BE.104 Spring
Biostatistics: Poisson Analyses and Power
J. L. Sherley

Outline

1) Poisson analyses
2) Power

What is a Poisson process?

Rare events
Values are observational (yes or no)
Random distributed over time or place
Observations do not affect the frequency of future observations (independent)
Much of the variance is due to statistical variation, sampling variation

Many natural processes can be fit to a Poisson distribution

Consider this case:

Incidence for leukemia in MA in 1996:
680 cases distributed over 351 towns & cities

Let us assume that

1) the cases are independent and randomly distributed
2) they are sufficiently infrequent as to not effect the total population (>5 million)

We can expect:

1) many towns & cities with no cases
2) many towns & cities with number of cases near the mean- "the expected number"
   # of cases: 680/351 ≈ 2.0 per town & city
3) few towns & cities with a number of cases that greatly exceed the mean.
4) As the number of cases increases, the number of towns & cities with that number will approach 0.0

Graphically- Poisson Distribution

<Graph>

If it were ideal:

Properties- 1) the most probably number of events = 1st integer < μ, the distribution mean, unless μ is an integer,
in which case there are two equally probable maxima at
μ and $\mu - 1$

2) $\mu = \sigma^2$, the variance; therefore $\sqrt{\mu} = \sigma$

So, μ is the parameter that completely defines the Poisson distribution

What questions can be addressed with Poisson stats?

1) You are notified of a town in MA with 12 cases of leukemia.

Is it significantly different than the mean for MA of 2?

Is it unique, or could 12 be expected by chance?

We calculate confidence intervals for **μ**, given an observed number of events = **x**, assuming a Poisson distribution:

95%CI for μ about $x = x + 1.92 \pm 1.960 \sqrt{x} + 1.0$

for x = 12, 95%CI = 6.8 to 21

Therefore, we have greater than 95% confidence that the Poisson distribution to which 12 events belongs is not equivalent to the Poisson distribution that has $\mu \approx 2.0$.

If we conclude that 12 events is a part of a different Poisson distribution (i.e., it is not expected by chance), we will be wrong < 5% of the time.

If our chance of error, in thinking that 12 is not expected by chance as a part of the Poisson distribution with $\mu \approx 2.0$, is > 5% then 2 will reside in the 95% CI about 12.

We can then say that our error for saying that "**something is going on**" in the town with 12 leukemias is less than 5%.

99%CI for μ about $x = x + 3.32 \pm 2.567 \sqrt{x} + 1.7$

for x = 12, 99%CI = 5.79 to 24.9

Based on the same reasoning as for the 95% CI, we have >99% confidence that 12 did not occur by chance when the observed population mean, μ (i.e., the most expected value of x), is approximately 2.

Poisson Probability Mass Function

Given a known Poisson distribution, we can estimate the probability of an event occurring…

$$\Pr\{X_{obs} = x\} = \frac{(e^{-\mu})\,\mu^{x}}{x!} \qquad x = 0, 1, 2, \ldots$$

Where X_{obs} = number observed
x = a given number possible
μ = mean number of events expected <estimated from sample data>

What is the probability of observing 12 cases of leukemia when a mean of ca. 2 is expected

$$\Pr\{X_{obs} = 12\} = \frac{(e^{-2})\,2^{12}}{12!}$$

$1/10^6 \Rightarrow$ unlikely, but not impossible $\Rightarrow p = 0.000001$

Given a Poisson mean of 2 cases per town, there is a one in a million chance that you will observe a town with 12 cases by chance.

Or…

There is a 1 in a million chance of being wrong if you conclude that 12 did not occur by chance, that something is going on in the town.

How Poisson are the data?

Two tests:

A quick and crude test:

Does xbar = σ^2? Remember: for Poisson distribution, $\mu = \sigma^2$

However not very specific test!

Better: Plot from the data: x_i *versus* ln(observed frequency of x_i) + ln(x_i!)

x_i = each observed number of events in the distribution

<graph>

If the plot is a line with positive slope as x_i increases, then the data are ideally Poisson distributed.

Consider another scenario:

Consider prostate cancer in MA: 1996 incidence = 7900

Avg. per city & town = 7900/351 = 23

Consider a town with 50 cases.

How likely is it that a town has 50 cases by chance?

How would you approach this problem?

Distributions Interrogation

Always try to interrogate the distribution of data
Why?

1) To estimate the form of the sample population's distribution
2) To look for informative features such as multiple populations, skew
3) To determine which are the most sensitive statistical methods to apply for analyses
4) To look for reasons other than chance for observations

Confidence Level

Why the 95% confidence level? $p \leq 0.05$

Meaning: ≤ a 5% chance that an observed numerical difference is occurring by chance, when in fact two compared populations are the same.

