## 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):

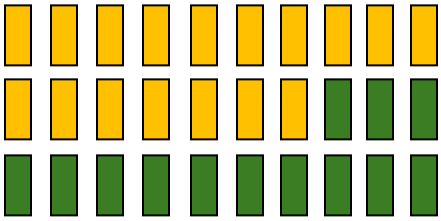
Explanatory variable (and type):

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

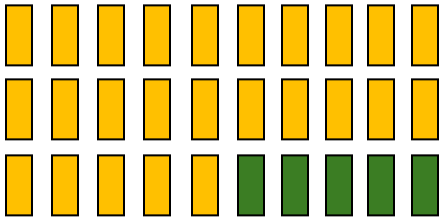
Based on the percentages reported at the end (and assuming more like 42 people per group), we can *approximate* the frequencies and place them in the following table

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
57% took risk



Happy Music Group
83% took risk

0.57 – 0.83 = _____

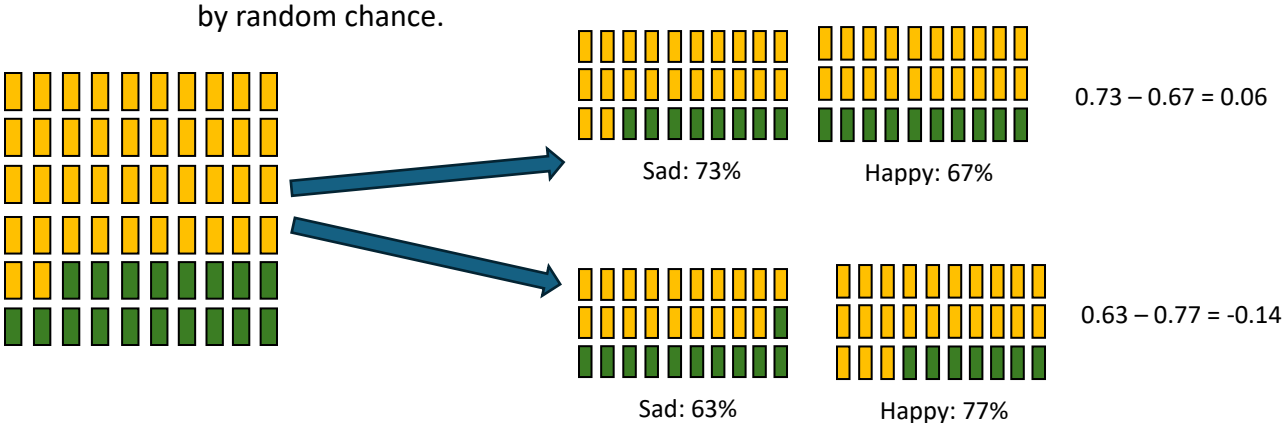
*But…could this difference be explained as _____ _____?*

**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?

- We'll again start with a **Permutation Test**, but this time, with a *categorical* response variable
    - If the null hypothesis is true (likelihood of taking risk is the same for each music condition), then we could again _____ the music labels randomly to see what sample proportions occur by random chance.




Let's use the **Association between Two Categorical Variables** sim from Art of Stat web apps to explore this idea further. We will do that by…

- Choosing the contingency table options and entering our data and labels.
- Visualizing the data more clearly with a stacked barplot.
- Switching to the "Permutation Distribution" tab to explore how often this difference could happen by random chance.

How often do we see the Sad music group at least 26% less willing to take the risk than the Happy music group when leaving it up to random chance to create our groups?

- There are several other **parametric** tests that might yield more precise results and in some cases, are better able to detect differences. The 10/10 rule is still a good rule of thumb here.
    - We might use a **z-test for two proportions** for the simple case of comparing two proportions.
    - A **Chi-Square Test for Independence** would yield the same result and has the advantage of being more flexible—it can be used to compare 3 or more proportions!
        - Note that this is what R will use when we complete `props.test`
        - One disadvantage is that it's a little more computationally involved.

- Completing these tests **computationally by hand** is **not** a goal of this course. But the calculations for a z-test for two proportions are included below for reference!

**Computation for z-test for two proportions (just for reference)**

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

**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:

Unit of observation:

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

Explanatory variable (and type):

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

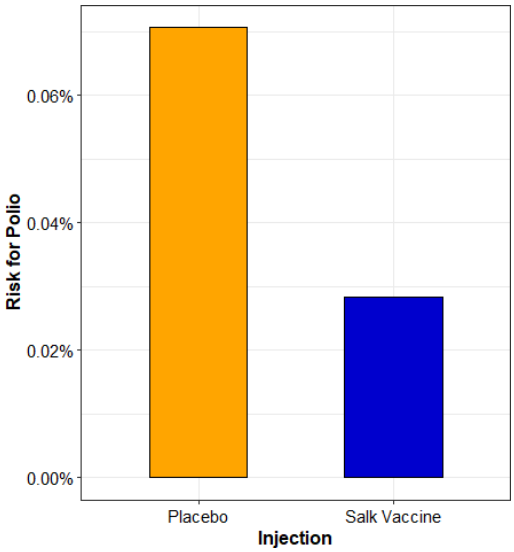
What about the proportion of children receiving the placebo?

