Confidence Intervals in Public Health

When public health practitioners use health statistics, sometimes they are interested in the actual number of health events, but more often they use the statistics to assess the true underlying risk of a health problem in the community. Observed health statistics, that is, those counts, rates or percentages that are computed or estimated from health surveys, vital statistics registries, or other health surveillance systems, are not always an accurate reflection of the true underlying risk in the population. Observed rates can vary from sample to sample or year to year, even when the true underlying risk remains the same.

Statistics based on samples of a population are subject to sampling error. Sampling error refers to random variation that occurs because only a subset of the entire population is sampled and used to estimate a finding for the entire population. It is often mis-termed "margin of error" in popular use. Even health events that are based on a complete count of an entire population, such as deaths, are subject to random variation because the number of events that occurred may be considered as one of a large series of possible results that could have arisen under the same circumstances. In general, sampling error or random variation gets larger when the sample, population or number of events is small.

Statistical sampling theory is used to compute a confidence interval to provide an estimate of the potential discrepancy between the true population parameters and observed rates. Understanding the potential size of that discrepancy can provide information about how to interpret the observed statistic. For instance, if the state infant death rate of 5.94 increased to 6.03 in a one-year period, is that increase something that should cause concern? If the smoking rate among teens decreased from 13% to 8%, is that cause for celebration?

Technically speaking, the 95% confidence interval indicates the range of values within which the statistic would fall 95% of the time if the researcher were to calculate the statistic (e.g., a percentage or rate) from an infinite number of samples of the same size, drawn from the same population. In less technical language, the confidence interval is a range of values within which the "true" value of the rate is expected to occur (with 95% probability). This document describes the most common methods for calculation of 95% confidence intervals for some rates and estimates commonly used in public health.

95% Confidence Interval for a Percentage From a Survey Sample:

To calculate a confidence interval for a percentage from a survey sample, one must first calculate the *standard error* of the percentage. A percentage is also known as the mean of a binomial distribution. The standard error of the mean is a measure of dispersion for the hypothetical distribution of means called the *sampling distribution of the mean*. This is a distribution of means calculated from an infinite number of samples of the same size drawn from the same population as the original sample.

Once you have calculated the standard error of the percentage, you must decide how large you want the confidence interval to be. The most common alternative is a 95% confidence interval. This is the width of the interval that includes the mean (the *sampling distribution of the mean*,

mentioned above) 95% of the time. In a little plainer language, a 95% confidence interval for a percentage is the range of values within which the percentage will be found at least 95% of the time if you went back and got a different sample of the same size from the same population.

Transforming the standard error into a 95% confidence interval is rather simple. Fortunately, the *sampling distribution of the mean* has a shape that is almost identical to what is known as the *normal distribution*.[1] You need only multiply the standard error by the $\underline{Z}$-score of the points in the normal distribution that exclude 2.5% of the distribution on either end (two-tailed). That $\underline{Z}$-score is 1.96. A $\underline{Z}$-score of 1.96 defines the 95% confidence interval. A $\underline{Z}$-score of 1.65 defines a 90% confidence interval.
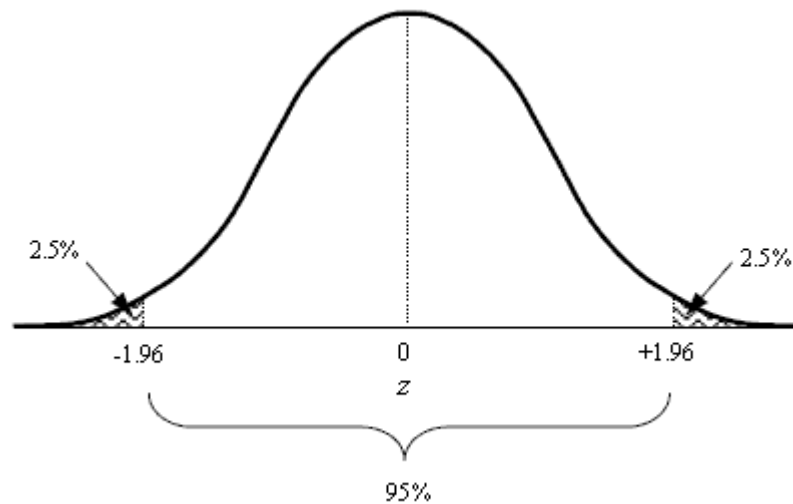
For a simple random sample, the standard error = $\sqrt{pq/n}$

where: p is the rate,
      q is 1 minus the rate, and
      n is the sample size.

