

CWRU DSCI351-451: Week06a Foundations of Inference

Roger H. French, JiQi Liu

07 October, 2018

Contents

6.1.1.1	Reading, Homeworks, Projects, SemProjects	1
6.1.1.2	Textbooks	1
6.1.1.3	Syllabus	1
6.1.1.4	Major Points for Distributions	3
6.1.1.4.1	Normal expectations	3
6.1.1.4.2	Skewness	3
6.1.1.4.3	Convenient measures for normal distributions	3
6.1.1.5	Next we'll see the following	3
6.1.1.5.1	Central Limit Theorem	3
6.1.1.5.2	Hypothesis Testing	3
6.1.1.5.3	Trials and Errors	3
6.1.1.6	Links	3

6.1.1.1 Reading, Homeworks, Projects, SemProjects

- Readings:
 - R4DS 7-8 Wrangle: Tibbles and readr for today
 - R4DS 9-16 More tidyverse Wrangling and then Programming for Thursday
- Homeworks
 -
- Data Science Projects:
 - Proj. 1 Due
- 451 SemProjects:
 -
- Friday Comm. Hour
 -

6.1.1.2 Textbooks

- [Peng: R Programming for Data Science](#)
- [Peng: Exploratory Data Analysis with R](#)
- [Open Intro Stats, v3](#)
- [Wickham: R for Data Science](#)
- [Hastie: Intro to Statistical Learning with R](#)

6.1.1.3 Syllabus

[Open Intro Stats, v3](#)

Day:Date	Foundation	Practicum	Reading	Due
w1a:Tu:8/28/18	ODS Tool Chain	R, Rstudio, Git		
w1b:Th:8/30/18	Setup ODS Tool Chain	Bash, Git, Twitter	PRP4-33	HW1
w2a:Tu:9/4/18	What is Data Science	OIS:Intro2R	PRP35-64	HW1 Due
w2b:Th:9/6/18	Data Analytic Style, Git	451SempProj, Git	PRP65-93, OI1-1.9	HW2
w3a:Tu:9/11/18*	Struct. of Data Analysis	ISLR:Intro2R, Loops	PRP94-116, OIS3	HW2 Due
w3b:Th:9/13/18*	OIS3 Intro to Data	GapMinder, Dplyr, Magrittr		
w4a:Tu:9/18/18	OIS3, Intro2Data part 2, Data	EDA: PET Degr.	EDA1-31	Proj1
w4b:Th:9/20/18	Hypothesis Testing	GGPlot2 Tutorial	EDA32-58	HW3
w5a:Tu:9/25/18	Distributions	SemProj RepOut1	R4DS1-3	HW3 Due
w5b:Th:9/27/18	Wickham DSCI in Tidyverse	SemProj RepOut1	R4DS4-6	SemProj1,
w6a:Tu:10/2/18	OIS Found. of Inference	Inference	R4DS7-8	Proj1 Due
w6b:Th:10/4/18		Midterm Review	R4DS9-16 Wrangle	
w7a:Tu:10/9/18*	Summ. Stats & Vis.	Data Wrangling		
w7b:Th:10/11/18*	MIDTERM EXAM			HW4
w8a:Tu:10/16/18	Numerical Inference	Tidy Check Explore	OIS4	HW4 Due
w8b:Th:10/18/18	Algorithms, Models	Pairwise Corr. Plots	OIS5.1-4	Proj 2, HW5
Tu:10/23	CWRU FALL BREAK		R4DS17-21 Program	
w9b:Th:10/25/18	Categorical Infer	Predictive Analytics	OIS6.1,2	
w10a:Tu:10/30/18	SemProj	SemProj	OIS7	SemProj2 HW5 Due
w10b:Th:11/1/18	Lin. Regr.	Lin. Regr.	OIS8	Proj.2 due
w11a:Tu:11/6/18	Inf. for Regression	Curse of Dim.	OIS8	Proj 3
w11b:Th:11/8/18	Model Accuracy	Training Testing	ISLR3	HW6
w12a:Tu:11/13/18	Multiple Regr.	Mul. Regr. & Pred.	ISLR4	HW6 due
w12b:Th:11/15/18	Classification		ISLR6	
w13a:Tu:11/20/18	Classification	Clustering	ISLR5	Proj 3 due
Th:11/22/18	THANKSGIVING			Proj 4
w14a:Tu:11/27/18	Big Data	Hadoop		
w14b:Th:11/29/18	InfoSec	VerisDB		SemProj3
w15a:Tu:12/4/18	SemProj Re- portOut3			
w15b:Th:12/6/18	SemProj Re- portOut3			Proj4
	FINAL EXAM	Monday12/17, 12:00-3:00pm	Olin 313	SemProj4 due

Figure 1: DSCI351-451 Syllabus

6.1.1.4 Major Points for Distributions

- Normal distribution is the basis of statistical expectations
- Geometric and Binomial Distributions are a form of expectations
- For two different way of posing questions
- Geometric: # of trials until success
- Binomial: P(given # of successes in given # of trials)

6.1.1.4.1 Normal expectations

- pnorm, gives us the expected probability of a given observed sample value
- for a given normal distribution

6.1.1.4.2 Skewness

- normal distribution is symmetrical
- if you have skewness (real data is “never” normal)
- check if a variable transformation can reduce skewness
- if so, then you statistical analysis will be better

6.1.1.4.3 Convenient measures for normal distributions

- normalize the mean and standard deviation
- using Z scores, so that you can cross-compare sample and population results
- and check your normal expectations against your data
- and
- All of these normal distribution concepts
- Are the foundation of statistical analysis
- And of defining statistical significance
- You’ll be using them in HWs, Projs. and SemProjs.

6.1.1.5 Next we’ll see the following

6.1.1.5.1 Central Limit Theorem

- -> With Standard Errors (SE)
- -> and Confidence Intervals

6.1.1.5.2 Hypothesis Testing

- -> test statistic
- -> p values

6.1.1.5.3 Trials and Errors

- -> Type I errors
- -> Type II errors

6.1.1.6 Links

Checkout the R documentation Project

- [R Doc Project](#)