



Mole Count Study in Colorado Children

Presentation by Ivy Chang

Table of contents

- 01 Research Background
- 02 About the Dataset
- 03 Research Questions
- 04 Hypotheses
- 05 How Will the Data Test the Hypotheses
- 06 How Will the Findings be Used
- 07 Presenting Analyses
- 08 Summary
- 09 Next Steps



Research Background

- Both melanoma and moles arise from melanocytes, and in many cases, moles are thought to be intermediate in the pathway from melanocyte to melanoma
- The presence of moles is the strongest predictor of melanoma risk, and individuals with high numbers of moles have a 6-7 fold increased risk of melanoma
- Most moles develop in childhood
- Sun exposure is the major determinant of mole prevalence
- An understanding of the genetic, environmental, and behavioral factors that contribute to mole growth and development is essential for understanding the pathways to melanoma

For reference, the research paper can be found [here](#).

About the Dataset

For reference, the original dataset can be found [here](#).

- The dataset being used for this analysis contains information on the mole count of Colorado children age 6 followed from baseline to age 10.
- The data is collected during the period from 2004 to 2008 .
- The data includes 472 records across 15 fields, but we will be focusing only on the fields used in the analysis, which are as follow:
 - molecount2008
 - number vacs birth thru 2007
 - gender
 - hispanic
 - Eyecolor
 - Baseskincolor
 - Haircolor

Research Questions

- **Question 1:** Is there a difference between the mole count for different gender, different ethnicity, different hair color, and different eye color?
- **Question 2:** Is there a relationship between the base skin color and the mole count, and is there a relationship between the number of vacations taken and the mole count?

Hypotheses for Question 1

Demographic Factors

Hypothesis 1

H_0 : There is no statistically significant difference in the mean of the mole count between male and female

H_a : There is a statistically significant difference in the mean of the mole count between male and female

Hypothesis 2

H_0 : There is no statistically significant difference in the mean of the mole count between Hispanic and Non-Hispanic groups

H_a : There is a statistically significant difference in the mean of the mole count between Hispanic and Non-Hispanic groups

Hypotheses for Question 1

Genetic Factors

Hypothesis 3

H_0 : There is no statistically significant difference in the mean of the mole count between light hair color and dark hair color groups

H_a : There is a statistically significant difference in the mean of the mole count between light hair color and dark hair color groups

Hypothesis 4

H_0 : There is no statistically significant difference in the mean of the mole count between blue or green eye color and brown or hazel eye color groups

H_a : There is a statistically significant difference in the mean of the mole count between blue or green eye color and brown or hazel eye color groups

Hypothesis 5

H_0 : There is no statistically significant difference in the mean of the mole count between lighter base skin color and darker base skin color groups

H_a : There is a statistically significant difference in the mean of the mole count between lighter base skin color and darker base skin color groups

Hypotheses for Question 2

Hypothesis 6

H_0 : There is no correlation between the base skin color and the mole count

H_a : There is some correlation between the base skin color and the mole count

Hypothesis 7

H_0 : There is no correlation between the number of vacations taken and the mole count

H_a : There is some correlation between the number of vacations taken and the mole count

How Will the Findings be Used

The parents of young children and the CDC will be interested in knowing:

- Does demographic backgrounds and genetic characteristics play a role in children's mole development
- Does the number of vacation taken have an impact on the number of moles

Knowing these things will be helpful for parents to understand their children's risk of getting melanoma. It will be helpful for the CDC to advise the general public in talking prevention efforts such as reducing sun exposure and taking preventive skin cancer screening

How Will the Data Test the Hypotheses

The data will be used to test the hypotheses in the following ways:

- t-tests will show the significance of mole count differences for different demographic groups and genetic groups
 - When $p \leq 0.5$, reject the null, statistical significance confirmed
 - When $p > 0.5$, fail to reject the null, no statistical significance found
- Visualizations will show correlation for:
 - mole count & base skin color
 - mole count & number of vacations taken.
- Pearson' correlation test will be used to test the likelihood of linear relationship in the overall population
 - Pearson correlation coefficient (r): rage between -1 and 1
 - $r = 1$: perfect positive correlation,
 - $r = -1$: perfect negative correlation
 - $r = 0$: no correlation

