STAT 526 HW 2

Code is in this R Notebook. Can use outline feature in the PDF to jump to different problems.

P0

Bowen Zheng

P1

> The dataset attitude in the faraway library contains survey responses from clerical employees of a large financial instiution. Approximately 35 employees scores were aggregated for each of 30 departments. Your goal is to determine the "best" linear model to predict rating given the six predictors. Please briefly describe your steps to derive the best model, what is your best model and any output and figures to support this model.

```r
# Load libraries
# datasets contains 'attitude', car provides VIF functions
if(!require(car)) install.packages("car")
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```r
library(faraway)
```

```
##
## Attaching package: 'faraway'
```

```
## The following objects are masked from 'package:car':
##
##     logit, vif
```

```r
library(MASS)
library(car)
library(GGally)
```

```
## Loading required package: ggplot2
```

```
##
## Attaching package: 'GGally'
```

```
## The following object is masked from 'package:faraway':
##
##     happy
```

```
library(corrplot)
```

```
## corrplot 0.95 loaded
```

Let us first have a look at our data.

```
data("attitude")
# 1. Fit the Full Model (no interactions for now)
full_model <- lm(rating ~ ., data = attitude)
summary(full_model)
```
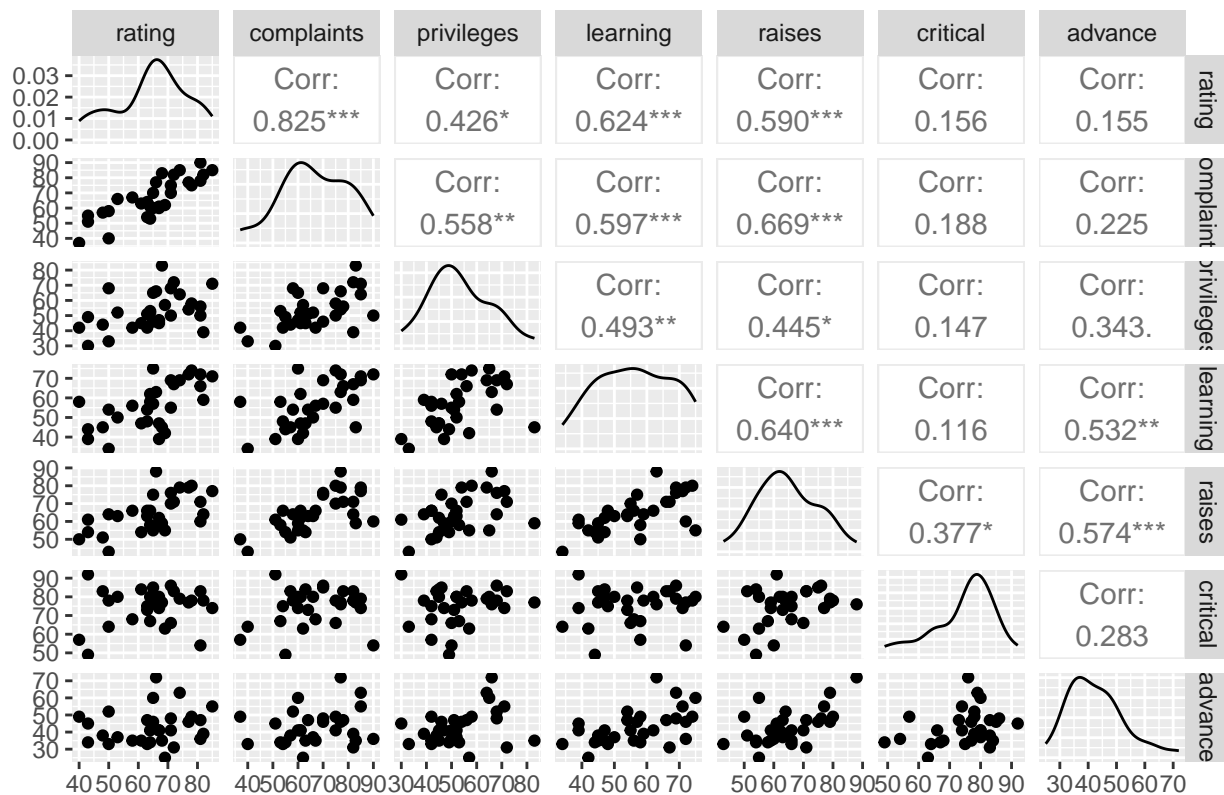
```
##
## Call:
## lm(formula = rating ~ ., data = attitude)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -10.9418  -4.3555   0.3158   5.5425  11.5990
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.78708   11.58926   0.931 0.361634
## complaints   0.61319    0.16098   3.809 0.000903 ***
## privileges  -0.07305    0.13572  -0.538 0.595594
## learning     0.32033    0.16852   1.901 0.069925 .
## raises       0.08173    0.22148   0.369 0.715480
## critical     0.03838    0.14700   0.261 0.796334
## advance     -0.21706    0.17821  -1.218 0.235577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.068 on 23 degrees of freedom
## Multiple R-squared:  0.7326, Adjusted R-squared:  0.6628
## F-statistic:  10.5 on 6 and 23 DF,  p-value: 1.24e-05
```

```
# Check for Multicollinearity
vif(full_model)
```
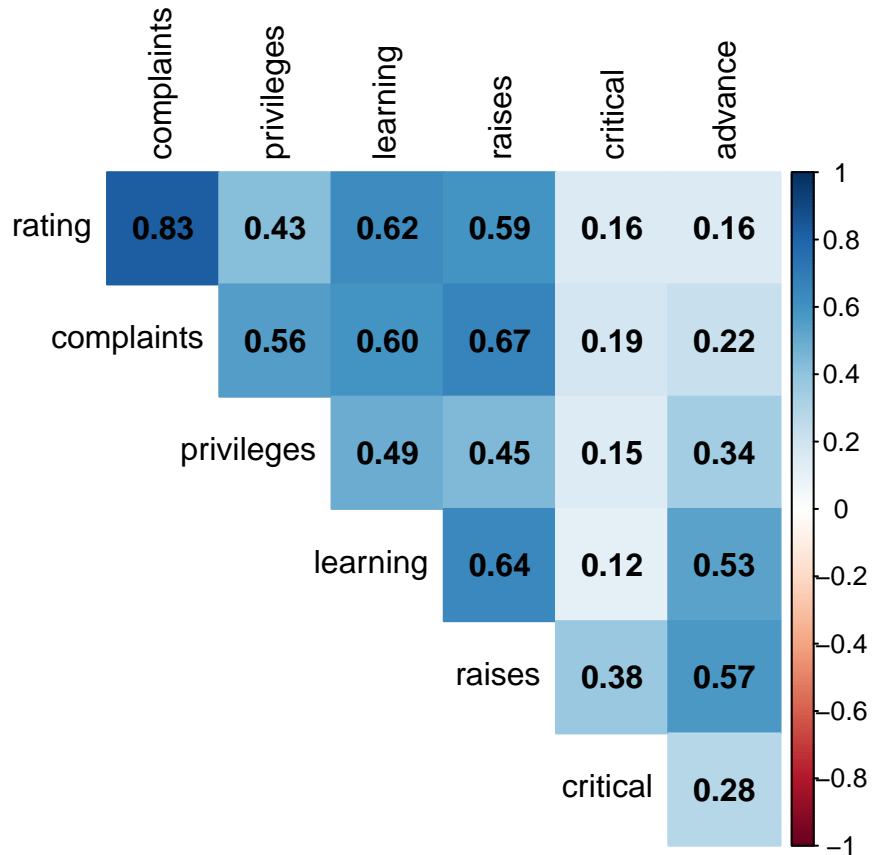
```
## complaints privileges   learning      raises   critical    advance
##   2.667060   1.600891   2.271043   3.078226   1.228109   1.951591
```

```
# Correlation
ggpairs(attitude, title="Scatterplot Matrix of Attitude Data")
```

