Chapter 9

Comparing Two Populations: Binomial and Poisson

9.1 Four Types of Studies

We will focus on the binomial in this chapter. In the last section we extend these ideas to the Poisson distribution.

When we have a dichotomous response we have focused on BT. The idea of finite population was introduced in Chapter 2 and presented as a special case of BT. In this section it is convenient to begin with finite populations.

The four in the title of this section is obtained by multiplying 2 by 2. When we compare two populations both populations can be trials or both can be finite populations. In addition, as we shall discuss soon, a study can be **observational** or **experimental**. Combining these two dichotomies, we get four types of study, for example an observational study on finite populations.

It turns out that the math results are (more or less) identical for the four types of studies, but the *interpretation* of the math results depends on the type of study.

We begin with an observational study on two finite populations. This example is in Chapter 7 of my textbook; the interested reader can go there for more details.

This was a real study performed over 20 years ago; it was published in 1988. The first finite population is undergraduate men at at the University of Wisconsin-Madison and the second population is undergraduate men at Texas A&M University. Each man's response is his answer to the following question:

> If a woman is interested in dating you, do you generally prefer for her: to ask you out;
> to hint that she wants to go out with you; or to wait for you to act.

The response 'ask' is labeled a success and either of the other responses is labeled a failure. The purpose of the study is to compare the proportion of successes at Wisconsin with the proportion of successes at Texas A&M.

The two populations obviously fit our definition of finite populations. Why is it called observational? The dichotomy of observational/experimental refers to the *control* available to the

Table 9.1: Responses to the Dating Study.

Population	Observed Frequencies Prefer Women to:			Row Proportions Prefer Women to:		
	Ask	Other	Total	Ask	Other	Total
Wisconsin	60	47	107	0.56	0.44	1.00
Texas A&M	31	69	100	0.31	0.69	1.00
Total	91	116	207			

researcher. Suppose that Matt is a member of one of these populations. As a researcher, I have control over whether I have Matt in my study, but I do *not* have control over the population to which he belongs. The variable that determines to which population a subject belongs, is often called a **study factor**. Thus, in the current study, the study factor is school attended and it has two **levels**: Wisconsin and Texas. This is an observational factor, sometimes called, for obvious reasons, a classification factor, b/c each subject is classified according to his school.

Table 9.1 are the data for this *Dating Study*.

Next, we have an example of comparing finite populations in an experimental study. This example also is discussed in my text in Chapters 1 and 7.

Medical researchers were searching for an improved treatment for persons with *Crohn's Disease*. They wanted to compare a new drug therapy, *cyclosporine*, to an inert drug, called a *placebo*.

Now we are at a hugely important distinction from the Dating Study. Below we are going to talk about comparing the 'cyclosporine population' to the 'placebo population.' But, as we shall see, and perhaps is already obvious, there is, in reality, neither a 'cyclosporine population' nor a 'placebo population.' Certainly not in the physical sense of there being a UW and Texas A&M.

Indeed, as I formulate a 'population approach' to this medical study, the only population I can imagine is one *superpopulation* of all persons, say in the US, who have Chron's Disease. This superpopulation gives rise to two imaginary populations: first, imagine that everybody in the superpopulation is given cyclosporine and, second, imagine that everybody in the superpopulation is given the placebo.

To summarize the differences between observational and experimental:

1. For observational, there exists two distinct finite populations.

    For experimental, there exists two **'treatments'** of interest and one superpopulation of subjects. The two populations are generated by imagining what would happen if each member of the superpopulation was assigned each treatment.

2. Here is a very important consequence of 1: For an observational study, the two populations consist of different subjects whereas for an experimental study, the two populations consist of the same subjects.

For the Dating study, the two populations are comprised of different men (Bubba, Bobby Lee, Tex, etc. for one; and Matt, Eric, Brian, etc. for the other). For the Crohn's study, both populations consist of the same persons, namely the persons in the superpopulation.

An experimental study also requires something called **randomization**. I will discuss it in the next section. Also, these ideas can and will be extended to BT that are for trials, not finite populations.

