$\theta = \frac{\mu_1}{\mu_0}$ $\mu_1 = \theta \times \mu_0 \xrightarrow{\log} \log(\mu_1) = \log(\theta) + \log(\mu_0) \rightarrow$ South

$\log(\mu) = \log(\mu_0) + \log(\theta) \cdot \text{South} \rightarrow$ general equations

Population regression equations

if you're in the north

$\mu_{north} = \mu_0$

if you're in the south

$\mu_{south} = \mu_0 + (\Delta\mu)$

the population mean depth is determined by μ_0, $\Delta\mu$

A deterministic equation

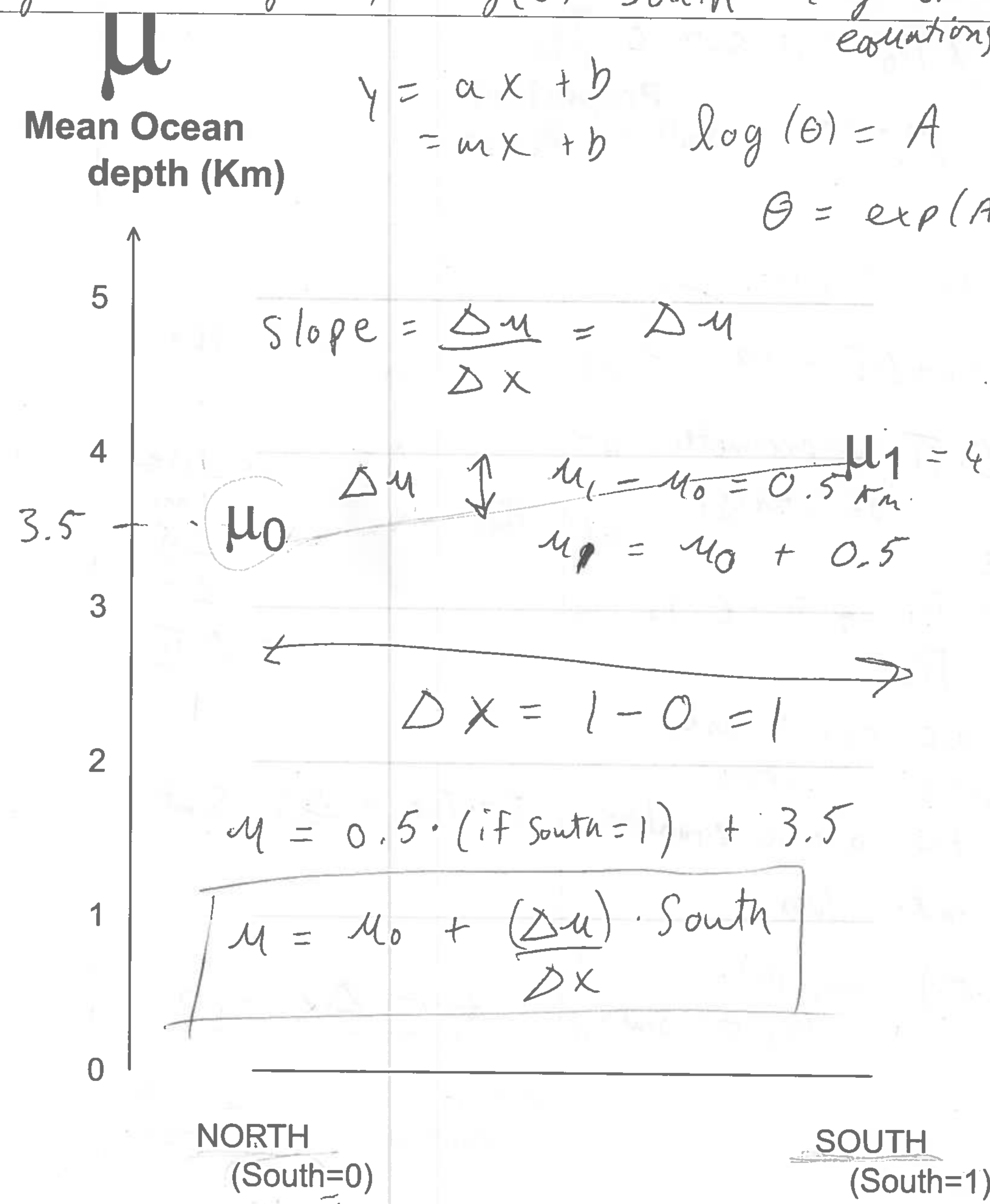
South=0 $\rightarrow$ $\mu_{north} = 0.5(0) + 3.5$

