CWRU DSCI351-451: Week06a Foundations of Inference

Roger H. French, JiQi Liu 29 September, 2018

Contents

6.1.1.1	Reading, Homeworks, Pr	ojects, SemProj	ects		 				. 1
6.1.1.2	Textbooks				 				. 1
6.1.1.3	Syllabus								
6.1.1.4	Major Points for Distribu	itions			 				. 2
6.3	.1.4.1 Normal expecta	tions			 				. 2
6.3	.1.4.2 Skewness				 				. 2
6.3	.1.4.3 Convenient mea	sures for norma	l distribu	tions	 				. 2
6.1.1.5	Next we'll see the following	ng			 				. 2
	.1.5.1 Central Limit T	_							
6.	.1.5.2 Hypothesis Test	ing			 				. 2
6.	.1.5.3 Trials and Error	·s			 				. 2
6.1.1.6	Cee-lo a good, no house	advantage game			 				. 4
6.1.1.7	Cee-lo dice game				 				. 4
	.1.7.1 Cee-lo without								
6.	.1.7.2 The combinatio	ns in Cee-lo			 				. 4
6.	.1.7.3 Probabilities[ed:	it]			 				. 5
6.	.1.7.4 dice, an R packs	age to calculate	dice game	es .	 				. 5
6.	.1.7.5 Rolling the Dice	on a Warm Ni	ght		 				. 5
6.1.1.8	Links				 				. 5

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

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 statitstical 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)
- \bullet -> and Confidence Intervals

