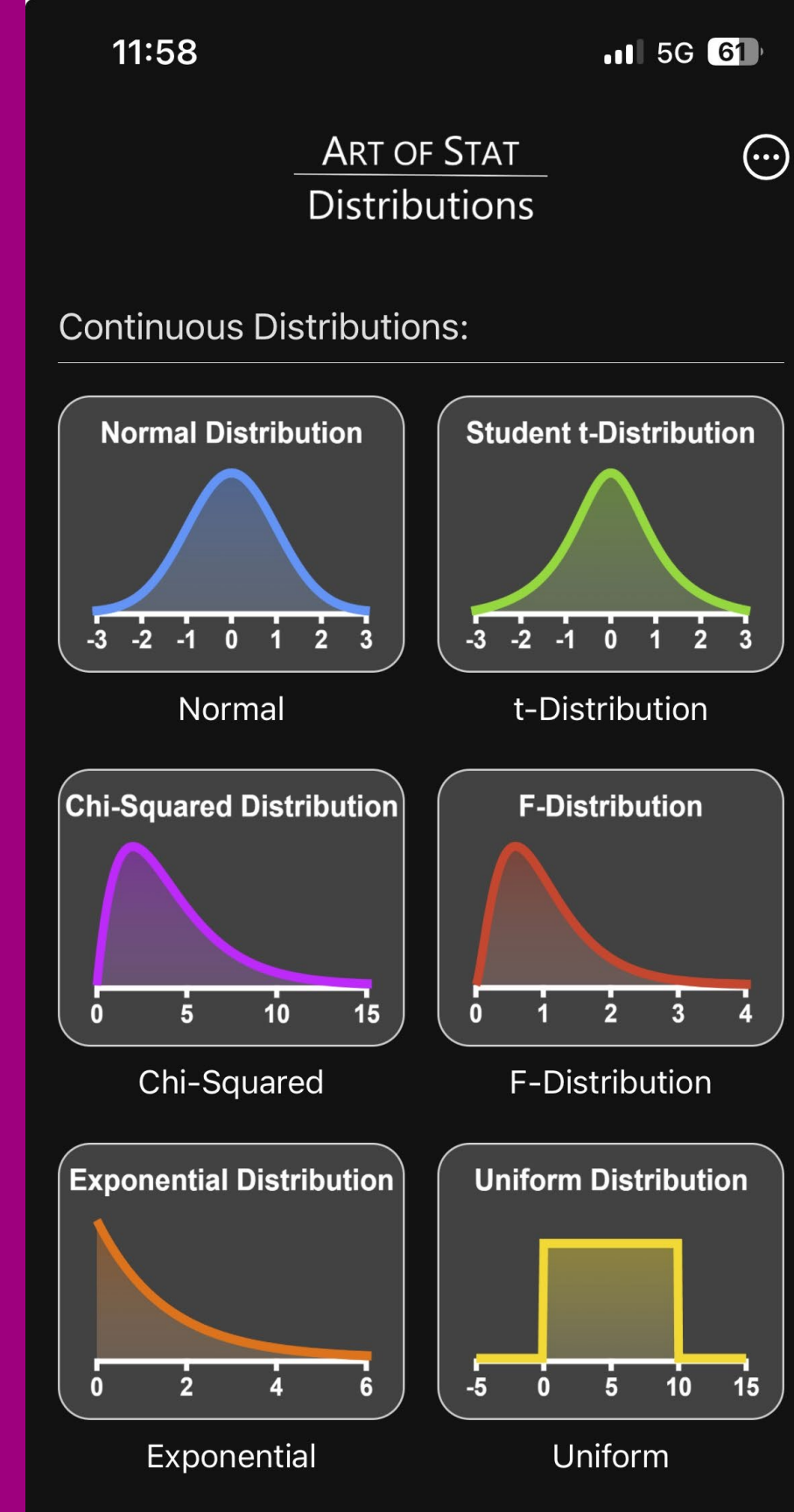


Art of Stat

Mobile Apps for Teaching & Learning Statistics

Bernhard Klingenberg
(bklingen@artofstat.com)





Agenda

1

The Six Art of Stat Mobile Apps

2

Technology: Screen Share, Videos, Zoom, Uploading Data, Airplane Mode

3

Case Studies:

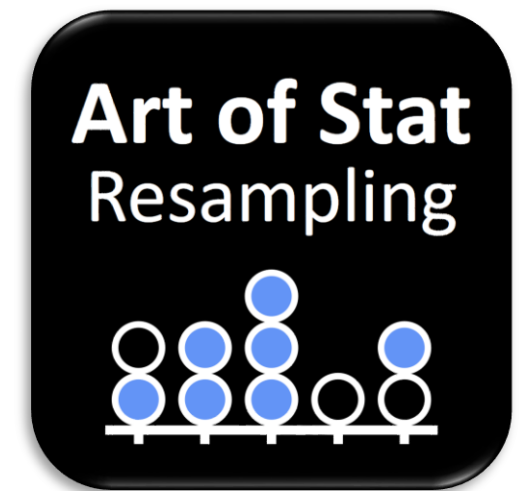
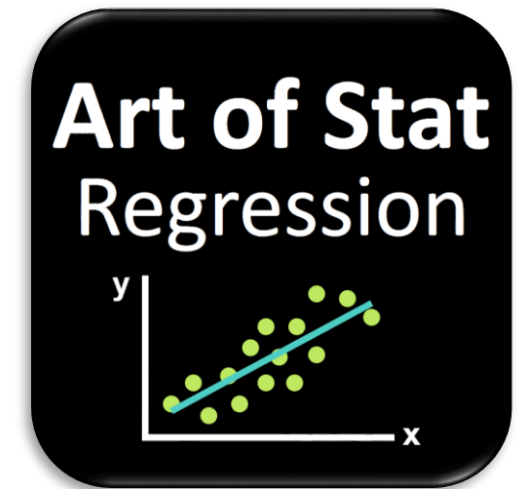
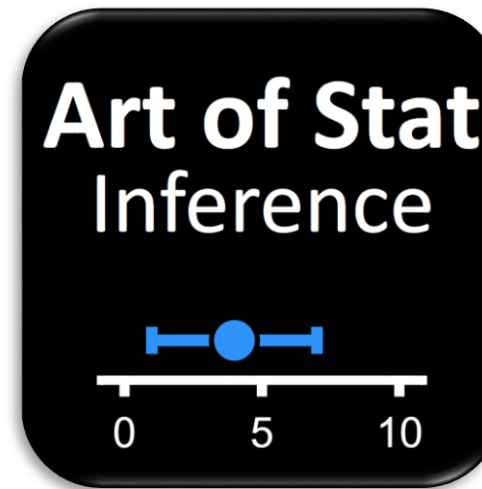
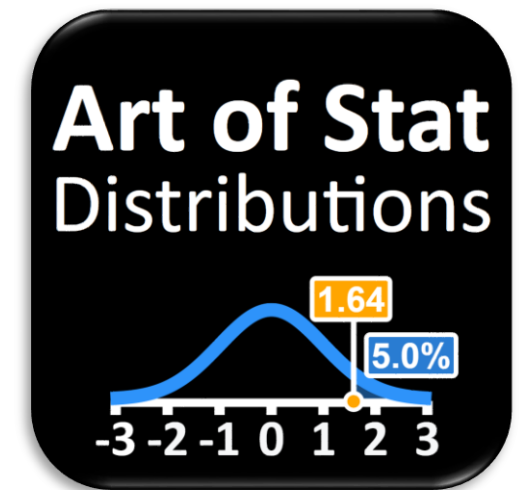
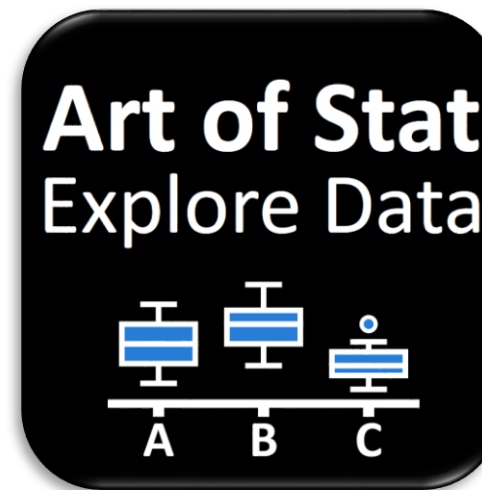
- Art of Stat: Explore Data & Distributions
- Art of Stat: Inference & Regression

4

Your Turn: Q&A or Creating Class Material

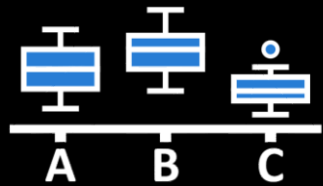
1) The Six Art of Stat Mobile Apps

- Explore Data
- Inference
- Concepts
- Distributions
- Regression
- Resampling



Art of Stat: Explore Data

Art of Stat
Explore Data



Descriptive Statistics &
Plots for Categorical
and Quantitative
Variables

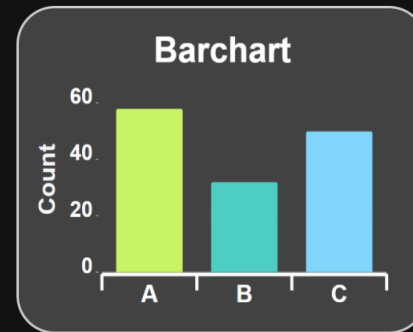
1:06

5G 58

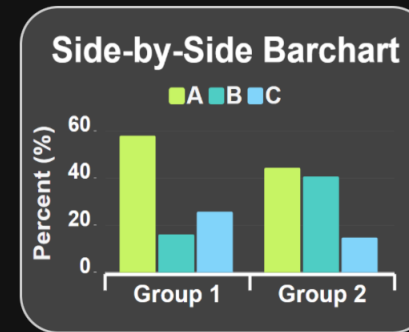
ART OF STAT Explore Data



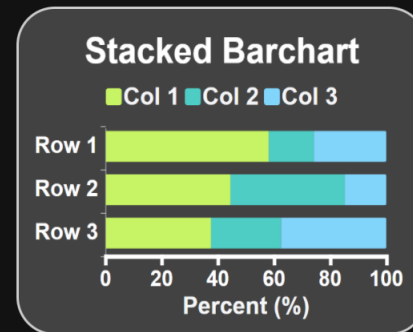
Categorical Variables:



One Categorical Variable

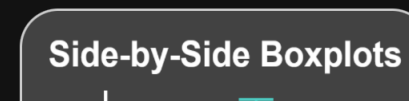


Compare Groups on
Categorical Variable



Relationship Between
Two Categorical
Variables

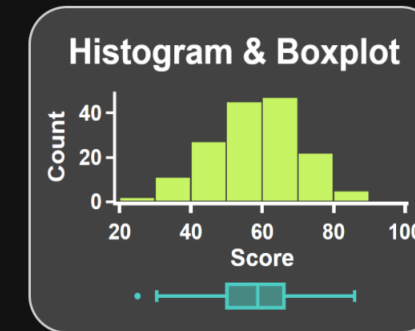
Quantitative Variables:



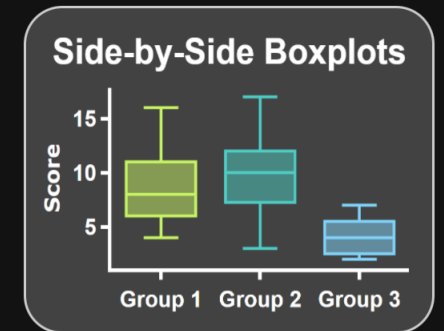
1:07

5G 58

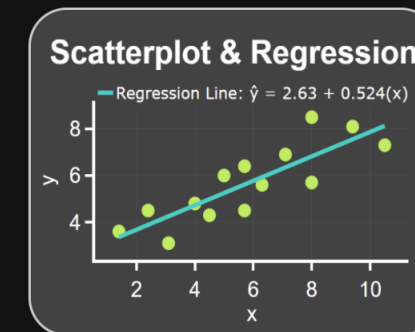
Quantitative Variables:



One Quantitative
Variable



Compare Groups on
Quantitative Variable



Relationship Between
Two Quantitative
Variables

Create Your Own Datasets

Data Editor

Obs.	Animal	Longevity
1	Wolf	17
2	Fox	13
3	Coyote	22
4	Cougar	

Art of Stat: Distribution



Explore & Visualize
Discrete and
Continuous Probability
Distributions

1:22 5G 57

ART OF STAT
Distributions

Continuous Distributions:

<p>Normal Distribution</p> <p>Normal</p>	<p>Student t-Distribution</p> <p>t-Distribution</p>
<p>Chi-Squared Distribution</p> <p>Chi-Squared</p>	<p>F-Distribution</p> <p>F-Distribution</p>
<p>Exponential Distribution</p> <p>Exponential</p>	<p>Uniform Distribution</p> <p>Uniform</p>

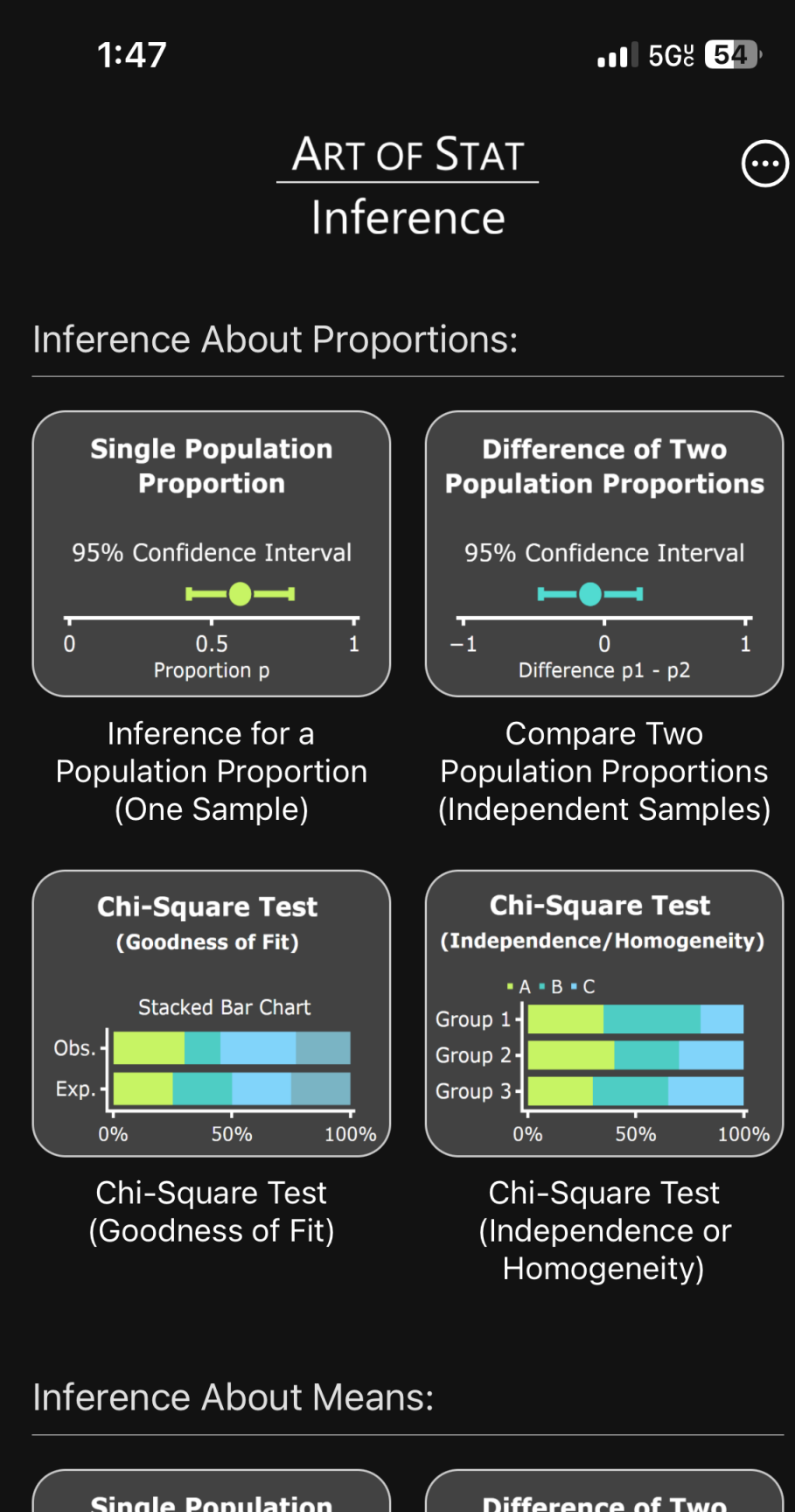
1:37 5G 55

<p>Gamma Distribution</p> <p>Gamma</p>	<p>Beta Distribution</p> <p>Beta</p>
Discrete Distributions:	
<p>Binomial Distribution</p> <p>Binomial</p>	<p>Geometric Distribution</p> <p>Geometric</p>
<p>Poisson Distribution</p> <p>Poisson</p>	<p>Discrete Distribution</p> <p>Discrete</p>

