

# Simple Linear Regression in R

# What is Regression?

- Allows you to predict  $y$  based on values of  $x$
- Both IV and DV can be continuous
- The basic statistic behind modeling
- + simple = only one IV
- + linear = data forms a straight line

# Code for Regression

```
modelName <- lm(DV ~ IV, data)  
summary(modelName)
```

# Interpreting Regression

- Text

call:

```
lm(formula = Assault ~ UrbanPop, data = USArrests)
```

Residuals:

Min	1Q	Median	3Q	Max
-150.78	-61.85	-18.68	58.05	196.85

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	73.0766	53.8508	1.357	0.1811
UrbanPop	1.4904	0.8027	1.857	0.0695 .

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 81.33 on 48 degrees of freedom

Multiple R-squared: 0.06701, Adjusted R-squared: 0.0475

8

F-statistic: 3.448 on 1 and 48 DF,

p-value: 0.06948

1. Overall model is  
significant if  $< .05$

# Interpreting Regression

- Text

Call:

```
lm(formula = Assault ~ UrbanPop, data = USArrests)
```

Residuals:

Min	1Q	Median	3Q	Max
-150.78	-61.85	-18.68	58.05	196.85

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	73.0766	53.8508	1.357	0.1811
UrbanPop	1.4904	0.8027	1.857	0.0695 .

Amount of variability in the DV accounted for in the IV

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Convert to a %

Residual standard error: 81.33 on 48 degrees of freedom

Multiple R-squared: 0.06701, Adjusted R-squared: 0.0475

8

F-statistic: 3.448 on 1 and 48 DF, p-value: 0.06948

# Interpreting Regression

- Text

Call:

```
lm(formula = Assault ~ UrbanPop, data = USArrests)
```

Residuals:

Min	1Q	Median	3Q	Max
-150.78	-61.85	-18.68	58.05	196.85

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	73.0766	53.8508	1.357	0.1811
UrbanPop	1.4904	0.8027	1.857	0.0695

Variable is significant if  $< .05$

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 81.33 on 48 degrees of freedom

Multiple R-squared: 0.06701, Adjusted R-squared: 0.0475

8

F-statistic: 3.448 on 1 and 48 DF, p-value: 0.06948

# Interpreting Regression

- Text

Call:

```
lm(formula = Assault ~ UrbanPop, data = USArrests)
```

Residuals:

Min	1Q	Median	3Q	Max
-150.78	-61.85	-18.68	58.05	196.85

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	73.0766	53.8508	1.357	0.1811
UrbanPop	1.4904	0.8027	1.857	0.0695 .

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 81.33 on 48 degrees of freedom

Multiple R-squared: 0.06701, Adjusted R-squared: 0.0475

8

F-statistic: 3.448 on 1 and 48 DF, p-value: 0.06948

A one-unit increase  
in this variable  
influences the DV  
by the estimate  
amount

# Making a Scatter Plot with Best Fit Line

```
ggplot(data, aes(x=column, y=column))  
  + geom_point()  
  + geom_smooth(method=lm, se=FALSE)
```