Inference

By

Frances Hinchliffe, WGC

Ivy Liu, VUW

Content

Secondary School	STAT193
• PPDAC • Problem • Plan • Data • Analysis • Conclusion	• Descriptive statistics • Bivariate statistics • Regression • Prob / Binomial r. v. • Normal distribution • CLT / Confidence intervals • Hypothesis testing • t-tests (small samples) • ANOVA • Contingency tables

Content

Secondary School	STAT193
• PPDAC • Problem • Plan • Data • Analysis • Conclusion	• Descriptive statistics • Bivariate statistics • Regression • Prob / Binomial r. v. • Normal distribution • **CLT / Confidence intervals !!!** • Hypothesis testing • t-tests (small samples) • ANOVA • Contingency tables

Differences: CLT

Secondary School

- Bootstrap confidence intervals for the difference between medians

STAT193

- Central Limit Theorem (CLT)
- Asymptotic distribution of $\bar{X}$

  Provided n is large, then the sample mean will be normal.

- Formulae

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

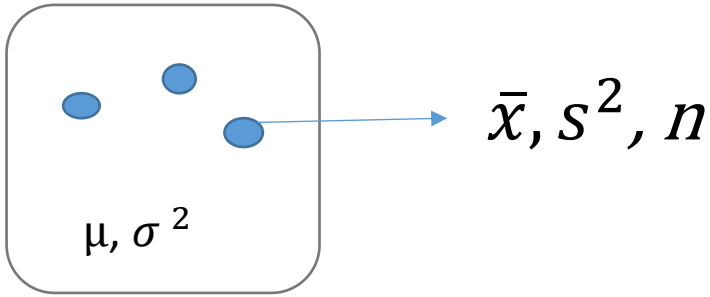
Differences: Confidence Intervals

Secondary School

- Our confidence in our interval comes from the fact that the method we are using almost always works.
- Refer to a bootstrap confidence interval

STAT193

- Population vs. Samples



- CI for μ is

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

- Confidence level and significance level

Content

Secondary School	STAT193
• PPDAC • Problem • Plan • Data • Analysis • Conclusion	• Descriptive statistics • Bivariate statistics • Regression • **Prob / Binomial r. v. !!!** • Normal distribution • CLT / Confidence intervals • Hypothesis testing • t-tests (small samples) • ANOVA • Contingency tables

Differences: Binomial

Secondary School

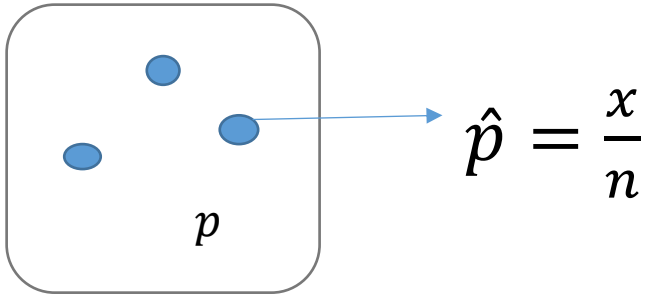
- No proportions

STAT193

- X counts the number of successes in n trials

$$X \sim Binomial\ (n, p)$$

- Population vs. Samples



- CLT: Sample proportion
  ⇨ Population proportion

Differences: Softwares, Calculators, Formulae

Secondary School

- All graphs created using iNZight
- Bootstrapping performed by iNZIght
- All statistics calculated by iNZight
- Confidence interval for difference between medians generated by iNZight
- No formulas

STAT193

- Use Graphics calculator
  Casio FX9750Gii
- No Statistical software
- Some formulae are required, e.g., confidence intervals.

Content

Secondary School	STAT193
• PPDAC • Problem • Plan • Data • Analysis • Conclusion	• Descriptive statistics • Bivariate statistics • Regression • Prob / Binomial r. v. • Normal distribution • CLT / Confidence intervals • **Hypothesis testing** • **t-tests (small samples)** • ANOVA • Contingency tables

Differences: Hypothesis testing or t-tests

Secondary School

- Interpretation of confidence interval using “It’s a fairly safe bet”

STAT193

- Hypothesis testing

  Eg. H_0: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$
- Hypothesis testing

  $\Leftrightarrow$ Confidence intervals
- The t-tests are for small samples

How little a student needs to achieve

Pose a comparison investigative question using a given multivariate data set

Comparative investigative question: comparing of the median % of body fat difference in female and male.