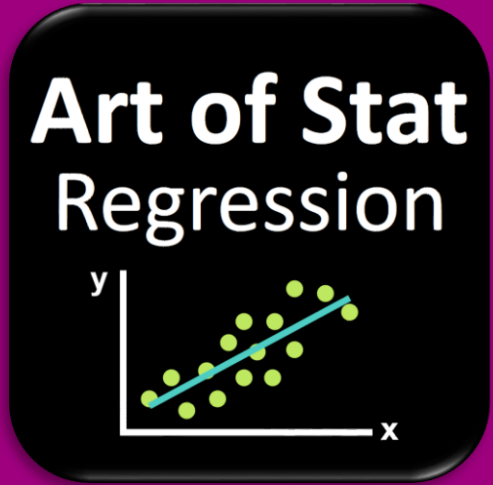
Art of Stat: Inference



Confidence Intervals &
Hypothesis Tests



Art of Stat: Regression



Simple Linear and
Logistic Regression &
Multiple Linear
Regression

2:46 5G 51

ART OF STAT Regression

Simple Regression (One Predictor)

Linear Regression

$$\hat{y} = 5.2 + 0.23(x)$$

A scatter plot showing a positive linear relationship between Gestation (x-axis, 40-100) and Longevity (y-axis, 10-30). A blue regression line is fitted to the data points.

Linear Regression

Exponential Regression

$$\hat{y} = 4.9 * 1.52^{(x)}$$

A scatter plot showing an exponential growth relationship between Days (x-axis, 0-10) and Population (y-axis, 0-300). A blue regression curve is fitted to the data points.

Exponential Regression

Logistic Regression

$$\hat{y} = 1 / (1 + \exp(7.1 - 0.23x))$$

A scatter plot showing a logistic relationship between Income (x-axis, 20k-60k) and P(Card) (y-axis, 0-1). A blue S-shaped regression curve is fitted to the data points.

Logistic Regression

Multiple Regression (Several Predictors)

Multiple Linear Regression

A scatter plot with three axes (x1, x2, y) showing data points and a regression surface.

2:47 5G 51

Multiple Regression (Several Predictors)

Multiple Linear Regression

A scatter plot with three axes (x1, x2, y) showing data points and a regression surface.

Multiple Linear Regression

Create Your Own Datasets

Data Editor

Obs.	Animal	Longevity
1	Wolf	17
2	Fox	13
3	Coyote	22
4	Cougar	

Data Editor

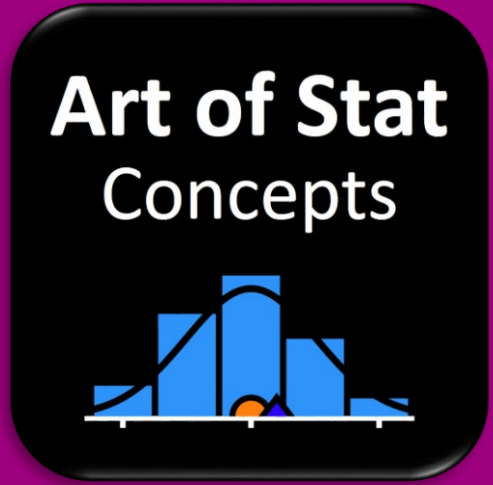
Check Out Other Apps:

Art of Stat Explore Data

Art of Stat Distributions

Art of Stat Inference

Art of Stat: Concepts



Central Limit Theorem,
Correlation,
Regression,
Coverage & Power

2:51 5G 50

ART OF STAT
Concepts

Central Limit Theorem

Distribution of the Sample Mean \bar{x}
Population (green), Sampling (blue)

Distribution of the Sample Proportion \hat{p}
Simulated (green bars), Normal (CLT) (blue curve)

CLT for Means CLT for Proportions

Correlation and Regression

Guess Correlation (Scatter plot, $r = ?$)
Explore Regression (Scatter plot with Regression Line)

Explore Correlation Explore Regression

Coverage, Errors and Power

Explore Coverage (Confidence intervals)

Errors and Power
 $-P(\text{Type I}) = 5\%$ $-P(\text{Type II}) = 21\%$

Explore Coverage of Confidence Intervals Explore Type I & II Errors and Power

2:52 5G 50

Correlation and Regression

Guess Correlation (Scatter plot, $r = ?$)
Explore Regression (Scatter plot with Regression Line)

Explore Correlation Explore Regression

Coverage, Errors and Power

Explore Coverage (Confidence intervals)

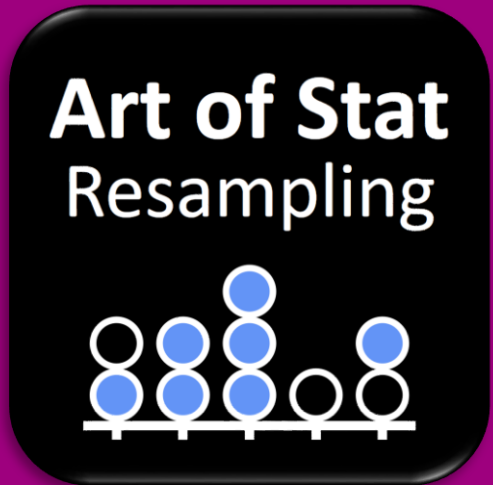
Errors and Power
 $-P(\text{Type I}) = 5\%$ $-P(\text{Type II}) = 21\%$

Explore Coverage of Confidence Intervals Explore Type I & II Errors and Power

Check Out Other Apps:

Art of Stat Distributions (1.64, 5.0%)
Art of Stat Explore Data
Art of Stat Inference

Art of Stat: Resampling



Bootstrap Confidence Intervals & Permutation Tests

3:00 5G% 49

ART OF STAT
Resampling

Bootstrap Confidence Intervals:

Population Mean, Median, Std. Dev.

Mean, Median, Std. Deviation

Population Proportion, Odds

Proportion & Odds

Correlation & Slope

Original: $r = 0.79$
Bootstrap: $r = 0.81$ $r = 0.81$

Correlation & Slope

Difference Between Two Means or Medians

$\bar{x}_1 - \bar{x}_2 = -2.7$

Differences in Means or Medians

Permutation Tests:

Permutation Test Mean & Median

$\bar{x}_{perm} = 19$

Permutation Test Two Means or Medians

$\bar{x}_1 - \bar{x}_2 = 1.4$

3:00 5G% 49

Permutation Tests:

Permutation Test Mean & Median

$\bar{x}_{perm} = 19$
 $\bar{x}_{obs} = 31$

Permutation Test for the Mean or Median

Permutation Test Two Means or Medians

$\bar{x}_1 - \bar{x}_2 = 1.4$

Permutation Test Comparing Two Groups

Chi-Square Test (Independence/Homogeneity)

Under Construction

Permutation Chi-Square Test

Create Or Edit Your Own Datasets

Data Editor

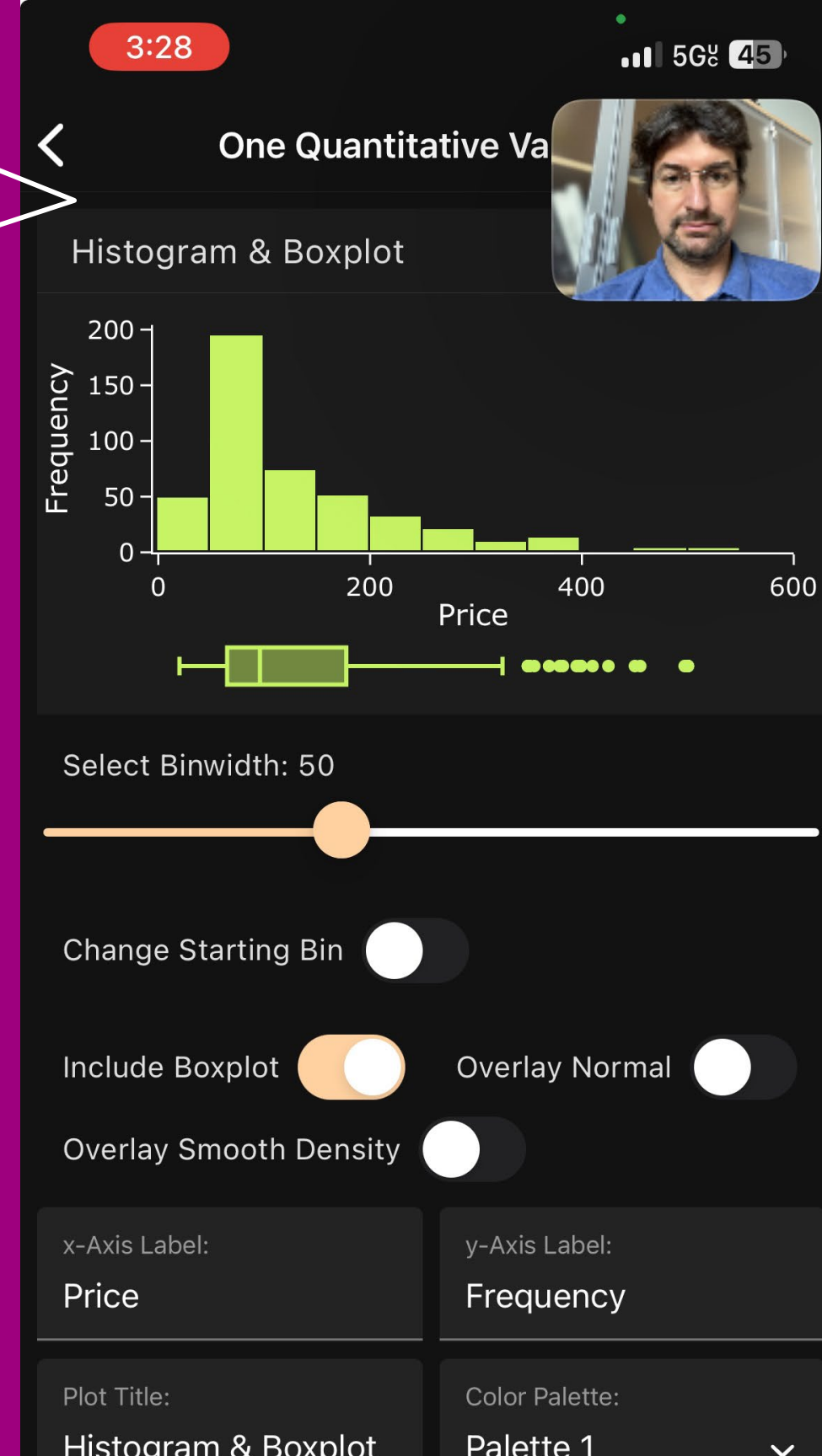
Obs.	Animal	Longevity
1	Wolf	17
2	Fox	13
3	Coyote	22
4	Cougar	

Data Editor

2) Technology

- Connect Cell Phone to Screen
- Screen Shots & Sharing
- Screen Recording (Videos)
- Join Zoom Call (Share Screen)
- Upload Data
- Price

Me in a Zoom call sharing my screen with the class



Technology: Connect to Screen

Connect Device to a
Screen/Projector via an
HDMI Adapter

- Connect to a Projector or Screen via HDMI

For older iPhone models



For Android Phones,
most tablets, iPads, etc.

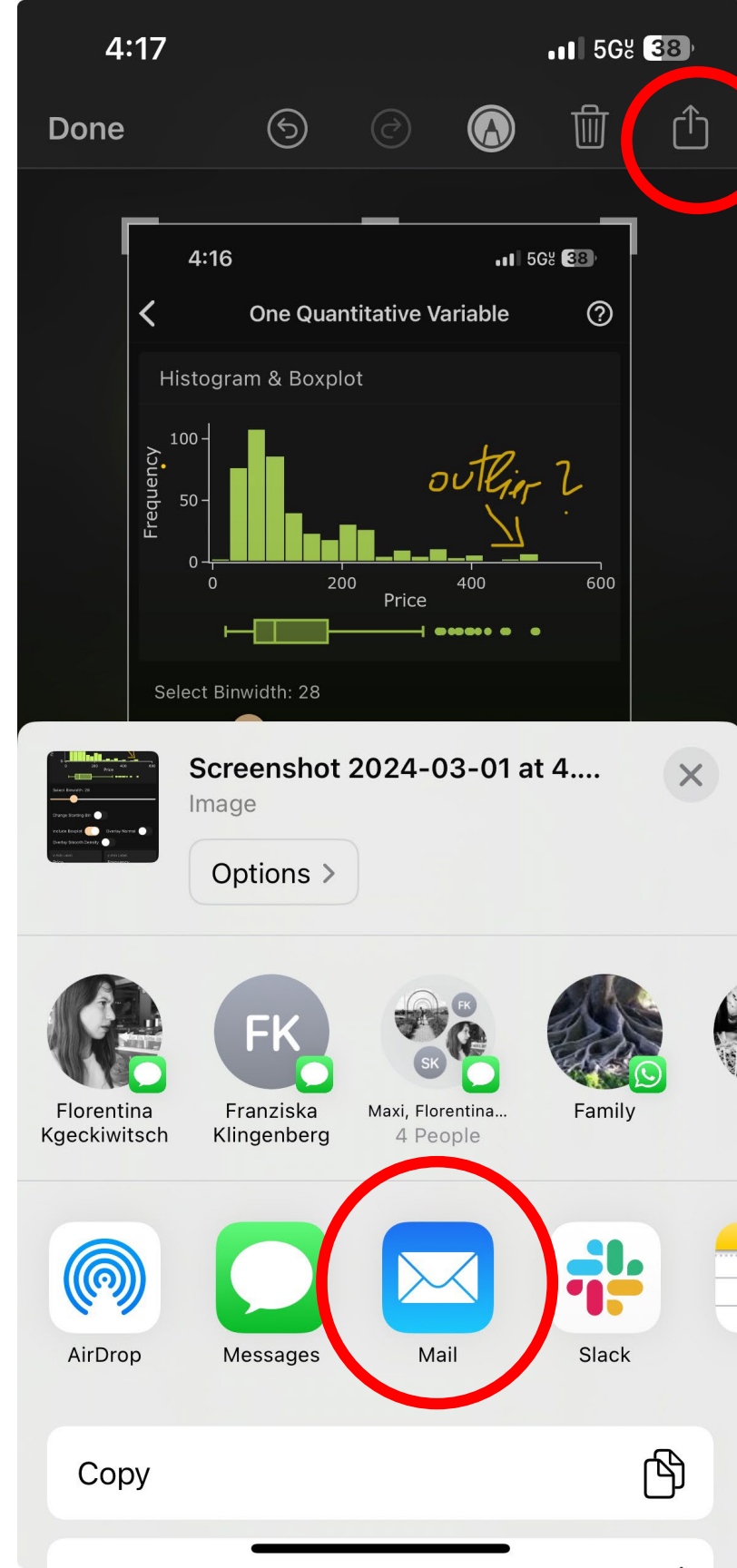


- Another Option: Screen Casting
or Mirroring to another screen



Technology: Screen Shots & Sharing

Connect Device to a
Screen/Projector via an
HDMI Adapter



- For both iOS and Android, easy to take screen shot
- Annotate screenshots
- Share via text message, email, Social Media, etc.

