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 100 Humans Risk and Music video to learn more about this experiment.

Unit of Observation:

Response variable (and type):

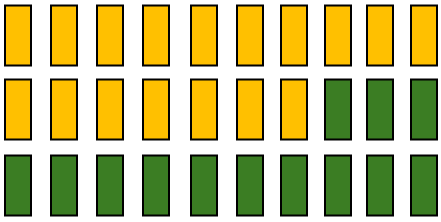
Explanatory variable (and type):

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

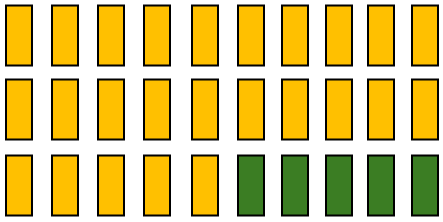
Based on the percentages reported at the end (and assuming about 30 people per group), we can *approximate* the frequencies and place them in the following table.

Table 1. Music and Risk Frequency Table

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



Sad Music Group
56.7% took risk



Happy Music Group
83.3% took risk

0.567 – 0.833 = _____

But…could this difference be explained as _____ _____?

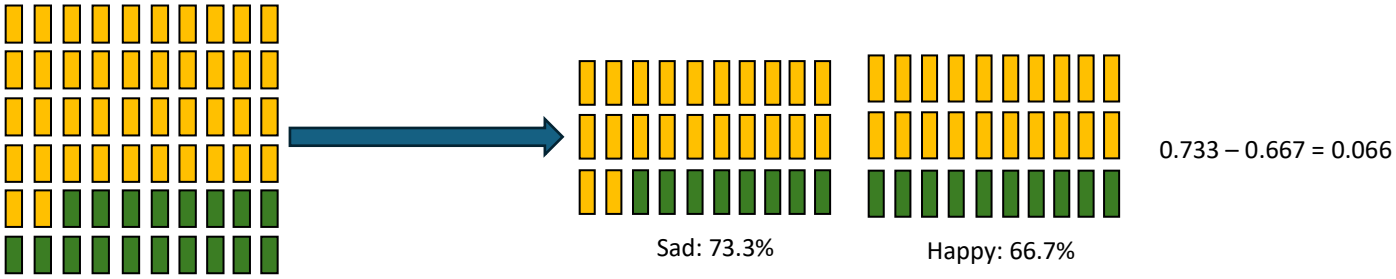
Hypothesis Testing for a Difference in Proportions

- When testing for a difference in proportions, we commonly test the null hypothesis that $\pi_1 = \pi_2$. Or to say it another way, $\pi_1 - \pi_2 = 0$.
- With that in mind, let's choose our Null Model to be the distribution of $\hat{p}_1 - \hat{p}_2$, and under the null hypothesis, this distribution should have a mean of __.

What did the researchers hypothesize about how music would affect risk-taking behavior? How might we write our null and alternative hypotheses?

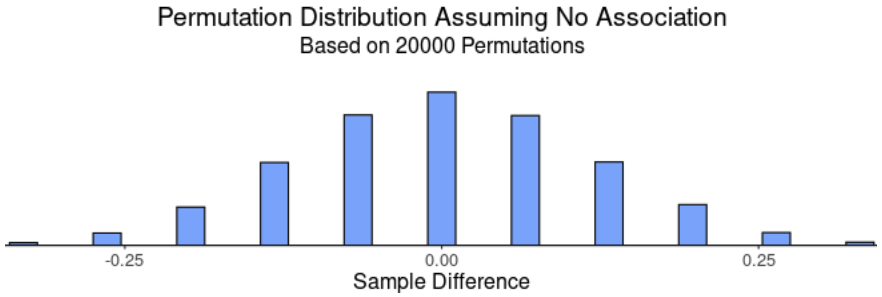
A **Permutation Test** is a non-parametric approach and would function similarly as when we had a numeric response outcome. It's valid to use at any sample size *(assuming your sampling method is representative!)*.