Example: 13% of surveyed respondents indicated that they smoked cigarettes. The sample consisted of 500 persons in a simple random sample.

The standard error $= \sqrt{(0.13*0.87)/500} = .015$

---

[1] A *distribution* is a tool that is used in statistics to associate a statistic (e.g., a percentage, average, or other statistic) with its probability. When researchers talk about a measure being "statistically significant," they have used a *distribution* to evaluate the probability of the statistic, and found that it would be improbable under ordinary conditions. In most cases, we can rely on measures such as rates, averages, and proportions as having an underlying *normal distribution*, at least when the sample size is large enough.

Then the 95% confidence interval is:

$$.13 \pm 1.96 * \text{standard error} = .13 \pm 1.96 * .015 = .13 \pm .0294 = .101 - .159$$

So the 95% confidence interval has a lower limit of 10.1% and an upper limit of 15.9%

The formula used above applies to a binomial distribution, which is the distribution of two complimentary values (e.g., heads and tails, for and against). If you are calculating a confidence interval for a different statistic, such as an average, you'll need to modify the equation. The quantity (pq) is the variance of a binomial distribution. If your measure is not a proportion, but, say, an average, you must modify the formula, substituting the pq quantity with the variance. The standard error can also be calculated as the standard deviation divided by the square root of the sample size:
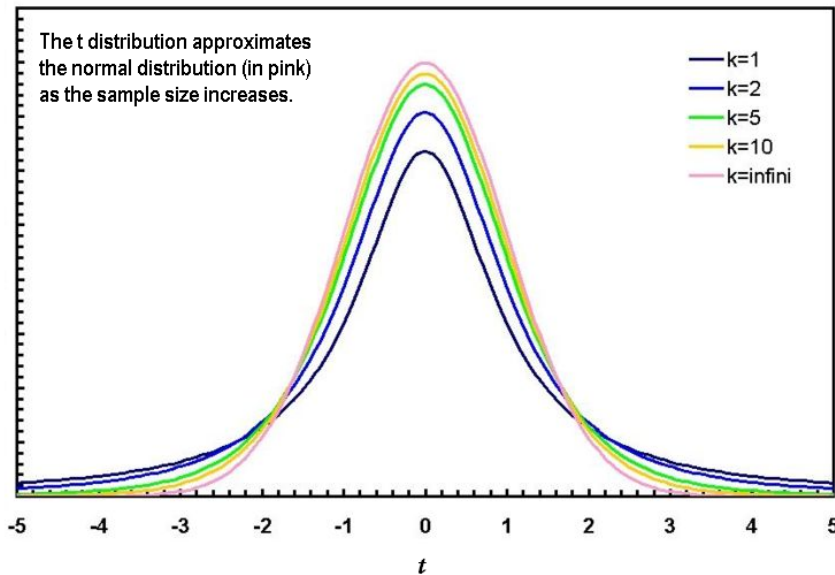
$$\frac{\sigma}{\sqrt{n}}$$

Small Samples

If the sample from which the percentage was calculated was rather small (according to central limit theorem we can define small as 29 or fewer) then the shape of the sampling distribution of the mean is not the same as the shape of the normal distribution. In this special case, we can use another distribution, known as the t distribution, that has a slightly different shape than the normal distribution.[2]

The procedures in this case are analogous to those above but the t-score comes from a family of distributions that depend on the "degrees of freedom." The number of degrees of freedom is defined as "n-1" where "n" is the size of the sample. For a sample of size=30 the degrees of freedom is equal to 29. So, for a 95% confidence interval, you must use the t-score associated with 29 degrees of freedom. That particular t-score is 2.045 (see Appendix 1.). So you would multiply the standard error by 2.045 instead of 1.96 to generate the 95% confidence interval.

If our sample were a different size, say 20, then the degrees of freedom would be 19, which is associated with a t-score of 2.093 for a 95% confidence interval. As you see the interval will get wider as our sample size is reduced. This reflects the uncertainty in our estimate of the variance in the population. For a 95% confidence interval with 9 degrees of freedom the t-score is 2.262. Table 1. lists the t-scores for specific degrees of freedom and sizes of confidence interval. For a 95% confidence interval, you would use the t-score that defines the points on the distribution that excludes the most extreme 5% of the distribution, which is 0.025 on either end of the curve.