9.2 Assumptions and Results

We begin with an observational study on finite populations. Assume that we have a random sample of subjects from each population and that the samples are independent of each other. Independence here is much the same idea as it was for trials. For our Dating study, independence means that the method of selecting subjects from Texas was totally unrelated to the method used in Wisconsin. 'Totally unrelated' is, of course, rather vague, but bear with me for now. Additionally and sadly, I will not at this time give you an example where independence fails to be true in a major way. Later when we consider **paired data** we will revisit this issue.

All this talk of independence should not make us forget that, just like for a single finite population, the biggest challenge is to actually get a random sample. Usually the sample is clearly not random and the researcher simply pretends that it is. This is too big of a topic for typing; I will discuss it in lecture.

The sample sizes are n_1 from the first population and n_2 from the second population. We define X to be the total number of successes in the sample from the first population and Y to be the total number of successes in the sample from the second population. Given our assumptions, $X \sim$ Bin(n_1, p_1) and $Y \sim$ Bin(n_2, p_2), where p_i is the proportion of successes in population i, $i = 1, 2$.

Always remember that you can study the populations separately using the estimation methods of Chapter 3. The purpose of this chapter is to *compare* the populations, or, more precisely, to compare the two p's. We will consider both estimation and testing.

For estimation, our goal is to estimate $p_1 - p_2$. Define $\hat{p}_1 = X/n_1$ and $\hat{p}_2 = Y/n_2$; these are out point estimators of the p's. The obvious, and correct, point estimator of $p_1 - p_2$ is $\hat{p}_1 - \hat{p}_2 = X/n_1 - Y/n_2 = W$. We will use the result of Chapter 7 to obtain the mean and variance of W:

$$\mu_W = \mu_X/n_1 - \mu_Y/n_2 = p_1 - p_2, \text{ and}$$

$$\sigma_W^2 = \sigma_X^2/n_1^2 + \sigma_Y^2/n_2^2 = p_1 q_1/n_1 + p_2 q_2/n_2.$$

Thus, it is easy to standardize W:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{(p_1 q_1)/n_1 + (p_2 q_2)/n_2}}. \tag{9.1}$$

It can be shown that if both n_1 and n_2 are large and neither p_i is too close to either 0 or 1, then probabilities for Z can be well approximated by using the snc. Slutsky's results also apply here. Define

$$Z' = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{(\hat{p}_1 \hat{q}_1)/n_1 + (\hat{p}_2 \hat{q}_2)/n_2}}. \tag{9.2}$$

Subject to the same conditions we had for Z, probabilities for Z' can be well approximated by using the snc. Thus, using the same algebra we had in Chapter 3, Formula 9.2 can be expanded to give the following two-sided confidence interval for $(p_1 - p_2)$:

$$(\hat{p}_1 - \hat{p}_2) \pm z\sqrt{(\hat{p}_1\hat{q}_1)/n_1 + (\hat{p}_2\hat{q}_2)/n_2}. \tag{9.3}$$

I will use this formula to obtain the 95% confidence interval for the Dating study. First, as before, 95% confidence gives $z = 1.96$. Using the summaries in Table 9.1 we get the following:

$$(0.56-0.31)\pm1.96\sqrt{(0.56)(0.44)/107 + (0.31)(0.69)/100} = 0.25\pm1.96(0.0666) = 0.25\pm0.13 =$$

$[0.12, 0.38]$. I can't say much about interpreting these endpoints b/c, fundamentally, I find this research question, while interesting to me, to be largely frivolous. I will discuss this more in lecture.

For a test of hypotheses, the null hypothesis is $H_0 : p_1 = p_2$. There are three choices for the alternative:

$$H_1 : p_1 > p_2; \ H_1 : p_1 < p_2; \ \text{or} \ H_1 : p_1 \neq p_2.$$

Recall that we need to know how to compute probabilities given that the null is true. First, note that if the null is true, then $(p_1 - p_2) = 0$. Making this substitution into Z in Equation 9.1, we get:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{(p_1q_1)/n_1 + (p_2q_2)/n_2}}. \tag{9.4}$$

We again have the problem of unknown parameters in the denominator. For estimation we used Slutsky's results and handled the two unknown p's separately. But for testing, we proceed a bit differently.

