

# Lecture 21

# Last time

Survey data from 77 college students on a dry campus (i.e., alcohol is prohibited) in the US. Survey asks students “How many alcoholic drinks did you consume last weekend?”

- `drinks`: number of drinks the student reports consuming
- `sex`: whether the student identifies as male
- `OffCampus`: whether the student lives off campus
- `FirstYear`: whether the student is a first-year student

Our goal: model the number of drinks students report consuming.

# Last time

```
1 library(pscl)
2
3 m1 <- hurdle(drinks ~ sex + FirstYear + OffCampus,
4             dist = "poisson", zero.dist = "binomial",
5             data = wdrinks)
6
7 m1$coefficients
```

\$count

(Intercept)	sexm	FirstYearTRUE	OffCampusTRUE
0.8132113	0.9706640	-0.2181068	0.3762608

\$zero

(Intercept)	sexm	FirstYearTRUE	OffCampusTRUE
0.1230510	0.3377969	-0.8554289	1.5803472

**Question:** I want to know whether there is a relationship between sex and the number of drinks a student reports consuming (after accounting for other variables). What hypotheses should I test?

# Hypothesis tests

```
1 m1 <- hurdle(drinks ~ sex + FirstYear + OffCampus,  
2             dist = "poisson", zero.dist = "binomial",  
3             data = wdrinks)  
4 m2 <- hurdle(drinks ~ FirstYear + OffCampus,  
5             dist = "poisson", zero.dist = "binomial",  
6             data = wdrinks)  
7  
8 2*(m1$loglik - m2$loglik)
```

```
[1] 29.8251
```

```
1 pchisq(2*(m1$loglik - m2$loglik), 2, lower.tail=F)
```

```
[1] 3.338586e-07
```

# Hypothesis tests

```
1 m1 <- hurdle(drinks ~ sex + FirstYear + OffCampus,  
2             dist = "poisson", zero.dist = "binomial",  
3             data = wdrinks)  
4  
5 m1$coefficients
```

\$count

(Intercept)	sexm	FirstYearTRUE	OffCampusTRUE
0.8132113	0.9706640	-0.2181068	0.3762608

\$zero

(Intercept)	sexm	FirstYearTRUE	OffCampusTRUE
0.1230510	0.3377969	-0.8554289	1.5803472

**Question:** I want to know whether there is a relationship between sex and whether a student reports consuming *any* drinks. What hypotheses should I test?

# Hypothesis tests

```
1 m1 <- hurdle(drinks ~ sex + FirstYear + OffCampus,  
2             dist = "poisson", zero.dist = "binomial",  
3             data = wdrinks)  
4  
5 m2 <- hurdle(drinks ~ sex + FirstYear + OffCampus | FirstYear + OffCa  
6             dist = "poisson", zero.dist = "binomial",  
7             data = wdrinks)  
8 m2$coefficients
```

\$count

(Intercept)	sexm	FirstYearTRUE	OffCampusTRUE
0.8132113	0.9706640	-0.2181068	0.3762608

\$zero

(Intercept)	FirstYearTRUE	OffCampusTRUE
0.2318016	-0.9249488	1.5599579

# Hypothesis tests

```
1 m1 <- hurdle(drinks ~ sex + FirstYear + OffCampus,  
2             dist = "poisson", zero.dist = "binomial",  
3             data = wdrinks)  
4  
5 m2 <- hurdle(drinks ~ sex + FirstYear + OffCampus | FirstYear + OffCa  
6             dist = "poisson", zero.dist = "binomial",  
7             data = wdrinks)  
8  
9 pchisq(2*(m1$loglik - m2$loglik), df=1, lower.tail=F)
```

```
[1] 0.5357824
```

# Hypothesis tests

```
1 m1 <- hurdle(drinks ~ sex + FirstYear + OffCampus,  
2             dist = "poisson", zero.dist = "binomial",  
3             data = wdrinks)  
4  
5 m1$coefficients
```

\$count

(Intercept)	sexm	FirstYearTRUE	OffCampusTRUE
0.8132113	0.9706640	-0.2181068	0.3762608

\$zero

(Intercept)	sexm	FirstYearTRUE	OffCampusTRUE
0.1230510	0.3377969	-0.8554289	1.5803472

**Question:** *Among students who report at least one drink, I want to know whether male students tend to drink *more*. What hypotheses should I test?*



# Hypothesis tests

```
1 m1 <- hurdle(drinks ~ sex + FirstYear + OffCampus,  
2             dist = "poisson", zero.dist = "binomial",  
3             data = wdrinks)  
4  
5 summary(m1)$coefficients
```

\$count

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.8132113	0.1586497	5.1258298	2.962302e-07
sexm	0.9706640	0.1854917	5.2329229	1.668504e-07
FirstYearTRUE	-0.2181068	0.3796621	-0.5744761	5.656457e-01
OffCampusTRUE	0.3762608	0.2111140	1.7822634	7.470629e-02

\$zero

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.1230510	0.3292870	0.3736893	0.7086355
sexm	0.3377969	0.5475895	0.6168798	0.5373140
FirstYearTRUE	-0.8554289	0.5836060	-1.4657645	0.1427124
OffCampusTRUE	1.5803472	1.1179974	1.4135518	0.1574936

```
1 pnorm(5.233, lower.tail=F)
```

```
[1] 8.339037e-08
```

# What's next

- Problem: excess zeros!
- Solution so far: hurdle model (Poisson, negative binomial, etc.)
- Alternative method: zero-inflated models

# Class activity

[https://sta712-f23.github.io/class\\_activities/ca\\_lecture\\_21.html](https://sta712-f23.github.io/class_activities/ca_lecture_21.html)