Technology: Screen Shots & Sharing

Upload screenshots
and annotate

Annotated screenshots from my Stat 101 class

Stat 101: CLT and Confidence Intervals

8 / 9

Share

You can do all computations with the Art of Stat: Inference app:

For one proportion, select this:

Enter the data:

Statistic	Value
Sample Size (n)	16
Number of Successes	5
Sample Proportion (\hat{p})	0.313
Standard Error (se) of the Sample Proportion	0.116
Confidence Level	95%
z-score ($\alpha = 5\%$)	1.960
Margin of Error (me)	± 0.227
Lower Bound for p	0.0854
Upper Bound for p	0.540

95% Confidence Interval (0.0854, 0.540)

Population Proportion p

Confidence Level: 95%

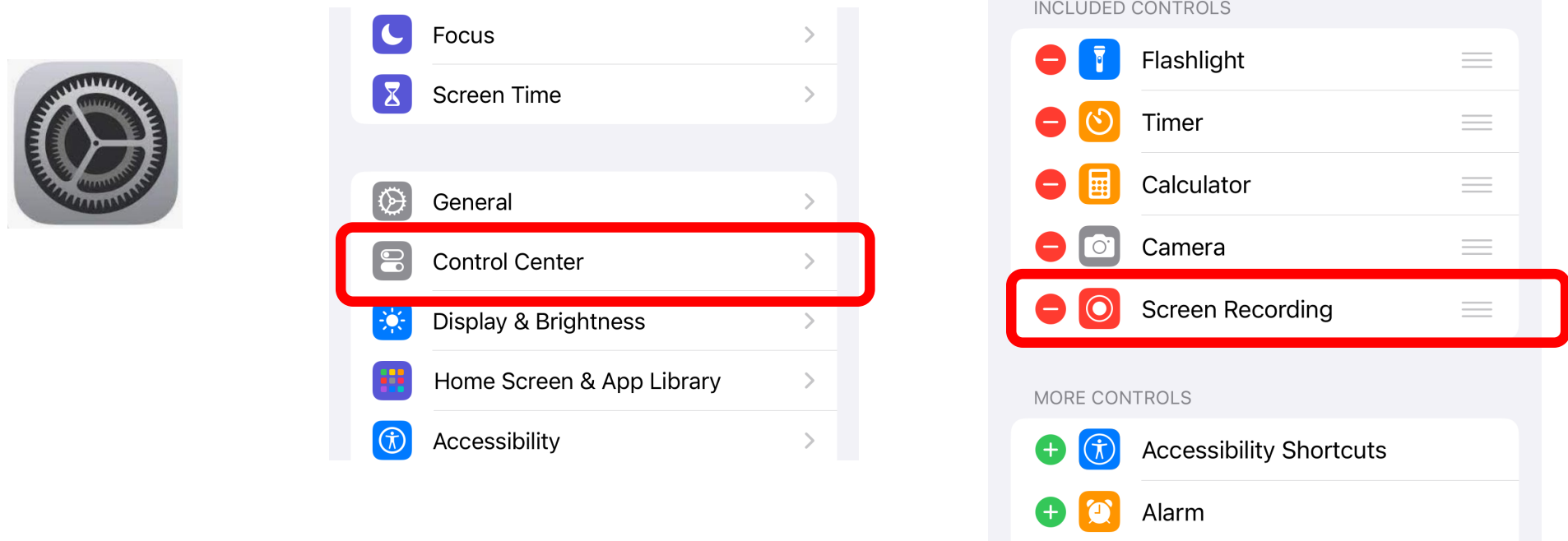
ENTER DATA CONFIDENCE INTERVAL HYPOTHESIS TEST

second tab

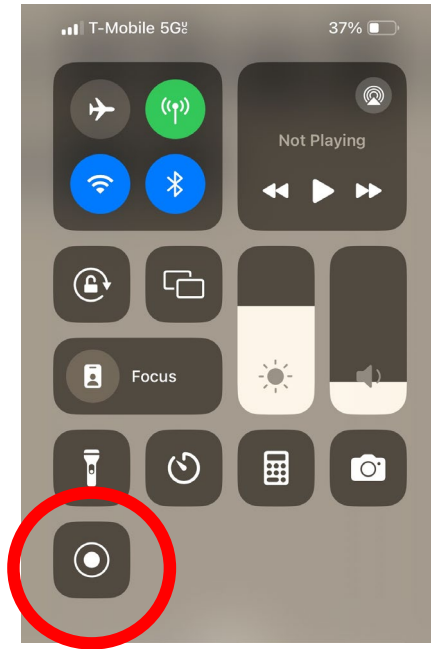
Technology: Screen Casting

Record your Screen to Produce a Short Video of a Concept or Analysis

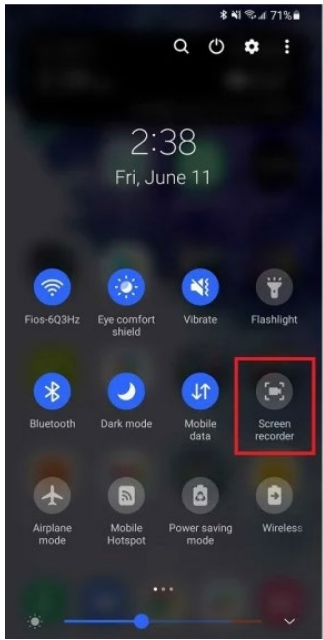
Settings > Control Center > Customize Controls



Swipe down, start recording



On Android, swipe down from the top once or twice, to reveal Screen Recorder

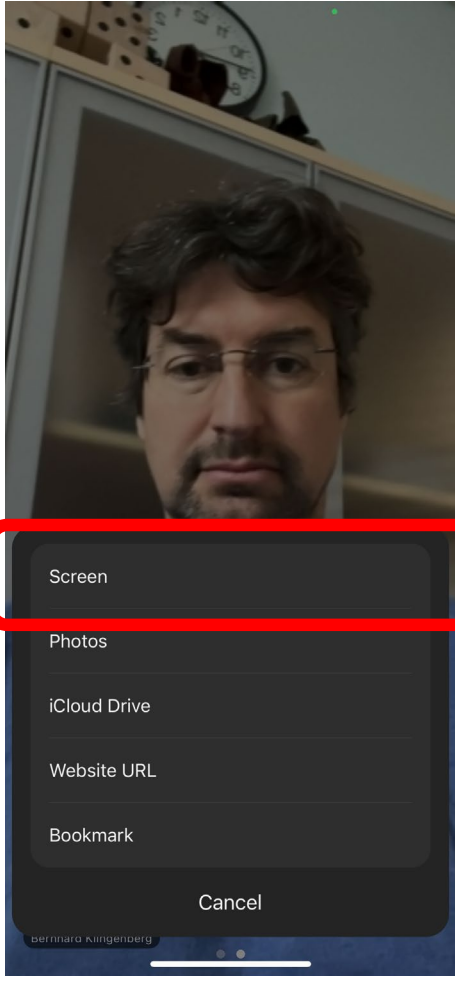


Technology: Zoom Calls

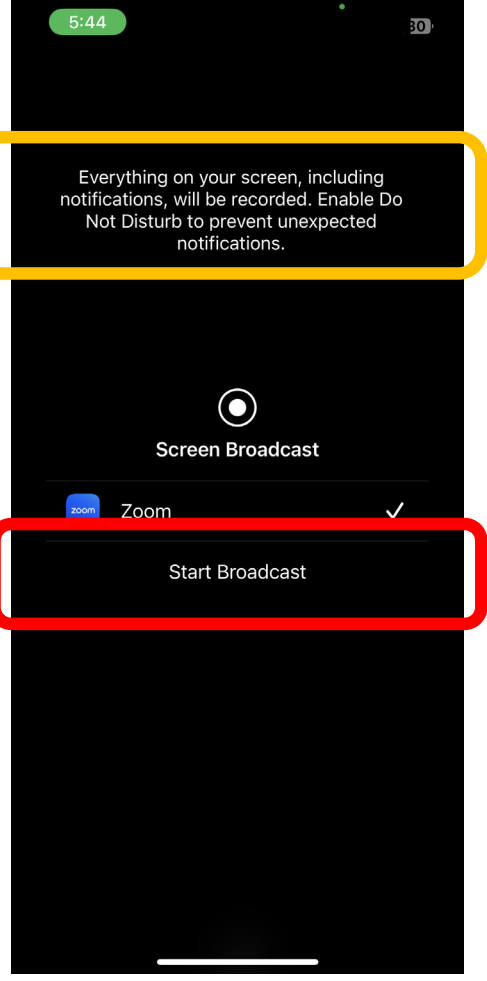
Start (or Join) a Zoom Call and Share your Screen to Teach Statistics



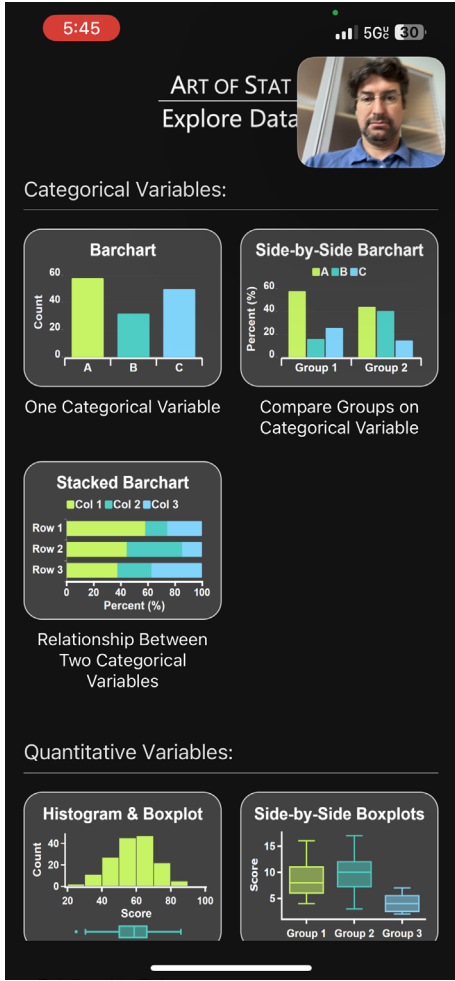
In the Zoom call, select "Share"



Select "Screen"



Enable *Do Not Disturb* on your phone and start broadcasting



Teach with the apps

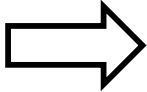
Technology: Upload Data

In every app, you can not only type in your own data, but also upload a CSV file and select columns from it

1) Create a dataset

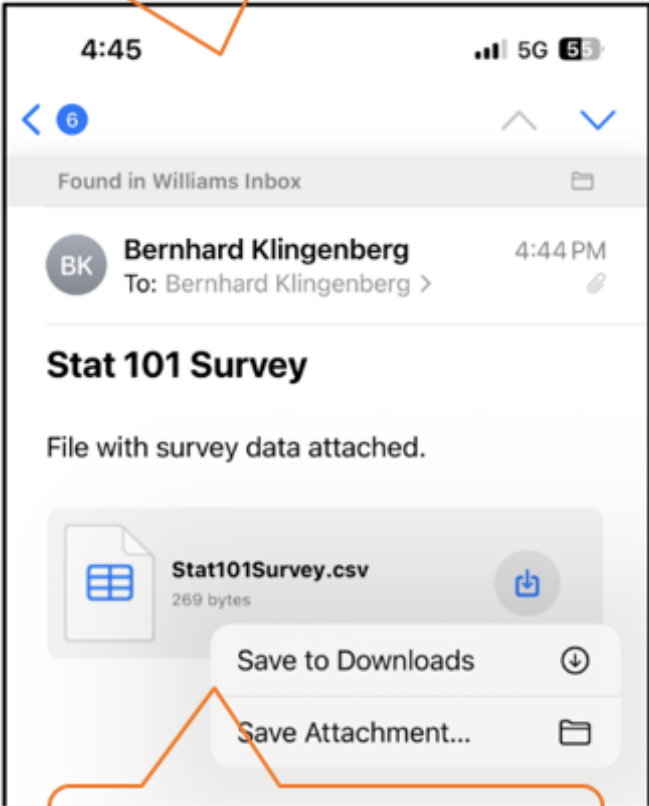
	A	B	C
1	Frosh	Preference	Cost
2	No	milk	0
3	No	dark	200
4	No	milk	0
5	Yes	white	500
6	No	dark	400
7	Yes	dark	80
8	No	milk	130