Scatterplot Matrix of Attitude Data

	rating	complaints	privileges	learning	raises	critical	advance	
0.03 0.02 0.01 0.00		Corr: 0.825***	Corr: 0.426*	Corr: 0.624***	Corr: 0.590***	Corr: 0.156	Corr: 0.155	rating
90 80 70 60 50 40			Corr: 0.558**	Corr: 0.597***	Corr: 0.669***	Corr: 0.188	Corr: 0.225	omplaint
80 70 60 50 40 30				Corr: 0.493**	Corr: 0.445*	Corr: 0.147	Corr: 0.343.	privileges
70 60 50 40					Corr: 0.640***	Corr: 0.116	Corr: 0.532**	learning
90 80 70 60 50						Corr: 0.377*	Corr: 0.574***	raises
90 80 70 60 50							Corr: 0.283	critical
70 60 50 40 30								advance
	40 50 60 70 80	40 50 60 70 80 90	30 40 50 60 70 80	40 50 60 70	50 60 70 80 90	50 60 70 80 90	30 40 50 60 70	

```r
cor_matrix <- cor(attitude)
corrplot(cor_matrix, method = "color", addCoef.col = "black",
         type = "upper", tl.col = "black", diag = FALSE)
```

	complaints	privileges	learning	raises	critical	advance
rating	0.83	0.43	0.62	0.59	0.16	0.16
complaints		0.56	0.60	0.67	0.19	0.22
privileges			0.49	0.45	0.15	0.34
learning				0.64	0.12	0.53
raises					0.38	0.57
critical						0.28

Rating and complaints seems to have rather high correlation. VIF here does not show to us any redundant predictors. We will keep this in mind as we go into variable selection.

```r
# 2. Automated Model Selection (Backward Elimination by AIC)
step_model <- step(full_model, direction = "backward")
```

```
## Start:  AIC=123.36
## rating ~ complaints + privileges + learning + raises + critical +
##     advance
##
##               Df Sum of Sq    RSS    AIC
## - critical     1      3.41 1152.4 121.45
## - raises       1      6.80 1155.8 121.54
## - privileges   1     14.47 1163.5 121.74
## - advance      1     74.11 1223.1 123.24
## <none>                     1149.0 123.36
## - learning     1    180.50 1329.5 125.74
## - complaints   1    724.80 1873.8 136.04
##
## Step:  AIC=121.45
## rating ~ complaints + privileges + learning + raises + advance
##
##               Df Sum of Sq    RSS    AIC
## - raises       1     10.61 1163.0 119.73
## - privileges   1     14.16 1166.6 119.82
## - advance      1     71.27 1223.7 121.25
```

```
## <none>                    1152.4 121.45
## - learning    1    177.74 1330.1 123.75
## - complaints  1    724.70 1877.1 134.09
##
## Step:  AIC=119.73
## rating ~ complaints + privileges + learning + advance
##
##              Df Sum of Sq    RSS    AIC
## - privileges  1     16.10 1179.1 118.14
## - advance     1     61.60 1224.6 119.28
## <none>                     1163.0 119.73
## - learning    1    197.03 1360.0 122.42
## - complaints  1   1165.94 2328.9 138.56
##
## Step:  AIC=118.14
## rating ~ complaints + learning + advance
##
##              Df Sum of Sq    RSS    AIC
## - advance     1     75.54 1254.7 118.00
## <none>                     1179.1 118.14
## - learning    1    186.12 1365.2 120.54
## - complaints  1   1259.91 2439.0 137.94
##
## Step:  AIC=118
## rating ~ complaints + learning
##
##              Df Sum of Sq    RSS    AIC
## <none>                     1254.7 118.00
## - learning    1    114.73 1369.4 118.63
## - complaints  1   1370.91 2625.6 138.16
```

```r
# Let us also look through full interaction model
# May be dealing with a limited data issue, 30 observations, 15 predictors
# The formula ( . )^2 means "all main effects and all their two-way interactions"
full_interaction_model <- lm(rating ~ (.)^2, data = attitude)
# Check the summary
summary(full_interaction_model)
```

```
##
## Call:
## lm(formula = rating ~ (.)^2, data = attitude)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -12.9113  -1.4119  0.0038  3.2482  8.9388
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -70.853887 204.807968  -0.346    0.738
## complaints          0.859653   4.723143   0.182    0.860
## privileges          1.760864   4.723723   0.373    0.719
## learning           -3.429585   4.623592  -0.742    0.479
## raises              0.959619   8.535406   0.112    0.913
## critical            1.755719   3.003720   0.585    0.575
```

```
## advance                1.853385    5.896320    0.314    0.761
## complaints:privileges  -0.014262    0.026615   -0.536    0.607
## complaints:learning     0.023325    0.057722    0.404    0.697
## complaints:raises      -0.016907    0.067765   -0.249    0.809
## complaints:critical    -0.012674    0.064163   -0.198    0.848
## complaints:advance      0.026233    0.041632    0.630    0.546
## privileges:learning    -0.005002    0.067428   -0.074    0.943
## privileges:raises       0.007165    0.100821    0.071    0.945
## privileges:critical     0.002843    0.054572    0.052    0.960
## privileges:advance     -0.029913    0.061901   -0.483    0.642
## learning:raises        -0.008942    0.056199   -0.159    0.878
## learning:critical       0.013085    0.069776    0.188    0.856
## learning:advance        0.048251    0.053595    0.900    0.394
## raises:critical         0.009436    0.091840    0.103    0.921
## raises:advance         -0.007778    0.052586   -0.148    0.886
## critical:advance       -0.058114    0.099434   -0.584    0.575
##
## Residual standard error: 10.08 on 8 degrees of freedom
## Multiple R-squared:  0.8108, Adjusted R-squared:  0.3141
## F-statistic: 1.633 on 21 and 8 DF,  p-value: 0.2425
```

```r
# We use step() to prune this massive model down
best_interaction_model <- step(full_interaction_model, direction = "backward", trace = 0)
```

