

Workshop 10.5a: Logistic regression

Murray Logan

September 5, 2016

Table of contents

1	Logistic regression	1
2	Worked Examples	5

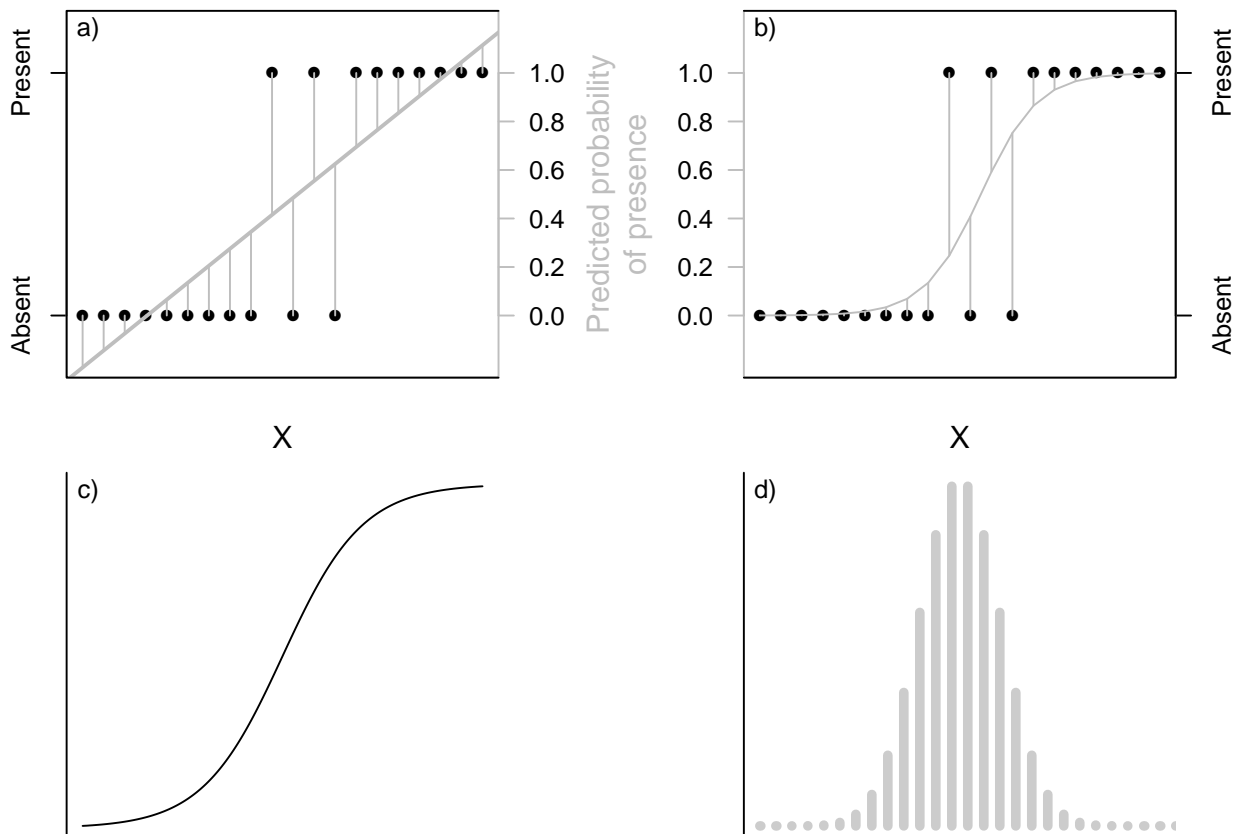
1. Logistic regression

1.1. Logistic regression

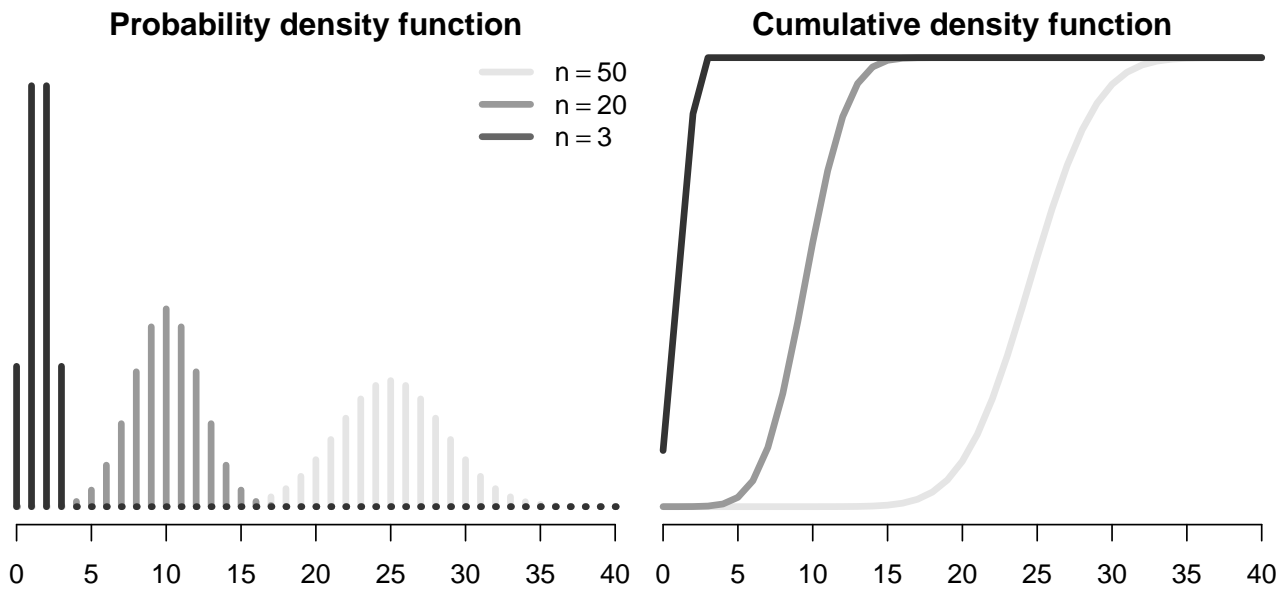
1.1.1. Binary data

Link: $\log\left(\frac{\pi}{1-\pi}\right)$ Transform scale of linear predictor $(-\infty, \infty)$ into that of the response $(0,1)$