On the assumption that the null is true, X and Y have the same p, so we should combine them to estimate the common value of p. In particular, define $\hat{p} = (X + Y)/(n_1 + n_2)$. We replace the unknown p_i's in Equation 9.4 with this $\hat{p}$ to get the following test statistic.

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{(\hat{p}\hat{q})(1/n_1 + 1/n_2)}}. \tag{9.5}$$

Assuming that n_1 and n_2 are both large and that the common value of p_1 and p_2 is not too close to either 0 or 1, probabilities for Z in this last equation can be well approximated with the snc.

If we let z denote the observed value of Z, then the approximate P-value is given below for each possible choice of the alternative.

- For $H_1 : p_1 > p_2$: The area under the snc to the right of z.

- For $H_1 : p_1 < p_2$: The area under the snc to the left of z. (Or, if you prefer, The area under the snc to the right of $-z$.)

- For $H_1 : p_1 \neq p_2$: Twice the area under the snc to the right of $|z|$.

I will demonstrate these ideas with the Dating study.

I would choose the alternative $p_1 > p_2$, but I will calculate the approximate P-value for all three possibilities. I begin by calculating $\hat{p} = (60 + 31)/(107 + 100) = 0.44$, giving $\hat{q} = 0.56$. The test statistic is

$$Z = \frac{0.25}{\sqrt{(0.44)(0.56)(1/107 + 1/100)}} = \frac{0.25}{0.0690} = 3.62.$$

Using the snc calculator, the P-value for $p_1 > p_2$ is 0.00015; for $p_1 < p_2$ it is 0.99985; and for $p_1 \neq p_2$ it is $2(0.00015) = .00030$.

The Fisher test site can be used to obtain the exact P-value. Read the number next to 'Right' for the alternative $p_1 > p_2$; read the number next to 'Left' for the alternative $p_1 < p_2$; and read the number next to '2-Tail' for the alternative $p_1 \neq p_2$. Using the site I get the following exact P-values: 0.00022, 0.99993 and 0.00043. The snc approximation is quite good.

We will return now to the Crohn's Disease study, our example of an experimental study on finite populations. B/c the two populations do not actually exist in the physical world, we modify our sampling a bit.

- Decide on the numbers n_1 and n_2, where n_i is the number of subjects who will be given treatment i. Calculate $n = n_1 + n_2$, the total number of subjects who will be in the study.

- Select a random sample of n subjects from the superpopulation.

- Divide the n subjects selected for study into two treatment groups by **randomization**. Assign n_1 subjects to the first treatment and n_2 subjects to the second treatment. Randomization is defined below.

It is easiest for me to give an example of how to randomize in a particular problem.

Suppose that I have $n = 20$ subjects to study and I want to place $n_1 = 8$ on the first treatment and $n_2 = 12$ on the second treatment. We randomize by completing the following steps.

1. Assign the numbers 1, 2, ..., 20 to the subjects in any manner.

2. Take 20 identical cards and mark them 1, 2, ..., 20, one number per card and use all numbers.

3. Place the 20 cards in the box and mix thoroughly.

4. Select $n_1 = 8$ cards at random from the box. The subjects assigned to the 8 selected numbers are given the first treatment; the remaining subjects are given the second treatment.

Of course, you live in the information age and don't want to care around lots of cards and a box. There is website that will perform steps 2–4 for you and it is linked to our course webpage under calculators. I went to the website, clicked on Randomize (there are several choices) and entered (this will make sense if you do it): 1, 8, 1 and 20 in the four spaces; kept the default 'Yes' for unique; selected 'Yes: Least to Greatest' on the sort option; and clicked on 'Randomize Now!' (Do these programmers really think that this will become your favorite class if they use exclamation points?) The result I obtained was:

Table 9.2: Responses to the Crohn's Disease Study.

| Population | Observed Frequencies Prefer Women to: | | | Row Proportions Prefer Women to: | | |
	Ask	Other	Total	Ask	Other	Total
Cyclosporine	22	15	37	0.595	0.405	1.000
Placebo	11	23	34	0.324	0.676	1.000
Total	33	38	71			

2, 3, 4, 6, 8, 11, 16, 18.

