3.11, 3.12, 3.13, 3.20.

3.11. a. For the model $\log \mu = \alpha + \beta x$, where $x = 1$ for treatment B and $x = 0$ for treatment A. Then

$\log \mu_B = \alpha + \beta$, and $\log \mu_A = \alpha$, it is obvious

$\beta = \log \mu_B - \log \mu_A = \log(\mu_B/\mu_A)$,

and $e^{\beta} = \mu_B/\mu_A$.

b. Fitting the model, we get $\log(\hat{\mu}) = 1.6094 + 0.5878x$.

$\exp(\hat{\beta}) = \exp(0.5878) = 1.80$. The estimated mean number of defects with treatment B is 80% higher than that with treatment A.

```
glm(formula = defects ~ trt, family = poisson, data = wafer)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5280  -0.7622  -0.1699   0.6938   1.5399

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.6094     0.1414  11.380  < 2e-16 ***
trtB          0.5878     0.1764   3.332 0.000861 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 27.857  on 19  degrees of freedom
Residual deviance: 16.268  on 18  degrees of freedom
AIC: 94.35

Number of Fisher Scoring iterations: 4
```

c. To test $H_0 : \beta = 0$ vs $H_a : \beta \neq 0$, the Wald test statistic is $z = 3.332$ with p-value = 0.0009. Reject H_0. The mean number of defects are significantly different in the two treatment groups.

d. A 95% CI for $log(\mu_B/\mu_A)$ (witch is just β) is:

$0.5878 \pm 1.960.1764 = (0.2421, 0.9335)$ and

A 95% CI for (μ_B/μ_A) is

$(\exp(0.2421), \exp(0.9335)) = (1.27, 2.54)$

The estimated mean number of defects in Treatment B is between 1.27 to 2.54 times the estimated mean number of defects in Treatment A.

3.12. Adding coating as a predictor, we get

```
glm(formula = defects ~ trt + coating, family = poisson, data = wafer)

Deviance Residuals:
```

```
    Min       1Q   Median       3Q      Max
-1.2952  -0.6785  -0.2688   0.6776   1.6307

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.7177     0.1602  10.719  < 2e-16 ***
trtB          0.5878     0.1764   3.332 0.000861 ***
coating1     -0.2296     0.1701  -1.349 0.177246
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 27.857  on 19  degrees of freedom
Residual deviance: 14.435  on 17  degrees of freedom
AIC: 94.517

Number of Fisher Scoring iterations: 4
```

The conditional effects of treatment is:

$\exp(0.5878) = 1.80$. Controlling for coating, the estimated mean number of defects in B is 80% higher than in A.

$\exp(-0.2296) = 0.79$. Controlling for treatment, the estimated mean number of defects is 21% lower with thick coating than with thin coating. But this effect is not significant.

3.13.

a. Change the weight unit to kg (this can be done by dividing the weight column by 1000), and fit the model we get

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.42841    0.17893  -2.394   0.0167 *
weigh        0.58930    0.06502   9.064   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 632.79  on 172  degrees of freedom
Residual deviance: 560.87  on 171  degrees of freedom
AIC: 920.16
```

The fitted equation is $\log(\hat{\mu}) = -0.4284 + 0.5893x$

b. $\exp(\log(\hat{\mu})) = \exp(-0.4284 + 0.5893 * 2.44) = 2.74$.

c. $\exp(0.5893 \pm 1.96 * 0.065) = \exp(0.4619, 0.7167) = (1.59, 2.05)$.

The estimated mean number of satellites increase between 59% and 105% for each additional kg increase in weight.

d. From the output, the z test statistic is 9.064 with p-value close to 0. The effect of weight is significant.

e. The LR test can be performed by comparing the null deviance and residual deviance. The difference is $632.79 - 560.87 = 71.92, d.f. = 1$,

p-value $= P(\chi^2 > 71.92) \approx 0$.

3.20.

```
     age personyears smoking deaths  deathrate
1  35-44       18793      no      2  0.1064226
2  35-44       52407     yes     32  0.6106055
3  45-54       10673      no     12  1.1243324
4  45-54       43248     yes    104  2.4047355
5  55-64        5710      no     28  4.9036778
6  55-64       28612     yes    206  7.1997763
7  65-74        2585      no     28 10.8317215
8  65-74       12663     yes    186 14.6884624
9  75-84        1462      no     31 21.2038304
10 75-84        5317     yes    102 19.1837502
```

a. The death rate per 1000 person-years are given in the deathrate columun above. We can see the death rate increases with age, for both smokers and nonsmokers.

b.

```
glm(formula = deaths ~ age + smoking, family = poisson, data = doctor,
    offset = log(personyears))

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -7.9194     0.1918 -41.298  < 2e-16 ***
age45-54      1.4840     0.1951   7.606 2.82e-14 ***
age55-64      2.6275     0.1837  14.301  < 2e-16 ***
age65-74      3.3505     0.1848  18.131  < 2e-16 ***
age75-84      3.7001     0.1922  19.250  < 2e-16 ***
smokingyes    0.3545     0.1074   3.302  0.00096 ***

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 935.091  on 9  degrees of freedom
Residual deviance:  12.134  on 4  degrees of freedom
AIC: 79.202
```

b. The main effects only model assumes constant ratio of nonsmokers to smokers death rates over levels of age. It also assumes the age effect is the same

over levels of smoking status. Based on a), the model might not be proper as the effect of smoking becomes less pronounced at older age groups.

c. Based on a, we can see the death rate strictly increases with age, therefore we can treat age as a quantitative variable. This model can be specified as $\log(\mu/t) = \alpha + \beta_1 age + \beta_2 smoking + \beta_3(smoking * age)$.

Now for smokers, $\log(\mu/t) = \alpha + \beta_2 + (\beta_1 + \beta_3)age$,

for nonsmokers, $\log(\mu/t) = \alpha + \beta_1 age$.

So for either smokers or nonsmokers, the log of the rate changes linearly with age.

d. Assign scores 1, 2, 3, 4, 5 to the age groups, we get

```
Call:
glm(formula = deaths ~ age2 * smoking, family = poisson, data = doctor,
    offset = log(personyears))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.8784  -2.1219  -0.2482   1.7184   3.5269

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)    -8.86716    0.30567 -29.009  < 2e-16 ***
age2            1.04685    0.07743  13.520  < 2e-16 ***
smokingyes      1.28369    0.32583   3.940 8.16e-05 ***
age2:smokingyes -0.24899    0.08359  -2.979  0.00289 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 935.091  on 9  degrees of freedom
Residual deviance:  59.895  on 6  degrees of freedom
AIC: 122.96

Number of Fisher Scoring iterations: 4
```

By comparing the deviance of the two models, the model in b) seems more appropriate. This model also has a smaller AIC value. The model in d does not fit well perhaps due to the fact the log of rate does not change linearly with age. In fact, adding an age^2 terms can significantly improve the fit of the model.