

GRADING EFFICIENCY BASED ON GAMMA DISTRIBUTION WITH SAS PROCEDURES

Dagan Zhang, Ph D
Program Administrator, Texas Workers' Compensation Commission

Abstract

Grading efficiency, accuracy, or quality of a same kind of jobs done by a group of people individually is necessary to **raise administration level**, e.g. cure efficiency of physicians, data accuracy of entries, judgment fairness of judges, etc. An illustrative example with faked data is given in this paper.

Let NTC be number of treatments/services provided by a DC (doctor of chiropractic) for a claimant (injured worker), ANTCD average of NTC for a DC, an index measuring cure efficiency of DCs. Many DCs service for only a few of claimants, i.e. many ANTCD values are means of small samples of the population constructed by NTC values. SAS proc corr, univariate, and chart are used. 1. We conclude that NTC is gamma distributed. 2. **A gamma test technology has been developed.** 3. With Chebyshev's Inequality for simplifying, the technology is utilized to grade cure efficiency of DCs. (Based on the gotten grades, of course, we must further analyze specialties of claimants cared by the DCs particularly concerned by us to avoid wrong comments on them.)

This paper would be interesting, particularly to those working for administration, and possessing knowledge in a college statistics course. (for any OS with SAS)

Keywords: gamma test for grading efficiency.

Working Draft

Introduction

To administrate medical cost for workers' compensation, a basic, intuitive measurement for cure efficiency of DCs (doctors of chiropractic) should be found out. Let NTC be number of treatments/services provided by a DC (doctor of chiropractic) for a claimant (injured worker), ANTCD average of NTC for a DC. ANTCD is a good index measuring cure efficiency of DCs not only by tuition, but also, using SAS proc CORR, we conclude that ANTCD is very closely relative to medical payment. (See attached Table 1 NTC, Table 2 ANTCD, and Program 1.)

To grade cure efficiency of DCs with ANTCD, distribution of NTC is analyzed with SAS proc univariate and proc chart in section I bellow. We conclude that NTC has a gamma distribution. (See attached Chart 1 and Program 2.)

Then a gamma test is developed in section II of this paper. To simplify calculation, how to utilize Chebyshev's inequality to calculate confidence intervals is showed in section III below.

Finally, grades of cure efficiency of DCs are gotten by program 3 based upon the skills developed in this paper.

This paper would show that statistical common sense is not enough even to an administrative agency. To raise administration level, SAS would not be replaced by another lower computer language.

I. NTC is Gamma Distributed

To grade ANTCD, we regard all values of NTC as a population, a value of ANTCD as mean of a sample of population NTC. Since numbers of claimants cared by different DCs are different generally, the sample sizes are not same. Some samples even are very small.

Using SAS proc chart we conclude that NTC is gamma distributed. See attached Program 2 and Chart 1.

Parameters of the distribution function $\Gamma(\alpha, \beta)$ can be deduced as follows. Since

$$\alpha/\beta = \mu = E(NTC) = 80.51, \text{ where } E(NTC) \text{ is expectation of NTC,}$$

$$\alpha/\beta^2 = \sigma^2 = V(NTC) = 138.04^2, \text{ where } V(NTC) \text{ is variance of NTC,}$$

we get

$$\alpha = 0.34,$$

$$\beta = 0.004225.$$

II. A Gamma Test

To grade ANTCD, the following lemma is produced.

Lemma. Suppose X , and $X_j, j=1, 2, \dots, n$, are $n+1$ identical independent random variables with distribution function $\Gamma(\alpha, \beta)$. Let

$$Y = (1/n)\sum X_j, \text{ where } \sum X_j \text{ is sum of } X_1, X_2, \dots, X_n.$$

Then Y is a gamma distributed variable with distribution function $\Gamma(n\alpha, n\beta)$.

Proof. At first, $\sum X_j$ is gamma distributed with gamma function $\Gamma(n\alpha, \beta)$ since theorem 7.3.2 or 8.4.4, [2]. On the second, because cX has the gamma distribution with parameter α and β/c for $c > 0$ (example 7.1.5, [2]), we get Y has the gamma distribution with parameters $n\alpha$, and $n\beta$. ♦

As a corollary of this lemma, the mean, Y , of all samples, X_j , with same size n of population NTC constructs a population Ω with distribution function $\Gamma(80.51, 138.04^2/n)$. Obviously, expectation $E(Y) = E(X) = 80.51$, and variance $V(Y) = (1/n)V(X) = 138.04^2/n$.

