Chapter 8: Comparing Proportions and Risk

Investigation: In the Netflix series "100 Humans," the research team asked: "**Can music influence how much risk we're willing to take?**" Let's watch the following video to learn more about this experiment.

https://mediaspace.illinois.edu/media/t/1_cycmg5c1

Unit of observation:

Response variable (and type):

Explanatory variable (and type):

Rather than a comparison of means, in this chapter we're going to focus on a comparison of proportions.

Based on the percentages reported at the end (and assuming 30 people per group), we can figure out the frequencies and place them in the following table.

	Take Risk	**Avoid Risk**	**Totals**
Sad music	17	13	30
Happy music	25	5	30

What proportion of sad music listeners were willing to take the risk?

0.57 (57%)

What proportion of happy music listeners were willing to take the risk?

0.83 (83%)

But…could this difference be explained as random chance?

- Exploring this investigation through a **Permutation Test**
  - IF the null is true and there is no underlying difference between these two populations on average**…**we can "shuffle" the group designation randomly and simulate sample proportion differences under the null hypothesis?

Let's complete a permutation test where we apply the null hypothesis that $\pi_S = \pi_H$

If that's the case, then we can randomly permute the labels to see how often we see differences in proportions as large as we did in our sample. https://istats.shinyapps.io/Association_Categorical/

- Choose the contingency table options and enter our data and labels
- Let's visualize our data more clearly with a stacked barplot
- Switch to the "Permutation Distribution" tab to explore how often this difference could happen by random chance

Hypothesis Testing for a Difference in Proportions

- When testing for a difference in proportions, we commonly test the null hypothesis that $\pi_1 = \pi_2$, or to say it another way…$\pi_1 - \pi_2 = \underline{0}$.
- With that in mind, our Null Model represents the distribution of $\underline{\hat{p}_1 - \hat{p}_2}$ under this assumption. This distribution should have a mean of 0 and a standard deviation of $SE_{\hat{p}_1-\hat{p}_2}$

Let's first consider our investigation using the **Permutation Test for Two Proportions**

The researchers hypothesized that happy music might make people more willing to take the risk. With that in mind, what are the null and alternative hypotheses?

H_0: $\underline{\pi_S \leq \pi_H}$ H_A: $\underline{\pi_S > \pi_H}$

According to our simulated null model, how often do we see permuted sample differences as low or lower than our particular sample difference?

- **Parametric Assumption.** As with a one-sample test for a proportion, this Null Model will be approximately <u>normally distributed</u> in many situations!
  - In a two-sample context, we might bump up our 10/10 rule to a 20/20 rule. We'd like to see at least 20 of each response cumulatively across both groups combined.
  - When this is true, it's reasonable to use a z-test to approximate the p-value
  - *When this is not true, we might stick with something like a permutation test.*

Now, let's consider how we'd approximate the p-value with a **Z-test for Two Proportions**

Computational Background for comparing proportions

$$z = \frac{\hat{p}_1 - \hat{p}_2}{SE_{\hat{p}_1-\hat{p}_2}} \quad \text{where} \quad SE_{\hat{p}_1-\hat{p}_2} = \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}$$

Our hypotheses are still the same. We just need to calculate the Standard Error and use that to find the standardized position (z-score) of our sample mean difference in this model.

According to the standard normal distribution, we should expect to see a z-score this low or lower about **1.2%** of the time. That is a small, but noticeable discrepancy from our permutation test!

But is the parametric test reliable to use in this example?

<u>We're actually just short of the 20/20 rule here. So it's not a bad fit, but it's not as reliable as we'd like. The permutation test would be better here!</u>

Framing Proportions in terms of Risk

Investigation: During the 1950s, the poliovirus posed a serious health threat with no highly effective treatments available. In response, Jonas Salk developed a vaccine that he hoped would minimize the risk of polio, especially of more severe outcomes such as paralysis or death.


Courtesy of Boston Children's Hospital Archive

After early success with small samples and little to no ill side-effects, the NFIP approved a large U.S. study enrolling nearly hundreds of thousands of children, ages 6 – 9. Children were randomly assigned to either receive the experimental Salk Vaccine or a placebo injection http://www.medicine.mcgill.ca/epidemiology/hanley/c622/salk_trial.pdf

	Total	Polio	Polio Paralysis	Polio Fatality
Salk Vaccine	200,745	57	33	0
Placebo	201,229	142	115	4

We'd like to know *if* the risk for polio might be lower with the vaccine than with the placebo, and by *how much*.

