HW 5 sol'ns

1.

Let $z = \frac{y-\mu}{\sqrt{\mu}}$, then

$$M_z(t) = E(e^{tz}) = E(e^{t\cdot\frac{y-\mu}{\sqrt{\mu}}}) = E(e^{\frac{ty}{\sqrt{\mu}}})e^{-\sqrt{\mu}t} = M_y(\frac{t}{\sqrt{\mu}})e^{-\sqrt{\mu}t}$$

Since the variable $y \sim Poison(\mu)$,

$$M_y(t) = \sum_{y=0}^{\infty} e^{ty}\cdot\frac{e^{-\mu}\mu^y}{y!} = e^{-\mu}\sum_{y=0}^{\infty}\frac{(e^t\mu)^y}{y!} = e^{-\mu}(e^{e^t\mu}) = e^{\mu(e^t-1)}$$

Thus,

$$M_z(t) = M_y(\frac{t}{\sqrt{\mu}})e^{-\sqrt{\mu}t} = \exp(\mu e^{\frac{t}{\sqrt{\mu}}} - \mu - \sqrt{\mu}t)$$

We have

$$K_z(t) = \mu e^{\frac{t}{\sqrt{\mu}}} - \mu - \sqrt{\mu}t$$

Hence,

$$K_1(0) = (\mu e^{\frac{t}{\sqrt{\mu}}}\frac{1}{\sqrt{\mu}} - \sqrt{\mu})|_{t=0} = \sqrt{\mu} - \sqrt{\mu} = 0$$

$$K_2(0) = \mu e^{\frac{t}{\sqrt{\mu}}}\frac{1}{\mu}|_{t=0} = 1$$

$$K_3(0) = \mu e^{\frac{t}{\sqrt{\mu}}}(\frac{1}{\sqrt{\mu}})^3|_{t=0} = \mu^{-\frac{1}{2}}$$

$$K_4(0) = \mu e^{\frac{t}{\sqrt{\mu}}}(\frac{1}{\sqrt{\mu}})^4|_{t=0} = \mu^{-1}$$

$$K_5(0) = \mu e^{\frac{t}{\sqrt{\mu}}}(\frac{1}{\sqrt{\mu}})^5|_{t=0} = \mu^{-\frac{3}{2}}$$

$$\vdots$$

$$K_n(0) = \mu e^{\frac{t}{\sqrt{\mu}}}(\frac{1}{\sqrt{\mu}})^n|_{t=0} = \mu^{1-\frac{n}{2}}$$

Since the cumulants for $N(0,1)$ are 0, 1, 0, 0, 0, $\cdots$, and z has cumulants 0, 1, $O_p(\mu^{-\frac{1}{2}})$, $O_p(\mu^{-\frac{1}{2}})$, $\cdots$, then we have

$$z = \frac{y-\mu}{\sqrt{\mu}} \sim N(0,1) + O_p(\mu^{-\frac{1}{2}})$$

2.

We need to use the Taylor series expansion for $g(y) = \sqrt{y}$ about μ

$$g(y) \approx g(\mu) + (y-\mu)g'(\mu) + (y-\mu)^2\frac{g''(\mu)}{2}$$
$$\sqrt{y} \approx \sqrt{\mu} + \frac{1}{2}(y-\mu)\frac{1}{\sqrt{\mu}} - \frac{1}{8}(y-\mu)^2\mu^{-\frac{3}{2}}$$
$$E(\sqrt{y}) \approx \sqrt{\mu} + \frac{1}{2}E(y-\mu)\frac{1}{\sqrt{\mu}} - \frac{1}{8}E(y-\mu)^2\mu^{-\frac{3}{2}}$$
$$E(\sqrt{y}) \approx \sqrt{\mu} - \frac{1}{8}\mu\mu^{-\frac{3}{2}}$$
$$\Rightarrow E(\sqrt{y}) \approx \sqrt{\mu} - \frac{1}{8}\mu^{-\frac{1}{2}} = \mu^{\frac{1}{2}}(1 - \frac{1}{8\mu})$$

3.