File: Stat101Survey.csv



2a) Email it to yourself and then long-press to save it on your device

Email Client on phone

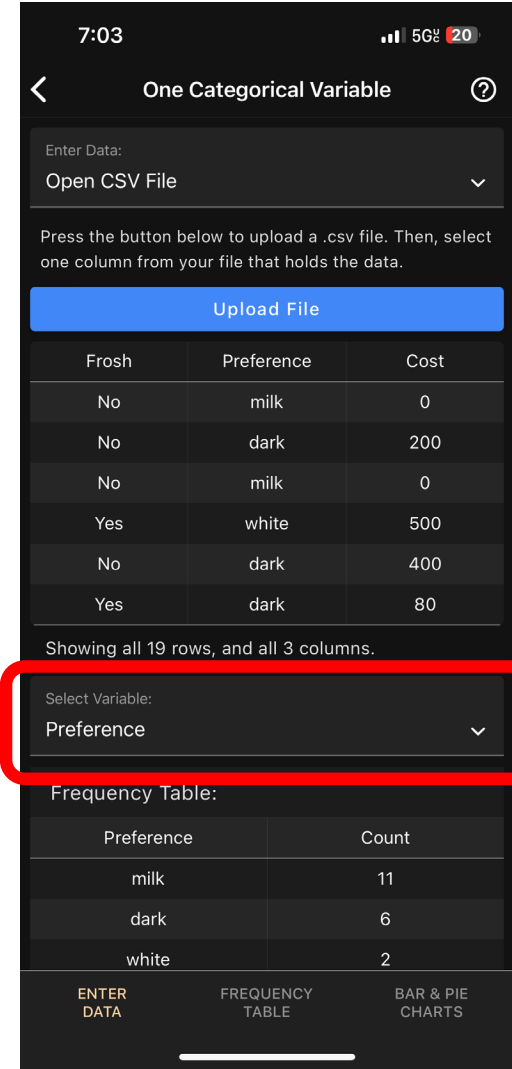
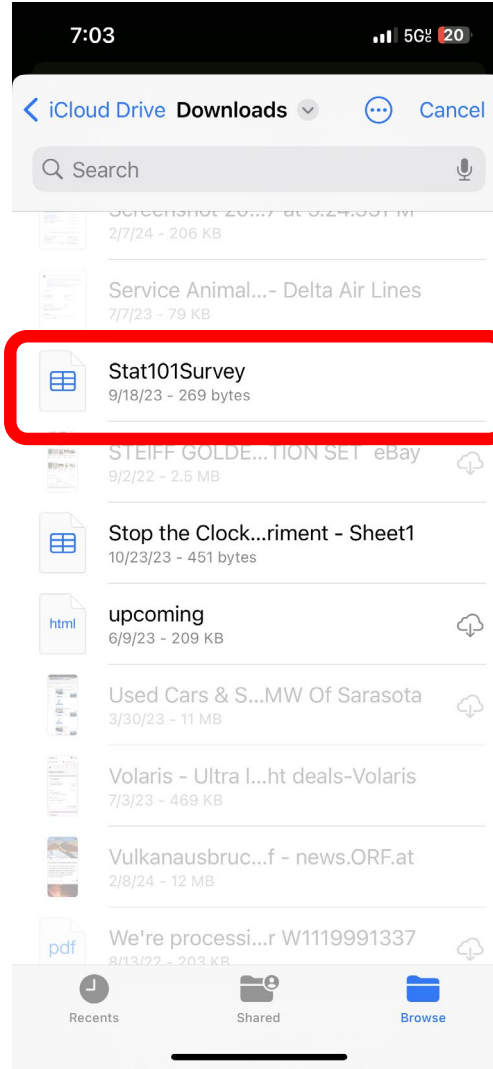
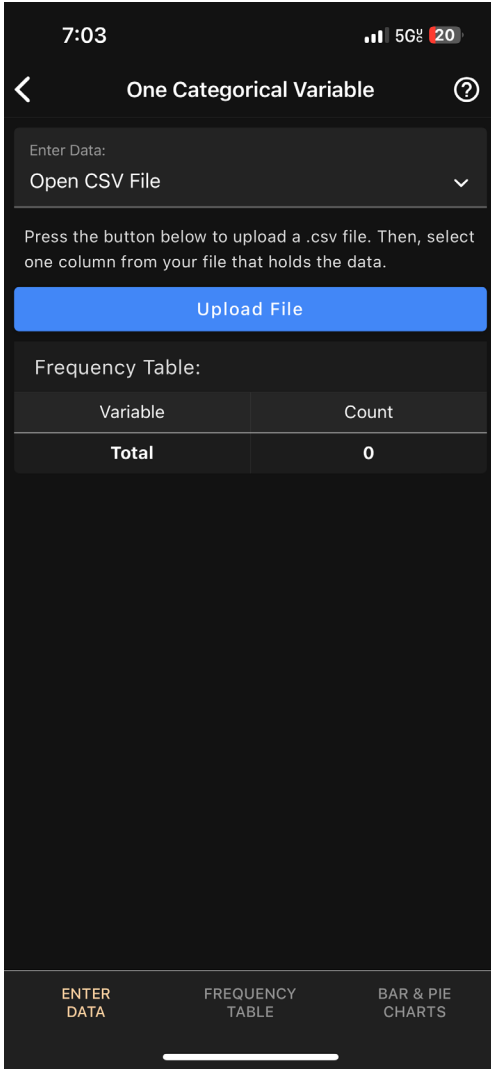
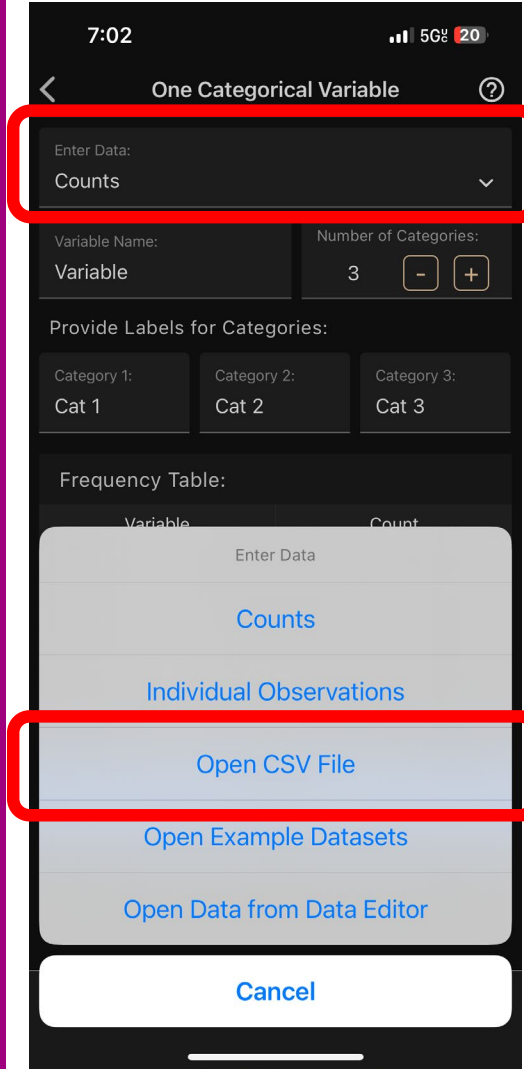


Save on your phone

2b) Alternatively, put it on iCloud or Google Drive so you can access it from your phone/tablet

Technology: Upload Data

In every app, you can upload a CSV file and select columns from it. Alternatively, you can create or edit datasets with the Data Editor.



In any app, under *Enter Data*, select *Upload CSV File*

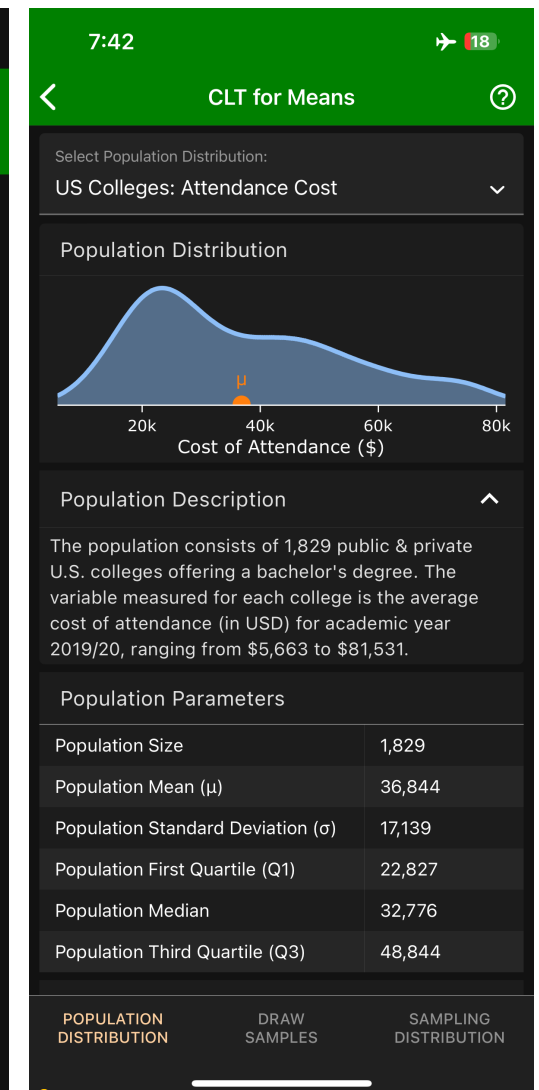
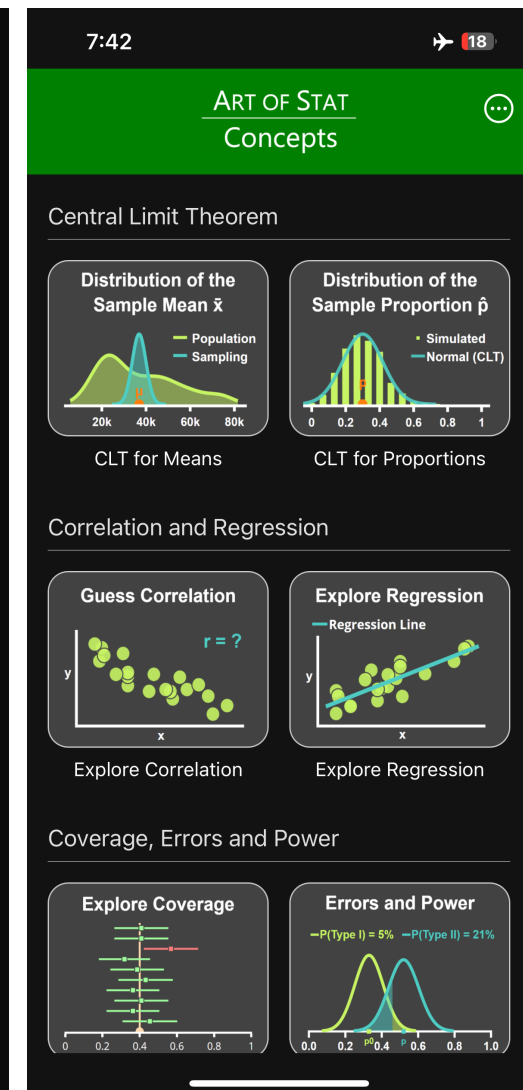
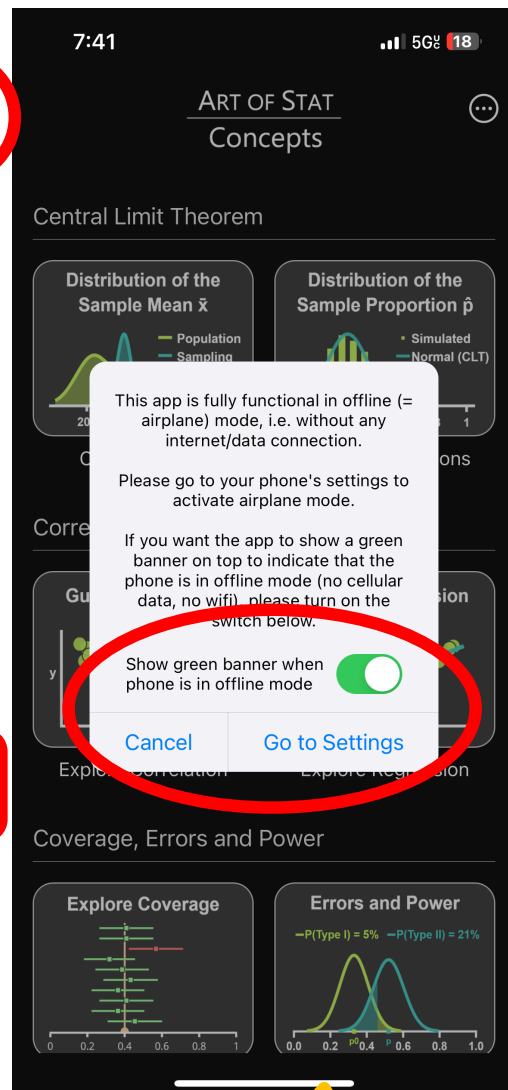
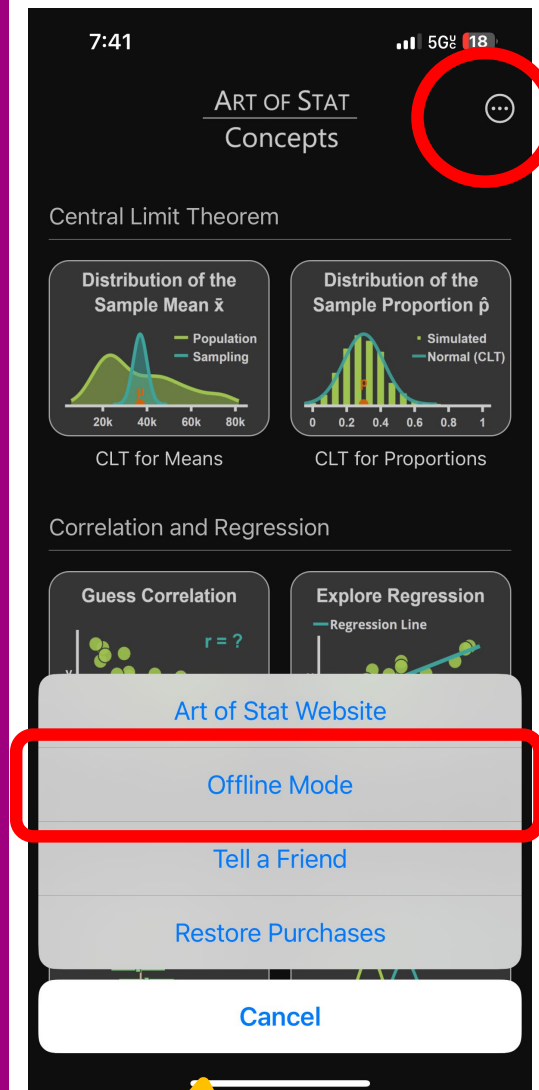
Press *Upload File* button

Navigate to folder and select file. File loads and is displayed

Select variable from file

Technology: Airplane Mode

All apps work in
Airplane Mode.
No wifi or cellular
connection necessary.