Population:

Unit of observation:

Response variable (and type):

Explanatory variable (and type):

What proportion of the children receiving the salk vaccine were eventually diagnosed with polio?

57/200,745

What about the proportion of children receiving the placebo?

142/201,229

Let's use the sim to double check our calculations and also visualize this difference. https://istats.shinyapps.io/Association_Categorical/

In contexts like this, we might reasonably refer to a proportion as a risk

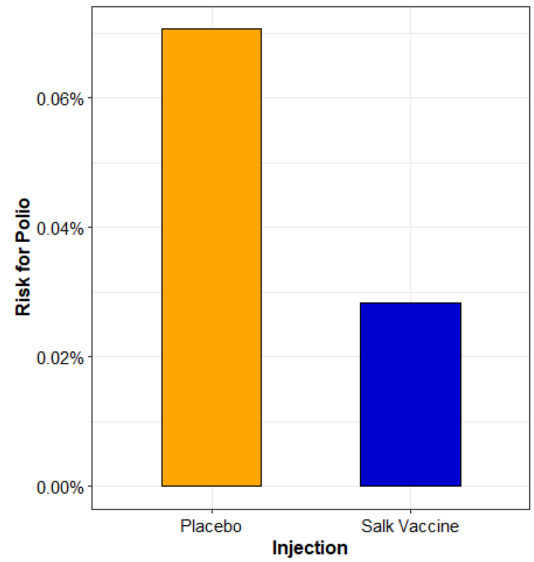
- The proportion of children who got polio while using the placebo represents the risk for polio if taking the placebo.

Absolute Risk *Reduction* reports the absolute value difference in risk between two groups (it's just the difference in proportions!). It is typically reported as an *absolute value.*

ABR = $|\text{Risk}_A - \text{Risk}_B|$.

This could be reported as a percentage: $|\text{Risk}_A - \text{Risk}_B| * 100\%$

Practice: What is the (estimated) absolute risk reduction for polio when taking the vaccine?



Relative Risk (RR) *(sometimes referred to as a "Risk Ratio")* represents the ratio of risk under one condition to another condition.

RR = $\frac{\text{Risk}_A}{\text{Risk}_B}$. *This could be reported as a percentage*: $\left(\frac{\text{Risk}_A}{\text{Risk}_B}\right) * 100\%$

Practice: What is the (estimated) relative risk for polio when taking the vaccine?

The estimated risk of contracting polio after taking the Salk Vaccine is ______ times the risk for polio after taking the placebo.

- If relative risk is **below 1**, that means risk is **reduced**.
- If relative risk is **above 1**, that means risk is **increased**.

Effectiveness: Represents the percent of individuals that would avoid the infection by taking part in the intervention. This is often reported for vaccines, but could be reported for other treatments as well.

Effectiveness = 1 – RR. *This could be reported as a percentage:* (1 – RR)*100%

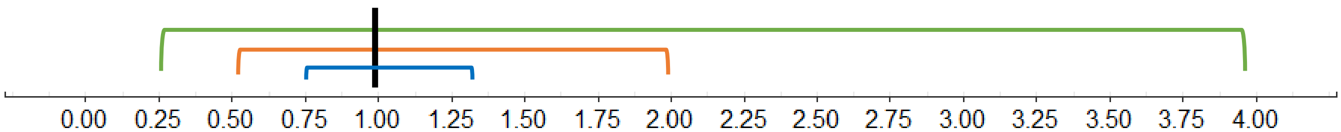
Practice: How effective is the Salk Vaccine at preventing polio?