---

[2] Student's t-distribution, downloaded on 2/13/09 from http://en.wikipedia.org/wiki/File:Student_densite_best.JPG

Student's *t* Distribution at Varying Degrees of Freedom (k)

The t distribution approximates the normal distribution (in pink) as the sample size increases.

k=1
k=2
k=5
k=10
k=infini

Appendix 1 lists the t-scores for specific degrees of freedom and sizes of confidence interval. For a 95% confidence interval, you would use the t-score that defines the points on the distribution that excludes the most extreme 5% of the distribution, which is 0.025 on either end of the distribution.

Finite Populations

If the survey sampled all or most of the members of the population, then using the finite population correction factor will improve (decrease the width of) the confidence interval.

The finite population correction factor = 1-*f*, where *f* is the sampling fraction

f = n/N where n is the size of the sample and N is the size of the population. The sampling fraction is simply the proportion of the population that was included in the sample.

The standard error of the mean for a binomial distribution for a finite sample

$$s.e._{percentage} = \sqrt{pq/n * (1-f)}$$

When the Percentage is Close to 0% or 100%

When the percentage is close to 0% or 100%, the formulas given above can result in illogical results - confidence limits that fall below 0% or above 100%. A special formula is used to calculate asymmetric confidence limits in these cases. Because survey estimates can be small percentages, the confidence intervals for the surveys on IBIS-Q are based on logit transformations. Logit transformations yield asymmetric interval boundaries that are more balanced with respect to

the probability that the true value falls below or above the interval boundaries than is the case for standard symmetric confidence intervals for small proportions.

The method used is as follows:
(1) Perform a logit transformation of the original percentage estimate:

      $f = \log(p) - \log(1-p)$
      where:
      p = the percentage estimate
      f = the logit transformation of the percentage

(2) Transform the standard error of the percentage to a standard error of the it's logit transformation:

      $se(f) = se(p)/(p*(1-p))$
      where:
      se = standard error

(3) Calculate the lower and upper confidence bounds of the logit transformation of the percentage:

      $Lf = f - t(alpha/2, df)*se$
      $Uf = f + t(alpha/2, df)*se$
      where:
      Lf = lower confidence bound of f
      Uf = upper confidence bound of f
      $t(alpha/2, df)$ = the value of the t-score corresponding to the desired alpha level (0.05 for a 95% Confidence Interval) and the degrees of freedom (degrees of freedom is defined as "n-1" where "n" is the sample size).

(4) Finally, perform inverse logit transformations to get the confidence bounds of p:

      $Lp = \exp(Lf)/(1+\exp(Lf))$
      $Up = \exp(Uf)/(1+\exp(Uf))$
      where:
      Lp = lower confidence bound of p
      Up = upper confidence bound of p

Complex Sample Designs

The above formulas assume that the survey sample was a *simple random sample*. If the survey used a complex sample design (such as clustering within households or disproportionate sampling from various geographic regions), special techniques must be used to calculate the standard error of the mean. Those techniques are accomplished using statistical software such as SAS®, STATA®, or SUDAAN®.

When the Rate is Equal to 0

When the percentage or rate is equal to zero, using the above calculation will yield a confidence interval of zero, which is incorrect. A simple method you can use to estimate the

confidence interval when the percentage or rate is zero is to assume the number of cases in the numerator of your rate is "3," then calculate the confidence interval using the population size in your original calculation.[3]

95% Confidence Intervals for Rare Events:

In the case of rare events distributed randomly across time, the normal distribution no longer applies. A different distribution, the Poisson distribution is used to model rare events that occur across time, such as the "100 year flood."[4] It is used to calculate confidence intervals for rare health events, such as infant mortality or cancer. This distribution is not symmetric about its mean and so the associated confidence intervals will not be symmetric (the upper limit is farther from the estimate than is the lower limit).

The Poisson distribution does, however, assume the shape of a normal distribution when there are 20 or more events in the numerator. So we use a Poisson distribution for rare events (when the number of events is less than 20), but we can use the normal distribution when the number of events is 20 or more.