We use backwards stepwise AIC model selection. One in the case of selecting on just the 6 predictors and the other one selecting on the 6 predictors and 9 interactions. It may not be very helpful to look at the second case because we only have 30 observations but 15 predictors.

```r
# 3. Final Refined Model
# Based on step() results and p-value inspection, we arrive at this model:
best_model <- lm(rating ~ complaints + learning, data = attitude)
summary(best_model)
```

```
##
## Call:
## lm(formula = rating ~ complaints + learning, data = attitude)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -11.5568  -5.7331   0.6701   6.5341  10.3610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.8709     7.0612   1.398    0.174
## complaints    0.6435     0.1185   5.432 9.57e-06 ***
## learning      0.2112     0.1344   1.571    0.128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.817 on 27 degrees of freedom
## Multiple R-squared:  0.708,  Adjusted R-squared:  0.6864
## F-statistic: 32.74 on 2 and 27 DF,  p-value: 6.058e-08
```

```r
summary(best_interaction_model)
```

```
##
## Call:
## lm(formula = rating ~ complaints + learning + advance + learning:advance,
##     data = attitude)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -13.5215  -3.9474   0.6536   3.7923  12.6782
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      63.53570   31.63292   2.009   0.0555 .
## complaints        0.63187    0.11474   5.507 1.01e-05 ***
## learning         -0.53005    0.54008  -0.981   0.3358
## advance          -1.44649    0.78852  -1.834   0.0785 .
## learning:advance  0.02055    0.01266   1.623   0.1171
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.532 on 25 degrees of freedom
## Multiple R-squared:  0.7518, Adjusted R-squared:  0.712
## F-statistic: 18.93 on 4 and 25 DF,  p-value: 2.836e-07
```

The first model seems promising. Lowest AIC and it also has predictors with statistical significance. The second one from the all interaction model seems to do much worse. Let us continue with the first model of only the two additive main predictors: complaints and learning.