Confidence Intervals for Relative Risk (RR)

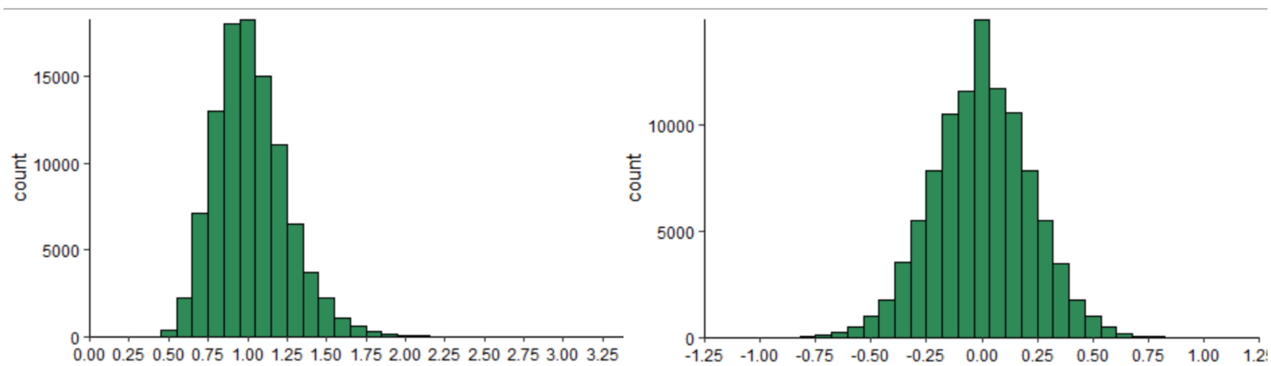
- We still need to acknowledge that we have only a sample of children in this study, and our calculations for ABR, RR, and Effectiveness are all sample statistics
- We acknowledge that when testing for a difference in proportions/risk.

H_0: $\pi_A = \pi_B$ which could also be written as... H_0: RR = 1

- But if our goal is to also estimate by how much the risk is reduced/increased, we might prefer to use a confidence interval.
- Typically in the context of risk, researchers report confidence intervals for the Relative Risk/Effectiveness as a way to measure proportional increase/decrease.
- What makes this trickier for RR is that it is an exponentially-scaled measure.



- However, the logarithm of the distribution of possible RR's will be symmetrically distributed!
  - The distribution on the left represents the sampling distribution for $\widehat{RR}$ for null RR = 1 (n_1 = 100, n_2 = 100)
  - The distribution on the right represents the sampling distribution for $\log(\widehat{RR})$ for null RR = 1.



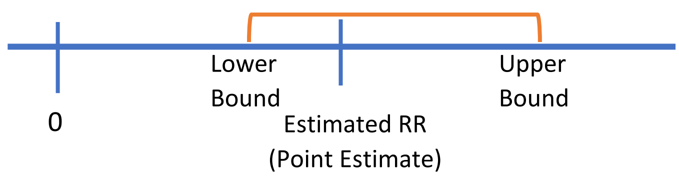
  - Mathematically, we can make a parametric assumption and find a z-interval for $\log(\widehat{RR})$. Then change the units back to report a confidence interval for RR!

$\log(\widehat{RR}) \pm z^* SE_{\log(\widehat{RR})}$ *...which after taking the exponent of both sides is...* $\widehat{RR} \pm e^{z^* SE_{\log(\widehat{RR})}}$

  - Don't worry—we'll just focus on using an online calculator for the 95% interval and then simply interpreting this interval. **No calculations by hand necessary!**

- **The confidence interval will be** assymetric about the point estimate—which might feel strange, but appropriately reflects the Relative Risk scale.

Asymmetric Confidence Interval



Practice: Let's again use our simulation for help. Change the "Statistic" option from "Difference of proportions" to "Ratio of proportions" and then check the 95% confidence interval option.
https://istats.shinyapps.io/Association_Categorical/

What is our 95% confidence interval for the **relative risk** of polio with the vaccine relative to the placebo?

How might we use this confidence interval to generate a 95% confidence interval for the true **effectiveness** of the Salk Vaccine in preventing polio (relative to the placebo)?

Interpreting Relative Risk and Effectiveness

Researchers found the Pfizer vaccine to be 95% effective in preventing the original strain of the SARS-CoV-2 virus. https://www.yalemedicine.org/news/covid-19-vaccine-comparison. This implies that the relative risk for SARS-CoV-2 when taking the vaccine was approximately 5%. Which statement below correctly interprets the **relative risk**? Which statement interprets **effectiveness**?