Foolish convention?

Why not a convention…at say the 99.99% level? $p \leq 0.0001$

As we shall see, this standard would require:

1) A study of larger sample size
2) More work to do the study
3) More expense
4) And…it might not be necessary at all in the end

Related to Statistical Error Ideas

Type 1 Error- Confidence requirement set too low
p is too large (e.g., $p < 0.2$ or $p > 0.05$)

High False Positive Rate- You think that an observed numerical difference is **not** occurring by chance, but it in fact is.
"Low specificity "

This is a common problem with the convention of setting $p \leq 0.05$, and accepting a 5% error rate for concluding that populations are different in some parameter.

When is this **not** acceptable? Depends on the consequences of the decision that follows. When consequence of the error of thinking there is a difference when

there is not, are dire for quality of life, health, or life, this type of statistical error is not acceptable.

E.g. Everyday you test your drinking water to determine whether its level for a toxin is less than the lethal dose. I.e., **there is a difference** between your water's level and the lethal dose

Would you accept a test with $p = 0.05$?

Remember, based on this, you will be accepting a 5% error rate.

1 in 20 times this type 1 error will result in your death. 5% of the time when you conclude that the observed difference does not occur by chance, it will have occurred by chance.

Which means that 5% of the time the water's level and the lethal dose will appear different by chance, when they are in fact equivalent. And that will give you a lethal dose.

Type 2 Error- Confidence requirement set too high
p too low (e.g., $p << 0.05$, $p = 0.00001$)

High False Negative rate Low sensitivity	You conclude that an observed difference is due to chance and error, when it is not I.e. you think there is no difference when there really IS.

When is this acceptable? When the potential loss of quality of life, health, life is considered minimal compared to the wasted costs of an unnecessary response.

So, you don't want to do anything.

However, realize that there are more human costs than just change in health! Quality of life measures are much harder to evaluate! Even if the garbage dump or chemical exhausts are not making you physically sick you still would have a higher quality of life if it were removed.

Power to test

Suppose you perform a study of exposed versus non-exposed and do not detect an effect or difference at the $p = 0.01$ level (99% confidence) when there really is one.

I.e., You conclude with a low degree of uncertainty that there is not a difference, when there really is one. (Type II error; e.g., Fisher's error about cigarette smoke and lung cancer)

Then, the *Power* of your test is too low.

Power = the sensitivity of your study design to detect differences that do not occur solely by chance and error

What could you do to increase your power to detect?

1) You are stuck with the small effect, but consider: Could give a higher dose? I.e., look for an example where exposure was greater.

2) Reduce Variance: You might improve your precision, but often hard to do because you my be stuck with your test

3) You might increase your sample size.

   By how much? ⇒ **Power to test analyses**

   (There are some limits here: a)costs; and b) statistical limits on the amount of improvement)

4) Lower your confidence <increase p towards 0.05>
   ARE you allowed to do this?

Note that this is a new statistical realm

We now assume that there is a difference. One sample set may show the difference, whereas another may not by chance, even though the samples are drawn from the same populations.

How we now consider again the two types of statistical errors:

β = Type II error: At a given level of significance accepting that there is no difference (i.e., accepting "the null hypothesis") when there is. That is concluding that observed numerical difference occur by chance and error, when in fact they occur because of other factors.

Vs

α – Type I error: At a given level of significance concluding that there is a difference (i.e. rejecting the null hypothesis), when the observed numerical difference is in fact due to chance and error.

$\alpha = p$ (When $p = 0.05$, 5% of the time we will conclude there is a difference, when there really isn't.)

The Key Question:

How often will a test give a t-statistic that causes us to conclude that there is a difference, when there is one <i.e., reject the null hypothesis when it is false>

What is the sensitivity of our test?

Power = $1 - \beta$ = fraction of the time that a test will detect a difference when there is one

β = Risk of missing a real effect

Therefore, when our type II error is small, β approaches 0, and Power approaches 100% ability to detect differences.