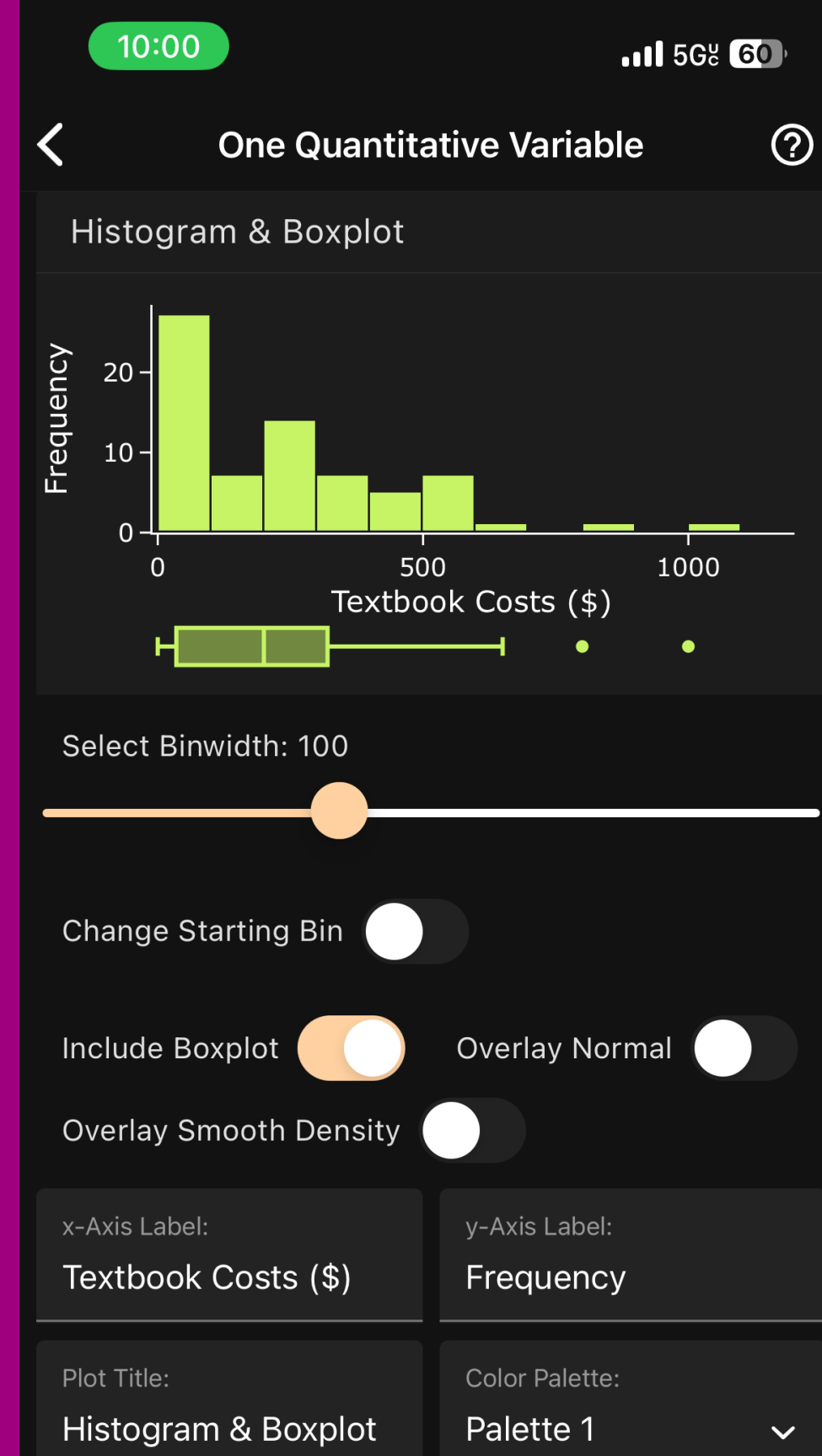
In any app, tab on the three dots in the upper right corner and select *Offline Mode*

Toggle switch for showing green banner, then go to Settings and enable airplane mode.

Every screen now shows green banner on top to indicate no internet/data connection

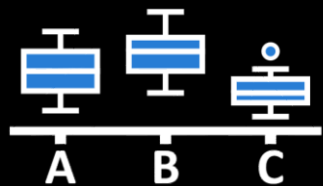
3) Case Studies

- Art of Stat: *Explore Data* app
 - One Quantitative Variable
 - Two Categorical Variables
- Art of Stat: *Distributions* app
 - Normal Distribution
 - Binomial Distribution
- Art of Stat: *Concepts* app
 - Central Limit Theorem

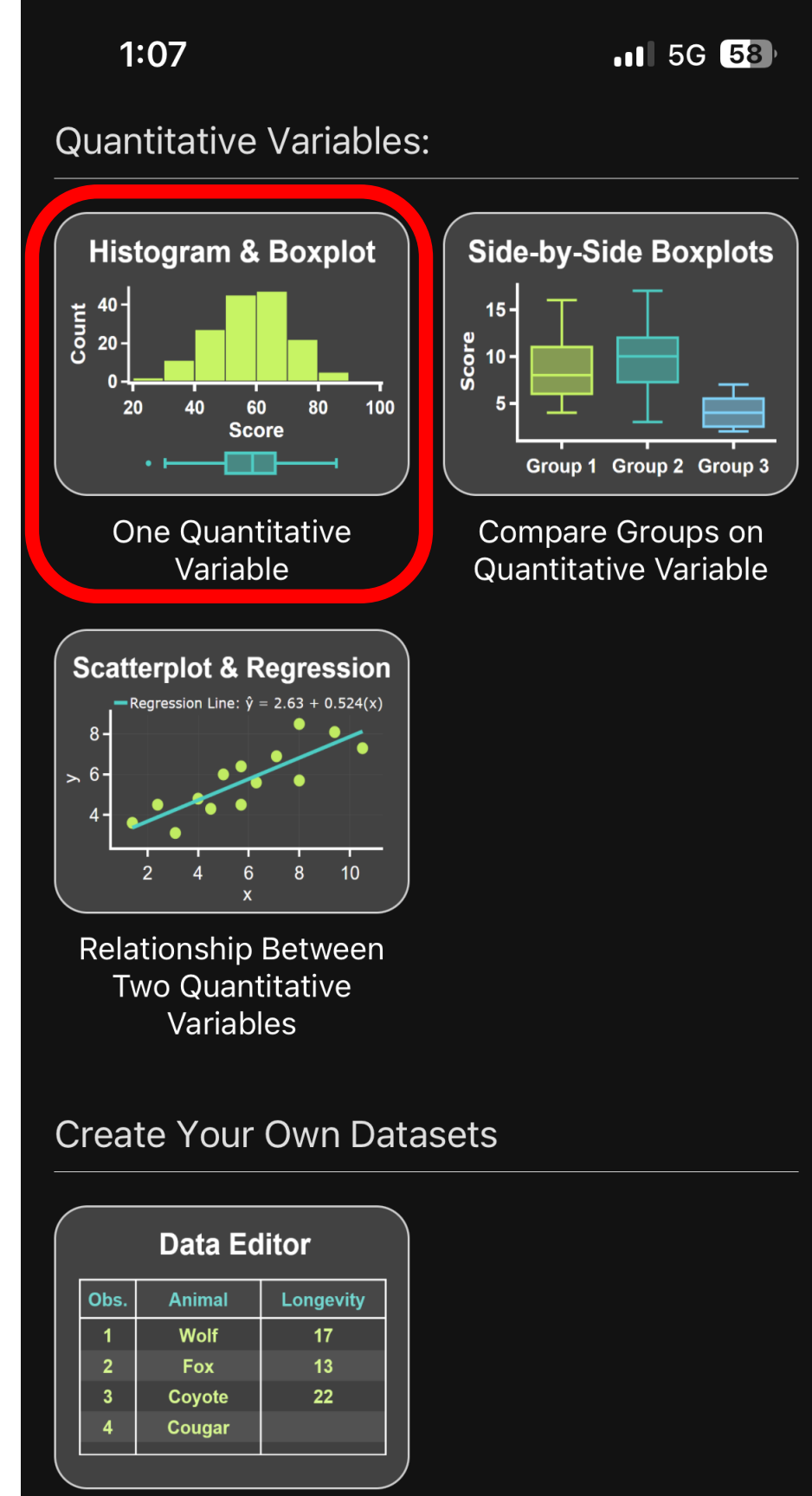


Art of Stat: Explore Data

Art of Stat
Explore Data

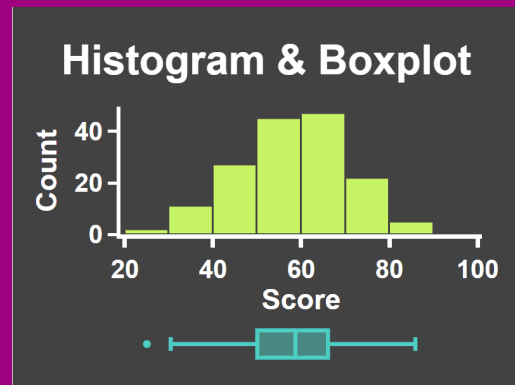


Descriptive Statistics &
Plots for Categorical
and Quantitative
Variables



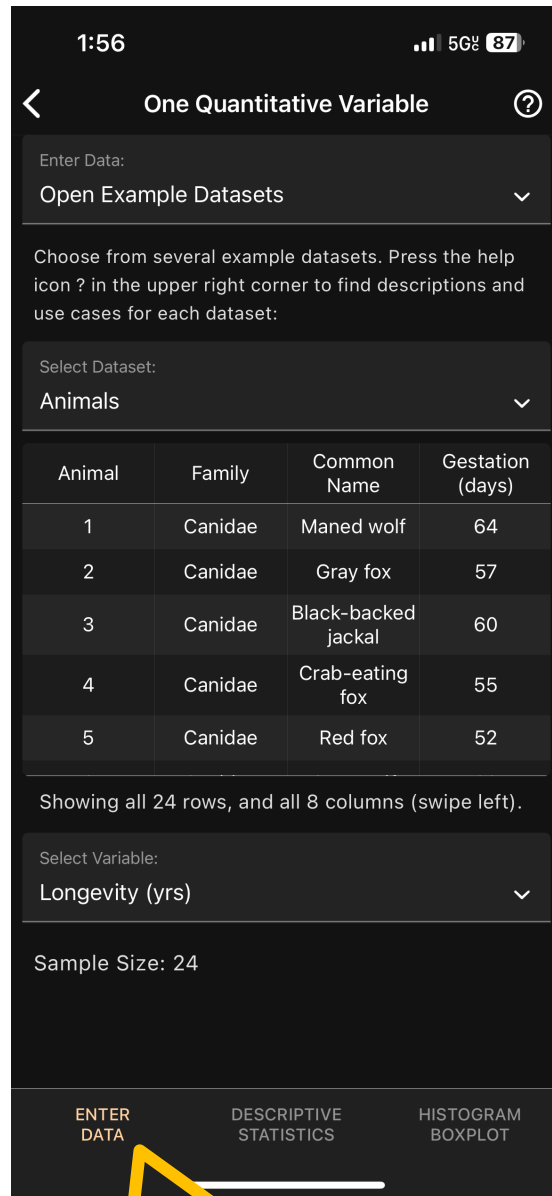
Art of Stat: Explore Data

One Quantitative Variable

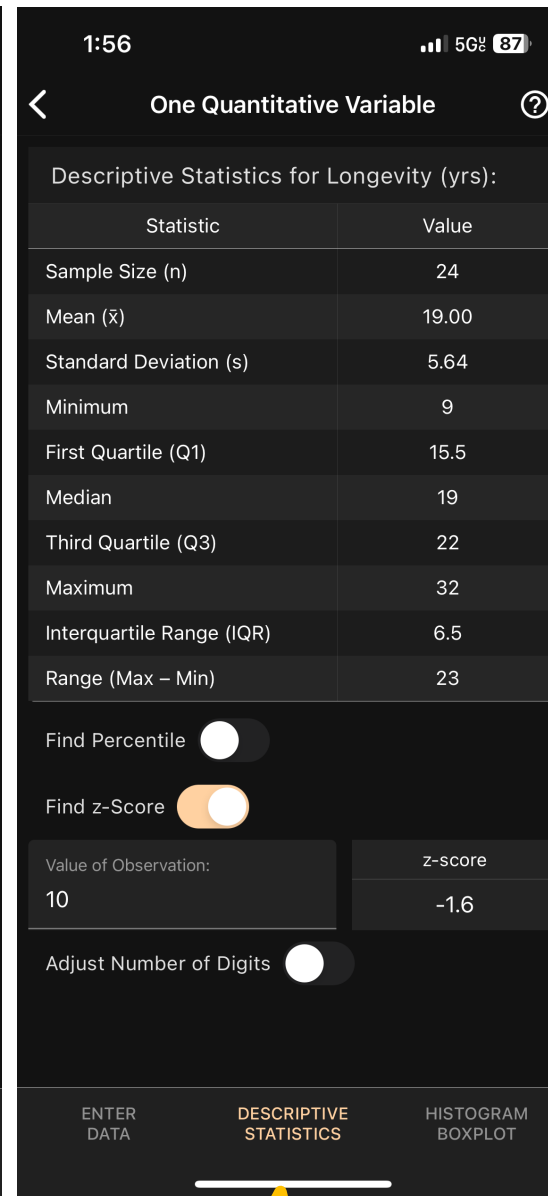


Screens:

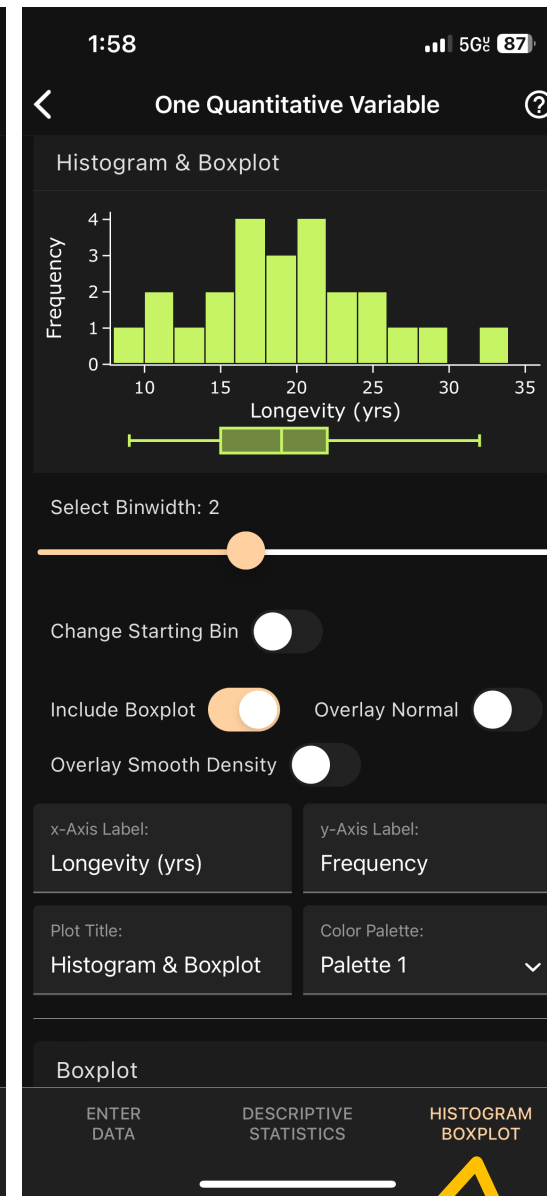
- Enter Data
- Descriptive Statistics
- Histogram, Boxplot



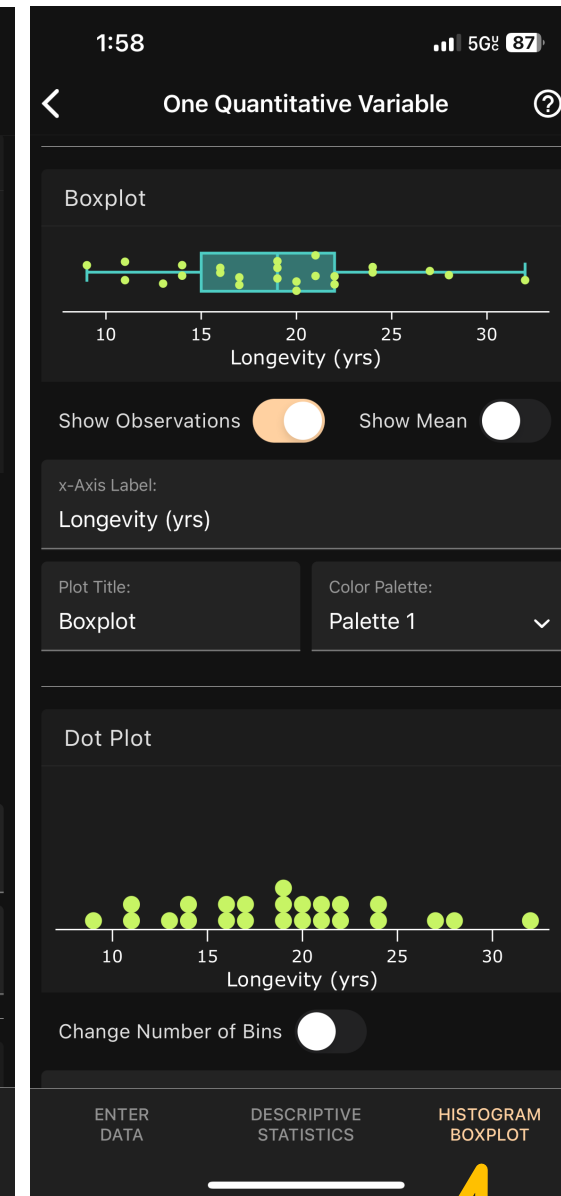
Enter data manually or copy and paste from somewhere else. You can also upload a CSV file, or use one of several example datasets.