Let us check for the interaction between complaints and learning.

```r
model_int <- lm(rating ~ complaints * learning, data = attitude)
anova(best_model, model_int)
```

```
## Analysis of Variance Table
##
## Model 1: rating ~ complaints + learning
## Model 2: rating ~ complaints * learning
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     27 1254.7
## 2     26 1232.1  1    22.553 0.4759 0.4964
```

```r
# Check the AIC for both
AIC(best_model, model_int)
```

```
##            df      AIC
## best_model  4 205.1387
## model_int   5 206.5945
```
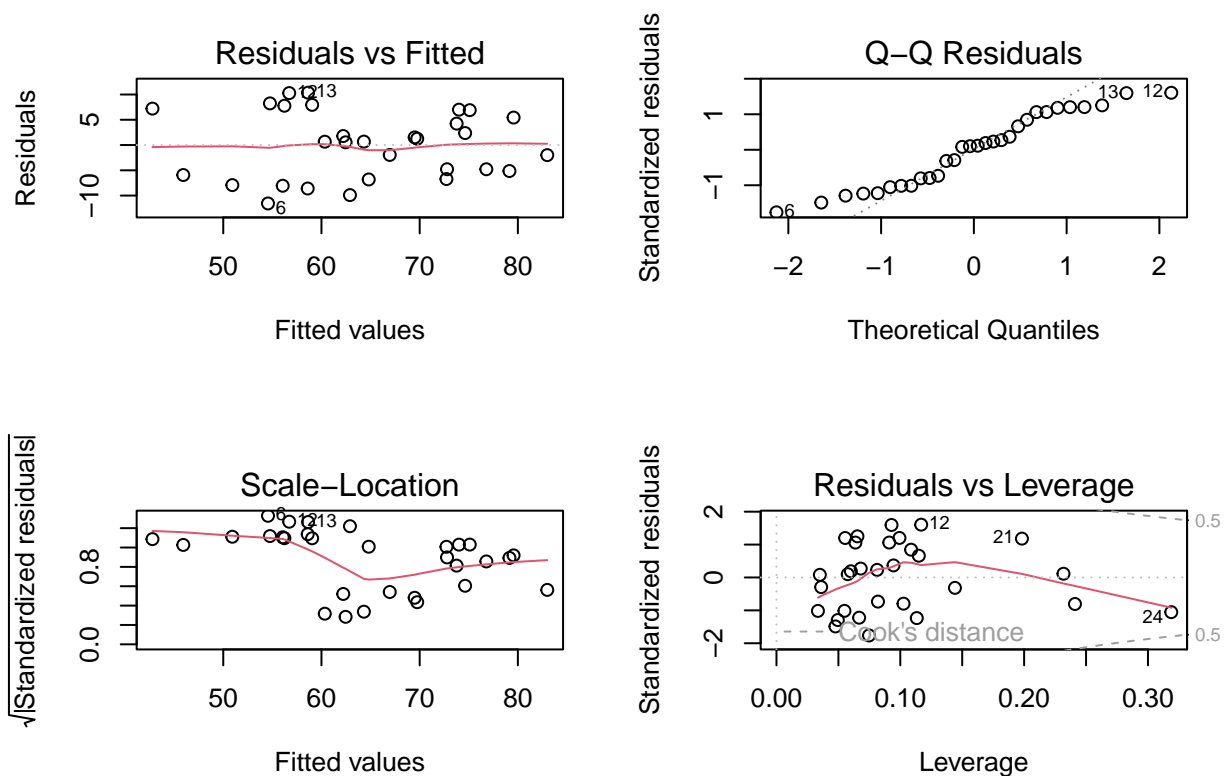
```r
# 5. Look at the summary of the interaction model
summary(model_int)
```

```
##
## Call:
## lm(formula = rating ~ complaints * learning, data = attitude)
##
## Residuals:
##     Min     1Q  Median      3Q     Max
## -11.804  -5.831   1.374   4.328  10.625
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          31.203424  31.733940   0.983    0.335
## complaints            0.306661   0.502732   0.610    0.547
## learning             -0.170646   0.569888  -0.299    0.767
## complaints:learning   0.005886   0.008531   0.690    0.496
##
## Residual standard error: 6.884 on 26 degrees of freedom
## Multiple R-squared:  0.7133, Adjusted R-squared:  0.6802
## F-statistic: 21.56 on 3 and 26 DF,  p-value: 3.174e-07
```

It seems like adding the interaction is not helpful. ANOVA comparing of the fit of the base model vs the interaction model shows that we shoudl not choose the interaction model. The interaction model also has higher AIC. We continue with our basse model again.

```
# 4. Diagnostic Plots
par(mfrow=c(2,2))
plot(best_model)
```

This generates 4 plots: Residuals vs Fitted, Normal Q-Q, Scale-Location, and Cook's distance. This is to make sure that our model properly satisfies the assumptions needed. The residuals vs fitted seems random, scale location and residuals vs leverage seem ok, the QQ plot is generally along the line, etc. Our diagnosis looks decent. Our final model is rating predicted by learning and complaints.

$$\text{rating} = 9.87 + 0.644(\text{complaints}) + 0.211(\text{learning})$$

While it is hard to interpret the final model given the lack of prior knowledge, we do know that learning is not statistically significant in the model. However, it does lower the AIC score so it does provide a meaningful improvement to the model so we choose to include it. It could be that more complaints and learning increase the rating and that the average rating is already quite high at 9.87.