1. The risk for SARS-CoV-2 with the vaccine was **5% lower** relative to the placebo injection

2. The risk for SARS-CoV-2 with the vaccine was **95% lower** relative to the placebo injection

Effectiveness

3. The risk for SARS-CoV-2 with the vaccine was **5% of** the risk with the placebo injection

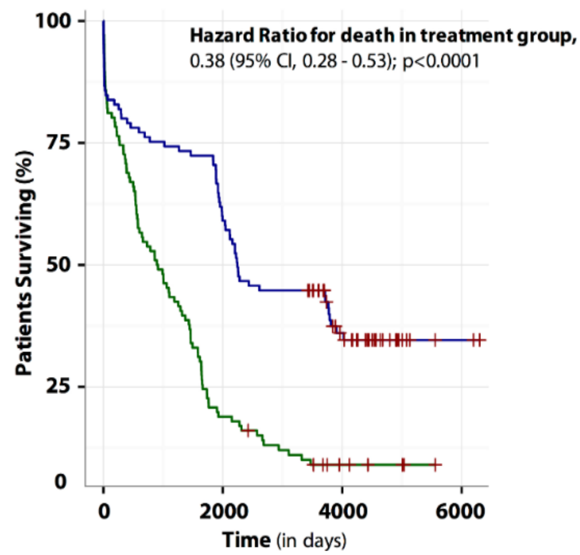
Relative Risk

4. The risk for SARS-CoV-2 with the vaccine was **95% of** the risk with the placebo injection

Introduction to Survival Curves

- What is Survival Analysis?
  - Before, we focused on identifying the risk of some event happening over a fixed period of time.
  - However, we could also bring time in as an additional variable by asking how the risk might change based on the time frame we are looking at.
  - As a method, we term this <u>Survival analysis</u>, but keep in mind that the outcome is not always death/survival—it may simply be risk of disease, hospitalization, impairment, or some other non-fatal outcome.

Investigation: Consider this fictional study (as described from the link below) designed to study the effectiveness of a new mediation to improve survival rates among patients with heart failure. Patients were either assigned to the new medication or to some control group receiving standard care. The results are presented in the following image.



Albarqouni (2016)
https://s4be.cochrane.org/blog/2016/04/05/tutorial-hazard-ratios/

Population of interest: <u>patients with heart failure</u>

Unit of observation: <u>1 patient</u>

Response variable: <u>Survival/Death</u>

Explanatory variable: <u>Receiving medication or not</u>

- **Reading Survival Curves (Kaplan-Meier Plots)**
  - The x-axis typically represents <u>time</u> since the study period or experiment began.
  - The y-axis typically represents the percentage of patients <u>surviving</u> to that time point
  - The groups in question are each represented as a separate <u>line</u>.
  - Some curves *may* also include markers to identify when patients are "censored" from the study
    - These may be patients who do not remain in the study and cannot be observed for the full study period. This may also be due to patients who joined late.

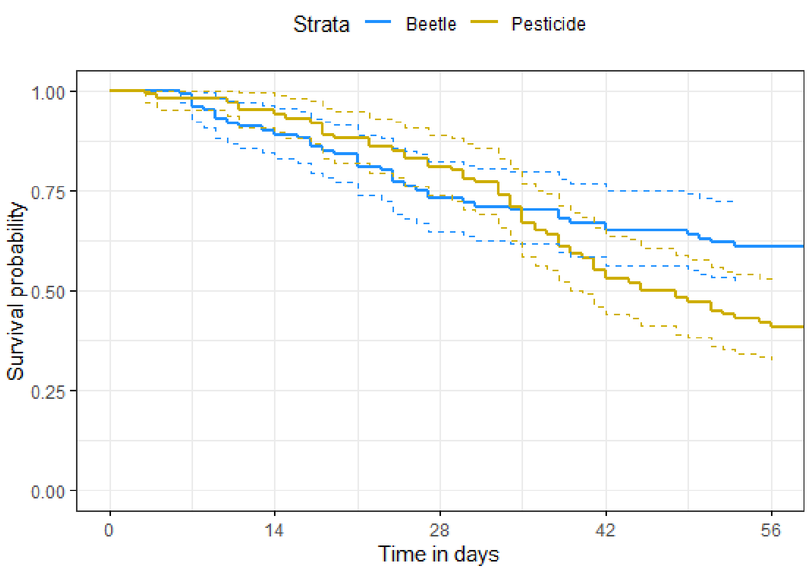
The data indicates that the medication group had higher survival/lower mortality rates than the control group. Which line would you expect to be the medication group?