1. With categorical response outcomes, the **difference in our sample proportions** would serve as an **estimator** for $\pi_1 - \pi_2$.
2. We would again **assume the Null hypothesis is true** and that _____. If that's the case, the grouping factor is arbitrary and the difference in our data is a result of random chance.
3. To determine how plausible that is, we would **take the outcomes** and **assign them to each group randomly** to see what difference in proportions might occur by _____ _____.



… We would simulate doing this many many times!

4. Create a **permutation sampling distribution** to represent the distribution of possible sample proportion differences that truly occur by random chance.



5. And finally, we would **calculate our p-value** as the proportion of simulations producing sample proportions at least as extreme as what we observe

Observed Value	# of Permutations ≤ Observed Value	Percentile (%)
-0.27	242	2.42

In addition to the permutation test approach, you may also see one of the following used.

- A **Z-test for two proportions** is a parametric test option. It's generally reliable when your sample size is ≥ 100 and there are at least 10 observations of each possible response.
- You will also commonly see scientists use a "_____ _____ **test**" when comparing proportions as well. In the case of two proportions, it's computationally equivalent to a **z-test**, but it can be flexibly applied to compare proportions from 3 or more groups for equality!

In this chapter, we will **not** bother with **completing these procedures by hand**. Instead, we will focus on recognizing when such tests may be appropriate and using computational tools to complete the analysis!

Framing Proportions in terms of Risk

Courtesy of Boston Children's Hospital Archive

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.

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. Here is a link to the Salk Trial Article.

Table 2. Outcomes due to Polio

	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: **All children in U.S. (at the time)**

Unit of observation: **One child**

Response variable (and type): **Whether or not participant has polio.**

Explanatory variable (and type): **vaccine vs. placebo (categorical)**

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

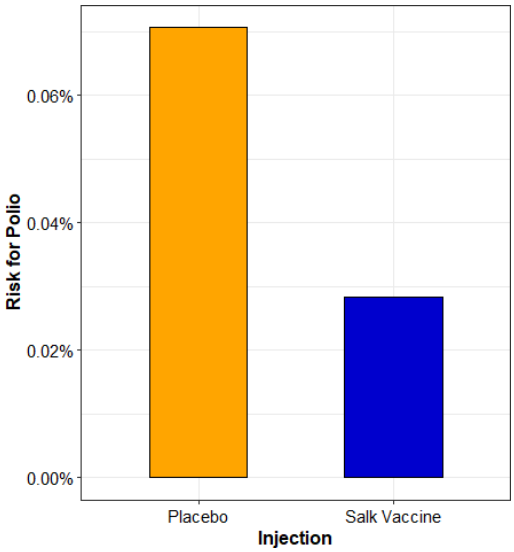
What about the proportion of children receiving the placebo?

If time: Let's input our results into a standard 2x2 contingency table.

Table 3. Contingency Table

Salk Vaccine		
Placebo		

In contexts like this, we might reasonably refer to a proportion as a _____. The proportion of children who got polio while using the placebo represents the absolute risk for polio.



Absolute Risk *Reduction* (ARR) reports the absolute value difference between risk for two different groups (it's just the difference in proportions!). *Typically reported as an absolute value.*

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

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

Practice: What is the 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 = $\frac{Risk_A}{Risk_B}$. *This can also be reported as a percentage*: $\left(\frac{Risk_A}{Risk_B}\right)*100\%$

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

- Relative risks below 1 mean that the risk for the numerator group is _____.
- Relative risks above 1 mean that the risk for the numerator group is _____.

Effectiveness (EFF): Represents the proportion of individuals that would ***avoid*** the infection by taking part in the intervention. *This will often be reported in clinical trials for vaccines and other treatments.*

EFF = 1 – RR. *This can also be reported as a percentage:* (1 – RR)*100%

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

If Time: **Number Needed to Treat (NNT)** would be a way to determine how many people would need to be treated before we would expect to prevent one adverse event.

NNT = 1/ARR

Practice: How many children would we need to treat in order to prevent one case of polio on average?

