

Lesson Plan 3/27

Tuesday, March 27, 2018 10:32 AM

Admin

- Midterm grades on elms, rubric and comments on gradescope see piazza@370
- Calendar update

HDSC

- Data Science Design Manual (via Brian G.)
<http://www.springer.com/gp/book/9783319554433>

Midterm I

- My questions (3,6,8,10,12)
- Your questions?

Project 2

- Go over description

Finish Estimation

- LLN
- CLT
- Normal distribution
- Continuous probability distributions
- CLT finalized

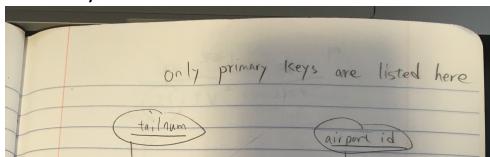
Hypothesis Testing



Midterm notes:

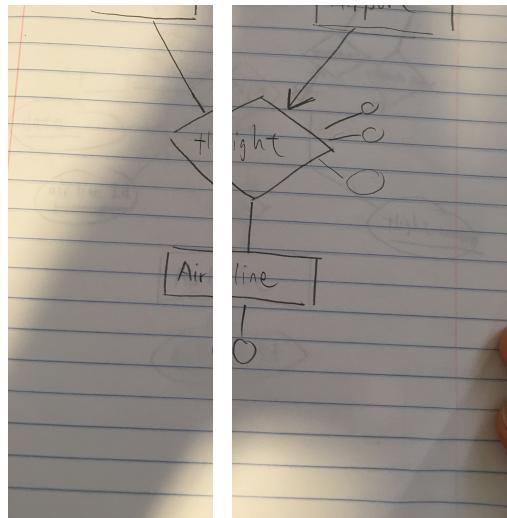
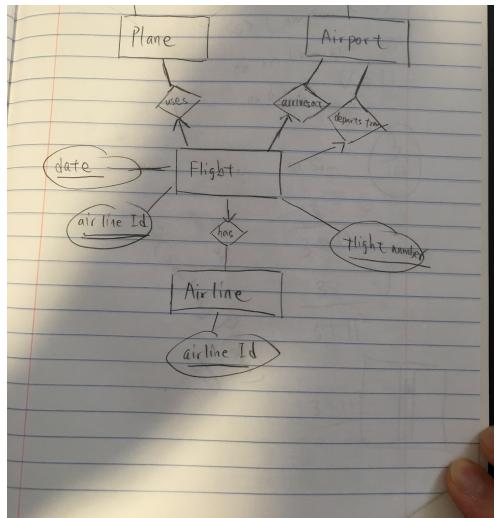
Q10:

Flights as entity



Flights as relationship





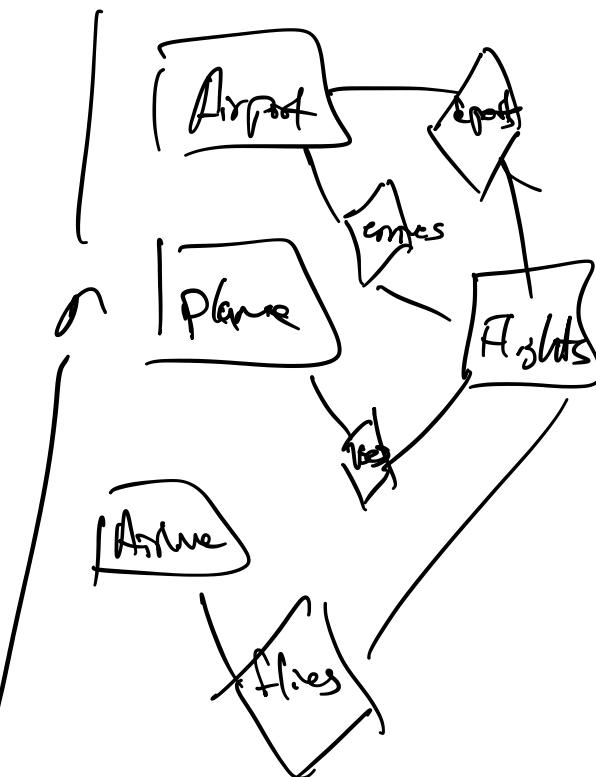
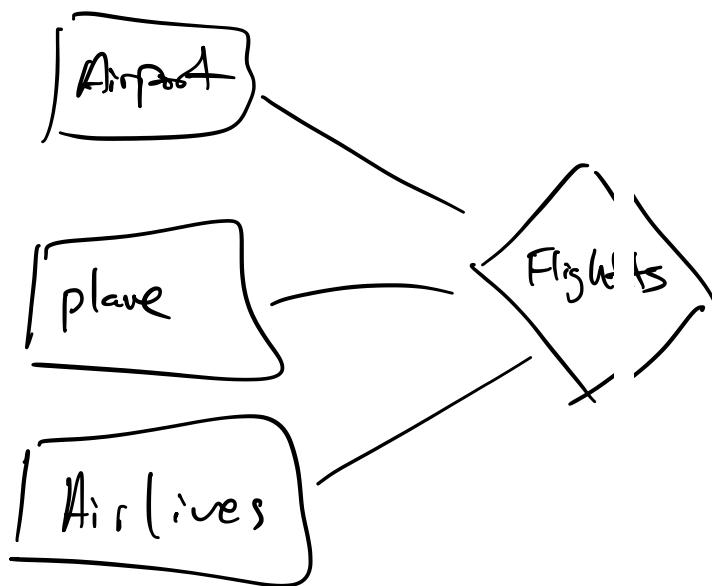
Q8

Airlines

		Airlines	
		Rating	Gross Number
Rating	>20		12
	≤ 20		4

Q10

Q10



Q12

teamID yearID DPW

E . sum |

select t.teamID, t.yearID, sum(s.salary) / w

from Teams as T join Salaries as S

on t.teamID = s.teamID and

t.yearID = s.yearID

group by t.teamID, t.yearID

where t.yearID >= 2000 and t.yearID <= 2015

Teams %>%

join (Salaries, by = c(teamID, yearID)) %>%

group_by (teamID, yearID) %>%

summarise (dpw = sum(salary) / w) %>%

filter (yearID >= 2000, yearID <= 2015)

Team % > %
filter (yearID > 2000, yearID ≤ 2015) % >%
join ---
groupby ---
summarize ---

Tweetbot example

- proportion bot-created tweets in some population

-

..

Random Variable

$$X \in \{0, 1\}$$

$$X_1, X_2, \dots, X_n$$

Estimate proportion of bot-generated tweets

$$E X = p$$

proportion
of bot-generated
tweets

Estimation procedure $\hat{p} \approx p$

Use the sample mean as estimate

$$\hat{p} = \frac{1}{n} \sum x_i$$

$\hat{P}_i \implies \bar{P} = \frac{1}{n} \sum_{i=1}^n p_i$

estimate
How good is that estimate?



Central Limit theorem

Given X_i there

$i = 1, \dots, n$

- independent
- identically distributed

$\bar{P} = \frac{1}{n} \sum_{i=1}^n p_i$ tends to

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

a normal distribution

with mean : $E\bar{X}$

variance : $\frac{1}{n} - \text{Var } \bar{X}$ ↗ natural, occurring variation

$$\text{Var } \bar{X} : E[(\bar{X} - E\bar{X})^2]$$

$$\sum_{\substack{\text{values} \\ X}} (X - E\bar{X})^2 P(X=x) =$$

= 2, r, 1, r, -1

$$(0 - \mathbb{E}X)^2 p(X=0) + (1 - \mathbb{E}X)^2 p(X=1) = r(1-p)$$

1