Study descriptive statistics, and find percentiles.



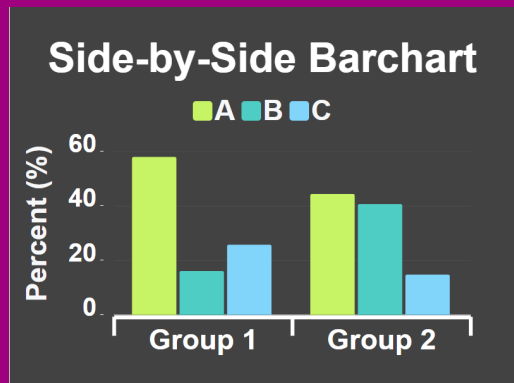
Obtain histogram, vary the bin-size or starting bin, include boxplot, overlay normal or smooth density.



Interact with boxplot, obtain dotplot.

Art of Stat: Explore Data

Two Categorical Variables



Screens:

- Enter Data
- Descriptive Statistics
- Histogram, Boxplot

2:24 5G% 86

Compare Groups: Categorical Response ?

Enter Data:
Open Example Datasets

Choose from several example datasets. Press the help icon ? in the upper right corner to find descriptions and use cases for each dataset:

Select Dataset:
Online Lending

Grade	Home	Income	FICO Score
C	rent	160000	710
C	mortgage	65000	675
B	rent	34000	690
C	mortgage	150000	740
C	own	70000	705
A	mortgage	100000	785

Showing first 100 rows of the 200 rows, and all 8 columns (swipe left).

Group Variable: Duration
Response Variable: Grade

Select one variable that indicates group membership, and select a response variable that holds the observations.

ENTER DATA COMPUTE PROPORTIONS BAR CHART

Datasets about online lending. Can sort categories of variable (not shown, used to sort loan grade categories)

2:24 5G% 86

Compare Groups: Categorical Response ?

Joint Distribution

Contingency Table (Counts):

Duration	Grade			
	A	B	C	D
36	55	31	36	13
60	7	19	14	11
Total	62	50	50	24

Display: Counts Digits: 0

Conditional Distribution

Conditional Distribution of Grade, given Duration:

Duration	Grade			
	A	B	C	D
36	0.393	0.221	0.257	0.093
60	0.117	0.317	0.233	0.183
Total	0.310	0.250	0.250	0.120

Type of Conditional Distribution: Grade, given Duration

ENTER DATA COMPUTE PROPORTIONS BAR CHART

Obtain contingency table, joint, conditional and marginal distributions. (Row or column percentages.)

2:25 5G% 86

Compare Groups: Categorical Response ?

Bar Chart of Conditional Distribution: Grade, given Duration

Stacked Horizontal Labels

x-Axis Label: Duration y-Axis Label: Percent (%)

Legend Title: Grade Color Palette: Palette 1

Plot Title: Bar Chart of Conditional Distribution: Grade, given Duration

ENTER DATA COMPUTE PROPORTIONS BAR CHART

Side-by-Side bar charts

2:25 5G% 86

Compare Groups: Categorical Response ?

Bar Chart of Conditional Distribution: Grade, given Duration

Stacked Horizontal Labels

x-Axis Label: Percent (%) y-Axis Label: Duration

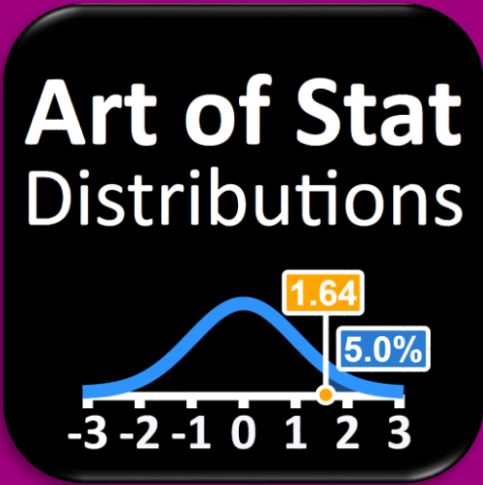
Legend Title: Grade Color Palette: Palette 1

Plot Title: Bar Chart of Conditional Distribution: Grade, given Duration

ENTER DATA COMPUTE PROPORTIONS BAR CHART

Stacked bar charts

Art of Stat: Distribution



Explore & Visualize
Discrete and
Continuous Probability
Distributions

1:22 5G 57

ART OF STAT
Distributions

Continuous Distributions:

- Normal Distribution**
Normal
- Student t-Distribution
t-Distribution
- Chi-Squared Distribution
Chi-Squared
- F-Distribution
F-Distribution
- Exponential Distribution
Exponential
- Uniform Distribution
Uniform

1:37 5G 55

Gamma Distribution
Gamma

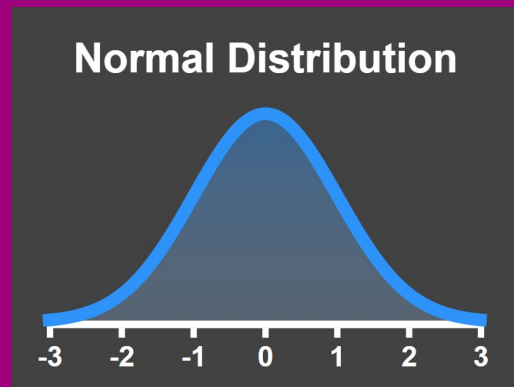
Beta Distribution
Beta

Discrete Distributions:

- Binomial Distribution**
Binomial
- Geometric Distribution
Geometric
- Poisson Distribution
Poisson
- Discrete Distribution
Discrete

Art of Stat: Distributions

Normal Distribution



Screens:

- Explore Distribution
- Find Probability
- Find Percentiles
- Histogram, Boxplot

6:29 5G 100

The Normal Distribution

Graph of Distribution: $\mu = 100, \sigma = 10$

68% in [90.0, 110.0], 95% in [80.0, 120.0]

Mean μ : 100 Standard Dev. σ : 10

Explore Meaning of μ and σ

Selected Percentiles ($\mu = 100, \sigma = 10$)

Percentile	Value	Percentile	Value
0.5%	74.2417	75.0%	106.7449
1.0%	76.7365	90.0%	112.8155
2.5%	80.4004	95.0%	116.4485
5.0%	83.5515	97.5%	119.5996
10.0%	87.1845	99.0%	123.2635

EXPLORE DISTRIBUTION FIND PROBABILITY FIND PERCENTILE SIMULATE NUMBERS

6:29 5G 100

Graph of Normal Distribution

$\mu = 115, \sigma = 8$

Mean μ : 115

Standard Deviation σ : 8

EXPLORE DISTRIBUTION FIND PROBABILITY FIND PERCENTILE SIMULATE NUMBERS

6:31 5G 99

The Normal Distribution

Mean μ : 100 Standard Dev. σ : 10

Normal Distribution with $\mu = 100, \sigma = 10$

$P(X \leq 83.55) + P(X \geq 116.4) = 0.1000$

Percentage: 10 % Type of Percentile: Two-Tailed: $P(|X| \geq x)$

Percentile: Lower Quantile = 83.55 Upper Quantile = 116.4

Selected Percentiles ($\mu = 100, \sigma = 10$)

Percentile	Value	Percentile	Value
0.5%	74.2417	75.0%	106.7449

EXPLORE DISTRIBUTION FIND PROBABILITY FIND PERCENTILE SIMULATE NUMBERS

6:32 5G 99

The Normal Distribution

Mean μ : 100 Standard Dev. σ : 10

Number to Simulate: 30 Simulate Reset

Histogram and Boxplot

30 Random Numbers Simulated from a Normal Distribution with $\mu = 100$ and $\sigma = 10$:

109.18	107.77	98.67	95.64	93.27
97.25	89.02	105.68	108.08	90.39
111.01	100.99	97.90	102.96	94.13
103.65	114.31	106.02	79.89	103.13
97.89	96.12	90.40	106.17	81.95

Copy Numbers

EXPLORE DISTRIBUTION FIND PROBABILITY FIND PERCENTILE SIMULATE NUMBERS

Enter values for the mean and standard deviation, and get an overview of all the essentials

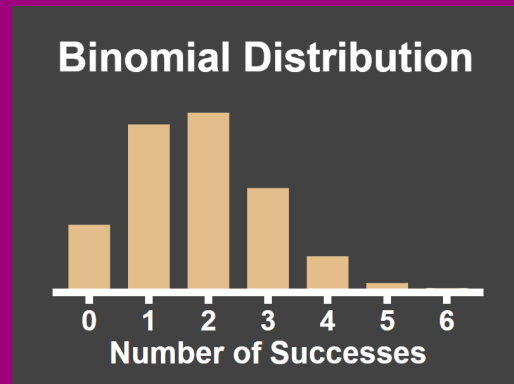
To understand the meaning of "mu" and "sigma", use sliders to see that these describe location and variability

You can find probabilities and percentiles, and confirm them visually

Simulate from a normal distribution

Art of Stat: Distributions

Binomial Distribution



Screens:

- Explore Distribution
- Find Probability
- Simulate Numbers

7:47

The Binomial Distribution

Graph of Distribution: $n = 10, p = 0.3$

Mean: 3, Standard Deviation: 1.449

Number of Bernoulli Trials (n): 10

Probability of Success (p): 0.3

Enter Value for n Enter Value for p

Table of Binomial Probabilities (n = 10, p = 0.3)

x	P(X = x)	P(X ≤ x)	P(X ≥ x)
0	0.0000	0.0000	1.0000
1	0.0000	0.0000	1.0000
2	0.0000	0.0000	1.0000
3	0.0000	0.0000	1.0000
4	0.0000	0.0000	1.0000
5	0.0000	0.0000	1.0000
6	0.0000	0.0000	1.0000
7	0.0000	0.0000	1.0000
8	0.0000	0.0000	1.0000
9	0.0000	0.0000	1.0000
10	0.0000	0.0000	1.0000

EXPLORE DISTRIBUTION FIND PROBABILITY SIMULATE NUMBERS

7:47

The Binomial Distribution

Number of Trials (n): 10 Probability p: 0.3

Binomial Distribution with $n = 10, p = 0.3$
 $P(X = 5) = 0.1029$

Mean: 3, Standard Deviation: 1.449

Value of x: 5 Type of Probability: Probability: P(X = x)

Probability:
 $P(X = 5) = 0.1029$

Find the number of possible ways of obtaining 5 successes in 10 trials

EXPLORE DISTRIBUTION FIND PROBABILITY SIMULATE NUMBERS

7:47

The Binomial Distribution

Number of Trials (n): 10 Probability p: 0.3

Binomial Distribution with $n = 10, p = 0.3$
 $P(2 ≤ X ≤ 4) = 0.7004$

Mean: 3, Standard Deviation: 1.449

Lower Bound x1: 2 Type of Probability: Interval: P(x1 ≤ X ≤ x2)

Upper Bound x2: 4

Probability:
 $P(2 ≤ X ≤ 4) = 0.7004$

EXPLORE DISTRIBUTION FIND PROBABILITY SIMULATE NUMBERS

7:48

The Binomial Distribution

Number of Trials (n): 10 Probability p: 0.3

Number to Simulate: 30

30 Random Numbers Simulated from a Binomial Distribution with $n = 10$ and $p = 0.3$:

1	4	3	1	2
2	1	2	5	3
5	7	4	4	3
4	1	3	3	2

Copy Numbers

EXPLORE DISTRIBUTION FIND PROBABILITY SIMULATE NUMBERS

Visualize the binomial distribution, see how it changes with p, get the probability table for any value of n and p.

Easily find and visualize binomial probabilities...

... of any type, individual, lower tail, upper tail or interval.

Simulate from a binomial distribution

Art of Stat: Concepts



Central Limit Theorem,
Correlation,
Regression,
Coverage & Power

2:51 5G 50

ART OF STAT
Concepts

Central Limit Theorem

Distribution of the Sample Mean \bar{x}
Population (green), Sampling (blue)
CLT for Means

Distribution of the Sample Proportion \hat{p}
Simulated (green), Normal (CLT) (blue)
CLT for Proportions

Correlation and Regression

Guess Correlation
 $r = ?$
Explore Correlation

Explore Regression
Regression Line
Explore Regression

Coverage, Errors and Power

Explore Coverage
Explore Coverage of Confidence Intervals

Errors and Power
 $-P(\text{Type I}) = 5\%$ $-P(\text{Type II}) = 21\%$

2:52 5G 50

Correlation and Regression

Guess Correlation
 $r = ?$
Explore Correlation

Explore Regression
Regression Line
Explore Regression

Coverage, Errors and Power

Explore Coverage
Explore Coverage of Confidence Intervals

Errors and Power
 $-P(\text{Type I}) = 5\%$ $-P(\text{Type II}) = 21\%$

Check Out Other Apps:

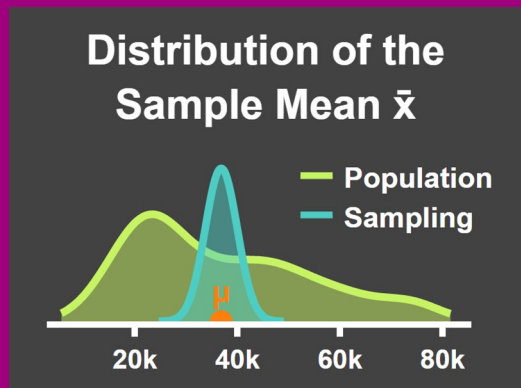
Art of Stat Distributions
1.64, 5.0%

Art of Stat Explore Data

Art of Stat Inference

Art of Stat: Concepts

Central Limit Theorem for Means



7:52

CLT for Means

Select Population Distribution:
US Colleges: Attendance Cost

Population Distribution

Population Description

The population consists of 1,829 public & private U.S. colleges offering a bachelor's degree. The variable measured for each college is the average cost of attendance (in USD) for academic year 2019/20, ranging from \$5,663 to \$81,531.