South=1 $\rightarrow$ $\mu_{south} = 0.5(1) + 3.5$

parameter contrast

$\mu_{south} - \mu_{north} = (\mu_0 + \Delta\mu) - \mu_0$

$= \Delta\mu$



$y = ax + b$

$= mx + b$

$\log(\theta) = A$

$\theta = \exp(A)$

$\text{slope} = \frac{\Delta\mu}{\Delta x} = \Delta\mu$

$\Delta\mu \updownarrow$ $\mu_1 - \mu_0 = 0.5$ km

$\mu_1 = 4$ km

$\mu_1 = \mu_0 + 0.5$

$\Delta x = 1 - 0 = 1$

$\mu = 0.5 \cdot (\text{if South}=1) + 3.5$

$$\boxed{\mu = \mu_0 + \frac{(\Delta\mu)}{\Delta x} \cdot \text{South}}$$

one-sample $\mu = \mu_0 \rightarrow$ intercept only

$$\pi = \begin{cases} \pi_0, & \text{if South}=0 \\ \pi_0 + \Delta\pi, & \text{if South}=1 \end{cases}$$

π
Proportion Water

$\pi_{south} - \pi_{north}$

$= \pi_0 + \Delta\pi - \pi_0$

$= \Delta\pi \rightarrow$ parameter of interest

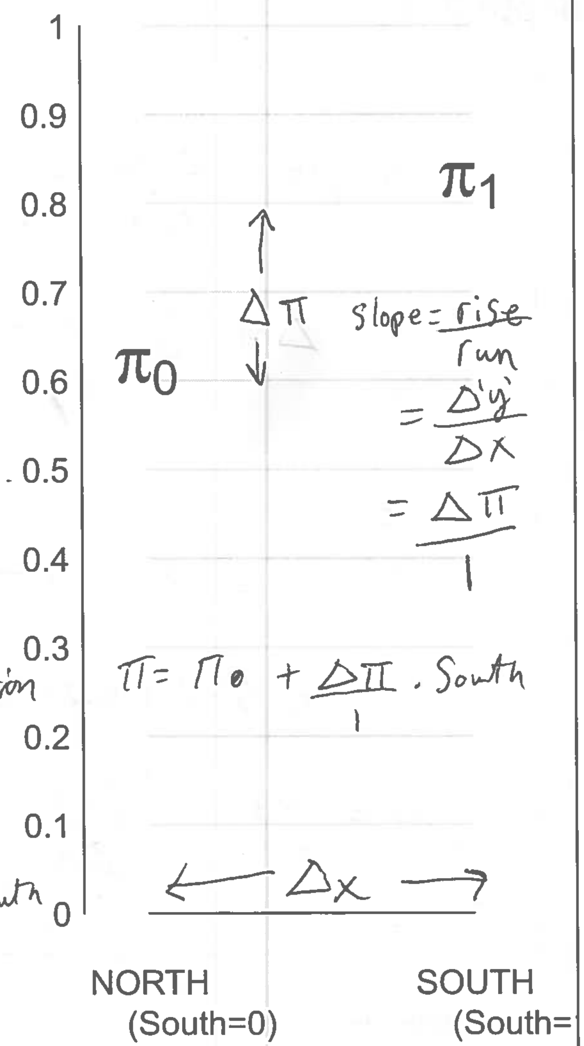
risk

$\theta = \frac{\pi_1}{\pi_0} \rightarrow \pi_1 = \theta \cdot \pi_0$

But we can't use regression routines for the above equation

A: take log

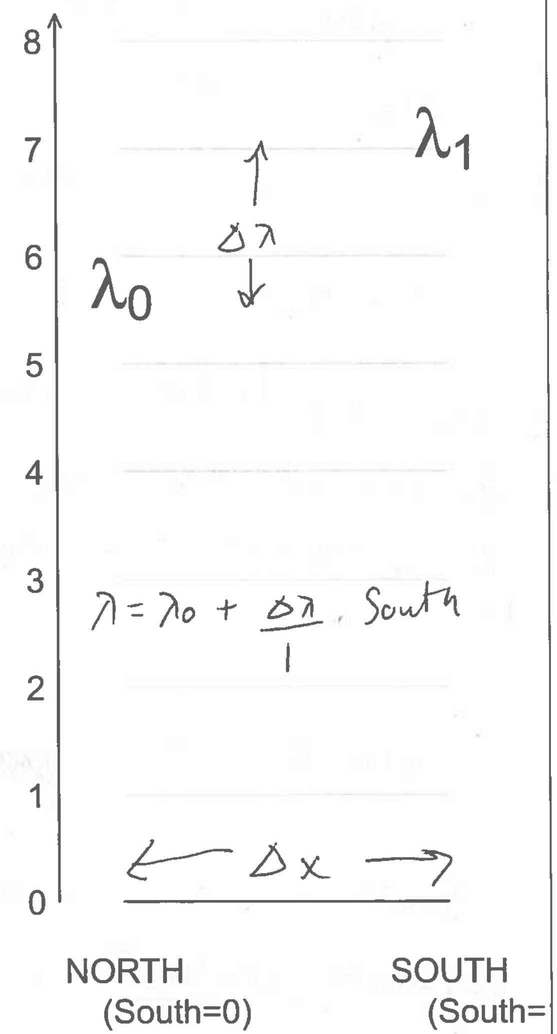
$\log(\pi) = \log(\pi_0) + \log(\theta) \cdot \text{South}$