The Analyses - Hypothesis 1

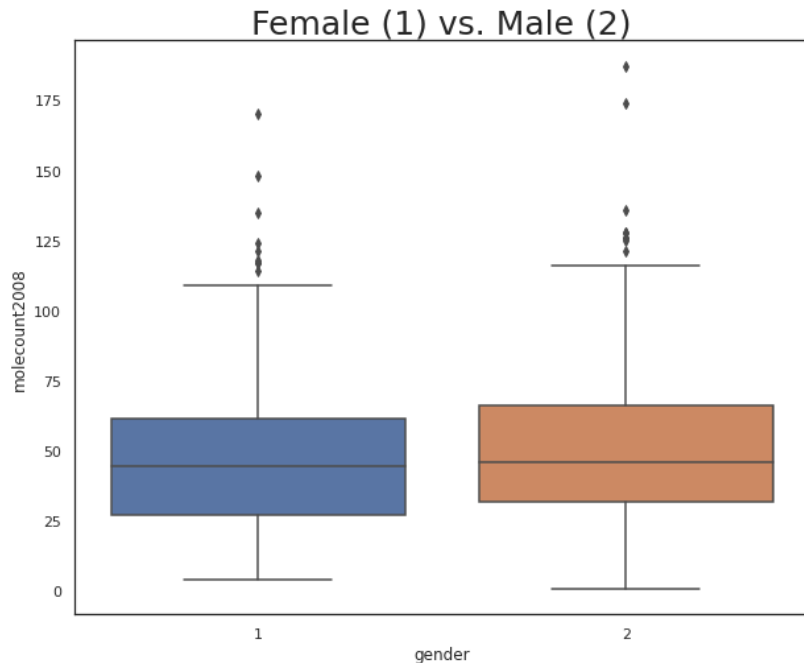
H_0 : There is no statistically significant difference in the mean of the mole count between **male** and **female**

H_a : There is a statistically significant difference in the mean of the mole count between **male** and **female**

Independent-samples t-test is used to test the hypothesis:

- p-value = 0.13715103022597452
- Fail to reject the null

There **is not** a statistically significant difference in the mean of the mole count between male and female



The Analyses - Hypothesis 2

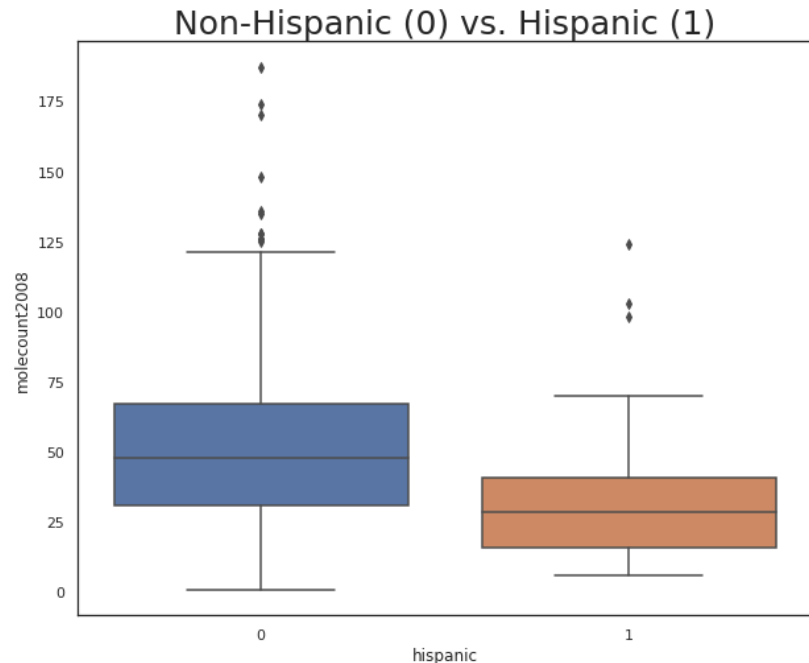
H_0 : There is no statistically significant difference in the mean of the mole count between between **Hispanic** and **Non-Hispanic** groups

H_a : There is a statistically significant difference in the mean of the mole count between between **Hispanic** and **Non-Hispanic** groups

Independent-samples t-test is used to test the hypothesis:

- p-value = 1.0401046671474124e-05
- Reject the null

There **is** a statistically significant difference in the mean of the mole count between Hispanic and Non-Hispanic groups



The Analyses - Hypothesis 3

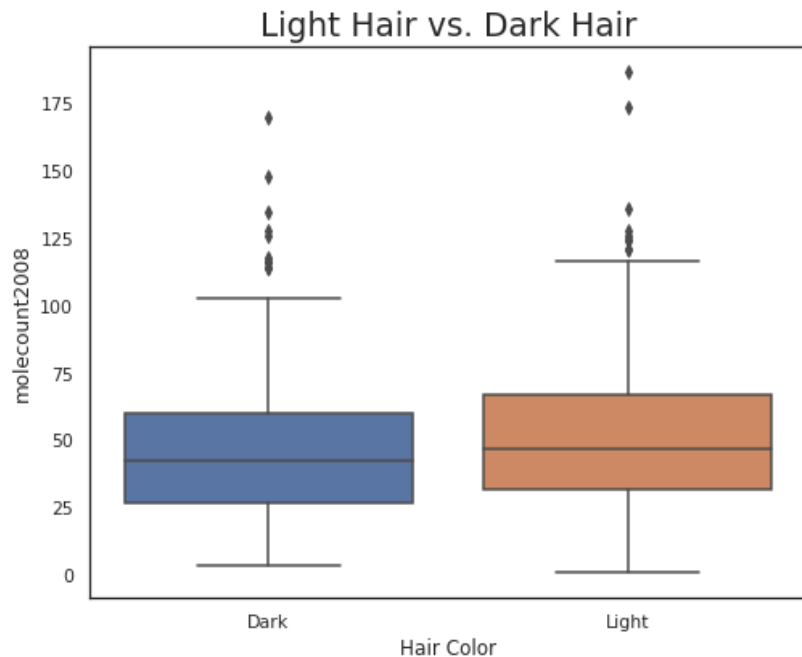
H_0 : There is no statistically significant difference in the mean of the mole count between between **light hair color** (blonde, red) and **dark hair color** (brown, black) groups