Thus, the subjects assigned these numbers are assigned to the first treatment and the remaining 12 subjects are assigned to the second treatment.

If we now turn to the two imaginary populations, we see that our samples are not quite independent. The reason is quite simple. Any member of the superpopulation, call him Ralph, *cannot* be given both treatments. Thus, if, for example, Ralph is given the first treatment he cannot be given the second treatment. Thus, knowledge that Ralph is in the sample from the first population tells us that he is not in the sample from the second population; i.e. the samples depend on each other. But if the superpopulation has a large number of members compared to n, which is usually the case in practice, then the dependence between samples is very weak and can be safely ignored, which is what we will do.

Ignoring the slight dependence between samples, we can use the same estimation and testing methods that we used for the Dating study. The details are now given for the Crohn's Disease study. First, Table 9.2 presents the data. Here is the 95% confidence interval for $p_1 - p_2$:

$$(0.595-0.324)\pm1.96\sqrt{(0.595)(0.405)/37 + (0.324)(0.676)/34} = 0.271\pm0.223 = [0.048, 0.494].$$

While one should be encouraged b/c this CI indicates that cyclosporine is superior to a placebo, the great width of the interval tells me that we really don't have much of an idea about *how much* cyclosporine is better.

For the test of hypotheses, I choose the first alternative, $p_1 > p_2$. Using the website, the exact P-value is 0.0198. To use the snc approximation, first we need $\hat{p} = 33/71 = 0.465$. Plugging this into Equation 9.5, we get

$$Z = \frac{0.271}{\sqrt{(0.465)(0.535)[1/37 + 1/34]}} = \frac{0.271}{0.1185} = 2.287.$$

From the snc calculator, the approximate P-value is 0.0111. The approximation is very bad; it is not close to the exact 0.0198.

9.3   Bernoulli Trials

This distinction between observational and experimental studies with BT confuses many people. It's a good indicator of whether somebody thinks like a mathematician or a statistician/scientist.

Table 9.3: Results for the 3-Point Basket Study.

	Observed Frequencies			Row Proportions		
Population	Basket	Miss	Total	Basket	Miss	Total
Front	21	29	50	0.42	0.58	1.00
Left Corner	20	30	50	0.40	0.60	1.00
Total	41	59	100			

Mathematicians think as follows: Given that we assume we have BT, p is constant and there is no memory. Thus, you can prove (and they are correct on this) that there is no reason to randomize.

Statistician/scientists think as follows: In a real problem we can never be certain that we have BT and frequently we have serious doubts about it. As a result, if we *can* randomize, it adds another 'level of validity' to our findings.

I will begin with an example of a student project performed by a former star basketball player in my class, Arnold (Clyde) Gaines. Clyde wanted to study his ability to shoot in basketball. He chose two locations from which to attempt shots, so the idea was to study whether he was better from either location. The two locations he selected were both behind the three-point line. Location (Treatment) 1 was shooting from in front of the basket and location (treatment) 2 was shooting from the left corner. Clyde decided to perform a total of $n = 100$ jump shots with $n_1 = n_2 = 50$; i.e. the same number of shots from each location. Clyde's data are in Table 9.3. First, I note, descriptively, that his performance was almost identical from the two locations. Using Fisher's test, the exact P-values are: 0.5000 for $>$; 0.6577 for $<$; and 1 for $\neq$. In the taxonomy of Chapter 8, this is a D study: there is neither statistical nor practical significance. The CI, however, provides an interesting insight.

The 95% CI for $p_1 - p_2$ is

$$(0.42 - 0.40) \pm 1.96\sqrt{(0.42)(0.58)/50 + (0.40)(0.60)/50} = 0.02 pm 0.19 = [-0.17, 0.21].$$

This CI indicates, IMHO, that, statistically, the study was a waste of time. The CI is so wide as to be useless. Here is what I mean. Using my 'expertise' as a basketball person, I would be amazed if the true p's for a highly skilled basketball player differed by more than 0.15. As a result, the CI tells me **less** than what I 'knew' before the data were collected. Clearly, to study this problem one needs many more than $n = 100$ shots.