slope $= \frac{\text{rise}}{\text{run}} = \frac{\Delta y}{\Delta x} = \frac{\Delta\pi}{1}$

$\pi = \pi_0 + \frac{\Delta\pi}{1} \cdot \text{South}$

λ
Magnitude 6 or higher Earthquakes/Month



$\lambda = \lambda_0 + \frac{\Delta\lambda}{1} \cdot \text{South}$

one-sided → $H_a: \mu_0 > 0$, p-value $= P(t_{stat} > t_{(n-1)} \mid H_0)$

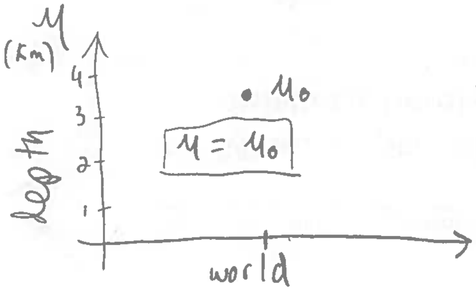
$H_a: \mu_0 < 0$, $= P(t_{stat} < t_{(n-1)} \mid H_0)$

1 Mean depth of the ocean

```
head(depths)
                                      y_i
##              X        lon       lat  alt water South
## 41995 41995  -87.21236 59.290367  190     1     0
## 11151 11151 -122.33034  5.554558 4167     1     0
## 43640 43640 -148.54790 36.237464 5447     1     0
## 8615   8615  -24.92364 21.625967 5063     1     0
## 8126   8126  177.18458 13.880370 5634     1     0
## 16548 16548   48.88215  3.229250 3691     1     0

dim(depths)
       n=
## [1] 400   6

fit <- lm(alt ~ 1, data = depths)
summary(fit)

##
## Call:
## lm(formula = alt ~ 1, data = depths)
##
## Residuals:
##     Min      1Q  Median     3Q     Max
## -3681.5  -584.8   405.5 1197.2  2827.5
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3683.52      78.71    46.8   <2e-16 ***
##                 1           2        3       5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##                  6      4
## Residual standard error: 1574 on 399 degrees of freedom
```

"~1" means intercept only

- there are no determinants of μ
- the true mean depth is equal to μ_0
- This model is the "mother of all regressions"
- Intercept only regression



① estimate of $\mu_0 \to \hat{\beta}_0$ is exactly equal to the sample mean $\bar{y}$

mean(depths\$alt) = 3683.52

② standard error of the mean → $s/\sqrt{n}$

sd(depths\$alt)$/\sqrt{400} = 78.71$

③ test statistic. $H_0: \mu_0 = 0$ $H_a: \mu_0 \neq 0$

$$t_{stat} = \frac{\bar{Y} - \mu_0}{s/\sqrt{n}} = \frac{3683.52 - 0}{78.71} = 46.8$$

④ and ⑤ p-value: $P(|t_{stat}| > t_{(n-1)} \mid H_0)$ → two sided

$= P(t_{stat} < t_{(n-1)} \mid H_0) + P(t_{stat} > t_{(n-1)} \mid H_0)$

pt(q = 46.8, df = 399, lower.tail = F) × 2

= pt(q = -46.8, 399, lower.tail = T) +

pt(q = 46.8, 399, lower.tail = F)