Let $y \sim Bin(m, p_m)$, $m = 1, 2, 3, \cdots$ such that $p_m \to 0$ and $mp_m \to \lambda$ as $m \to \infty$. The moment generating function of y is

$$M_y(t) = (e^t p_m + 1 - p_m)^m = (1 + p_m e^t - p_m)^m = \left(1 + \frac{mp_m(e^t-1)}{m}\right)^m$$
$$\xrightarrow{m\to\infty} e^{\lambda(e^t-1)}$$

Thus,

$$\lim_{m\to\infty} M_y(t) = e^{\lambda(e^t-1)},$$

which is the moment generating function of the poison distribution with mean λ. Therefore, the distribution of y will approach a poison distribution with mean λ as $m \to \infty$.

4. We first fit the model using indicator functions and compare it to the model inputting the predictors linearly. We then see that we can eliminate two of the predictors, so our final model is

$$\log(\text{death}) = \beta_0 + \log(\text{exp.death}) + \beta_1 \text{EXP} + \beta_2 \text{TFE}.$$

We calculate $\hat{\beta}_1 = 0.519$ and $\hat{\beta}_2 = -0.552$. This indicates that for a one unit increase in exposure category (roughly corresponding to 4 unit increase in exposure) we expect the number of deaths to increase by 67.7%. Also for a unit increase in employment time ($\approx$ 10 years) we expect the number of deaths to decrease by a factor of 0.576 (for a fixed exposure category).

Clearly, we are able to conclude higher levels of nickel exposure indicate a larger death toll from lung cancer.

Because the model deviance is less than the degrees of freedom, we do not require an overdispersion parameter.

R code for # 4

```
> nickel.df <- read.table('nickel.txt')
> nickel.df <- nickel.df[c(1:4,6,10)]
> colnames(nickel.df) <- c('AFE','YFE','EXP','TFE','death','exp.death')
> nickel.df$exp.death <- nickel.df$exp.death/1000
>
> nickel.full <- glm(death ~ factor(AFE) + factor(YFE)+ factor(EXP) + factor(TFE)
+      + offset(log(exp.death)), family=poisson, data=nickel.df)
> nickel.linear <- glm(death ~ AFE + YFE + EXP + TFE + offset(log(exp.death)), family=poisson, data=nickel.df)
> ( deviance(nickel.linear)-deviance(nickel.full) < qchisq(.95,10) )
[1] TRUE
>
> summary(nickel.linear)

Call:
glm(formula = death ~ AFE + YFE + EXP + TFE + offset(log(exp.death)),
    family = poisson, data = nickel.df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5163  -0.8275  -0.3791   0.1153   2.9845

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.38806    0.65157   5.200 1.99e-07 ***
AFE         -0.04800    0.11379  -0.422    0.673
YFE         -0.14833    0.09468  -1.567    0.117
EXP          0.42507    0.09272   4.584 4.56e-06 ***
TFE         -0.60044    0.09244  -6.495 8.28e-11 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 298.77  on 241  degrees of freedom
Residual deviance: 205.38  on 237  degrees of freedom
AIC: 396.88

Number of Fisher Scoring iterations: 5

>
> nickel.et <- glm(death ~ EXP + TFE + offset(log(exp.death)), family=poisson, data=nickel.df)
> ( deviance(nickel.et)-deviance(nickel.linear) < qchisq(.95,2) )
[1] TRUE
>
> summary(nickel.et)

Call:
glm(formula = death ~ EXP + TFE + offset(log(exp.death)), family = poisson,
    data = nickel.df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7456  -0.8241  -0.3710   0.1881   3.1774

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.49732    0.30151   8.283  < 2e-16 ***
```

```
EXP          0.51695    0.07248   7.132 9.90e-13 ***
TFE         -0.55242    0.08069  -6.847 7.57e-12 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 298.77  on 241  degrees of freedom
Residual deviance: 208.07  on 239  degrees of freedom
AIC: 395.57

Number of Fisher Scoring iterations: 5

>
> exp(nickel.et$coeff[2:3])
      EXP       TFE
1.6769053 0.5755538
>
```