Reflection Questions

8.1. Can you explain how a permutation test works? How would we use it to determine whether the difference in sample proportions between our two groups could reasonably be explained as random chance?

8.2. Besides the permutation test, what other test(s) might be used to compare two (or more) proportions for equality?

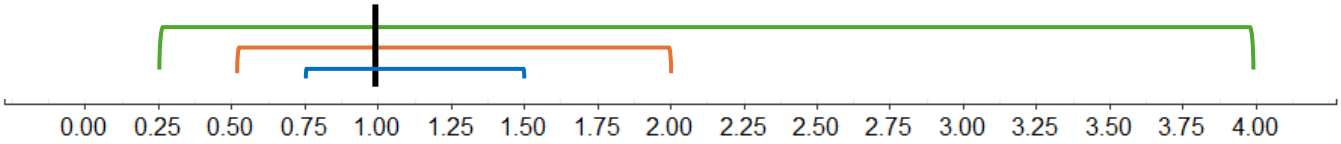
8.3. In a particular population, 4% of residents are diagnosed with Parkinson's disease by age 60. A clinical trial finds that those who complete a preventative treatment have only a 1% probability of Parkinson's by age 60. News A reports this vaccine reduced the risk by 3%, while News B reports this vaccine reduced the risk by 75%. What names would you give the measures reported by News A and News B?

8.4. This same vaccine for Parkinson's has a number needed to treat of approximately 33. What does that mean? How might you have used the previous information to calculate this number on your own?

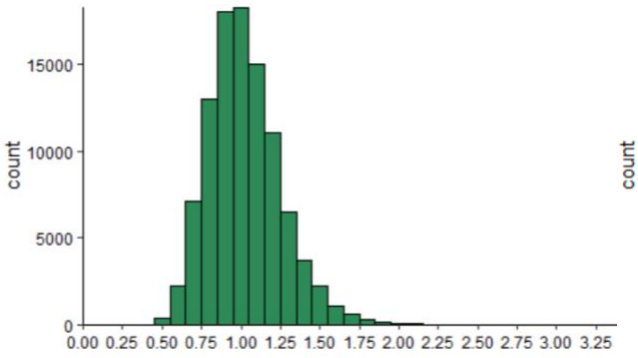
Confidence Intervals for Relative Risk (RR)

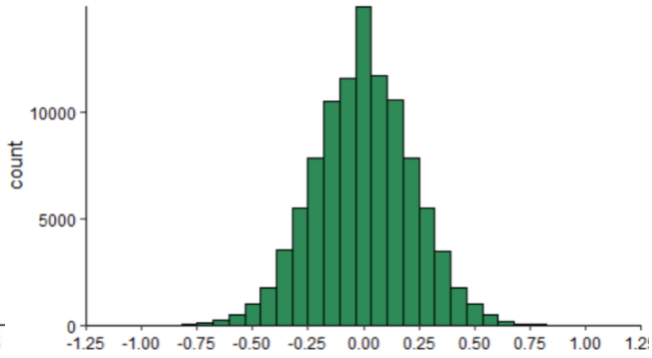
- We still need to acknowledge that we have only a sample of children in this study, and our calculations for ABR, RR, and Effectiveness are all sample statistics.
- We acknowledge that when testing for a difference in proportions/risk, we could write the null hypothesis in one of two ways:

- 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 the 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 $\widehat{RR}$ when the true parameter RR = 1 and each group size = 100.
  - The distribution on the right represents the sampling distribution for the log($\widehat{RR}$) when RR = 1 and each group size = 100.