⑥ Residual = observed − predicted

= depths\$alt − 3683.52

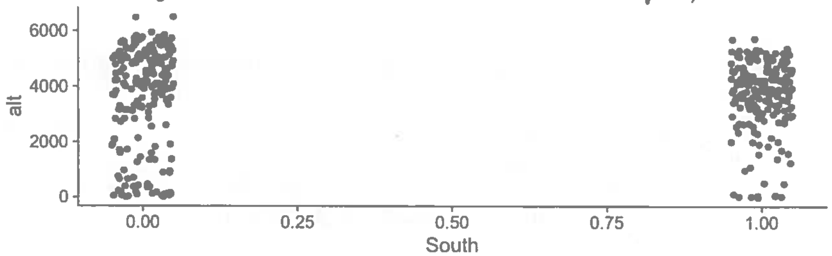
$e_i = y_i - \hat{y}_i$, $i = 1, \ldots, n$

residual standard error $= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} e_i^2}$

s = residual standard error ONLY in the intercept only model $= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$

2 Mean depth of the ocean in northern and southern hemisphere

ggformula :: gf_jitter (alt ~ south, data=depths, width= 0.05)




```
fit <- lm(alt ~ South, data = depths)
summary(fit)
```

confint (fit)

```
##
## Call:
## lm(formula = alt ~ South, data = depths)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3722.0  -608.5   401.5  1200.4  2867.9
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3643.08     111.42  32.698   <2e-16 ***
## South          80.88     157.56   0.513    0.608
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1576 on 398 degrees of freedom
## Multiple R-squared:  0.0006617,Adjusted R-squared:  -0.001849
## F-statistic: 0.2635 on 1 and 398 DF,  p-value: 0.608

t.test(alt ~ South, data = depths, var.equal = TRUE)

##
##  Two Sample t-test
##
## data:  alt by South
## t = -0.51334, df = 398, p-value = 0.608
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -390.6487  228.8787
## sample estimates:
## mean in group 0 mean in group 1
##         3643.080        3723.965
```

(1) (Intercept) → estimate of μ_0 → $\hat{\beta}_0 = 3643.08$

mean(depths_north \$alt) = 3643.08

South → estimate of $\Delta\mu$ → $\hat{\beta}_1 = 80.88$

estimate of μ_1 → $\hat{\beta}_0 + \hat{\beta}_1 = 3723.965$

(2) Std. error for (Intercept) → complicated formula

$= \sqrt{Var(\hat{\beta}_0)} = \sqrt{s^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)} = 111.42$

Std. error for South → $SE_{\bar{Y}_1-\bar{Y}_0} = \sqrt{\frac{s_0^2}{n_0} + \frac{s_1^2}{n_1}} = 157.56$

s_0^2 → var(depths_north \$alt) s_1^2 → var(depths_south \$alt)

$n_0 = 200$ $n_1 = 200$

(3) $H_0: \mu_0 = \mu_1$ or $\mu_1 - \mu_0 = 0$

$H_a: \mu_1 - \mu_0 \neq 0$ $t_{stat} = \frac{(\bar{y}_1 - \bar{y}_0) - (\mu_1 - \mu_0)}{SE_{\bar{Y}_1-\bar{Y}_0}} = \frac{80.88}{157.56} = 0.5$

(4) and (5) p-value $= P(|t_{stat}| > t_{(n_0-1+n_1-1)} \mid H_0)$

$df = n-p$, where p is the # of parameters

$= pt(q= 0.513, df= 398, lower.tail=F) \times 2$

$= 0.608$

(6) residual std. error $= \sqrt{\frac{1}{n-p}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} = 1576$

$\hat{y}_i$ → predicted value → vector of length 400

y_i → observed data.

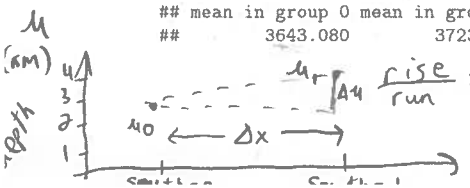
(7) 95% CI : $\mu_0 - \mu_1$

$(\bar{y}_0 - \bar{y}_1) + qt(p=c(0.025, 0.975), df=n-p) \times SE_{\bar{Y}_0-\bar{Y}_1}$

$= -80.88 + 1.96 \times 157.56 = [-390.64, 228.87]$

$\mu = \mu_0 + \Delta\mu \cdot South$

$= \begin{cases} \mu_0, & \text{if South}=0 \\ \mu_0 + \Delta\mu, & \text{if South}=1 \end{cases}$



$\frac{rise}{run} = \frac{\Delta\mu}{\Delta x} = \Delta\mu$

3 Ratio depth of the ocean in northern and southern hemisphere

```
# note: we are now using glm
fit <- glm(alt ~ South, data = depths, family = gaussian(link=log))
summary(fit)
```

```
##
## Call:
## glm(formula = alt ~ South, family = gaussian(link = log), data = depths)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3722.0   -608.5    401.5   1200.4   2867.9
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.20058    0.03058 268.144   <2e-16 ***
## South        0.02196    0.04278   0.513    0.608
##                 1          5        2        3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 2482673)
##                    6
##     Null deviance: 988758010  on 399  degrees of freedom
## Residual deviance: 988103771  on 398  degrees of freedom
## AIC: 7029.1               7             4
##
## Number of Fisher Scoring iterations: 5
```