I end this section with an example of an observational study with BT.

Having lived in Michigan or Wisconsin my entire life, I had noted that weather seems to be less predictable in Spring than in Summer. In 1988, I collected some data to investigate this issue.

Every day the morning Madison newspaper would print a predicted high temperature for that day and the actual high for the previous day. Using these data over time, one could evaluate how well the predictions performed. I arbitrarily decided that a prediction that came within two degrees of the actual was a success and all other predictions were failures. Thus, for example, if the predicted high was 60 degrees, then if the actual high was between 58 and 62 degrees, inclusive, then the prediction was a success; o.w. it was a failure. Table 9.4 presents the data that I collected.

Table 9.4: Results for the High-Temperature Forecast Study.

| Population | Observed Frequencies | | | Row Proportions | | |
	S	F	Total	S	F	Total
Spring	46	43	89	0.517	0.483	1.00
Summer	50	39	89	0.562	0.438	1.00
Total	96	82	178			

The Summer predictions were better (descriptively), but not by much. I found this surprising. My choice of alternative is $<$, which has exact P-value equal to 0.3260. The other exact P-values are 0.7739 for $>$; and 0.6520 for $\neq$.

The 95% CI for $p_1 - p_2$ is

$$(0.517-0.562)\pm1.96\sqrt{(0.517)(0.483)/89 + (0.562)(0.438)/89} = -0.045\pm0.146 = [-0.191, 0.101].$$

I am not an expert at weather forecasting, so I cannot really judge whether this CI is useful scientifically. But I doubt it.

9.4 Simpson's Paradox

The most important difference between an observational and experimental study is in how we interpret our findings. Let us compare and contrast the Dating and Crohn Disease studies.

In both studies we concluded that the populations had different p's. Of course these conclusions can be wrong, but let's ignore that issue for now. We have concluded that the populations are different, so it is natural to wonder **why?**

In the Dating Study we don't know why. Let me be clear. We have concluded that Wisconsin men and Texas men have very different attitudes, but we don't know why. Is it b/c the groups differ on:

Academic major? Ethnicity? Religion? Liberalism/Conservatism? Wealth?

We don't know why. Indeed, perhaps the important differences are in the **women** mentioned in the question. Perhaps it means something really very different to be asked out by a Texas woman versus a Wisconsin woman.

The above comments (not all of which are silly!) are examples of what is true for any observational study. We can conclude that the two populations are different, but we don't know why.

Let us contrast the above with the situation for the Crohn's Disease study. In this case, the two populations consist of exactly the same subjects! Thus, the only possible explanation for the difference between populations is that cyclosporine is better than the placebo. (This is a good time to remember that our conclusion that the populations differ could be wrong.)

Simpson's Paradox (no, not named for Homer, Marge, Bart, Maggie, Lisa or even O.J.) provides another, more concrete, way to look at this same issue.

Table 9.5: Hypothetical Observational Data.

Sex	Released? Yes	No	Total	$\hat{p}$
Female	60	40	100	0.60
Male	40	60	100	0.40
Total	100	100	200	

Table 9.6: Hypothetical Observational Data with Background Factor: **Case 1**.

Job A					Job B				
	Released?					Released?			
Sex	Yes	No	Total	$\hat{p}$	Sex	Yes	No	Total	$\hat{p}$
Female	30	20	50	0.60	Female	30	20	50	0.60
Male	20	30	50	0.40	Male	20	30	50	0.40
Total	50	50	100			50	50	100	

Years ago I worked as on expert witness in several cases of workplace discrimination. As a result of this work, I was invited to make a very brief presentation at a continuing education workshop for State of Wisconsin administrative judges. (In Wisconsin, the norm was (is?) to have workplace discrimination cases settled administratively rather than by a jury of citizens.) Below I am going to show you what I presented in my 10 minutes.

These are totally and extremely hypothetical data. A company with 200 employees decides it must reduce its work force by one-half. Table 9.5 reveals the relationship between sex and outcome. The table shows that the proportion of women who were released was 20 percentage points larger than the proportion of men who were released. But this is an observational study (the researcher did not assign Sally to be a woman by randomization). This means that we do not know *why* there is a difference. In particular, it would be wrong to say it is b/c of discrimination.

