D/(n-p)$ value is 14.4041, which is quite far from unity.

The next step is to do a partial deviance test on each of the regressors. The result can be seen in the table below. The corresponding critical value from the χ^2 is 3.8415. With the exception of β_2, they are all much larger than this critical value. However, they are all in fact larger than the critical value, and so for each regressor we reject the null that $\beta_j = 0$.

Table 7 The table below shows the partial deviances for each of the regressors in the model.

	β_1	β_2	β_3
$D(\boldsymbol{\beta}_2 \vert \boldsymbol{\beta}_1)$	28.2395	4.1800	35.1665

We can also perform a similar test, by looking at the Wald statistics. The Wald statistics can be seen below in Table 8. The hypothesis test is as follows,

$$H_0: \beta_j = 0, H_1: \beta_j \neq 0$$

where the critical value at $\alpha = 0.05$ is approximately 1.9600 from the standard normal distribution. Therefore, the decision rule is to reject the null for all the regressors. This could indicate that all regressors contribute significantly to the model, given that the other regressors are held constant.

Table 8 The table below shows the Wald statistics for each of the regressors from the model.

	β_0	β_1	β_2	β_3
Z_0	14.0849	-5.2266	-2.0339	-5.9113

We can lastly analyze the confidence intervals for each of the regressors based on the previously calculated Wald statistics. The 95% Wald confidence intervals can be seen below in Table 9.

Table 9 The below table shows the 95% confidence intervals for the model parameters.

	95% Confidence Interval
β_0	$(7.2481, 9.5914)$
β_1	$(-0.4048, -0.1840)$
β_2	$(-0.2983, -0.0055)$
β_3	$(-0.0376, -0.0189)$