The parameter of interest is the ratio

$\theta = \frac{\mu_1}{\mu_0} \Rightarrow \mu_1 = \mu_0 \times \theta$

log both sides $\Rightarrow \log(\mu_1) = \log(\mu_0) + \log(\theta$

$\log(\mu) = \log(\mu_0) + \log(\theta) \cdot South$

$\log(\mu) = \begin{cases} \log(\mu_0), & \text{if } South = 0 \\ \log(\mu_0) + \log(\theta), & \text{if } South = 1 \end{cases}$

1 (Intercept) Estimate $= \log(\bar{y}_0) \rightarrow$ an estimate of the parameter $\log(\mu_0)$

$\log(\bar{y}_0) = 8.20058 \Rightarrow \bar{y}_0 = \exp(8.20058)$
$= 3643.08$

South Estimate $= \log(\hat{\theta}) = 0{,}02196$

$\hat{\theta} = \exp(0{,}02196) = 1{,}022$

2 $H_0: \log(\theta) = 0$ $\quad H_a: \log(\theta) \neq 0$

$t_{stat} = \frac{\widehat{\log(\theta)} - \log(\theta)}{SE\ \widehat{\log(\theta)}} = \frac{0.02196}{0{,}04278} = 0.513$

3 and 4 p-value $= P(|t_{stat}| > t_{(398)} \mid H_0)$

$pt(q = 0{,}513,\ df = 398,\ lower.tail = F) \times 2 = 0{,}608$

95% CI for $\log(\theta)$: $0{,}02196 + qt(c(0{,}025, 0.975), 398) \times 0{,}04278$
$= [-0{,}062,\ 0.106]$

95% CI for $\theta = [\exp(-0.062), \exp(0.106)]$
$= [0.94,\ 1.11]$

what is the 95% CI for the % difference?

$[-0{,}06,\ 0{,}11] = [-6\%,\ 11\%]$

Point estimate = 2.2%

7 $R^2 = 1 - \frac{\text{Residual deviance}}{\text{Null deviance}} = \frac{988103771}{988758010}$ 5

$= 1 - 0{,}99$

$= 0{,}01$

4 Student drinking

```
fit <- lm(drinks ~ gender, data = drinks)
summary(fit)
```

```
##
## Call:
## lm(formula = drinks ~ gender, data = drinks)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.5185 -1.7947 -0.2947  1.4815  9.4815
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.2947     0.2837  15.138  < 2e-16 ***
## gender        2.2238     0.4182   5.318  3.2e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.765 on 174 degrees of freedom
## Multiple R-squared:  0.1398,Adjusted R-squared:  0.1348
## F-statistic: 28.28 on 1 and 174 DF,  p-value: 3.197e-07
```

```
fit <- glm(drinks ~ gender, data = drinks, family = gaussian(link=log))
summary(fit)
```

```
##
## Call:
## glm(formula = drinks ~ gender, family = gaussian(link = log),
##     data = drinks)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -5.5185  -1.7947  -0.2947   1.4815   9.4815
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.45739    0.06606  22.062  < 2e-16 ***
## gender       0.41726    0.08115   5.142 7.27e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 7.646385)
##
##     Null deviance: 1546.7  on 175  degrees of freedom
## Residual deviance: 1330.5  on 174  degrees of freedom
## AIC: 861.48
##
## Number of Fisher Scoring iterations: 5
```

Model

Expected # of cases (or events) = Rate x Person time

$$\mu = \boxed{\lambda} \times PT \quad (1)$$

→ we need a model for λ

we focus on the ratio $\theta = \frac{\lambda_1}{\lambda_0}$

$$\Rightarrow \lambda_1 = \lambda_0 \cdot \theta, \quad NBF = 1 \quad (2)$$

$$\Rightarrow \lambda_0 = \lambda_0 \cdot 1, \quad NBF = 0 \quad (3)$$

How can we combine Eqn. (2) and (3) into a single equation?

$$\boxed{\lambda = \lambda_0 \cdot \theta^{NBF}} \quad (4) \Rightarrow NBF = 1 \rightarrow \lambda_1 = \lambda_0 \cdot \theta^1$$

$$NBF = 0 \rightarrow \lambda_0 = \lambda_0 \, \theta^0$$

multiplicative model

Substitute Eqn. (4) into (1) we get

$$\mu = (\lambda_0 \cdot \theta^{NBF}) \times PT \quad (5)$$

log both sides of Eqn. (5)

$$\log(\mu) = \log(\lambda_0) + \log(\theta) \cdot NBF + 1 \cdot \log(PT)$$

Specifying link='log' means fit the $\log(\mu)$ model.
Specifying that $\log(PT)$ is an offset sets its accompanying regression coefficient to 1.