A value of ANTCD with n claimants is an item, or individual, or element of this population Ω . To grade ANTCD, a 95% confidence interval (a, b) for a DC with n cared claimants is used. The grade is G, or N, or P if the value of ANTCD $\leq a$, or locates between a and b , or $\geq b$, correspondingly.

III. Calculating Confidence Intervals

As we know, it is very complicated to calculate values of gamma functions. To calculate confidence intervals, hence, Chebyshev's Inequality

$$P(|Y - E(Y)| \geq k) \leq V(Y)/k^2$$

is utilized. Fortunately, the loosened intervals since the approximation is "better" to our grading: not too severe to grade P!

Let $V(Y)/k^2 = 0.95$. Then we get $k = 617.33/\sqrt{n}$. The confidence interval for Ω constructed from means of all samples with same size n of population NTC is $(80.51 - 617.33/\sqrt{n}, 80.51 + 617.33/\sqrt{n})$.

IV. Grading ANTCD

Grades of cure efficiency of DCs are calculated with the technologies developed in sections II and III above. See attached Program 3 and Table 3 Grades of ANTCD.

Attached Tables, Programs and Corresponding Outputs

Table 1 NTC
Number of Treatments/Services for a Claimant by a DC

Claim ID	Doctor License	Number of treatments/services
2000001	C11112	900
21000003	C00010	5
.....		
21000003	C00096	2
.....		
22039876	C21105	71
220111111	C01019	3

Note 1: If a claimant saw m DCs, then the claim ID appear m times in this table.

Note 2: Number of data rows in this table is 35,976.

Table 2 ANTCD
Averages of NTC for DCs

Dr License	Number, Claims	Aver., Pay per Claim	Aver., Charge per Claim	Pay/Charge	ANTCD
C11112	1	\$25000	\$36230	69%	668
C00010	2	\$8300	\$24000	34.58%	571.50
C21105	95	\$1300	\$173.33	75%	11.74
.....					

```

Program 1 Corr
PROC CORR DATA=SASUSER.TABLE2;
  VAR APC ACHC PAYOCHG ANTC;
  WEIGHT NCLM;
RUN;

```

The output is as follows.

The CORR Procedure

4 Variables: APC ACHC PAYOCHG ANTC
Weight Variable: NCLM

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
APC	4416	2416	6132	86917532	0	38110
ACHC	4416	3711	9781	133522038	0	68511
PAYOCHG	4413	0.69020	0.69216	24828	0	19.15152
ANTC	4413	80.31343	216.500	2889115	0.83333	908.00000

Pearson Correlation Coefficients

Prob > |r| under H0: Rho=0

Number of Observations

	APC	ACHC	PAYOCHG	ANTC
APC	1.00000	0.93817	-0.01772	0.85663
		<.0001	0.2391	<.0001
	4416	4416	4413	4413
ACHC	0.93817	1.00000	-0.17542	0.91483
	<.0001		<.0001	<.0001
	4416	4416	4413	4413
PAYOCHG	-0.01772	-0.17542	1.00000	-0.15731
	0.2391	<.0001		<.0001
	4413	4413	4413	4413
ANTC	0.85663	0.91483	-0.15731	1.00000
	<.0001	<.0001	<.0001	
	4413	4413	4413	4413

```

Program 2 UnivChar
PROC UNIVARIATE DATA=SASUSER.TABLE1;
  VAR NTC;
  OUTPUT OUT=STAT0 MEAN=ANTC STDMEAN=STDDEV MIN=MINNTC MAX=MAXNTC;
PROC CHART DATA=SASUSER.TABLE1;
  VBAR NTC /TYPE=FREQ MIDPOINTS= 20 to 2220 by 20;
RUN;

```

The output is as follows.