Code Appendix

```
### Problem 1
df <- MPV::p13.2
n <- 15
set.seed(1); chosen_rows <- sort(sample(seq(1, nrow(df)), n))
df <- df[chosen_rows,]

# part a
df_fit <- glm(formula = y~x, family = binomial(), data = df)

# part b
# Reference: https://stats.stackexchange.com/questions/108995/interpreting-residual-and-null-deviance-in-g
lm-r
summary(df_fit)
anova(df_fit, test = 'Chi')
df_fit$null.deviance - df_fit$deviance > qchisq(p = 0.95, df = 1)

deviance_D <- 2 *
  sum(df$y * log(df$y / df_fit$fitted.values),
      (1 - df$y) * log((1 - df$y) / (1 - df_fit$fitted.values)),
      na.rm = TRUE)
deviance_D > qchisq(p = 0.95, df = n - 2)
deviance_D / (n - 2)
```

```
# part d
df$x2 <- df$x^2
df_fit2 <- glm(formula = y ~ x + x2, family = binomial(), data = df)
summary(df_fit2)
anova(df_fit2, test = 'Chi')
df_fit2$null.deviance - df_fit2$deviance >
  qchisq(p = 0.95, df = df_fit2$df.null - df_fit2$df.residual)
deviance_D <- 2 *
  sum(df$y * log(df$y / df_fit2$fitted.values),
      (1 - df$y) * log((1 - df$y) / (1 - df_fit2$fitted.values)),
      na.rm = TRUE)
deviance_D > qchisq(p = 0.95, df = n - 3)
deviance_D / (n - 3)

### Problem 2
df <- MPV::p13.4
n <- 9
set.seed(1); chosen_rows <- sort(sample(seq(1, nrow(df)), n))
df <- df[chosen_rows,]
df$y <- df$r / df$n

# part a
df_fit <- glm(formula = y~x, family = binomial(), data = df)

# part b
summary(df_fit)
anova(df_fit, test = 'Chi')
df_fit$null.deviance - df_fit$deviance > qchisq(p = 0.95, df = 1)

deviance_D <- 2 *
  sum(df$y * log(df$y / df_fit$fitted.values),
      (1 - df$y) * log((1 - df$y) / (1 - df_fit$fitted.values)),
      na.rm = TRUE)
deviance_D > qchisq(p = 0.95, df = n - 2)
deviance_D < qchisq(p = 0.95, df = n - 2, lower.tail = FALSE)
deviance_D / (n - 2)

# part c
plot(df$x, df_fit$fitted.values, ylim = c(0,1), pch = 3,
     main = 'Fitted and True Values against X',
     xlab = 'x', ylab = 'y, y-hat')
points(df$x, df$y, pch=2)
legend("right", legend = c('Fitted Values', 'True Values'),
       pch = c(3,2))

# part d
df$x2 <- df$x^2
df_fit2 <- glm(formula = y ~ x + x2, family = binomial(), data = df)
summary(df_fit2)
anova(df_fit2, test = 'Chi')
df_fit2$null.deviance - df_fit2$deviance >
  qchisq(p = 0.95, df = df_fit2$df.null - df_fit2$df.residual)

deviance_D <- 2 *
  sum(df$y * log(df$y / df_fit2$fitted.values),
      (1 - df$y) * log((1 - df$y) / (1 - df_fit2$fitted.values)),
      na.rm = TRUE)

deviance_D > qchisq(p = 0.95, df = n - 3)
deviance_D < qchisq(p = 0.95, df = n - 3, lower.tail = FALSE)

deviance_D / (n - 3)

# part e
plot(df$x, df_fit$fitted.values, ylim = c(0,1), pch = 3,
     main = 'Fitted and True Values against X',
     xlab = 'x', ylab = 'y, y-hat')
```

```r
points(df$x, df$y, pch=2, col = 'blue')
points(df$x, df_fit2$fitted.values, pch=1, col = 'red')
legend("right", legend = c('Model 1', 'Model 2', 'True Values'),
       pch = c(3,1,2), col = c('black', 'red', 'blue'))

# part f
summary(df_fit2)
std_errors <- sqrt(diag(summary(df_fit2)$cov.unscaled))
wald_statistics <- df_fit2$coefficients / std_errors

# part g
round(df_fit2$coefficients + c(1.96 * std_errors),4)
round(df_fit2$coefficients - c(1.96 * std_errors),4)

### Problem 3
df <- MPV::p13.7
n <- 30
set.seed(1); chosen_rows <- sort(sample(seq(1, nrow(df)), n))
df <- df[chosen_rows,]

# part a
df_fit <- glm(formula = y~., family = poisson(), data = df)
summary(df_fit)
anova(df_fit)

# part b
deviance_D <- summary(df_fit)$deviance
deviance_D > qchisq(p = 0.95, df = n - 5)
deviance_D / (n - 5)

# part c
df_fit1 <- glm(formula = y~x2+x3+x4, family = poisson(), data = df)
df_fit1$deviance - df_fit$deviance > qchisq(p = 0.95, df = 1)

df_fit2 <- glm(formula = y~x1+x3+x4, family = poisson(), data = df)
df_fit2$deviance - df_fit$deviance > qchisq(p = 0.95, df = 1)

df_fit3 <- glm(formula = y~x1+x2+x4, family = poisson(), data = df)
df_fit3$deviance - df_fit$deviance > qchisq(p = 0.95, df = 1)

df_fit4 <- glm(formula = y~x1+x2+x3, family = poisson(), data = df)
df_fit4$deviance - df_fit$deviance > qchisq(p = 0.95, df = 1)

# part d
summary(df_fit)
std_errors <- sqrt(diag(summary(df_fit)$cov.unscaled))
wald_statistics <- df_fit$coefficients / std_errors
alpha <- 0.05
abs(round(wald_statistics,4)) > qnorm(p = 1 - alpha / 2)

# part e
round(df_fit$coefficients + c(qnorm(p = 1 - alpha / 2) * std_errors),4)
round(df_fit$coefficients - c(qnorm(p = 1 - alpha / 2) * std_errors),4)

### Problem 5
df <- read.csv('p13_26.csv')
n <- 10
set.seed(1); chosen_rows <- sort(sample(seq(1, nrow(df)), n))
df <- df[chosen_rows,]

# Fit a model
df_fit <- glm(formula = y~., family = poisson(), data = df)
summary(df_fit)
anova(df_fit)

# Calculate the deviance
deviance_D <- summary(df_fit)$deviance
deviance_D > qchisq(p = 0.95, df = n - 4)
```

```
deviance_D / (n - 4)

# partial deviance
df_fit1 <- glm(formula = y~Oil+Time, family = poisson(), data = df)
df_fit1$deviance - df_fit$deviance > qchisq(p = 0.95, df = 1)

df_fit2 <- glm(formula = y~Temperature+Time, family = poisson(), data = df)
df_fit2$deviance - df_fit$deviance > qchisq(p = 0.95, df = 1)

df_fit3 <- glm(formula = y~Temperature+Oil, family = poisson(), data = df)
df_fit3$deviance - df_fit$deviance > qchisq(p = 0.95, df = 1)

# Wald statistics
summary(df_fit)
std_errors <- sqrt(diag(summary(df_fit)$cov.unscaled))
wald_statistics <- df_fit$coefficients / std_errors
alpha <- 0.05
abs(round(wald_statistics,4)) > qnorm(p = 1 - alpha / 2)

# Wald CI
round(df_fit$coefficients + c(qnorm(p = 1 - alpha / 2) * std_errors),4)
round(df_fit$coefficients - c(qnorm(p = 1 - alpha / 2) * std_errors),4)
```