P2

Faraway Chapter 8 Exercise #1.

Data is generated from the exponential distribution with density $f(y) = \lambda \exp(-\lambda y)$ where $\lambda, y > 0$.

a.)

Identify the specific form of λ, ϕ,a(),b() and c() for the exponential distribution.

$$
\begin{aligned}
f(y; \theta, \phi) &= \lambda e^{-\lambda y} \\
&= \exp(\ln(\lambda e^{-\lambda y})) \\
&= \exp(-\lambda y + \ln \lambda) \\
&= \exp(\frac{y(-\lambda) - (-\ln\lambda)}{1} + 0)
\end{aligned}
$$

This gives us $\theta = -\lambda$, $b(\theta) = -ln(-\theta) = -ln(\lambda)$, $\phi = 1$, $a(\phi) = 1$, $c(y, \phi) = 0$.

b.)

What is the canonical link and variance function for a GLM with a response following the exponential distribution?

The canonical link function for Exponential distribution is that of Gamma. Exponential distribution is a special form of Gamma. $\eta = \mu^{-1}$ according to the Faraway textbook page 154.

Technically, however, we know that the actual canonical form is $\mu = E[y]$. Since $\theta = -\lambda$ and $\mu = \dfrac{1}{\lambda} = \dfrac{-1}{\theta}$.
Then $\theta = \dfrac{-1}{\mu}$. Then $g(\mu) = -\mu^{-1}$ is our link function.

We know that $Var(Y) = b''(\theta)a(\phi)$.

$$
b''(\theta) = \frac{d}{d\theta}\left(-\frac{1}{\theta}\right) = \frac{1}{\theta^2}
$$

Then $Var(Y) = \frac{1}{\theta^2} = \mu^2$.

c.)

Identify a practical difficulty that may arise when using the canonical link in this instance.

When trying to predict $\eta = \frac{-1}{\mu} = X\beta$, we may find that we can get data such that $X\beta$ is positive and thus the mean is negative. Exponential is always semi-positive mean, so this would be a big issue if that were to be the case.

d.)

When comparing nested models in this case, should an F or X^2 test be used? Explain.

This is not a regular SSE model. Since we are using GLM on exponential, we know ϕ and we should use Pearson X^2 test. The F test with $\phi = 1$ is just a scaled version of the Chi-squared test anyways. However, it is likely that for small data sets, we cannot meet the assumptions of the chi-squared test.

e.)

Express the deviance in this case in terms of the responses y_i and the fitted values $\hat{\mu}_i$

From the textbook, for Gamma distributions, of which exponential is, we find

$$D = 2 \sum_{i=1}^{n} \left[\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} - \ln\left(\frac{y_i}{\hat{\mu}_i}\right) \right]$$

We can derive this relatively simply by plugging into our deviance formula.

Deviance is defined:

$$D = 2 \sum [\ell(y_i, y_i) - \ell(\hat{\mu}_i, y_i)]$$

We start with the log likelihood of $f(y) = \lambda \exp(-\lambda y)$. By plugging in $\lambda = 1/\mu$, we find

$$f(y; \mu) = \frac{1}{\mu} \exp\left(-\frac{y}{\mu}\right)$$

$$\ln f(y; \mu) = -\ln(\mu) - \frac{y}{\mu}$$

We have two log likelihoods:

$$\ell(y_i, y_i) = -\ln(y_i) - \frac{y_i}{y_i} = -\ln(y_i) - 1$$

$$\ell(\hat{\mu}_i, y_i) = -\ln(\hat{\mu}_i) - \frac{y_i}{\hat{\mu}_i}$$

We can simply plug these values into our deviance formula:

$$D = 2 \sum [\ell(y_i, y_i) - \ell(\hat{\mu}_i, y_i)]$$
$$= 2 \sum_{i=1}^{n} \left[(-\ln(y_i) - 1) - \left(-\ln(\hat{\mu}_i) - \frac{y_i}{\hat{\mu}_i}\right) \right]$$
$$= 2 \sum_{i=1}^{n} \left[\frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} - \ln\left(\frac{y_i}{\hat{\mu}_i}\right) \right]$$

This is the same as the one in the textbook.

P3

Faraway Chapter 8 Exercise #2.