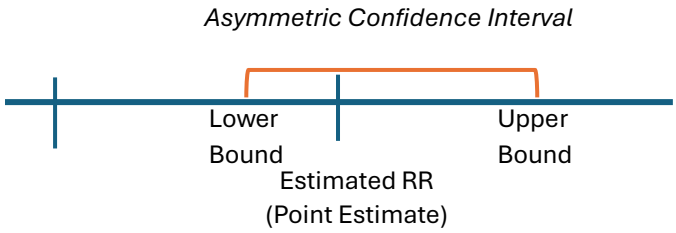


  - By finding a confidence interval for log($\widehat{RR}$), and then converting back, we have a confidence interval that works. In this class, we will only focus on **interpreting** these calculations, ***not on doing the calculations by hand.***

Computation for Relative Risk Confidence Interval (just for reference)

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

- The confidence interval will be asymmetric about the point estimate—which might feel strange, but appropriately reflects the relative risk scale.



Practice: Rather than do this calculation by hand, let's use the **MedCalc online statistical calculator** to help us. Choose Relative risk and enter our frequency counts to complete the analysis.

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

Interpreting Relative Risk and Effectiveness

The Pfizer vaccine was found to be 95% effective in preventing the original strain of the SARS-CoV-2 virus according to Yale Medicine Report. This implies that the relative risk for SARS-CoV-2 when taking the vaccine, was approximately 5%. Which statement correctly interprets **relative risk**? Which interprets **effectiveness**?

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

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

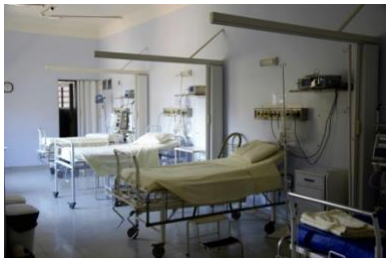
3. The risk for SARS-CoV-2 with the vaccine was **5% of** the risk relative to the placebo injection.

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

Basic Introduction to Survival Analysis

What is survival analysis?

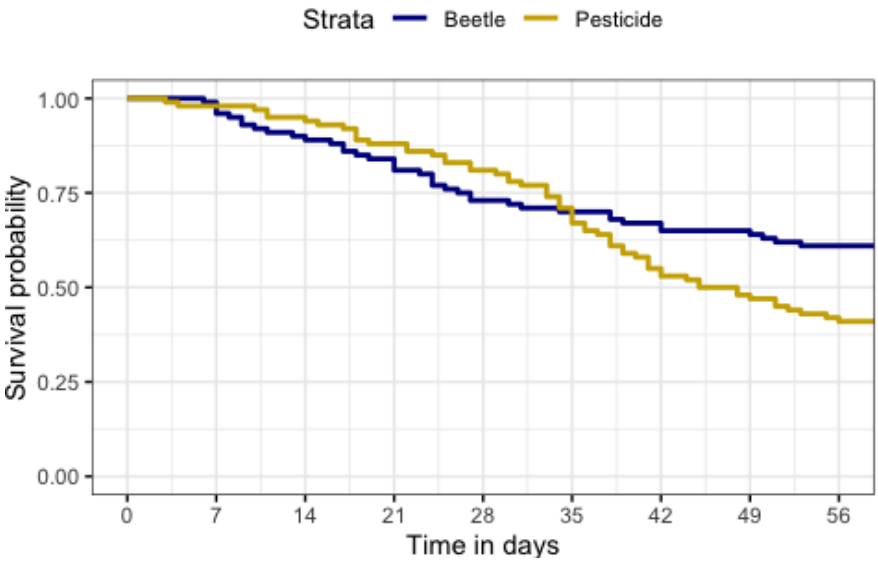
- Before, we focused on identifying the risk of some event happening over a _____ period of time. But comparing risks after a fixed point in time may not tell the whole story!
- Bringing in time as an additional variable allows us to see how the risks may vary at different time points!
- As a method, we term this Survival analysis, but keep in mind that the outcome is **not** always death/survival—it may be time until disease contraction, hospitalization, etc.



- **Reading Survival Curves (Kaplan-Meier Plots)**
  - The x-axis typically represents _____ since the study period or experiment began.
  - The y-axis typically represents the percentage of patients _____ to that time point
  - The groups in question are each represented with a line.
  - The response variable will no longer be "whether or not event occurred," but rather "time until event occurs."

