Logistic Regression Inference

Julie Deeke
Statistics with Python Course Developer

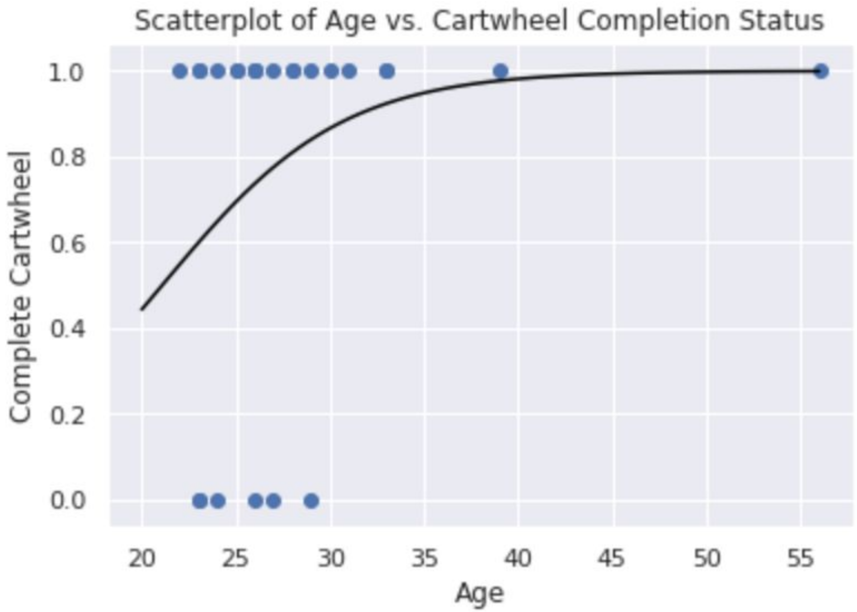
Cartwheel Data

Random sample of 25 adults attempted cartwheels

Primary Variable of interest: Cartwheel completion

Based on age, can we predict whether a cartwheel is completed?

Logistic Regression Line



Generalized Linear Model Regression Results

Dep. Variable:	CompleteGroup	**No. Observations:**	25
Model:	GLM	**Df Residuals:**	23
Model Family:	Binomial	**Df Model:**	1
Link Function:	logit	**Scale:**	1.0
Method:	IRLS	**Log-Likelihood:**	-12.534
Date:	Tue, 27 Nov 2018	**Deviance:**	25.068
Time:	16:11:20	**Pearson chi2:**	22.4
No. Iterations:	6		

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	-4.4213	4.429	-0.998	0.318	-13.101	4.259
Age	0.2096	0.171	1.225	0.221	-0.126	0.545

Confidence Interval

Best Estimate ± Margin of Error

Sample slope ± "a few" · estimated standard error

Confidence Interval

Best Estimate ± Margin of Error

Sample slope ± "a few" · estimated standard error

$b_1 \pm t^* \, se(b_1)$

Confidence Interval

Best Estimate ± Margin of Error

Sample slope ± "a few" · estimated standard error

$b_1 \pm t^* \, se(b_1)$

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.4213	4.429	-0.998	0.318	-13.101	4.259
Age	0.2096	0.171	1.225	0.221	-0.126	0.545

Confidence Interval

Best Estimate ± Margin of Error

Sample slope ± "a few" · estimated standard error

$b_1 \pm t^* \, se(b_1)$

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.4213	4.429	-0.998	0.318	-13.101	4.259
Age	0.2096	0.171	1.225	0.221	-0.126	0.545

Confidence Interval

Best Estimate ± Margin of Error

Sample slope ± "a few" · estimated standard error

df = n-2

$b_1 \pm t^* \, se(b_1)$

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.4213	4.429	-0.998	0.318	-13.101	4.259
Age	0.2096	0.171	1.225	0.221	-0.126	0.545

Confidence Interval

Best Estimate ± Margin of Error

Sample slope ± "a few" · estimated standard error

df = n-2

$b_1 \pm t^* \, se(b_1)$

$0.2096 \pm 2.069\ (0.171)$

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.4213	4.429	-0.998	0.318	-13.101	4.259
Age	0.2096	0.171	1.225	0.221	-0.126	0.545