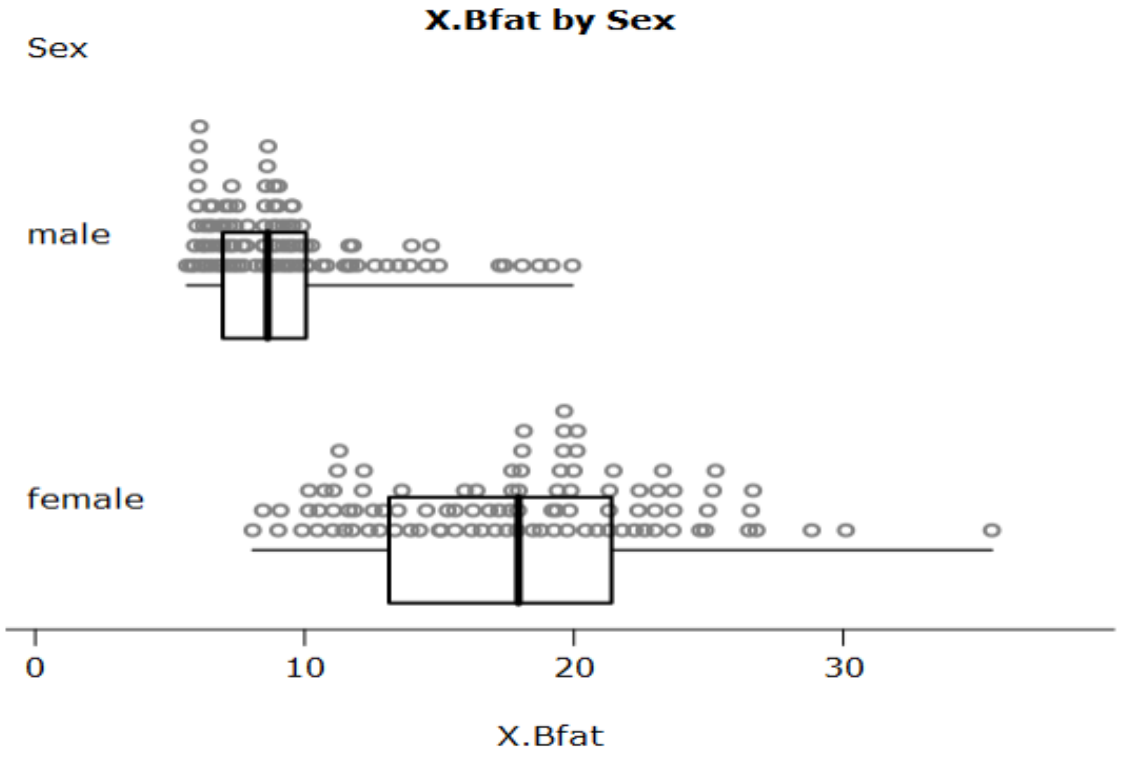
1

Variables to be investigated: in this data of the different sports from the Australian institution of sport (AIS) I will be comparing the % body fat between the female and male in the sports listed by AIS.