The Conway–Maxwell–Poisson distribution has probability function: $P(Y = y) = \frac{\lambda^y}{(y!)^\nu} \frac{1}{Z(\lambda,\nu)}$.
Place this in exponential family form, identifying all the relevant components necessary for use in a GLM.

We start with the pmf

$$
\begin{aligned}
P(Y = y) &= \frac{\lambda^y}{(y!)^\nu} \frac{1}{Z(\lambda,\nu)} \\
&= \exp\left[\ln\left(\frac{\lambda^y}{(y!)^\nu Z(\lambda,\nu)}\right)\right] \\
&= \exp\left(y\ln\lambda - \nu\ln(y!) - \ln Z(\lambda,\nu)\right)
\end{aligned}
$$

OK, so now we fit it in the form.

$$
f(y; \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)
$$

So this gives us

$$
\begin{aligned}
\theta &= ln\lambda \\
\phi &= v \\
a(\phi) &= 1 \\
b(\theta) &= \ln Z(e^\theta, v) \\
c(y,\phi) &= -v\ln(y!)
\end{aligned}
$$

P4

Consider the following distributions and determine whether they can be put in exponential family form. If it is not possible, consider whether there is a special case which follows the form.

Uniform

$$f(y; \alpha, \beta) = \frac{1}{\beta - \alpha} I(y \in [\alpha, \beta])$$
$$= \exp(-\ln(\beta - \alpha)) \cdot I_{[\alpha, \beta]}(y)$$

I do not think it is possible to absorb the indicator variable into theta or b(theta) terms. The special case would be if we know alpha and beta, but that would not be an interesting distribution and we would not have to use GLM for if that were the case.

Weibull

$$f(y; \alpha, \beta) = \alpha \beta^{-\alpha} y^{\alpha-1} \exp\left[\left(\frac{y}{\beta}\right)^\alpha\right]$$
$$= \exp\left(\ln\left(\frac{\alpha}{\beta^\alpha}\right) + (\alpha - 1)\ln(y) + \left(\frac{y}{\beta}\right)^\alpha\right)$$

There is no way to make a $y\theta$ term here, necessary for determining that it is EFD. The general Weibull distribution is not an exponential family.

However, we know that, more generally, if we fix α then the Weibull distribution is a one parameter EFD. Even so, it still is unable to be explicitly stated in the way that we need for EFD for GLM. We can reach the form required when $\alpha = 1$. Then we find

$$f(y; 1, \beta) = \frac{1}{\beta} \exp\left(\frac{y}{\beta}\right)$$

By letting $\theta = -1/\beta$, $b(\theta) = \ln(-1/\theta)$, $c = 0$ and $a(\phi) = 1$, we get:

$$f(y; \theta) = \exp(y\theta - \ln(-1/\theta))$$

For general case of fixed α, we can transform $Z = y^\alpha$, and use Z instead. We can simplify our earlier formula and find

$$f(y) = \exp\left(\frac{y^\alpha}{\beta^\alpha} + \ln(\alpha) - \alpha \ln(\beta) + (\alpha - 1)\ln(y)\right)$$

Where $z = y^\alpha$, $a(\phi) = 1$, $\theta = \beta^{-\alpha}$, $b(\theta) = ln(\theta^{-1/\alpha})$, and c would contain the remaining terms.

Gumbel distribution

$$f(y; \alpha, \beta) = \frac{1}{\beta} \exp\left(\frac{y - \alpha}{\beta} - \exp\left(\frac{y - \alpha}{\beta}\right)\right)$$
$$= \exp\left(\frac{y}{\beta} - \frac{\alpha}{\beta} - \exp\left(\frac{y}{\beta}\right)\exp\left(-\frac{\alpha}{\beta}\right) - \ln(\beta)\right)$$

Again, we cannot find a general $y\theta$ because for any θ we choose, we have $\exp(y/\beta) \cdot \exp(-\alpha/\beta)$ also that would mess it up. This does not fit the usual EFD form for GLM.

For fixed β we still cannot have the form we want. However, we can have a transformation like earlier in the Weibull distribution AND fix/know β. Let $Z = \exp(y/\beta)$. Then $\theta = -\exp(-\alpha/\beta)$, $b(\theta) = -ln(-\theta)$, and c contains our remaining terms.