Confidence Interval

Best Estimate ± Margin of Error

Sample slope ± "a few" · estimated standard error

df = n-2

$b_1 \pm t^* \, se(b_1)$

$0.2096 \pm 2.069\ (0.171)$

$(-0.1442,\ 0.5634)$

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.4213	4.429	-0.998	0.318	-13.101	4.259
Age	0.2096	0.171	1.225	0.221	-0.126	0.545

Confidence Interval

Best Estimate ± Margin of Error

Sample slope ± "a few" · estimated standard error

df = n-2

$b_1 \pm t^* \, se(b_1)$

$0.2096 \pm 2.069\,(0.171)$

$(-0.1442, 0.5634)$

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.4213	4.429	-0.998	0.318	-13.101	4.259
Age	0.2096	0.171	1.225	0.221	-0.126	0.545

0 is included

IVQ

Does it seem reasonable that there is a significant slope?

No

We'll continue working through the hypothesis testing framework before coming back to the answer of the IVQ.

Hypothesis Testing

Hypothesis Testing

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$

Hypothesis Testing

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$

null hypothesis

Test statistic: $\dfrac{b_1 - 0}{\text{s.e.}(b_1)}$

Hypothesis Testing

H_0: $\beta_1 = 0$
H_a: $\beta_1 \neq 0$

Test statistic: $$\frac{b_1 - 0}{s.e.\ (b_1)}$$

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.4213	4.429	-0.998	0.318	-13.101	4.259
Age	0.2096	0.171	1.225	0.221	-0.126	0.545

Hypothesis Testing

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$

Test statistic: $$\frac{b_1 - 0}{\text{s.e. }(b_1)} = \frac{0.2096}{0.171}$$

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.4213	4.429	-0.998	0.318	-13.101	4.259
Age	0.2096	0.171	1.225	0.221	-0.126	0.545

Hypothesis Testing

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$

Test statistic: $\dfrac{b_1 - 0}{s.e.(b_1)} = \dfrac{0.2096}{0.171} = 1.225$

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.4213	4.429	-0.998	0.318	-13.101	4.259
Age	0.2096	0.171	1.225	0.221	-0.126	0.545

Hypothesis Testing

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$

Test statistic: $$\frac{b_1 - 0}{\text{s.e.}(b_1)} = \frac{0.2096}{0.171} = 1.225$$

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.4213	4.429	-0.998	0.318	-13.101	4.259
Age	0.2096	0.171	$\boxed{1.225}$	0.221	-0.126	0.545

Hypothesis Testing

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$

Test statistic: $\dfrac{b_1 - 0}{\text{s.e. }(b_1)} = \dfrac{0.2096}{0.171} = 1.225$

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.4213	4.429	-0.998	0.318	-13.101	4.259
Age	0.2096	0.171	1.225	0.221	-0.126	0.545

Hypothesis Testing

need significance level

$H_0: \beta_1 = 0$
$H_a: \beta_1 \neq 0$

With our p-value of about 0.221, we would fail to reject the null hypothesis and cannot conclude that we have a significantly linear relationship between age and the log odds of the probability of successfully completing a cartwheel.

Test statistic: $\frac{b_1 - 0}{\text{s.e.}(b_1)} = \frac{0.2096}{0.171} = 1.225$

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-4.4213	4.429	-0.998	0.318	-13.101	4.259
Age	0.2096	0.171	1.225	0.221	-0.126	0.545

Summary

Logistic regression for a predicted variable with two options

~correct/incorrect or success/failure

Summary

Logistic regression for a predicted variable with two options

~correct/incorrect or success/failure

Confidence Intervals and Hypothesis Tests follow similar format as linear regression

~changing interpretation to the logistic context

Summary

Logistic regression for a predicted variable with two options

~correct/incorrect or success/failure

Confidence Intervals and Hypothesis Tests follow similar format as Linear Regression

~changing interpretation to the logistic context

Coming up, we'll look at an example from NHANES using blood pressure and smoking