1. $\widehat{\log(\lambda_0)} = 1.220832 \Rightarrow \hat{\lambda}_0 = \exp(1.220832) = 3.39$

$\widehat{\log(\theta)} = 0.087505 \Rightarrow \hat{\theta} = \exp(0.87505) = 1.091$

$1.091 \pm 1.96 \cdot 0.003012 \rightarrow [L, U] \rightarrow [\exp(L), \exp(U)]$

$1.091 \pm 1.96 \cdot \exp(0.003012)$

$0.087 \pm 1.96 \cdot 0.003012 \rightarrow [L, U] \rightarrow [\exp(L), \exp(U)]$

McGill

2 Breastfeeding and respiratory infection II

Calculate the crude incidence rate ratio and 95% CI comparing infants who were not breastfed with those who were.

```
fit <- glm(cases ~ not_breastfed + offset(log(PT)), family = poisson(link = log))
summary(fit)
```

```
##
## Call:
## glm(formula = cases ~ not_breastfed + offset(log(PT)), family = poisson(link = log))
##
## Deviance Residuals:
## [1]  0  0
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.220832   0.001395  875.46   <2e-16 ***
## not_breastfed 0.087505   0.003012   29.05   <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: 8.3002e+02  on 1  degrees of freedom
## Residual deviance: 1.1533e-10  on 0  degrees of freedom
## AIC: 32.678
##
## Number of Fisher Scoring iterations: 2
```

Model

Expected # of cases (or events) = Rate x Person time

$\mu = \boxed{\lambda} \times PT$ (1)

we need a model for λ

We focus on the ratio $\theta = \frac{\lambda_1}{\lambda_0}$

$\Rightarrow \lambda_1 = \lambda_0 \cdot \theta$, $NBF = 1$ (2)

$\Rightarrow \lambda_0 = \lambda_0 \cdot 1$, $NBF = 0$ (3)

How can we combine Eqn. (1) and (2) into a single equation?

$\boxed{\lambda = \lambda_0 \cdot \theta^{NBF}}$ (4)

Substitute Eqn (4) into (1) we get

$\mu = \lambda_0 \cdot \theta^{NBF} \cdot PT$ → multiplicative model.

log both sides → $\log(\mu) = \log(\lambda_0) + \log(\theta) \cdot NBF + 1 \cdot \log(PT)$

specifying 'link = log' means fit the $\log(\mu)$ model. Specifying that $\log(PT)$ is an offset sets its accompanying regression coefficient to 1.

1 $\log(\hat{\lambda}_0) = 1.220832 \Rightarrow \hat{\lambda}_0 = \exp(1.220832) = 3.39$

$\log(\hat{\theta}) = 0.087505 \Rightarrow \hat{\theta} = \exp(0.087505) = 1.091448$

3 Malaria control with bednets

See the 2018 Lancet article *Efficacy of Olyset Duo, a bednet containing pyriproxyfen and permethrin, versus a permethrin-only net against clinical malaria in an area with highly pyrethroid-resistant vectors in rural Burkina Faso: a cluster-randomised controlled trial* (`Bednets.pdf` in A9 folder of my-Courses) by Tiono et. al. Reproduce the Rate ratio (95% CI) in Table 2. Calculate the rate difference and 95% CI comparing PPF-treated to Standard long-lasting insecticidal nets.

1 Breastfeeding and respiratory infection I

A total of 189,612 person-years of follow up were accumulated over the course of the study: 151,690 among infants who were being breastfed and 37,922 among infants not being breastfed. Over the course of follow up the investigators identified 514,230 incident cases of respiratory infection among breastfeeding infants and 140,312 among non-breastfeeding infants. Calculate the crude incidence rate difference and 95% CI comparing infants who were not breastfed with those who were.

```
fit <- glm(cases ~ -1 + PT + PT:not_breastfed, family = poisson(link = identity))
summary(fit)
```

```
##
## Call:
## glm(formula = cases ~ -1 + PT + PT:not_breastfed, family = poisson(link = identity))
##
## Deviance Residuals:
## [1]  0  0
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## PT               3.390006   0.004727  717.10   <2e-16 ***
## PT:not_breastfed 0.310010   0.010951   28.31   <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:        Inf  on 2  degrees of freedom
## Residual deviance: 1.1195e-10  on 0  degrees of freedom
## AIC: 32.678
##
## Number of Fisher Scoring iterations: 2
```