The idea we will pursue is: What else do we know about these employees? In particular, do we know anything other than their sexes? Let us assume that we know their job classifications and that, for simplicity, there are only two jobs, denoted by A and B. We might decide to incorporate the job classification into our description of the data. I will show you four possibilities for what could occur. As will be obvious, this is not an exhaustive listing of possibilities.

My first possibility is shown in Table 9.6; it shows that bringing job into the analysis might have no effect whatsoever. The proportions in each sex and each job match exactly what we had in Table 9.5. Henceforth we will refer to our original table as the collapsed table and tables such as the two in Table 9.6 as the component tables.

My next possibility is in Table 9.7. In this Case 2 we find that job does matter and it matters in the sense that women are doing even worse in both jobs than they are doing in the collapsed table.

Our next possibility, Case 3 in Table 9.8 shows that if we incorporate job into the description,

Table 9.7: Hypothetical Observational Data with Background Factor: **Case 2**.

| | **Job A** | | | | | **Job B** | | | |
| | Released? | | | | | Released? | | | |
Sex	Yes	No	Total	$\hat{p}$	Sex	Yes	No	Total	$\hat{p}$
Female	30	10	40	0.75	Female	30	30	60	0.50
Male	30	30	60	0.50	Male	10	30	40	0.25
Total	60	40	100			40	60	100	

Table 9.8: Hypothetical Observational Data with Background Factor: **Case 3**.

| | **Job A** | | | | | **Job B** | | | |
| | Released? | | | | | Released? | | | |
Sex	Yes	No	Total	$\hat{p}$	Sex	Yes	No	Total	$\hat{p}$
Female	60	15	75	0.80	Female	0	25	25	0.00
Male	40	10	50	0.80	Male	0	50	50	0.00
Total	100	25	125			0	75	75	

the difference between the experiences of the sexes can disappear.

Finally, Case 4 in Table 9.9 shows that if we incorporate job into the description, the difference between the experiences of the sexes can **be reversed**! This reversal is called Simpson's Paradox.

So, what is going on? I will look at Cases 1 and 4 in detail.

We have a response (released or not), study factor (sex) and background factor (job). In the collapsed table we found an association between response and study factor and in Case 1 the association remained unchanged when we took into account background factor. To see why this is so, examine Table 9.10. We see that the background factor has no association (the row $\hat{p}$'s are identical) with either the study factor or the response. Hence, incorporating it into the analysis has no effect.

By contrast, in Case 4, putting the background factor into the analysis had a huge impact. We

Table 9.9: Hypothetical Observational Data with Background Factor: **Case 4: Simpson's Paradox**.

| | **Job A** | | | | | **Job B** | | | |
| | Released? | | | | | Released? | | | |
Sex	Yes	No	Total	$\hat{p}$	Sex	Yes	No	Total	$\hat{p}$
Female	56	24	80	0.70	Female	4	16	20	0.20
Male	16	4	20	0.80	Male	24	56	80	0.30
Total	72	28	100			28	72	100	

Table 9.10: Relationships With Background Factors for Study Factor and Response in Case 1.

Sex	A	B	Total	$\hat{p}$	Released?	A	B	Total	$\hat{p}$
		Job					Job		
Female	50	50	100	0.50	Yes	50	50	100	0.50
Male	50	50	100	0.50	No	50	50	100	0.50
Total	100	100	200			100	100	200	

Table 9.11: Relationships With Background Factors for Study Factor and Response in Case 4.

Sex	A	B	Total	$\hat{p}$	Released?	A	B	Total	$\hat{p}$
		Job					Job		
Female	80	20	100	0.80	Yes	72	28	100	0.72
Male	20	80	100	0.20	No	28	72	100	0.28
Total	100	100	200			100	100	200	

can see why in Table 9.11. In this case, the background factor is strongly associated with both the study factor and the response; in particular, women are disproportionately in Job A and persons in Job A are disproportionately released. It can be shown that 'something like Cases 2–4' can occur only if the background factor is associated with both the study factor and the response. But here is where randomization becomes relevant. If subjects are assigned to study factor level by randomization, then there should be either no or only a weak association between study factor and background factor. Thus, the message in the collapsed table will be pretty much the same as the message in any component tables.