Practice: Botanists are trying to stop 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 plants are randomly assigned to each treatment. The research team examines the plants each day to note whether flower thrip invasion has occurred or not by that day.

David Cappaert, bugwood.org



Unit of observation:

Response variable:

Explanatory variable:

What was the median survival time for plants under each condition?

What does the plot suggest about the long-term effectiveness of each treatment in preventing flower thrips?

- **Censored Data**
  - Participants that have "survived" (i.e., avoided the event of interest) for as long as we have observation for them, but who have not been observed for the _____ study period.
  - This is especially common in human studies where people's data may not be completed for one of many reasons! Most commonly:
    - Patients _____/no longer stay in contact before study concludes.
    - Patient dies from a _____ cause than what we are tracking.
    - Patients are recruited into a study or start treatment _____ than others, which means we have fewer days of observation.
  - Censored data is represented as vertical tick marks, representing each patient that survived to at least that point, but whose status is unknown afterwards.

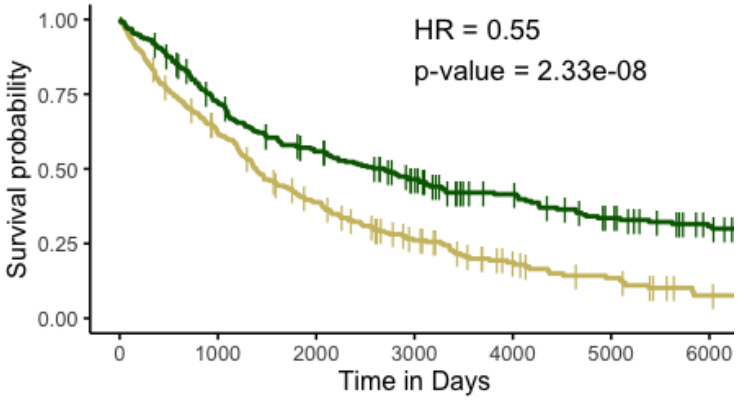
Investigation: Consider this fictional study to examine the effectiveness of a new medication to increase life expectancy among patients after stage 4 colon cancer diagnosis. Patients were either assigned to the new medication or to the control group to receive standard care.

Population of interest:

Explanatory variable:

Response variable:

The data indicates that the medication group had the longer life expectancy. Which line would that be?



- **Hazard Ratio (HR) vs. Relative Risk (RR)**
  - A hazard ratio (HR) and relative risk (RR) both compare two values as a ratio, with 1 therefore representing no difference. The distinction is that hazard ratios compare the _____ at which this event occurs between groups. It can also incorporate _____ data!
  - It is the probability ratio for experiencing the event of interest for the "survivors" at any particular period of time between the two comparison groups.
  - A p-value or confidence interval is common to report to determine evidence of a difference.

What does the **hazard ratio** communicate in this example? What does the **p-value** communicate?

Reflection Questions

8.5. If a 95% confidence interval for relative risk includes the value 1, can we claim evidence for a difference in risks with much confidence? What is special about 1?

8.6. What is censored data? What are common reasons why there might be censored data in a study?

8.7. How is a hazard ratio similar to a relative risk? In what data contexts would I use a hazard ratio rather than relative risk?

Chapter 8 Additional Practice (Videos available in the Ch 8 module on Canvas!)

Practice: In this Marijuana and Birth Defects Article appearing in the *American Journal of Obstetrics and Gynecology*, 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.

Table 4. Marijuana and Birth Defects Findings

Possible Outcomes	Relative Risk and 95% CI	P-value
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 <25th percentile	1.44 (0.82, 2.53)	0.202
Birthweight <25th 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 using $\alpha = 0.05$? Could we have made the same determinations using only the confidence intervals?

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 just the standard medication. Results are shown in the table below showing the mortality rate of patients after 6 months.

Table 5. TBI Study Frequencies

	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.

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? *Or more simply, would we expect the p-value from a test for a difference in proportions to be above or below 0.05?*