$\theta = \frac{\lambda_1}{\lambda_0}$

1 Malaria control with bednets

See the 2018 Lancet article *Efficacy of Olyset Duo, a bednet containing pyriproxyfen and permethrin, versus a permethrin-only net against clinical malaria in an area with highly pyrethroid-resistant vectors in rural Burkina Faso: a cluster-randomised controlled trial* (`Bednets.pdf` in `A9` folder of my-Courses) by Tiono et. al. Reproduce the Rate ratio (95% CI) in Table 2. Calculate the rate difference and 95% CI comparing PPF-treated to Standard long-lasting insecticidal nets. Check the goodness of fit.

```
##
## Call:
## glm(formula = cases ~ exposure + offset(log(years)), family = poisson(link = log),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -16.682   -4.732    1.497    3.984   12.024
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.68314    0.02432  28.092  < 2e-16 ***
## exposure    -0.26687    0.03286  -8.121 4.62e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1381.2  on 23  degrees of freedom
## Residual deviance: 1316.0  on 22  degrees of freedom
## AIC: 1476.7
##
## Number of Fisher Scoring iterations: 5
```

Model

Expected # cases of malaria = Rate × PT

$\mu = \boxed{\lambda} \times PT$

→ model for rate → $\lambda = \lambda_0 \cdot \theta^{\text{exposure}}$

multiplicative model

$\log(\lambda) = \log(\lambda_0) + \log(\theta) \cdot \text{Exposure}$

$\widehat{\log(\lambda_0)} = 0.68314 \Rightarrow \hat\lambda_0 = \exp(0.68314) = 1.98$ cases/child year

$\widehat{\log(\theta)} = -0.26687 \Rightarrow \hat\theta = \exp(-.26687) = 0.765$ "

Goodness of fit

we need to compare observed # of cases to expected # of cases.

$\hat\mu = \boxed{\hat\lambda} \cdot PT$

$= \hat\lambda_0 \cdot \hat\theta^{\text{Exposure}} \cdot PT$

$\hat\mu = \begin{cases} 1.98 \times PT & \text{if Exposed} = 0 \\ 1.98 \times 0.765 \times PT & \text{if Exposed} = 1 \end{cases}$

What does the data look like? → stop ask

month	exposure	cases	years	expected
June 2014	0	33	79 → stop ask	1.98 × 79 = 156.43
July 2014	0	454	123	1.98 × 123 = 243.54
August 2014	0	244	103	1.98 × 103 = 203.95
August 2014	1	43	23	1.98 × 0.765 × 23 = 34.87
Sept 2014	0	177	79	⋮ 1
Sept 2014	1	66	39	⋮

H₀: no lack of fit   Hₐ: lack of fit

observed ≃ expected

$X^2_{stat} = \sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(K-1)}$

$= \frac{(33-156.43)^2}{156.43} + \frac{(454-243.54)^2}{243.54} + \dots +$

Compare to a $\chi^2_{(K-1)}$

p-value = pchisq(... df = K-1, lower.tail = F)

2 Population mortality rates in Denmark

We can fit the following simple (multiplicative) rate ratio model to the patterns of mortality rates for 1980-1984 and 2000-2004. The reference cell is females 70-74, 1980-84. R = rate. M = multiplier.

Year	Age	Female (F)			Male (M)			
1980-1984	70-74	R_F			R_F		$\times M_M$	
	75-79	R_F	$\times M_{75}$		R_F	$\times M_{75}$	$\times M_M$	
	80-84	R_F	$\times M_{80}$		R_F	$\times M_{80}$	$\times M_M$	
	85-89	R_F	$\times M_{85}$		R_F	$\times M_{85}$	$\times M_M$	
2000-2004	70-74	R_F		$\times M_{20y}$	R_F		$\times M_M$	$\times M_{20y}$
	75-79	R_F	$\times M_{75}$	$\times M_{20y}$	R_F	$\times M_{75}$	$\times M_M$	$\times M_{20y}$
	80-84	R_F	$\times M_{80}$	$\times M_{20y}$	R_F	$\times M_{80}$	$\times M_M$	$\times M_{20y}$
	85-89	R_F	$\times M_{85}$	$\times M_{20y}$	R_F	$\times M_{85}$	$\times M_M$	$\times M_{20y}$