Poisson Distribution
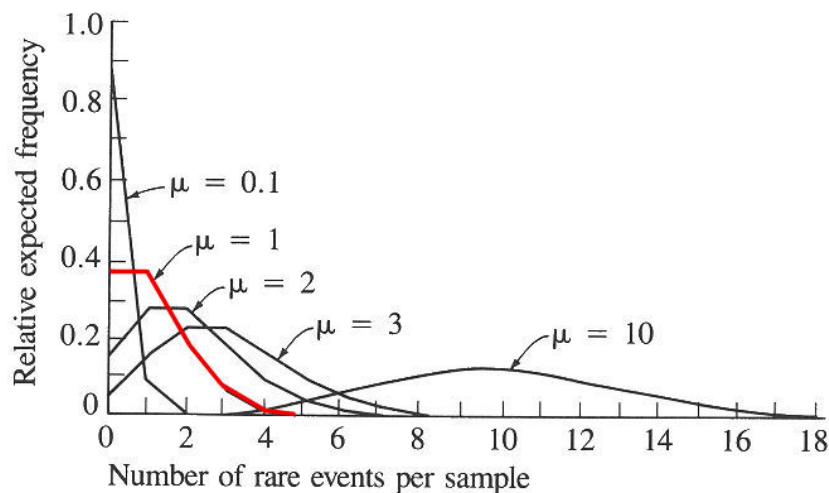
In Appendix 2 you will find lower and upper confidence factors for use in calculating a 95% confidence interval for a rate based on a specified number of events, from 1 to 20. To calculate the confidence interval, multiply the estimated rate by the confidence factor associated with the number of events on which the rate is based.

---

[3] Lilienfeld, DE and Stolley, PD Foundations of Epidemiology (3rd Ed.). Oxford University Press, 1994.

[4] "In probability theory and statistics, the Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate and independently of the time since the last event." Poisson Distribution, downloaded on 2/13/09 from http://en.wikipedia.org/wiki/Poisson_distribution.

For example, in a given geographic area, there were 722 births in a single year, and seven infant deaths. The infant mortality rate in was 9.7 per 1,000 live births, calculated as [(7/722)*1,000]. The lower and upper confidence limits are calculated using the confidence factors found on Appendix 2. The factors for seven events are .4021 and 2.0604 for the lower and upper limits of the confidence interval, respectively. The lower limit of the confidence interval = 9.7*.4021 = 3.90, and the upper limit = 9.7*2.0604 = 19.99, for a rate of 9.7 and a 95% confidence interval from 3.90 to 19.99. If this same rate had been based on 100 deaths then the confidence factors would be .8136 and 1.2163. The lower limit would be 9.7*.8136, and the upper limit 9.7*1.2163 for an estimate of 9.7 with a confidence interval from 7.89 to 11.80. This interval is much smaller due to the greater number of deaths on which the rate is based.

In the Utah IBIS-PH query system, starting in March 2010, the confidence factors are obtained by using SAS® software that requires specification of the percentage of the inverse gamma distribution to be excluded on either end of the distribution (2.5% for a 95% confidence interval), and the two parameters associated with the distribution function: the mean and the variance. In the case of the crude rate, where the variance and mean are equal, this is the special case of the gamma family of distributions known as the Poisson distribution.

Directly Age-Adjusted Rates

When comparing across geographic areas, some method of age adjusting is typically used to control for area-to-area differences in health events that can be explained by differing ages of the area populations. For example, an area that has an older population will have higher crude (not age-adjusted) rates for cancer, even though its exposure levels and cancer rates for specific age groups are the same as those of other areas. One might incorrectly attribute the high cancer rates to some characteristic of the area other than age. Age-adjusted rates control for age effects, allowing better comparability of rates across areas. Direct standardization adjusts the age-specific rates observed in the small area to the age distribution of a standard population (Lilienfeld & Stolley, 1994)[5].

The directly age-adjusted death rate is a weighted average of the age-specific death rates where the age-specific weights represent the relative age distribution of the standard population.