**If time:** We could also use the Association for Categorical Variables simulation to double check our calculations and visualize these results! But we'll need to rewrite our values into a standard contingency table format.

Table 3. Contingency Table

| | | |
|---|---|---|
| **Salk Vaccine** | | |
| **Placebo** | | |

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 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.

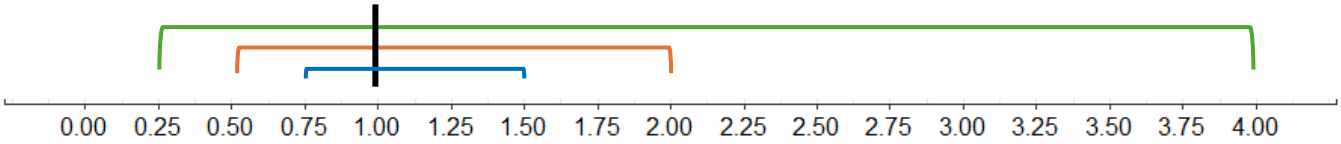
**NNT = 1/ARR**

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

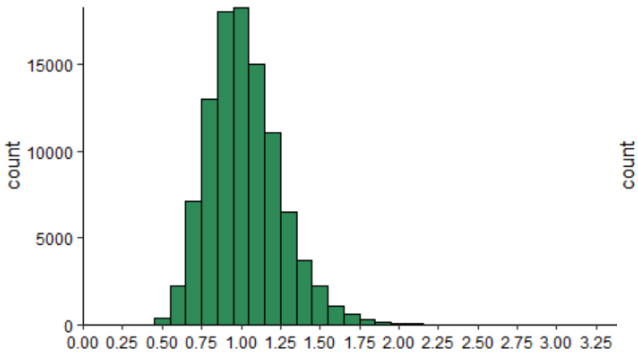
**Confidence Intervals for Relative Risk (RR)**

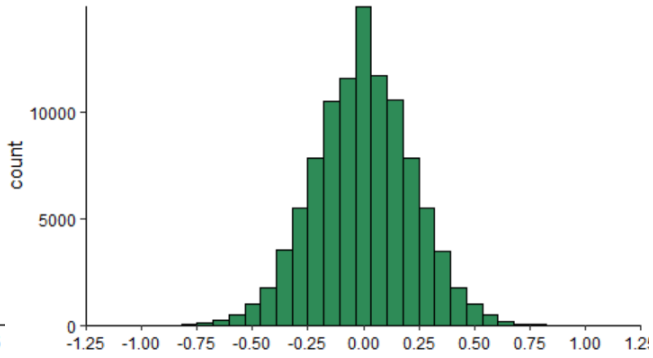
- We still need to acknowledge that we have only a sample of children in this study, and our calculations for ABR, RR, and Effectiveness are all sample statistics.
- We acknowledge that when testing for a difference in proportions/risk, we could write the null hypothesis in one of two ways:

- 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.



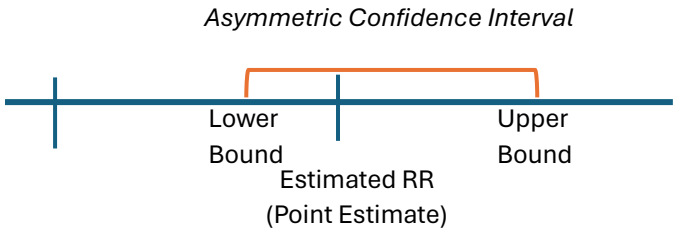




  - By finding a confidence interval for log($\widehat{RR}$), and then converting back, we have a confidence interval that works. In this class, we will only focus on **interpreting** these calculations, ***not on doing the calculations by hand.***

**Computation for Relative Risk Confidence Interval (just for reference)**

$\log(\widehat{RR}) \pm z^* SE_{\log(\widehat{RR})}$ *...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:** Let's again use our simulation for help. Change the statistic option from "Difference of proportions" to "ratio of proportions" to get our relative risk value. Then check the 95% confidence interval option.
Association for Categorical Variables simulation

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 **effectiveness** of the Salk Vaccine in preventing polio?