Year	Age	Female_deaths	Female_PT	Female_rate	Male_deaths	Male_PT	Male_rate
1980-1984	70-74	15989	586882.8	0.0272439	23810	456908.21	0.0521111
1980-1984	75-79	20838	454142.7	0.0458843	24707	300318.92	0.0822692
1980-1984	80-84	24073	297678.6	0.0808691	20319	167303.51	0.1214499
1980-1984	85-89	20216	147771.7	0.1368057	13524	74295.83	0.1820291
2000-2004	70-74	13912	521561.9	0.0266737	17360	436994.92	0.0397259
2000-2004	75-79	19731	471945.5	0.0418078	22477	341362.82	0.0658449
2000-2004	80-84	25541	369989.9	0.0690316	22992	217929.72	0.1055019
2000-2004	85-89	27135	226798.1	0.1196439	17444	104009.58	0.1677153
2005-2009	70-74	12179	540568.6	0.0225300	15782	472012.84	0.0334355
2005-2009	75-79	17273	444474.2	0.0388616	19547	344351.34	0.0567647
2005-2009	80-84	23513	363534.1	0.0646789	21781	230530.24	0.0944822
2005-2009	85-89	26842	237877.3	0.1128397	17811	114485.04	0.1555749

$$\text{Rate} = \underline{0.0272} \times \underset{\text{if } 75\text{-}79}{\underline{1.62}} \times \underset{\text{if } 80\text{-}84}{\underline{2.62}} \times \underset{\text{if } 85\text{-}89}{\underline{4.36}} \times \underset{\text{if male}}{\underline{1.52}} \times \underset{\text{if } 2000\text{-}04}{\underline{0.88}}$$

$$\log[\text{Rate}] = \underline{-3.603} + \underset{\text{if } 75\text{-}79}{\underline{0.482}} + \underset{\text{if } 80\text{-}84}{\underline{0.967}} + \underset{\text{if } 85\text{-}89}{\underline{1.472}} + \underset{\text{if male}}{\underline{0.419}} + \underset{\text{if } 2000\text{-}04}{\underline{-0.128}}$$

$$\log[\text{Rate}] = \underline{\beta_0} + \underline{\beta_{75}} \times I_{75-79} + \underline{\beta_{80}} \times I_{80-84} + \underline{\beta_{85}} \times I_{85-89} + \underline{\beta_m} \times I_{male} + \underline{\beta_{20year}} \times I_{2000-04}$$

where each 'I' is a (0/1) indicator of the category in question. By using both the 0 and 1 values of each I, this 6-parameter equation produces a fitted value for each of the $4 \times 2 \times 2 = 16$ cells.

For a closed-book exam, You are given 10 minutes to learn the mortality rates for males and females from 1980-2004. (16 numbers with 5 decimals, ~~closed-book exam to reproduce these with good accuracy~~. What is the smallest # of constants you need to remember in order to reproduce these with good accuracy

Age multipliers

years	Age	Female (F)	Male (M)
1980-1984	70-74	1	1
	75-79	0.0458/0.027 = 1.69	0.0822/0.052 = 1.58
	80-84	0.0808/0.027 = 2.97	0.121/0.052 = 2.33
	85-89	0.1368/0.027 = 5.02	0.182/0.052 = 3.49
2000-2004	70-74	1	1
	75-79	0.0418/0.0266 = 1.57	0.065/0.039 = 1.66
	80-84	0.069/0.0266 = 2.58	0.105/0.039 = 2.66
	85-89	0.1196/0.0266 = 4.49	0.167/0.039 = 4.22

A best estimate for M_{75} might be median(1.69, 1.58, 1.57, 1.66) = 1.62

M_{80} = median(2.97, 2.33, 2.58, 2.66) = 2.62

M_{85} = median(5.02, 3.49, 4.49, 4.22) = 4.36

Male Multiplier

years	Age	Female	Male
1980-1984	70-74	1	0.052/0.027 = 1.91
	75-79	1	1.79
	80-84	1	1.50
	85-89	1	1.33
2000-2004	70-74	1	1.49
		1	1.58
	75-79	1	1.53
	80-84	1	1.40
	85-89	1	

M_M = median(1.91, 1.79, ..., 1.40) = 1.52 (what does this mean?)

20-year multiplier

years	Age	Female	Male
1980-1984	70-74	1	1
	75-79	1	1
	80-84	1	1
	85-89	1	1
2000-2004	70-74	0.026/1.027 = 0.98	0.76
	75-79	0.91	0.80
	80-84	0.85	0.87
	85-89	0.88	0.92

M_{20y} = median(0.98, 0.76, ..., 0.921) = 0.88 (what does this mean?)