9.5 Poisson Populations

This will be a bit trickier notationally than what we had above for the binomial. Here is the idea. Suppose that we have two Poisson Processes. The first one has rate λ_1 and the second has rate λ_2. For concreteness, let's say the rate is 'per hour.' We could very well wonder whether the rates are the same and we could want to investigate this with testing or estimation. But now the problem arises. To be general, we must allow for each PP to be observed for any length of time.

Thus, suppose that the first PP is observed for t_1 hours and we obtain X successes. And suppose that the second PP is observed for t_2 hours and we obtain Y successes. It follows that $X \sim \text{Poisson}(\theta_1 = t_1\lambda_1)$ and $Y \sim \text{Poisson}(\theta_2 = t_2\lambda_2)$. Thus, our interest is in, for example, knowing whether

$$\frac{\theta_1}{t_1} = \frac{\theta_2}{t_2}, \text{ or } \frac{t_2}{t_1}\theta_1 = \theta_2, \text{ which means } \lambda_1 = \lambda_2.$$

Without loss of generality, we will set up our problem as follows. $X \sim \text{Poisson}(\theta_1)$ and $Y \sim \text{Poisson}(\theta_2$ are independent random variables. We want to estimate $(c\theta_1 - \theta_2)$ for some known

number c. And we want to test the null hypothesis that $(c\theta_1 = \theta_2)$ for some known number c. Remember that for two PP's $c = t_2/t_1$.

Define $W = cX - Y$. From our result in Chapter 7 and basic properties of the Poisson, W has mean and variance given as follows:

$$\mu_W = c\mu_X - \mu_Y = c\theta_1 - \theta_2 \text{ and } \sigma_W^2 = c^2\theta_1 + \theta_2.$$

Thus, it is easy to standardize W:

$$Z = \frac{(cX - Y) - (c\theta_1 - \theta_2)}{\sqrt{c^2\theta_1 + \theta_2}}.$$

If both $c\theta_1$ and θ_2 are large, then the snc will give good approximate probabilities for Z. Slutsky's results apply too. Define

$$Z' = \frac{(cX - Y) - (c\theta_1 - \theta_2)}{\sqrt{c^2X + Y}}.$$

If both $c\theta_1$ and θ_2 are large, then the snc will give good approximate probabilities for Z'. Thus, we are able to proceed with estimation and testing. The results are below.

Formula 9.6 is the snc approximate CI for $(c\theta_1 - \theta_2)$.

$$(cx - y) \pm z\sqrt{c^2x + y} \tag{9.6}$$

For testing, the null is $c\theta_1 = \theta_2$. There are three possible alternatives: replace the $=$ in the null with $>$, $<$ or $\neq$.

The test statistic is

$$Z = \frac{cX - Y}{\sqrt{c^2X + Y}}. \tag{9.7}$$

I will do one example to illustrate these methods.

The first PP is observed for 20 hours and yields $X = 100$ successes. The second PP is observed for 50 hours and yields $Y = 225$ successes. Do the PP's have the same rate?

Following the earlier presentation, $t_1 = 20$, $t_2 = 50$ and $c = t_2/t_1 = 50/20 = 2.5$. The 95% CI for $(2.5\theta_1 - \theta_2)$ is

$$(2.5(100) - 225) \pm 1.96\sqrt{(2.5)^2(100) + 225} = 25 \pm 1.96\sqrt{850} = 25 \pm 57.1 = [-32.1, 82.1].$$

Now, you can show that

$$\lambda_1 - \lambda_2 = (2.5\theta_1 - \theta_2)/50.$$

Thus,

$$[-32.1/50, 82.1/50] = [-0.642, 1.642]$$

is the 95% CI for $(\lambda_1 - \lambda_2)$.

For testing,

$$Z = \frac{25}{\sqrt{850}} = 0.858.$$

The approximate P-values are 0.1954 for $>$; 0.8046 for $<$; and $2(0.1954) = 0.3908$ for $\neq$.

94