6.1.1.5.2 Hypothesis Testing

- -> test statistic
- \rightarrow p values

6.1.1.5.3 Trials and Errors

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

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 SemProj w15a:Tu:12/4/18 SemProj ReportOut3 w15b:Th:12/6/18 SemProj ReportOut3 w15b:Th:12/6/18 SemProj ReportOut3 Proj4 FINAL EXAM Monday12/17, Olin 313 SemProj4 due	Day:Date	Foundation	Practicum	Reading	Due
Chain	w1a:Tu:8/28/18	ODS Tool Chain	R, Rstudio, Git		
w2b:Th:9/6/18	w1b:Th:8/30/18	-	Bash, Git, Twitter	PRP4-33	HW1
W3a:Tu:9/11/18*	w2a:Tu:9/4/18	ence	OIS:Intro2R		
Analysis	w2b:Th:9/6/18	Git	• •	PRP65-93, OI1-1.9	
w3b:Th:9/13/18* OIS3 Intro to Data GapMinder, Dplyr, Magrittr w4a:Tu:9/18/18 OIS3, Intro2Data EDA: PET Degr. EDA1-31 Proj1 part 2, Data 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 SemProj RepOut1 R4DS1-3 HW3 Due w5b:Th:9/27/18 OIS Found. of Inference R4DS7-8 Proj1 Due w6a:Tu:10/2/18 OIS Found. of 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 HW4 w7b:Th:10/11/18* MIDTERM EXAM HW4 Due w8b:Th:10/18/18 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 R4DS17-21 Program BREAK w9b:Th:10/25/18 Categorical Infer Predictive Analytics OIS6.1,2 w10a:Tu:10/30/18 SemProj SemProj OIS7 SemProj2 HW5 I 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.2 due w11a:Tu:11/3/18 Multiple Regr. Mul. Regr. & Pred. ISLR4 HW6 due w12b:Th:11/18/18 Model Accuracy Training Testing ISLR3 HW6 w12a: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 w15b:Th:12/6/18 SemProj ReportOut3 W15b:Th:12/6/18 SemProj ReportOut3 W15b:Th:12/6/18 SemProj ReportOut3 W15b:Th:12/6/18 Reproj.4 Radoup Reproj.4 R4DS1-217, Olin 313 SemProj4 due Radoup Radoup	w3a:Tu:9/11/18*	Analysis	Loops	PRP94-116, OIS3	HW2 Due
w4b:Th:9/20/18 hypothesis Testing GGPlot2 Tutorial EDA32-58 HW3 W5a:Th:9/25/18 Distributions SemProj RepOut1 R4DS1-3 HW3 Due w5b:Th:9/27/18 Wickham DSCI in Tidyverse SemProj RepOut1 R4DS4-6 SemProj1, Tidyverse W6a:Th:10/2/18 OIS Found. of 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 HW4 EXAM HW4 EXAM 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 R4DS17-21 Program BREAK w9b:Th:10/25/18 Categorical Infer Predictive Analytics OIS6.1,2 W10a:Tu:10/30/18 SemProj SemProj OIS7 SemProj2 HW5 I w10b:Th:11/4/18 Inf. for Regression Curse of Dim. OIS8 Proj.2 due w11a:Tu:11/6/18 Inf. for Regression Curse of Dim. OIS8 Proj.2 due w12a:Tu:11/13/18 Multiple Regr. Mul. Regr. & Pred. ISLR4 HW6 due w12b:Th:11/21/18 Classification Clustering ISLR5 Proj.3 due Th:11/22/18 Big Data Hadoop W14a:Tu:11/27/18 Big Data Hadoop W14a:Tu:12/4/18 SemProj ReportOut3 W15b:Th:12/6/18	w3b:Th:9/13/18*	OIS3 Intro to Data	GapMinder, Dplyr, Magrittr		
W5a:Tu:9/25/18	, ,	part 2, Data			
w5b:Th:9/27/18 Wickham DSCI in Tidyverse SemProj RepOut1 R4DS4-6 SemProj1,	w4b:Th:9/20/18	Hypothesis Testing	GGPlot2 Tutorial	EDA32-58	HW3
Tidyverse	/ /				HW3 Due
w6b:Th:10/4/18	, ,	Tidyverse	•		
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 BREAK W9b:Th:10/25/18 Categorical Infer Predictive Analytics OIS6.1,2 w10a:Tu:10/30/18 SemProj SemProj OIS7 SemProj2 HW5 I OIS8 Proj.2 due w11a:Tu:11/6/18 Inf. for Regression Curse of Dim. OIS8 Proj.2 due w11a:Tu:11/6/18 Model Accuracy Training Testing ISLR3 HW6 w12a:Tu:11/13/18 Multiple Regr. Mul. Regr. & Pred. ISLR4 HW6 due w13a:Tu:11/20/18 Classification Clustering ISLR5 Proj 3 due Th:11/22/18 Big Data Hadoop Hadoop W14a:Tu:12/27/18 Big Data Hadoop w15b:Th:12/6/18 SemProj ReportOut3 ReportOut3 <td>, ,</td> <td></td> <td></td> <td></td> <td>Proj1 Due</td>	, ,				Proj1 Due
W7b:Th:10/11/18* MIDTERM EXAM	w6b:Th:10/4/18		Midterm Review	R4DS9-16 Wrangle	
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	w7a:Tu:10/9/18*	Summ. Stats & Vis.	Data Wrangling		
W8b:Th:10/18/18 Algorithms, Models Pairwise Corr. Plots OIS5.1-4 Proj 2, HW5	w7b:Th:10/11/18*				HW4
Tu:10/23			Tidy Check Explore		
BREAK w9b:Th:10/25/18 Categorical Infer Predictive Analytics OIS6.1,2	w8b:Th:10/18/18	Algorithms, Models	Pairwise Corr. Plots	OIS5.1-4	Proj 2, HW5
w10a:Tu:10/30/18 SemProj SemProj OIS7 SemProj2 HW5 I 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 USLR6 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 ReportOut3 ReportOut3 Proj.4 Proj	,	BREAK			
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 ISLR6 w13a:Tu:11/20/18 Classification Clustering ISLR5 Proj 3 due Th:11/22/18 THANKSGIVING Proj 4 Proj 4 w14a:Tu:11/27/18 Big Data Hadoop Hadoop w15a:Tu:12/4/18 SemProj ReportOut3 w15b:Th:12/6/18 SemProj ReportOut3 FINAL EXAM Monday12/17, Olin 313 SemProj4 due	, ,	4.0	•	•	
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 ISLR6 w13a:Tu:11/20/18 Classification Clustering ISLR5 Proj 3 due Th:11/22/18 THANKSGIVING Proj 4 Proj 4 w14a:Tu:11/27/18 Big Data Hadoop Hadoop w14b:Th:11/29/18 InfoSec VerisDB SemProj3 w15a:Tu:12/4/18 SemProj ReportOut3 w15b:Th:12/6/18 SemProj ReportOut3 FINAL EXAM Monday12/17, Olin 313 SemProj4 due		5			SemProj2 HW5 Du
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 Proj 4 w14a:Tu:11/27/18 Big Data Hadoop SemProj 3 w15a:Tu:12/4/18 SemProj ReportOut3 Proj4 w15b:Th:12/6/18 SemProj ReportOut3 Proj4 FINAL EXAM Monday12/17, Olin 313 SemProj4 due		0			
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 Proj 4 w14a:Tu:11/27/18 Big Data Hadoop SemProj 3 w15b:Th:11/29/18 InfoSec VerisDB SemProj 3 w15a:Tu:12/4/18 SemProj ReportOut3 Proj4 w15b:Th:12/6/18 SemProj ReportOut3 Proj4 FINAL EXAM Monday12/17, Olin 313 SemProj4 due				1	
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 ReportOut3 w15b:Th:12/6/18 SemProj ReportOut3 FINAL EXAM Monday12/17, Olin 313 SemProj4 due					
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 w15a:Tu:12/4/18 SemProj ReportOut3 w15b:Th:12/6/18 SemProj ReportOut3 FINAL EXAM Monday12/17, Olin 313 SemProj4 due	, ,		Mul. Regr. & Pred.		HW6 due
Th:11/22/18 THANKSGIVING Proj 4		Classification		ISLR6	
w14a:Tu:11/27/18 Big Data Hadoop w14b:Th:11/29/18 InfoSec VerisDB w15a:Tu:12/4/18 SemProj ReportOut3 w15b:Th:12/6/18 SemProj ReportOut3 FINAL EXAM Monday12/17, Olin 313 SemProj4 due		Classification	Clustering	ISLR5	Proj 3 due
w14b:Th:11/29/18 InfoSec VerisDB SemProj3 w15a:Tu:12/4/18 SemProj ReportOut3 w15b:Th:12/6/18 SemProj ReportOut3 FINAL EXAM Monday12/17, Olin 313 SemProj4 due	Th:11/22/18	THANKSGIVING			Proj 4
w15a:Tu:12/4/18 SemProj ReportOut3	w14a:Tu:11/27/18	Big Data	Hadoop		
portOut3 Proj4			VerisDB		SemProj3
w15b:Th:12/6/18 SemProj ReportOut3 Proj4 FINAL EXAM Monday12/17, Olin 313 SemProj4 due	w15a:Tu:12/4/18				
portOut3 FINAL EXAM Monday12/17, Olin 313 SemProj4 due					
FINAL EXAM Monday12/17, Olin 313 SemProj4 due	w15b:Th:12/6/18				Proj4
		•			
12:00-3:00pm		FINAL EXAM	. , .	Olin 313	SemProj4 due