Approximate the median survival time for each group. *In other words, we estimate that 50% of people will survive for how long under each condition?*

- **Hazard Ratio (HR) vs. Relative Risk (RR)**
  - A hazard ratio (HR) is essentially the same as a relative risk/risk ratio (RR), but refers to the relative risk over a specified period of time.

Notice the HR reported on the previous plot, along with the 95% confidence interval and p-value. This refers to the hazard ratio after the first 1500 days. What might we say about the effectiveness of the medication in the first 1500 days relative to standard care?

Investigation: Botanists are trying to stop lettuce plant damage due to Western Flower Thrips, a plant-eating bug that rots the plant. The new Treatment is to embed the natural predator Rove Beetle and the standard treatment is to use the standard pesticide. 200 lettuce plants are randomly assigned to each treatment.



David Cappaert, bugwood.org

Unit of observation: One lettuce plant

Response variable: Whether plant has been infected by thrip

Explanatory variable: Treatment type

What is the median survival rate among plants treated with the pesticide and among plants treated with the beetle embedding?

Around 48 days for the pesticide group, and somewhere above 56 days for beetle group

Is one treatment consistently more effective than the other, or does it depend on the time frame?

Chapter 8 Additional Practice

Practice: A study was conducted to examine the effectiveness of an experimental brain stimulation treatment on patients with a traumatic brain injury (TBI). Of 143 patients recovering from TBI, 72 were randomly assigned to a brain stimulation treatment in addition to standard medication while the other 71 were assigned to the standard medication treatment. Results are shown in the table below showing the mortality rate of patients after 6 months.

	Death	Survival	Totals
Brain Stimulation	21	51	72
Standard Medication	28	43	71

What is the absolute risk reduction in death by taking the brain stimulation treatment rather than standard medication?

Estimate the relative risk for death for those in the brain stimulation intervention relative to the standard medication intervention. (Try it by hand, then check with the simulation!)

If we were testing whether or not there was a difference in risk for death between the brain stimulation and standard medication interventions, how might we write our null and alternative hypotheses?

The 95% confidence interval for relative risk is calculated to be (0.466, 1.170). What does this tell you about how confident we are in the brain stimulation treatment being better at preventing death? *In other words, do we expect the p-value from a z-test for two proportions to be above 0.05 or below 0.05?*

Practice: A research article appearing in the *American Journal of Obstetrics and Gynecology* published findings on the association between marijuana use and pregnancy-related issues. The researchers examined the records of 12,069 pregnancies and compared the likelihood for several adverse outcomes among marijuana users and non-users. The following 5 outcomes were higher for marijuana-users in those comparisons.

Possible Outcomes	Relative Risk and 95% CI	P-value from HT
Maternal-related Asthma	3.30 (1.52, 7.17)	0.003
2 or more mental health issues	5.97 (3.01, 10.78)	<0.001
Head circumference <25^{th} percentile	1.44 (0.82, 2.53)	0.202
Birthweight <25^{th} percentile	1.09 (0.61, 1.95)	0.763
Hypertension	1.30 (0.68, 2.50)	0.42

Which outcomes are we **not** especially confident concluding are truly higher for the marijuana group? What indicates that?

Chapter 8 Learning Goals

After this chapter, you should be able to…

- Distinguish a comparison of proportions from a comparison of means (is the response variable categorical or numeric?)
- Recognize when a z-test for a difference in proportions is appropriate
- Be able to interpret a test result for a difference in proportions
  - Identify the null and alternative hypotheses
  - Identify the p-value as probability of observing a difference in sample proportions at least this large by random chance
- Calculate and interpret absolute risk reduction as the absolute difference in risk between two conditions
- Calculate and interpret relative risk as the proportion of risk still present under one condition/treatment relative to some comparative condition/treatment.
- Calculate and interpret effectiveness as the proportion of cases for which the risk is avoided as a result of some treatment.
- Recognize the null hypothesis under a risk comparison test as equivalent to RR = 1 (which is functionally the same as $\pi_1 = \pi_2$)
- Interpret a p-value or confidence interval in the context of Relative Risk, including the relationship between an interval including/not including 1 and the p-value we would expect from a z-test
- Interpret a Kaplan-Meier Plot
  - Identify the survival rate of one condition at any particular time
  - Identify the median survival time under a particular condition
  - Approximate the hazard ratio at any particular time