Directly age-adjusted death rate (DAADR) = $\sum W_{si} * D_i / P_i = \sum W_{si} * R_i$

Where.....
W_{si} = the weight for the ith age group in the standard population (the proportion of the standard population in the ith age group) = $P_{si} / \sum P_{si}$
P_{si} = the population in age group i in the standard population
D_i = number of deaths (or other event) in age group i of the study population
P_i = the population in age group i in the study population
R_i = the age-specific rate in the ith age group

---

[5] Lilienfeld, DE and Stolley, PD (1994)

Using the properties of the Poisson distribution, the variance of the age-specific death rate is given by

$$\text{var}(R_i) = \text{var}(D_i / P_i) \;=\; 1/P_i^{\,2} \; \text{var}(D_i) = D_i / P_i^{\,2} = R_i^{\,2}/D_i$$

The variance of a directly age-adjusted death rate can then be computed as follows:

$$\text{var}(DAADR) = \sum W_{si}^{\,2} * \text{var}(R_i) \;=\; \sum W_{si}^{\,2} * R_i^{\,2}/D_i$$
$$\text{se}(DAADR) = \text{square root } (\text{var}(DAADR))$$

Where…..
var(DAADR) = the variance of the directly standardized rate
W_{si} = the weight for the ith age group in the standard population
R_i = the age-specific rate in the ith age group
var (R_i) = the variance of the age-specific death rate in the ith age group of the
      study population = $R_i^{\,2}/D_i$
D_i = number of deaths (or other event) in age group i of the study population
se(DAADR) = standard error of the directly standardized rate

The age-adjusted death rate is a linear combination of independent Poisson random variables and therefore is not a Poisson random variable itself. It can be placed in the more general family of gamma distributions of which the Poisson is a member. Statistical packages such as SAS have a function to calculate factors that may be applied to age-adjusted death rates to calculate 95-percent confidence intervals. These factors are derived from a standard gamma distribution.

In the IBIS-PH query system, starting in March 2010, the confidence intervals are obtained by using SAS® software that requires specification of the percentage of the distribution to be excluded on either end of the distribution (2.5% for a 95% confidence interval), and the two parameters associated with the distribution function: the mean and the variance. In the case of the age-adjusted rate, the variance and mean are not equal and this fact is accounted for in the calculation of the confidence factors used to compute the gamma confidence intervals for age-adjusted rates. [6] Gamma intervals perform well even when the number in any specific age-adjustment age group cell is small. [7]

Indirectly Age-Adjusted Rates

The direct method can present problems when population sizes are particularly small. Calculating directly standardized rates requires calculating age-specific rates, and for small areas these age-specific rates may be based on one or two events. The general rule of thumb is that if

[6] Anderson RN, Rosenberg HM, Age Standardization of Death Rates: Implementation of the Year 2000 Standard. National vital statistics reports; vol 47 no.3. Hyattsville, Maryland: National Center for Health Statistics. 1998
[7] Fay MP, Feuer EF, Confidence Intervals for Directly Standardized Rates: A Method based on the Gamma Distribution. Statistics in medicine, vol 16, 791-801 (1997)

there are fewer than 20 (some say 25) cases in the index population, indirect standardization of rates should be used.

Indirectly standardized rates are based on the Standardized Mortality Ratio (SMR) and the crude rate for a standard population. Indirect standardization adjusts the overall standard population rate to the age distribution of the small area (Lilienfeld & Stolley, 1994)[8]. Strictly speaking, it is valid to compare indirectly standardized rates only with the rate in the standard population, not with each other.[9]

An indirectly standardized death or disease rate (ISR) can be computed as:

$$ISR = SMR*Rs$$

$$SMR = \frac{\text{observed deaths/disease in the small area}}{\text{expected deaths/disease in the small area}} = \frac{D}{e} = \frac{D}{\sum (R_{si} * n_i)}$$

Where...

SMR = observed deaths in the small area/expected deaths in the small area

D = observed number of deaths in the small area

$e = \sum (R_{si} * n_i)$ = expected number of deaths in small area

R_s = the crude death rate in the standard population

R_{si} = the age-specific death rate in age group i of the standard population (# deaths / population count, before applying the constant)

n_i = the population count in age group i of the small area

For indirectly standardized rates based on events that follow a Poisson distribution and for which the ratio of events to total population is small (<.3) and the sample size is large, the following two methods can be used to calculate confidence interval (Kahn & Sempos, 1989)[10].

(1) When the number of events $\geq$20:

$$CI_{ISR} = \pm 1.96 \ \sqrt{(SMR/e)} * R_s * K$$

[8] Lilienfeld & Stolley (1994)
[9] Rothman, Kenneth J. and Greenland, Sander (1998) Modern Epidemiology (2nd Ed.). Philadelphia, PA: Lippincott.
[10] Harold A. Kahn and Christopher T. Sempos (1989) Statistical Methods in Epidemiology. New York: Oxford University Press.