Selecting and using appropriate displays and summary statistics

Summary of X.Bfat by Sex

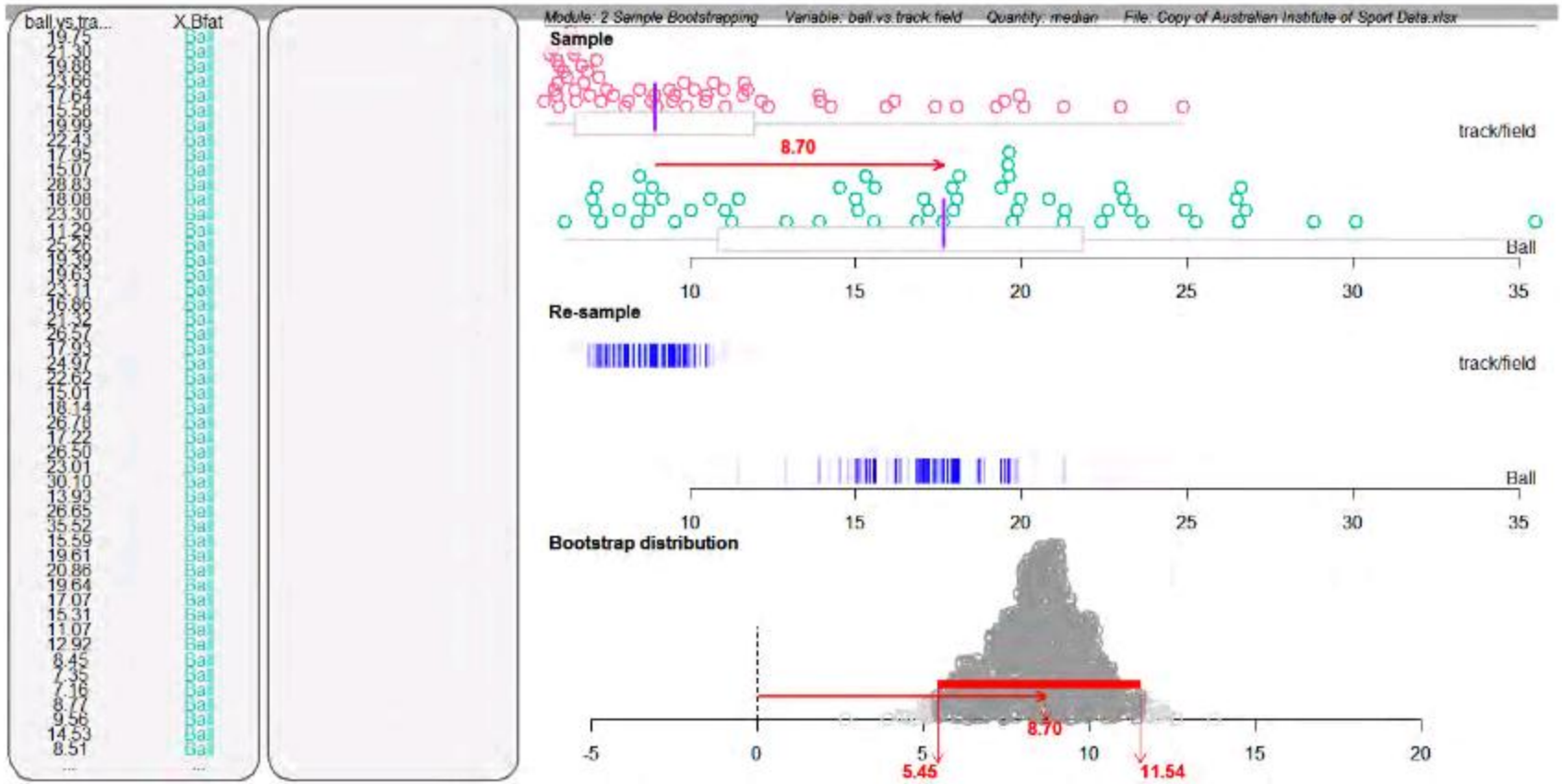
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std.dev	Sample.Size
female	8.07	13.240	17.940	17.850	21.36	35.52	5.453	100
male	5.63	6.967	8.625	9.251	10.01	19.94	3.185	102



- Discussing sample distributions

- Discuss sampling variability, including the variability of estimates

This can be implied with the use of bootstrapping


ball.vs.tra...
X.Bfat
19.75
21.30
19.88
23.66
17.64
15.58
19.99
22.43
17.95
15.07
28.83
18.08
23.30
11.29
19.39
19.63
23.11
16.86
21.32
26.57
17.93
24.97
22.62
15.01
18.14
26.70
17.22
26.50
23.01
30.10
13.93
26.65
35.52
15.59
19.61
20.86
19.64
17.07
15.31
11.07
12.92
8.45
7.35
7.16
8.77
9.56
14.53
8.51
Module: 2 Sample Bootstrapping
Variable: ball.vs.track.field
Quantity: median
File: Copy of Australian Institute of Sport Data.xlsx
Sample
track/field
8.70
Ball
10
15
20
25
30
35
Re-sample
track/field
Ball
10
15
20
25
30
35
Bootstrap distribution
8.70
-5
0
5
10
15
20
5.45
11.54

Making an appropriate formal statistical inference

Statistical reference: the data does suggest that it is fairly safe to suggest that back in the population that comparing % body fat of males and females that there is quite a large difference between them. In fact, the data suggests that the median % body fat for males is lower than the median % body fat in females, which is totally what I expected to find.

Looking at my bootstrap confidence interval I can fairly state that there is a difference in the female and male median % of body fat in the different sports from the AIS and ages 18 up to about 40. This is clearly evident that in the fact that my bootstrap confidence interval does not include 0.

- communicating findings in a conclusion