1.2. Logistic regression



1.3. Logistic regression



$$E(Y) = \binom{n}{x} p^x (1-p)^{n-x}$$

Spread assumed to be equal to mean. ($\phi = 1$)

1.4. Dispersion

1.4.1. Over-dispersion

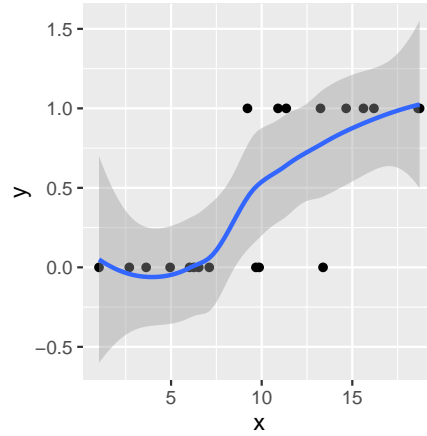
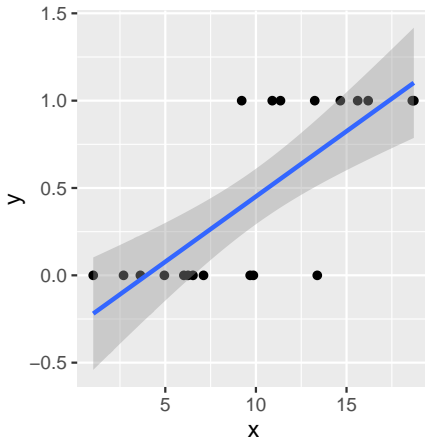
Sample more varied than expected from its mean