Factors that affect Power

1) size of $\bar{x}_1 - \bar{x}_2$
2) n, sample size
3) *t* level

t for $\bar{x}_1 - \bar{x}_2$

↓ estimates

t' for $\mu_1 - \mu_2 = \delta$

$$t' = \delta/\sigma \sqrt{n/2}$$

power $\propto \delta/\sigma$, non-centrality parameter = ϕ, phi

As stated earlier:

As δ increases, power increases
As σ decreases, power increases
As n increases, power increases

Given a level of significance, δ, and σ (and therefore ϕ), there are several methods for arriving at power to detect:

1) Software packages
2) Mathematical estimation (Schork and Remington)
3) Graphically

Logistics of the graphical method

1) Have an expected difference based on: $\bar{x}_1 - \bar{x}_2$
which is an estimate of $\mu_1 - \mu_2$

2) Have $s_1 + s_2$
calculate pool population variance s^2

$s^2 = 0.5\ (s^2_1 + s^2_2)$
s^2 is and estimate of σ^2
Therefore, $\sqrt{s^2} \Rightarrow \sigma; \Rightarrow \phi = \delta/\sigma$

3) Find power table or graph for t-test of given α (p)

E.g. 6-9, $\alpha = 0.05$
Given value of ϕ, you can then determine n, sample size, needed for a given level of power.

E.g. @ $\phi = 1$, n = 20 gives power to detect of 90%
$\Rightarrow$ B <type II error> = 10%
10% of the time true differences will be attributed to chance.

The Bonferroni Inequality

Judging "statistical success"

Let's say that you performed 3 tests for three replicates of the same comparison. In each case *t* indicated 95% confidence that you had not detected a difference by chance. What is the probability, α_T, that in at least one case the observed difference does in fact occur by chance?

$0.05 + 0.05 + 0.05 = 0.15$, 15%

Mathematically, where K= 3, number of tests

α = type I error, error of concluding a difference when due to chance

$\alpha_T = K\alpha$

In general, we want to set $\alpha_T < K\alpha$ "Bonferroni Inequality"
So for multiple tests, we set

$$\alpha \leq \frac{\alpha_T}{K}$$

E.g., If we desired that our probability of detecting one difference by chance in 4 trials to be ≤5% (i.e., α_T), then we must require for each trial $p = \alpha = \frac{0.05}{4} = 0.0125$ or less.

In particular, the Bonferroni Inequality should be considered when an effect is observed in only one of several trials. As the number of negative trials increases, the significance of an observed effect must be qualified by the Bonferroni Inequality. Given many opportunities to observe an effect at a low level of confidence…you are MORE likely to have observed it by chance. Take this into consideration when evaluating meta-analyses in which there are many trials reviewed, but only a few are "positive" at a low level of confidence.

In a similar fashion, when all three trials show a difference at the 5% level we can say that in our example the probability that all three observed differences occurred by chance = $(0.05)^3 = 1.25 \times 10^{-4} = (\alpha)^K$

Final Words on Statistical Analyses

1) When the hypothesis of a chemical → Δ health is supported, the p-value should match the envisioned consequences of the conclusion:

   Radon gas in basements ⇒ lung cancer?

   $p = 0.05 \Rightarrow$ Type I errors: conclude radon gas causes lung cancer, when it doesn't
   Lots of unnecessary expense to ventilate basements?

   $p = 0.0001 \Rightarrow$ Type II errors: conclude radon doesn't case lung cancer, when it does
   Lots of avoidable lung cancers?

   Public Health Policy & Government Regulations must balance cost to society versus health of individuals...who make up society…segmentally.
   Evaluations based on ATTRIBUTABLE RISK for radon.

2) When the hypothesis of a health effect is not supported, it does not mean that the hypothesis is wrong. Just that it can't be supported on statistical grounds.

   "Probably due to chance" ≠ "due to chance."

   In this case evaluation of biological mechanism information can be crucial- why?

   A poor test can lead to $p > 0.05$
   A small sample size

   ∴ Statistical Analysis must be tempered by scientific reasoning to decide what to do next:

   a) discontinue study
   b) Intensify efforts

   Statistics: Quantitative method is a tool of science, not the science!

Statistical Analyses are the beginning of scientific investigations of potential chemical exposure/health effect relationships. They are not the final word. EHS address the following difficult issues:

1) Something may be going on.
2) Probably nothing is going on.
3) Main concern- Yes, something is going on and it's terrible and it's preventable!

Toxicology & Biology allow us to evaluate the plausibility of an agent → Δhealth once it is detected by statistical methods. Toxicology & Biology can sometimes lead to continued analyses when the stats suggest nothing is going on.

Final Thoughts

1) Stats are not the final word. They allow us to quantify uncertainty. They do not allow us to establish facts.

2) Small effects could be biologically important although not statistically significant. Observed effects must also be integrated with independent experiments.

3) Large effects could be erroneous. Don't forget that $p(\alpha)$ means that there is a finite possibility that the difference does occur by chance.

4) Set p at a meaningful value in your work. If p is too low, it may prohibit a study (e.g. too expensive to achieve n needed). If p is too high, it may lead to errors that are also costly in the long run.