Where...

SMR = observed deaths in the small area/expected deaths in the small area

e = expected deaths in the small area = $\sum(R_{si} * n_i)$

R_s = the crude death rate in the standard population

R_{si} = the age-specific death rate in age group i of the standard population (# deaths / population count)

n_i = the population count in age group i of the small area

K = a constant (e.g., 100,000) that is being used to communicate the rate

(2) When the number of events <=20:

LL_{ISR} = (Lower limit for parameter estimate from Poisson table/e)) * R_s * K

UL_{ISR} = (Upper limit for parameter estimate from Poisson table/e)) * R_s * K

Where…

LL is the lower confidence interval limit, and

UL is the upper confidence interval limit.

Revision History

1/15/2002	Document created	Brian Paoli, Lois M. Haggard, Gulzar Shah, Office of Public Health Assessment, Utah Department of Health
8/11/2009	Images of sampling distributions, error correction, when rate=zero.	Lois M. Haggard, Community Health Assessment Program, New Mexico Department of Health
12/02/2009	Asymmetric confidence intervals for percentages close to 0% or 100%.	Michael Friedrichs, Chronic Disease Bureau, Utah Department of Health, Kathryn Marti, Brian Paoli, Office of Public Health Assessment, Utah Department of Health
4/23/2010	Paragraph on CI for non-binomial distributions	Lois M. Haggard, Community Health Assessment Program, New Mexico Department of Health
4/30/2010	Added text in '95% Confidence Intervals for Rare Events' and 'Directly Age-adjusted Rates' re: using SAS procedure to calculate confidence intervals for rare events.	Kathryn Marti, Brian Paoli, Office of Public Health Assessment, Utah Department of Health

Appendix 1. Upper critical values of Student's t distribution with *K* degrees of freedom

Probability of exceeding the critical value

K	0.10	0.05	0.025	0.01	0.001
1	3.078	6.314	12.706	31.821	318.313
2	1.886	2.920	4.303	6.965	22.327
3	1.638	2.353	3.182	4.541	10.215
4	1.533	2.132	2.776	3.747	7.173
5	1.476	2.015	2.571	3.365	5.893
6	1.440	1.943	2.447	3.143	5.208
7	1.415	1.895	2.365	2.998	4.782
8	1.397	1.860	2.306	2.896	4.499
9	1.383	1.833	2.262	2.821	4.296
10	1.372	1.812	2.228	2.764	4.143
11	1.363	1.796	2.201	2.718	4.024
12	1.356	1.782	2.179	2.681	3.929
13	1.350	1.771	2.160	2.650	3.852
14	1.345	1.761	2.145	2.624	3.787
15	1.341	1.753	2.131	2.602	3.733
16	1.337	1.746	2.120	2.583	3.686
17	1.333	1.740	2.110	2.567	3.646
18	1.330	1.734	2.101	2.552	3.610
19	1.328	1.729	2.093	2.539	3.579
20	1.325	1.725	2.086	2.528	3.552
21	1.323	1.721	2.080	2.518	3.527
22	1.321	1.717	2.074	2.508	3.505
23	1.319	1.714	2.069	2.500	3.485
24	1.318	1.711	2.064	2.492	3.467
25	1.316	1.708	2.060	2.485	3.450
26	1.315	1.706	2.056	2.479	3.435
27	1.314	1.703	2.052	2.473	3.421
28	1.313	1.701	2.048	2.467	3.408
29	1.311	1.699	2.045	2.462	3.396
30	1.310	1.697	2.042	2.457	3.385

Appendix 2. 95% Confidence Interval Factors for Poisson-Distributed Events

number of events	95% Confidence Interval, Lower Limit Factor	95% Confidence Interval, Upper Limit Factor
0	0.0000	3.7000
1	0.0253	5.5716
2	0.1211	3.6123
3	0.2062	2.9224
4	0.2725	2.5604
5	0.3247	2.3337
6	0.3670	2.1766
7	0.4021	2.0604
8	0.4317	1.9704
9	0.4573	1.8983
10	0.4795	1.8390
11	0.4992	1.7893
12	0.5167	1.7468
13	0.5325	1.7100
14	0.5467	1.6778
15	0.5597	1.6493
16	0.5716	1.6239
17	0.5825	1.6011
18	0.5927	1.5804
19	0.6021	1.5616
20	0.6108	1.5444