- variability due to other unmeasured influences
 - quasi- model
- due to more zeros than expected
 - zero-inflated model

1.5. Logistic regression

Example data

y	x
1	0.1024733
2	2.696719
3	3.626263
4	4.948643
5	6.024718
6	6.254113



1.6. Logistic regression

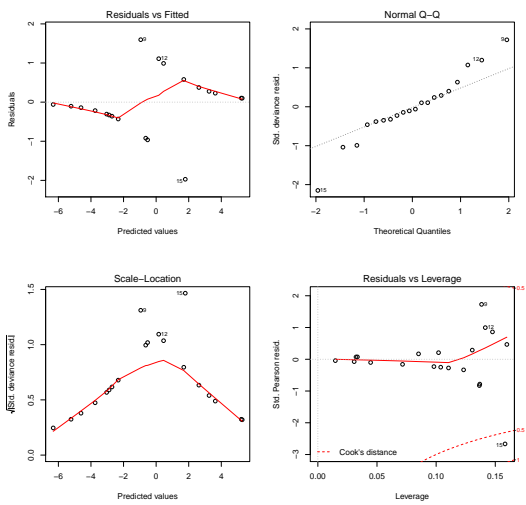
- Fit model

```
> dat.glmL <- glm(y ~ x, data = dat, family = "binomial")
```

1.7. Logistic regression

- Explore residuals

```
> par(mfrow=c(2,2))
> plot(dat.glmL)
```



1.8. Logistic regression

- Explore goodness of fit
- Pearson's χ^2 residuals

```
> dat.resid <- sum(resid(dat.glmL, type = "pearson")^2)
> 1 - pchisq(dat.resid, dat.glmL$df.resid)
```

[1] 0.8571451

- Deviance (G^2)



```
> 1-pchisq(dat.glmL$deviance, dat.glmL$df.resid)
```

[1] 0.8647024

1.9. Logistic regression

- Explore model parameters

Slope parameter is on log odds-ratio scale

```
> summary(dat.glmL)
```

Call:
glm(formula = y ~ x, family = "binomial", data = dat)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.97157	-0.33665	-0.08191	0.30035	1.59628

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-6.9899	3.1599	-2.212	0.0270 *
x	0.6559	0.2936	2.234	0.0255 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 27.526 on 19 degrees of freedom
Residual deviance: 11.651 on 18 degrees of freedom
AIC: 15.651

Number of Fisher Scoring iterations: 6

1.10. Logistic regression

- Quasi R²

$$quasiR^2 = 1 - \left(\frac{deviance}{null\ deviance} \right)$$

```
> 1-(dat.glmL$deviance/dat.glmL$null)
```

[1] 0.5767057

1.11. Logistic regression

- LD50

$$LD50 = - \frac{intercept}{slope}$$

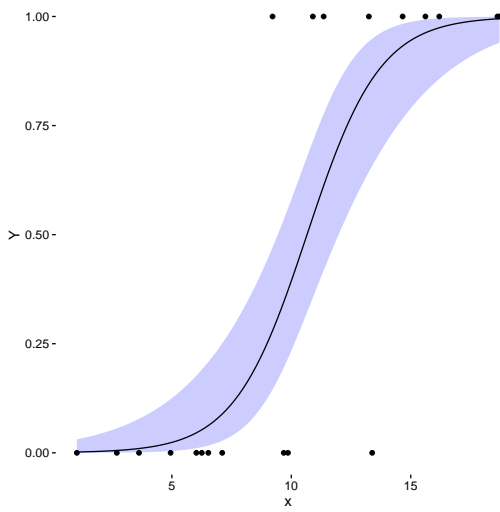
```
> -dat.glmL$coef[1]/dat.glmL$coef[2]
```

(Intercept)
10.65781



1.12. Logistic regression

- summary figure



2. Worked Examples

2.0. Worked Examples