Figure 1: DSCI351-451 Syllabus

6.1.1.6 Cee-lo a good, no house advantage game

- Cee-lo Dice Game
- Cee-lo Probabilities
- Rules and probabilities in readings cee-lo.txt
- Inference (Predicting the Future)

6.1.1.7 Cee-lo dice game

6.1.1.7.1 Cee-lo without a bank (winner take all)

In this version of the game,

• each round involves two or more players of equal status.

A bet amount is agreed upon and

• each player puts that amount in the pile or pot.

Each player then has to roll all three dice at once and

• must continue until a recognized combination is rolled.

Whichever player rolls the best combination

• wins the entire pot, and a new round begins.

In cases where two or more players tie for the best combination,

• they must have a shoot out to determine a single winner.

6.1.1.7.2 The combinations in Cee-lo

The combinations are similar to those described above, and can be ranked from best to worst as:

- 4-5-6
 - The highest possible roll. If you roll 4-5-6, you automatically win.
- Trips
 - Rolling three of the same number is known as rolling "trips".
 - Higher trips beat lower trips,
 - so 4-4-4 is better than * 3-3-3.
 - Any trips beats any established point.
- Point
 - Rolling a pair, and another number,
 - establishes the singleton as a "point".
 - A higher point beats a lower point,
 - so 2-2-6 is better than 5-5-2.
- 1-2-3
 - The lowest possible roll.
 - If you roll 1-2-3, you automatically lose.
- Any other roll is a meaningless combination and
 - must be rerolled until one of the above combinations occurs.

6.1.1.7.3 Probabilities[edit]

- With three six-sided dice there are $6 \times 6 \times 6$ or 216 possible permutations.
 - -4-5-6:6/216 = 2.77777778% (Automatic Win)
 - Trips: 6/216 = 2.777777778%
 - Point: 90/216 = 41.66666667%
 - -1-2-3:6/216 = 2.777777778% (Automatic Loss)
 - Meaningless permutations: 108/216 = 50%

6.1.1.7.4 dice, an R package to calculate dice games

dice

6.1.1.7.5 Rolling the Dice on a Warm Night

- Human mystical thinking
- And beware the bank

6.1.1.8 Links

Checkout the R documentation Project

• R Doc Project