Population Parameters

Population Size	1,829
Population Mean (μ)	36,844
Population Standard Deviation (σ)	17,139
Population First Quartile (Q1)	22,827
Population Median	32,776
Population Third Quartile (Q3)	48,844

POPULATION DISTRIBUTION DRAW SAMPLES SAMPLING DISTRIBUTION

7:54

CLT for Means

Population Distribution: $\mu = 36,844$, $\sigma = 17,139$

Distribution of Sample: $n = 25$, $\bar{x} = 34,304$, $s = 13,408$

Sampling Distribution of \bar{x} (5633 Simulations)
Mean = 36,801, Std. Dev. = 3,409

Sample Size n: 25 Draw 1 Sample Draw 100 Samples

Reset

Normal Distribution Zoom In

POPULATION DISTRIBUTION DRAW SAMPLES SAMPLING DISTRIBUTION

9:49

CLT for Means

Population Distribution: $\mu = 36,844$, $\sigma = 17,139$

Distribution of Sample: $n = 100$, $\bar{x} = 34,135$, $s = 15,933$

Sampling Distribution of \bar{x} (5633 Simulations)
Mean = 36,843, Std. Dev. = 1,721

Sample Size n: 100 Draw 1 Sample Draw 100 Samples

Reset

Normal Distribution Zoom In

POPULATION DISTRIBUTION DRAW SAMPLES SAMPLING DISTRIBUTION

7:56

CLT for Means

Sampling Distribution of \bar{x} ($n = 80$)

Histogram of Simulated Sample Means

Simulated Sampling Distribution of \bar{x} (smooth)

Sampling Distribution of \bar{x} (if CLT applies)

Population Distribution

Distribution of the Sample Mean \bar{x} ($n = 80$)

Statistic	Population Distribution	Theoretical Distribution*	Simulated Distribution**
Mean	36,844	36,844	36,845
Std. Dev.	17,139	1,916	1,910
Q1	22,827	35,552	35,537
Median	32,776	36,844	36,817
Q3	48,844	38,136	38,150

POPULATION DISTRIBUTION DRAW SAMPLES SAMPLING DISTRIBUTION

Look at a real population distribution, and its parameters (population mean, standard deviation). Several examples available.

Draw samples of a given size from this population distribution, and keep track of the sample means

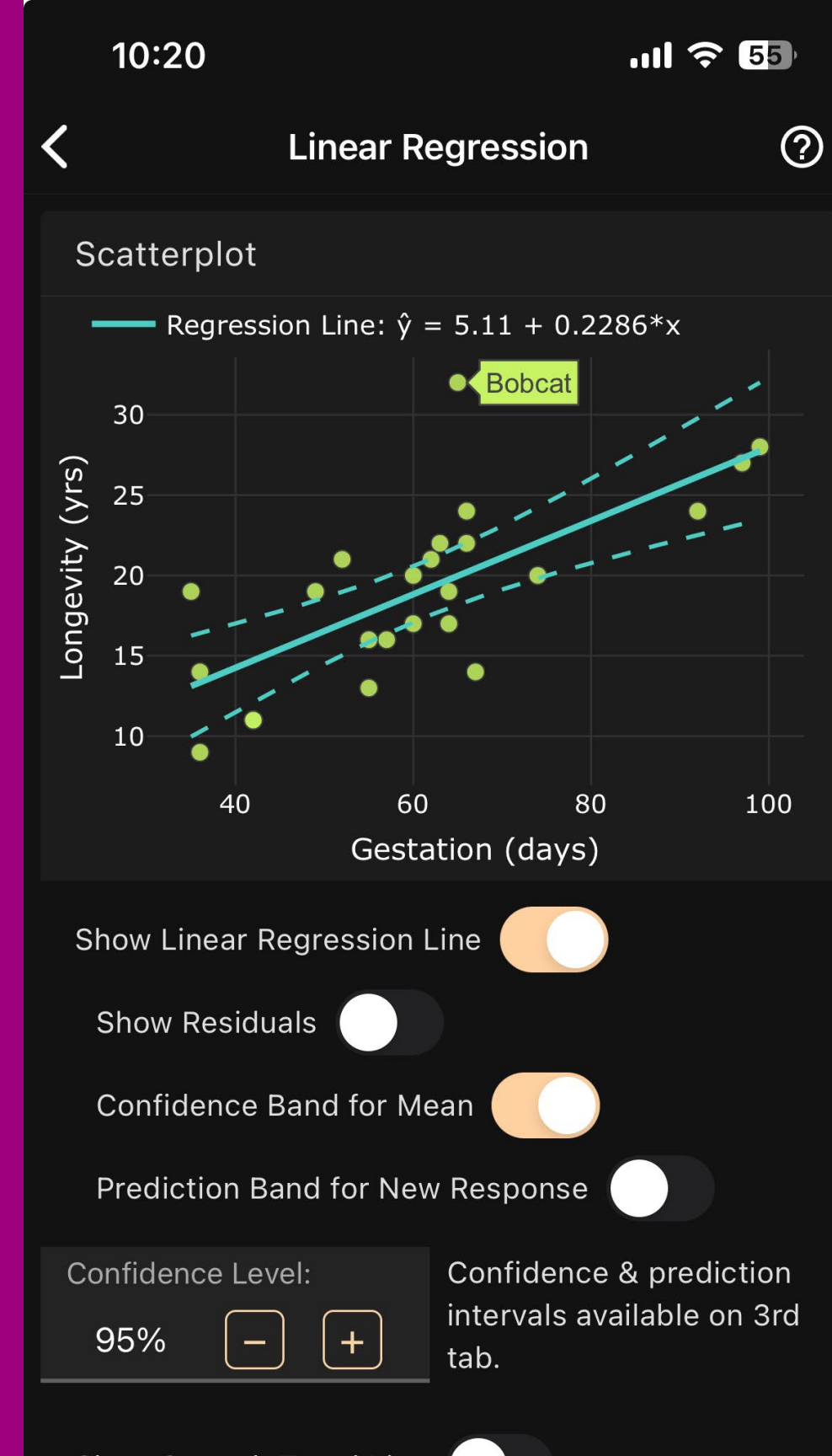
See what happens as the sample size increases

Show (simulated) sampling distribution, compare to population distribution

- Screens:
- Population Distribution
 - Draw Samples
 - Sampling Distribution

3) Case Studies (cont.)

- Art of Stat: *Inference* app
 - Inference for a Population Mean
 - Inference Comparing Two Population Proportions
- Art of Stat: *Regression* app
 - Linear Regression



Art of Stat: Inference



Confidence Intervals &
Hypothesis Tests

1:47 5G% 54

ART OF STAT Inference

Inference About Proportions:

- Single Population Proportion**
95% Confidence Interval
0 0.5 1
Proportion p
Inference for a Population Proportion (One Sample)
- Difference of Two Population Proportions** (highlighted with a red border)
95% Confidence Interval
-1 0 1
Difference $p_1 - p_2$
Compare Two Population Proportions (Independent Samples)
- Chi-Square Test (Goodness of Fit)**
Stacked Bar Chart
Obs. Exp.
0% 50% 100%
Chi-Square Test (Goodness of Fit)
- Chi-Square Test (Independence/Homogeneity)**
A B C
Group 1
Group 2
Group 3
0% 50% 100%
Chi-Square Test (Independence or Homogeneity)

Inference About Means:

- Single Population**
- Difference of Two**

1:48 5G% 54

Inference About Means:

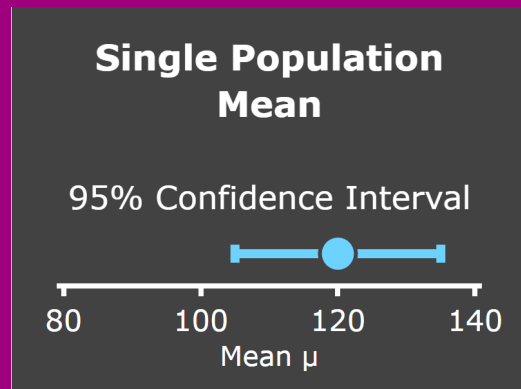
- Single Population Mean** (highlighted with a red border)
95% Confidence Interval
80 100 120 140
Mean μ
Inference for a Population Mean (One Sample)
- Difference of Two Population Means**
95% Confidence Interval
0 10 20 30
Difference $\mu_1 - \mu_2$
Compare Two Population Means (Independent Samples)
- One-Way ANOVA**
Group 1
Group 2
Group 3
0 5 10
One-Way ANOVA

Inference in Linear Regression:

- Linear Regression**
Regression Line: $\hat{y} = 2.6 + 0.5(x)$
y
x
2 4 6 8 10

Art of Stat: Inference

Inference for a Population Mean



- Screens:
- Enter Data
 - Confidence Interval
 - Hypothesis Testing

10:35

Inference for a Mean

Enter Data:
Open Sample Datasets

Select Dataset:
Online Lending

Loan	Loan Amount	Duration	Interest Rate
1	35000	36	13.6
2	12000	60	16.1
3	15000	36	11.8
4	3025	36	15.1
5	9000	36	14.5
6	20000	36	6.5
7	12500	60	7.8

Showing all 200 rows, and all 8 columns (swipe left).

Select Variable:
Loan Amount

Enter Population Standard Deviation

Note: Only switch on if you know the population standard deviation σ and want to use it for inference.

Descriptive Statistics for Loan Amount:

ENTER DATA CONFIDENCE INTERVAL HYPOTHESIS TEST

Type in or copy & paste data, load CSV file, or use example dataset

8:09

Inference for a Mean

Descriptive Statistics for Loan Amount:

Statistic	Value
Sample Size (n)	200
Sample Mean (\bar{x})	16604
Sample Standard Deviation (s)	9802
Minimum	1000
First Quartile (Q1)	10000
Median	15000
Third Quartile (Q3)	21437.5
Maximum	40000
Interquartile Range (IQR)	11437.5
Range (Max - Min)	39000

Adjust Number of Digits

Histogram

Frequency

Loan Amount

ENTER DATA CONFIDENCE INTERVAL HYPOTHESIS TEST

Immediately get descriptive statistics and graphs to check assumptions

8:10

Inference for a Mean

Confidence Interval for the Population Mean μ :

Statistic	Value
Sample Size (n)	200
Sample Mean (\bar{x})	16604
Sample Standard Deviation (s)	9802
Standard Error (se) of the Sample Mean (s/\sqrt{n})	693
Confidence Level	95%
t-score (df = 199, $\alpha = 5\%$)	1.972
Margin of Error (me)	± 1367
Lower Bound for μ	15237
Upper Bound for μ	17971

μ is the population mean of Loan Amount.

95% Confidence Interval (15237, 17971)

Population Mean μ

Confidence Level: 95%

ENTER DATA CONFIDENCE INTERVAL HYPOTHESIS TEST

Obtain confidence interval for mean, including all intermediate steps. Change slider for confidence coefficient.

8:11

Inference for a Mean

Null Value: 15000 Alternative Hypothesis: Two-Sided

Hypothesis Test for the Population Mean μ :

Null Hypothesis	$H_0: \mu = 15000$
Alternative Hypothesis	$H_a: \mu \neq 15000$
Sample Mean (\bar{x})	16604
Sample Standard Deviation (s)	9802
Standard Error (se) of the Sample Mean (s/\sqrt{n})	693
Test Statistic (t, df = 199)	2.31
P-value	0.0217

μ is the population mean of Loan Amount.

t-Distribution with df = 199
Test Statistic: t = 2.31, P-value = 0.0217

ENTER DATA CONFIDENCE INTERVAL HYPOTHESIS TEST

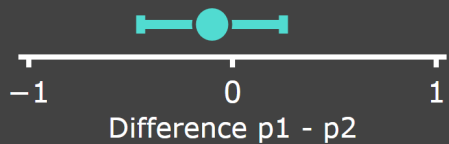
Carry out t-test, and obtain all intermediate steps leading to P-value. Visualize P-value on t-distribution.

Art of Stat: Inference

Inference for a Population Mean

Difference of Two Population Proportions

95% Confidence Interval



8:04

Inference for Two Proportions

Enter Data:
Open Sample Datasets

Select Dataset:
Online Lending

Loan Amount	Duration	Interest Rate	Grade
35000	36	13.6	C
12000	60	16.1	C
15000	36	11.8	B
3025	36	15.1	C
9000	36	14.5	C
20000	36	6.5	A
12500	60	7.8	A

Showing all 200 rows, and all 8 columns (swipe left).

Group Variable: Duration
Response Variable: Grade

Select a grouping variable that indicates group membership, and select a response variable that holds the observations. Then, select which category of the response variable you want to compute the proportions for.

Group 1: Group 2:

ENTER DATA CONFIDENCE INTERVAL HYPOTHESIS TEST

8:04

Inference for Two Proportions

Group 1: 36
Group 2: 60

Compute Proportion For:
A

	36	60
Statistic		
Number of Successes	55	7
Sample Size (n)	140	60
Sample Proportion (\hat{p})	0.393	0.117
Standard Error (se)	0.041	0.041

Difference in Sample Proportions: $\hat{p}_1 - \hat{p}_2 = 0.28$

Bar Chart

Modify Plot Title
Modify x-Axis Label

ENTER DATA CONFIDENCE INTERVAL HYPOTHESIS TEST

8:04

Inference for Two Proportions

Confidence Interval for the Difference $p_1 - p_2$ Between Two Population Proportions:

Statistic	Value
Difference Between the Two Sample Proportions ($\hat{p}_1 - \hat{p}_2$)	0.2762
Standard Error (se) of the Difference	0.0585
Confidence Level	95%
z-score ($\alpha = 5\%$)	1.9600
Margin of Error (me)	± 0.1146
Lower Bound for $p_1 - p_2$	0.1615
Upper Bound for $p_1 - p_2$	0.3908

p_1 and p_2 are population proportions for Group 1 and Group 2, respectively.

95% Confidence Interval (0.1615, 0.3908)

Confidence Level: 95%

ENTER DATA CONFIDENCE INTERVAL HYPOTHESIS TEST

8:05

Inference for Two Proportions

Hypothesis Test for the Difference $p_1 - p_2$ Between Two Population Proportions:

Null Hypothesis	$H_0: p_1 - p_2 = 0$
Alternative Hypothesis	$H_a: p_1 - p_2 \neq 0$
Difference Between the Two Sample Proportions ($\hat{p}_1 - \hat{p}_2$)	0.2762
Null Standard Error (se0) of the Difference	0.0714
Test Statistic (z)	3.87
P-value	0.0001

p_1 and p_2 are population proportions for Group 1 and Group 2, respectively.

Standard Normal Distribution
Test Statistic: $z = 3.87$, P-value = 0.0001

ENTER DATA CONFIDENCE INTERVAL HYPOTHESIS TEST

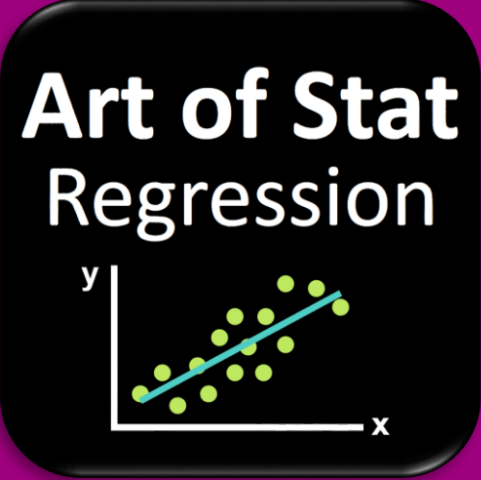
Type in number of successes and trials (not shown), or load CSV file, or use example dataset

Immediately get sample proportions in each group and bar graph to describe data

Obtain confidence interval for difference of proportions, including intermediate steps.

Carry out hypothesis test. Visualize P-value on normal distribution graph.

