

Part 3

Assumptions of binary logistic regression with one categorical predictor and things that can go wrong

Assumptions

1. Independence of errors

- Cases of data should not be related
- For instance, each cases should represent data from a different person

We can't really test for this – we should just know this is true based on the methodology



Things that can go wrong...

1. Quasi-complete separation

- You can perfectly predict the outcome for some (but not all) levels of the predictor variable

	Happiness - No	Happiness - Yes	Total
Hamster - No	23	0	23
Hamster - Yes	54	60	114

You can perfectly predict the outcome if Hamster = "No" (if Hamster = No, Happiness = No)



Creates issues for logistic regression

2. Complete separation

- You can perfectly predict the outcome for all levels of the predictor variable

	Happiness - No	Happiness - Yes	Total
Hamster - No	23	0	23
Hamster - Yes	0	54	54

You can perfectly
predict the outcome
for BOTH levels of
“Hamster”:

When Hamster = No,
Happiness = No

When Hamster = Yes,
Happiness = Yes

Creates issues
for logistic
regression

3. A failure to converge

- When you run a binary logistic regression model, R starts by estimating the parameters with a best guess
- It then attempts to estimate the parameters more accurately
- It stops when on each new attempt, the parameters are very similar (it “converges”)
- Sometimes it doesn’t converge:

```
Warning messages:  
1: glm.fit: algorithm did not converge
```
- **Ignore the output – not accurate!!**