Avery Simon
April 28, 2020

Problems for Data Scientists: Results

1.2

(a) The mean of [a] for the data is 1.253901934365926
The standard error of [a] for the data is 0.40367612799661906

(b) The Pearson correlation coefficient between [b,c] is 0.6838747512215381
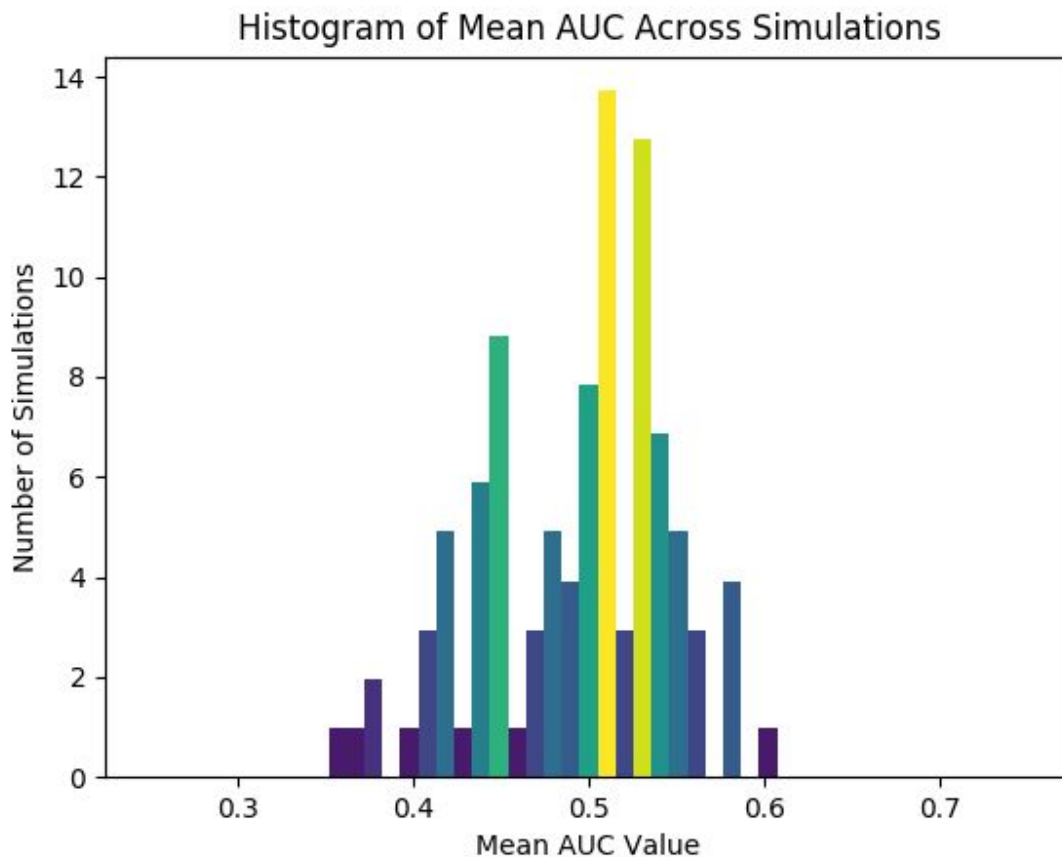The empirical p-value of these statistics:

[3.7135480807792836e-30, 1.434770386734372e-28, 3.396938434912199e-30,
9.29706857575344e-31, 1.2144933534646085e-27, 1.731702189978708e-32,
1.1578446156916315e-28, 1.5388239654929024e-24, 1.2833873405699404e-28,
5.279634601821537e-30, 5.840139014530645e-29, 2.699377631453385e-29,
1.010321895112508e-29, 3.233547576496417e-28, 2.709031498204324e-29,
3.8189460597935844e-28, 7.536824925398328e-30, 4.5839663743984427e-29,
7.7372120354633e-28, 5.804140562113273e-29, 7.049010460328536e-31,
1.9847646336441976e-29, 1.5356976953964692e-28, 5.245357679928484e-29,
2.4049099755018636e-27, 7.891077742004613e-29, 1.369405771091273e-32,
2.3326631403981205e-28, 1.6344041197833776e-30, 3.420151965049295e-27,
1.5912337873494734e-32, 2.8354240723374255e-29, 6.533250829409987e-28,
6.657246552143986e-29, 2.983284130823624e-35, 2.769481548124996e-28,
3.72738918443443e-28, 3.7643888581057656e-30, 8.058408529608733e-28,
7.896566877419778e-27, 3.952835217548001e-27, 6.535969157828926e-30,
1.6858166870386647e-32, 1.3505270353397622e-28, 4.7719053883883075e-32,
2.9677918137804065e-28, 1.3409950199250688e-32, 1.986960950817035e-30,
7.631656555680315e-28, 2.733858175616571e-29, 2.7747704159630492e-28,
5.164535723107687e-29, 2.1546509237905995e-31, 1.0396268319487932e-30,
3.38677103218811e-30, 1.3800228800230561e-30, 1.3773930016065121e-27,
1.4351824549051071e-28, 8.393483350739854e-28, 2.5450609498886023e-31,
2.908427532407331e-26, 1.8592640140492546e-29, 3.081396660098582e-31,
2.1992056929953722e-30, 2.12092926328376e-31, 2.221097817912625e-27,
3.368770624118774e-29, 4.399878259576048e-29, 4.9563442501383374e-30,
8.050263337136389e-28, 1.0730261882119253e-30, 1.3218911684535773e-29,
7.70646409944624e-28, 6.183914478526757e-27, 8.874718929460974e-28,
5.840806928018295e-28, 7.858586616089137e-30, 5.939252679690783e-29,
1.5558946755088276e-27, 2.56013177442756e-30, 6.803373292620724e-28,

2.7247650111466437e-31, 1.0979985536775665e-31, 5.04307032420344e-31,
1.8811095441105748e-29, 2.6157207987831315e-32, 4.610874258662116e-27,
5.375215003848629e-30, 3.133939294439346e-31, 9.022980390223426e-27,
3.711007520348434e-31, 2.19999715268027e-27, 9.560488431032578e-29,
2.2524076903118294e-28, 1.3200053679442158e-29, 1.9659402627430204e-25,
1.8776573485369664e-29, 7.879591012774375e-30, 6.638243532159533e-29,
2.1500351805519556e-30])

1.3

(a) The mean ROC-AUC across the 5 folds is 0.5266340167851371
(b)

Histogram of Mean AUC Across Simulations

(c) The mean AUC of this distribution is 0.4927221526462857
(d) The standard deviation of this distribution is 0.05210670748504408