The UNIVARIATE Procedure
Variable: NTC

Moments			
N	35883	Sum Weights	35883
Mean	80.5148678	Sum Observations	2889115
Std Deviation	138.041873	Variance	19055.5586
Skewness	3.7792587	Kurtosis	24.997576
Uncorrected SS	916368265	Corrected SS	683751553
Coeff Variation	171.448922	Std Error Mean	0.7287297

Basic Statistical Measures			
Location Variability			
Mean	80.51487	Std Deviation	138.04187
Median	20.00000	Variance	19056
Mode	1.00000	Range	2466
Interquartile Range		103.00000	

Tests for Location: $\mu_0=0$

Test	-Statistic-	p Value
Student's t	t 110.4866	Pr > t <.0001
Sign	M 17941.5	Pr >= M <.0001
Signed Rank	S 3.2191E8	Pr >= S <.0001

Quantiles (Definition 5)

Quantile	Estimate
100% Max	2467
99%	641
95%	343
90%	233
75% Q3	105
50% Median	20
25% Q1	2
10%	1
5%	1
1%	1
0% Min	1

The UNIVARIATE Procedure
Variable: NTC

Extreme Observations			
---Lowest---		---Highest---	
Value	Obs	Value	Obs
1	35879	2006	1593
1	35874	2017	34703
1	35872	2040	4176
1	35863	2229	26484
1	35854	2467	8710

Chart 1 Chart of NTC
Density of NTC (main piece)

NTC Midpoint		Freq	Cum. Freq	Percent	Cum. Percent
20	*****	19531	19531	54.43	54.43
40	*****	2649	22180	7.38	61.81
60	*****	2044	24224	5.70	67.51
80	***	1678	25902	4.68	72.18
100	***	1333	27235	3.71	75.90
120	**	1175	28410	3.27	79.17
140	**	1007	29417	2.81	81.98
160	**	898	30315	2.50	84.48
180	*	693	31008	1.93	86.41
200	*	658	31666	1.83	88.25
220	*	550	32216	1.53	89.78
240	*	468	32684	1.30	91.08
260	*	360	33044	1.00	92.09
280	*	344	33388	0.96	93.05
300	*	325	33713	0.91	93.95
320		242	33955	0.67	94.63
340		213	34168	0.59	95.22
360		210	34378	0.59	95.81
380		174	34552	0.48	96.29
400		146	34698	0.41	96.70
420		140	34838	0.39	97.09
440		103	34941	0.29	97.37
460		100	35041	0.28	97.65
480		77	35118	0.21	97.87

Program 3 Grading

```

DATA W;
  SET SASUSER.TABLE2;
  LEND = 80.31- 617.33/SQRT(NCLM);
  REND =80.31+ 617.33/SQRT(NCLM);
  RLEND = ROUND(LEND,1);
  RREND = ROUND(REND,1);
  GRADE='N';
  IF ANTCD LE RLEND THEN GRADE='G';
  IF ANTCD GE RREND THEN GRADE='P';
  DROP LEND REND;
PROC SORT;
  BY DESCENDING ANTCD;
PROC REPORT DATA=W SPLIT="*" CENTER HEADSKIP;
  COLUMN PHYSLIC NCLM APC ACHC PAYOCHAR ANTCD GRADE;
  DEFINE PHYSLIC/DISPLAY FORMAT=$8. WIDTH=8 "Dr*License";
  DEFINE NCLM/ DISPLAY FORMAT=4. WIDTH=5 "Number,*Claims";
  DEFINE APC/DISPLAY FORMAT=DOLLAR9.2 WIDTH=9 "Aver., Pay*per Claim";
  DEFINE ACHC/DISPLAY FORMAT=DOLLAR9.2 WIDTH=12 "Aver., Charge*per Claim";
  DEFINE PAYOCHAR/DISPLAY FORMAT=PERCENT=9.2 WIDTH=10 "Pay/Charge";
  DEFINE ANTCD/DISPLAY FORMAT=COMMA9.0 WIDTH=7 "ANTCD";
  DEFINE GRADE/DISPLAY FORMAT=$1. WIDTH=6 "GRADE";
RUN;

```

The part of output is listed as follows.

Table 3 Grades of Cure Efficiency of DCs

Dr License	Number, Claims	Aver., Pay per Claim	Aver., Charge per Claim	Pay/Charge	ANTCD	Grade
C11112	1	\$25000	\$36230	69%	668	N
C00010	2	\$8300	\$24000	34.58%	571.50	P
C21105	95	\$1300	\$1733.33	75%	11.74	G
.....						...

References

- [1] George W. Snedecor, William G. Cochran, "Statistical Methods," eighth edition, Iowa State University Press /Ames, 1995.
- [2] Michael Woodroffe, "Probability with Application," MCGRAW-HILL BOOK COMPANY, New York, 19XX.