Art of Stat: Regression



Simple Linear and
Logistic Regression &
Multiple Linear
Regression

2:46 5G 51

ART OF STAT
Regression

Simple Regression (One Predictor)

Linear Regression

$\hat{y} = 5.2 + 0.23(x)$

Longevity vs Gestation

Linear Regression

Exponential Regression

$\hat{y} = 4.9 * 1.52^{(x)}$

Population vs Days

Exponential Regression

Logistic Regression

$\hat{y} = 1 / (1 + \exp(7.1 - 0.23x))$

P(Card) vs Income

Logistic Regression

Multiple Regression (Several Predictors)

Multiple Linear Regression

Multiple Linear Regression

2:47 5G 51

Multiple Regression (Several Predictors)

Multiple Linear Regression

Multiple Linear Regression

Create Your Own Datasets

Data Editor

Obs.	Animal	Longevity
1	Wolf	17
2	Fox	13
3	Coyote	22
4	Cougar	

Data Editor

Check Out Other Apps:

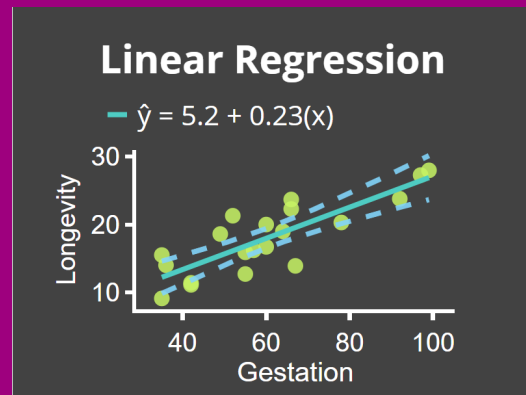
Art of Stat
Explore Data

Art of Stat
Distributions

Art of Stat
Inference

Art of Stat: Inference

Inference for a Population Mean



8:14

Linear Regression

Enter Data:
Open Sample Datasets

Select Dataset:
Palmer Penguins

Species	Island	Bill Length (mm)	Bill Depth (mm)
Adelie	Torgersen	39.1	18.7
Adelie	Torgersen	39.5	17.4
Adelie	Torgersen	40.3	18
Adelie	Torgersen	36.7	19.3
Adelie	Torgersen	39.3	20.6
Adelie	Torgersen	38.9	17.8

Showing first 250 rows of the 342 rows, and all 7 columns (swipe left).

Select x-Variable: Bill Length (mm) | Select y-Variable: Flipper Length (mm)

ID Variable | Color Points

Descriptive Statistics:

	Bill Length (mm)	Flipper Length (mm)
Mean	43.92	187.61
Std. Error	1.96	10.81
Lower Bound	40.98	176.79
Upper Bound	46.86	198.43

ENTER DATA | SCATTERPLOT AND REGRESSION MODEL | FITTED VALUES AND RESIDUALS

Type in or copy & paste data, load CSV file, or use example dataset.

8:18

Linear Regression

Scatterplot

Regression Line: $\hat{y} = 126.68 + 1.69*x$

Obs: 169

Show Linear Regression Line | Show Residuals | Confidence Band for Mean | Prediction Band for New Response | Show Smooth Trend Line

x-Axis Label: Bill Length (mm) | y-Axis Label: Flipper Length (mm)

Plot Title: Scatterplot | Color Palette: Palette 1

ENTER DATA | SCATTERPLOT AND REGRESSION MODEL | FITTED VALUES AND RESIDUALS

Obtain scatterplot and superimpose linear regression line. Identify observations.

8:16

Linear Regression

Estimates of Parameters α and β in Linear Regression Model: $\mu = \alpha + \beta*x$

	Intercept α	Slope β (Bill Length (mm))
Estimate	126.68	1.690
Std. Error (se)	4.67	0.105
Null Hypothesis	$H_0: \alpha = 0$	$H_0: \beta = 0$
Test statistic (t)	27.16	16.03
P-value	< 0.0001	< 0.0001

Standard Errors & P-values | Confidence Interval for the Slope

Confidence Level: 95%

95% Confidence Interval for the Slope β :

Estimate	Std. Error	Lower Bound	Upper Bound
1.690	0.105	1.483	1.897

Correlation and Model Statistics:

Pearson Correlation Coefficient (r): 0.6562

ENTER DATA | SCATTERPLOT AND REGRESSION MODEL | FITTED VALUES AND RESIDUALS

Get estimates of intercept and slope, their standard errors, and P-values. Get R^2 .

8:19

Linear Regression

Fitted Values & Residuals:

ID	Bill Length (mm)	Flipper Length (mm)	Fitted Values (\hat{y})
1	39.1	181	192.8
2	39.5	186	193.4
3	40.3	195	194.8
4	36.7	193	188.7
5	39.3	190	193.1

Showing first 250 rows of the 342 rows, and all columns (swipe left).

Download Dataset of Fitted Values

Find Predicted Value

Value for Bill Length (...): 44 | Predicted Value: 201.0

Confidence & Prediction Interval | Show Residual Plot

x-Axis: Bill Length (mm) | y-Axis: Studentized

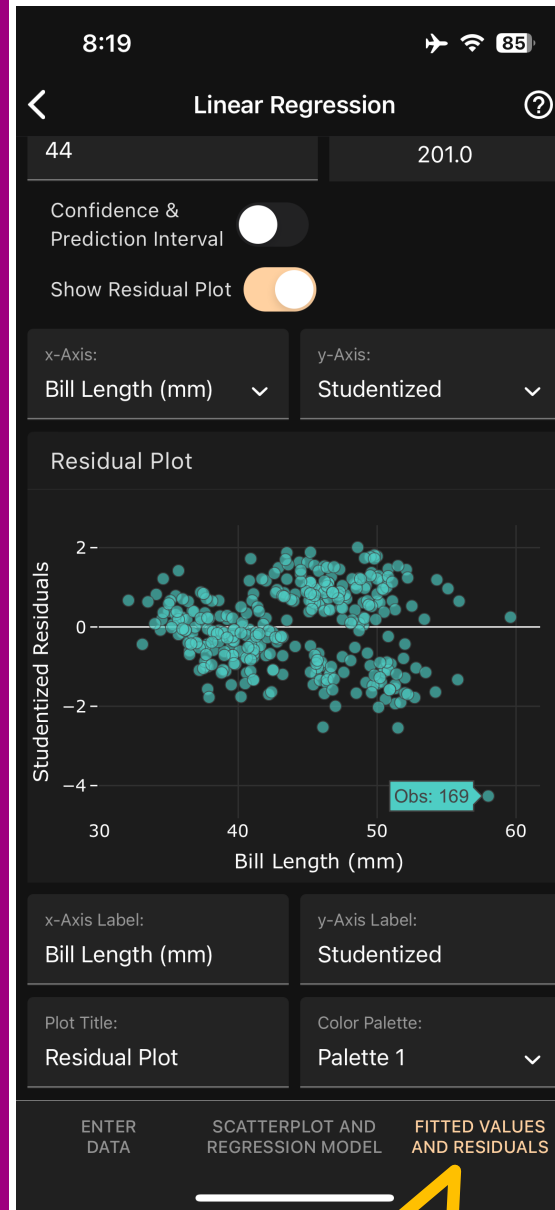
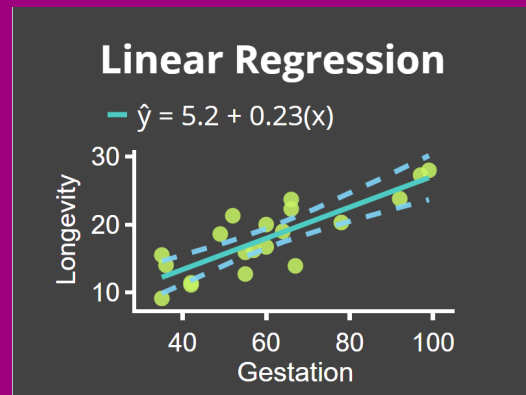
Residual Plot

ENTER DATA | SCATTERPLOT AND REGRESSION MODEL | FITTED VALUES AND RESIDUALS

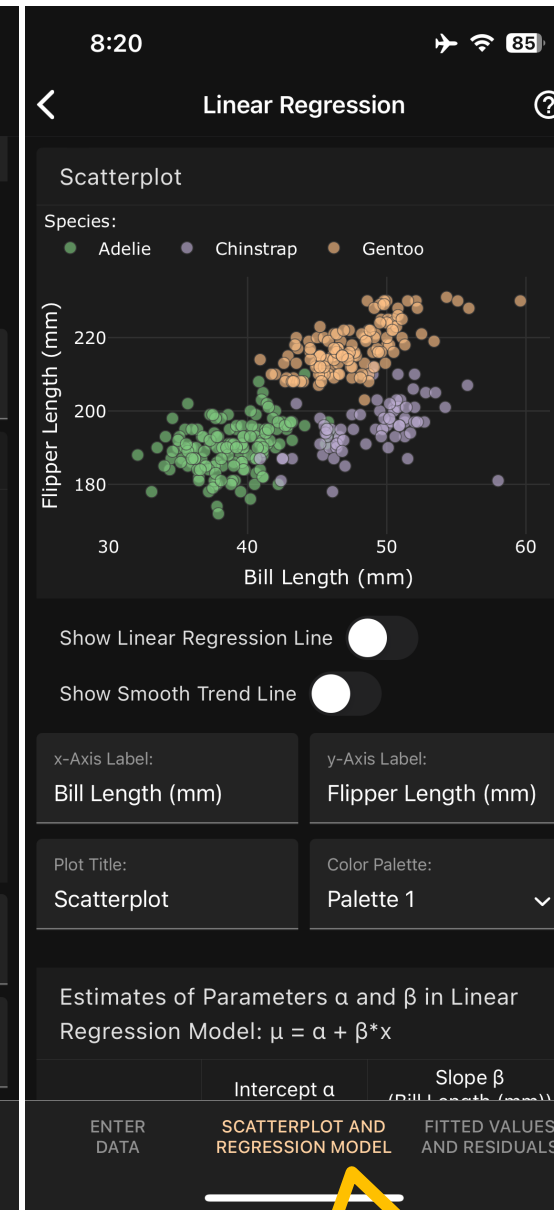
Obtain fitted values, and make predictions for new x values.

Art of Stat: Inference

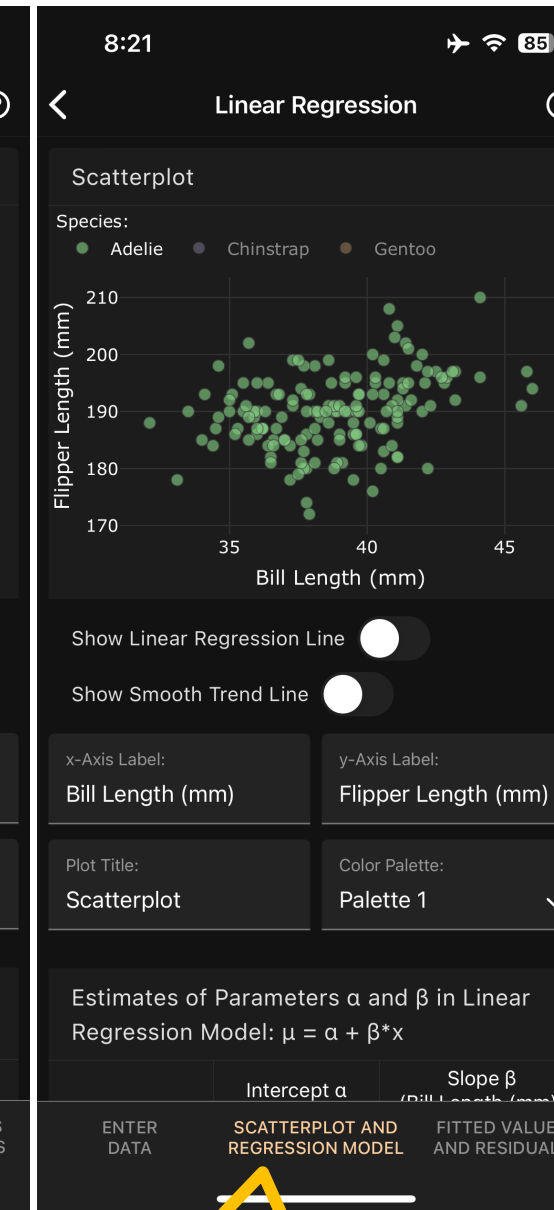
Inference for a Population Mean



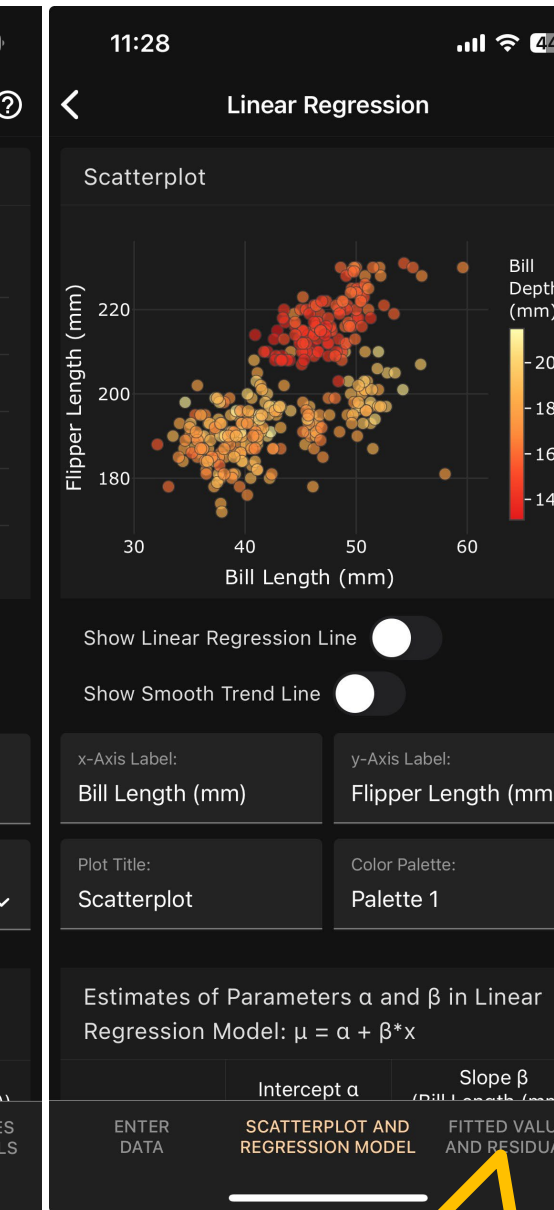
Obtain residual plot, using standardized residuals.



Use a third variable to color the dots (selected in Enter Data screen) to reveal group structures.



Tab on legend to look at only one group.



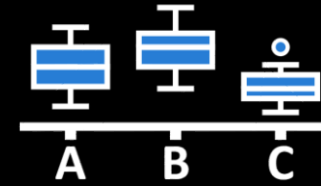
Can also color dots according to a third continuous variable

4) Q&A

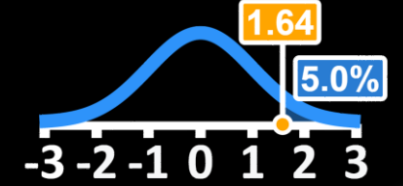
Thank You!

Any Questions?

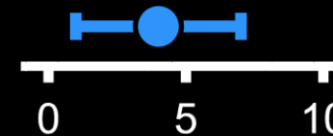
Art of Stat
Explore Data



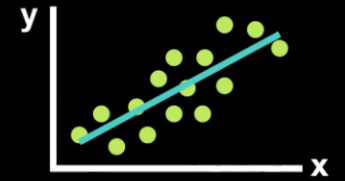
Art of Stat
Distributions



Art of Stat
Inference



Art of Stat
Regression



Art of Stat
Concepts



Art of Stat
Resampling