**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 according to this Yale Medicine Report. This implies that the relative risk for SARS-CoV-2 when taking the vaccine, was approximately 5%. Which statement below correctly interprets **relative risk**? Which correctly 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.

**Introduction to Survival Curves**

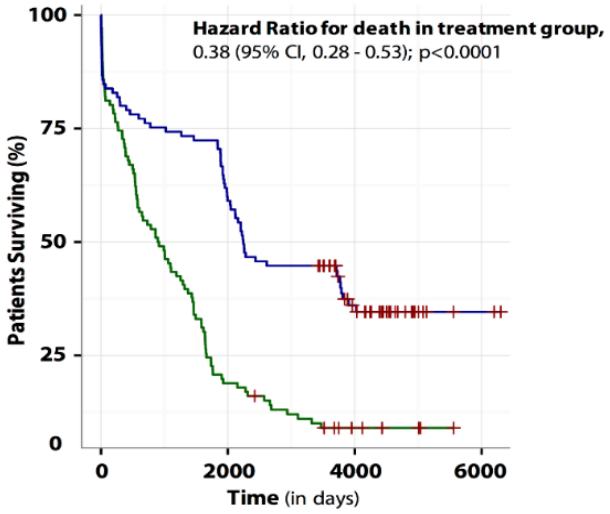
- What is survival analysis?
    - Before, we focused on identifying the risk of some event happening over a fixed period of time.
    - However, we could also bring in time as an additional variable by asking how the rates of incidence compare between two different conditions.
    - As a method, we term this Survival analysis, but keep in mind that the outcome is not always death/survival—it may be disease contraction, hospitalization, impairment, etc.

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

Population of interest:

Explanatory variable:

Response variable:




Albarqouni (2016)

- **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.
    - Some curves *may* also include markers to identify when patients are "censored" from the study
        - Patients drop out, or they joined late and cannot be observed for length of study. These situations require additional analytical methods that we will forego in this class

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

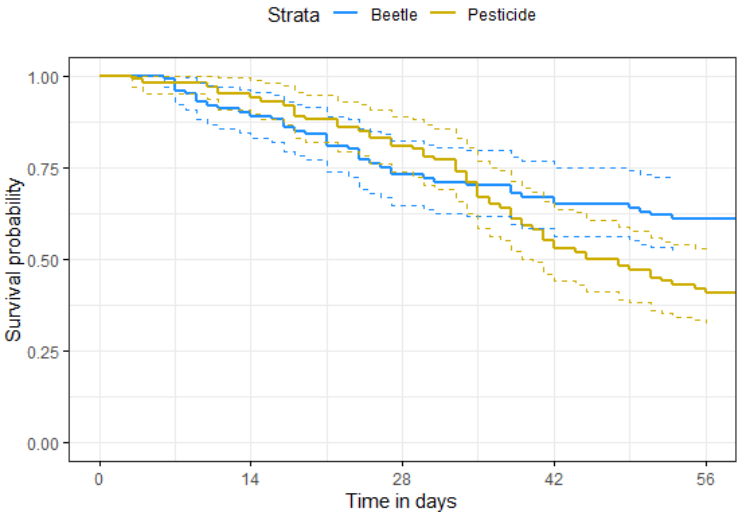
Approximate the median survival time for each group. *In other words, we estimate that 50% of people will survive for how long under each condition?*

- **Hazard Ratio (HR) vs. Relative Risk (RR)**
  - A hazard ratio (HR) and relative risk (RR) both are comparing two conditions as a ratio.
  - The difference is that a hazard ratio incorporates time and is effectively comparing the _____ at which some event happens under each condition.

Notice that the HR reported on the previous plot, along with the 95% confidence interval and p-value. Let's rewrite them here.

What might we say about the effectiveness of the medication relative to standard care?

**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.




**David Cappaert, bugwood.org**

Unit of observation:

Response variable:

Explanatory variable:

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

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

## Chapter 8 Reflection Questions

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?

What are the names of the two *parametric* tests we learned about that we might use when testing for a difference in two proportions? *One of them looks like the name of a tea, but is pronounced differently!*

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?

This same vaccine for Parkinson's has a number needed to treat of 33. What does that mean?

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?

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

How would you use a survival curve to estimate the median "survival" time (median time until the outcome of interest) for a particular group?

## Chapter 8 Additional Practice (if you need it!)

**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 4. 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? In other words, do we expect the p-value from a test for a difference in proportions to be above or below 0.05?

**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 5. Marijuana and Birth Defects Findings

| Possible Outcomes | Relative Risk and 95% CI | P-value from Hyp. Test |
|---|---|---|
| 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 using $\alpha = 0.05$? Could we have made the same determinations using only the confidence intervals?