H_a : There is a statistically significant difference in the mean of the mole count between between **light hair color** (blonde, red) and **dark hair color** (brown, black) groups

Independent-samples t-test is used to test the hypothesis:

- p-value = 0.07550330806948502
- Fail to reject the null

There **is not** a statistically significant difference in the mean of the mole count between light hair color and dark hair color groups



The Analyses - Hypothesis 4

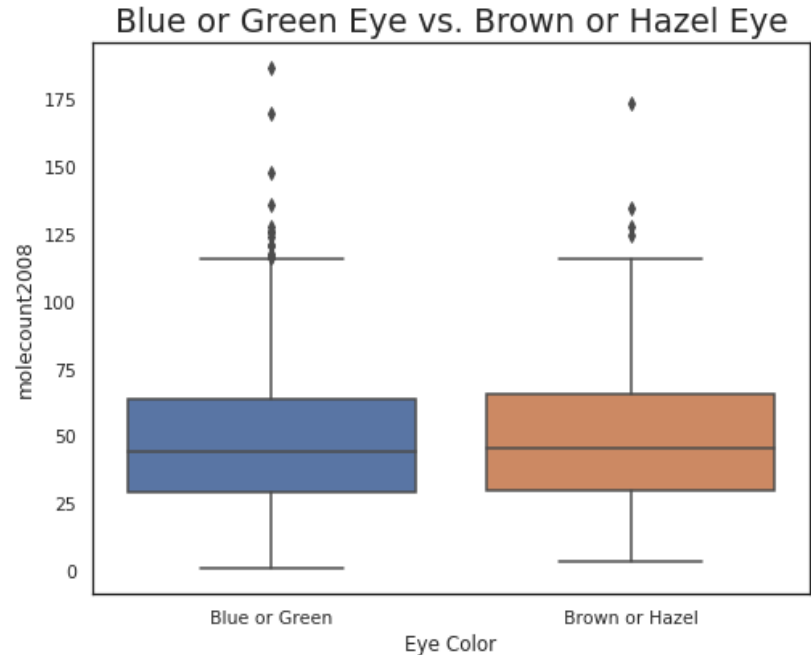
H_0 : There is no statistically significant difference in the mean of the mole count between **blue or green eye color** and **brown or hazel eye color** groups

H_a : There is a statistically significant difference in the mean of the mole count between **blue or green eye color** and **brown or hazel eye color** groups

Independent-samples t-test is used to test the hypothesis:

- p-value = 0.003132578990205146
- Reject the null

There **is** a statistically significant difference in the mean of the mole count between blue or green eye color and brown or hazel eye color groups



The Analyses - Hypothesis 5

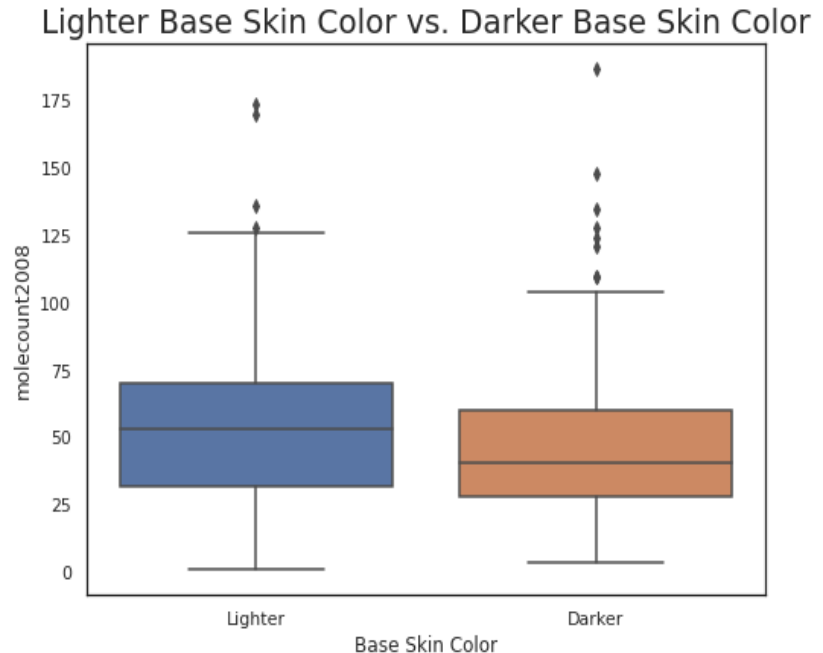
H_0 : There is no statistically significant difference in the mean of the mole count between **lighter base skin color** and **darker base skin color** groups

H_a : There is a statistically significant difference in the mean of the mole count between **lighter base skin color** and **darker base skin color** groups

Independent-samples t-test is used to test the hypothesis:

- p-value = 0.0012442142500233677
- Reject the null

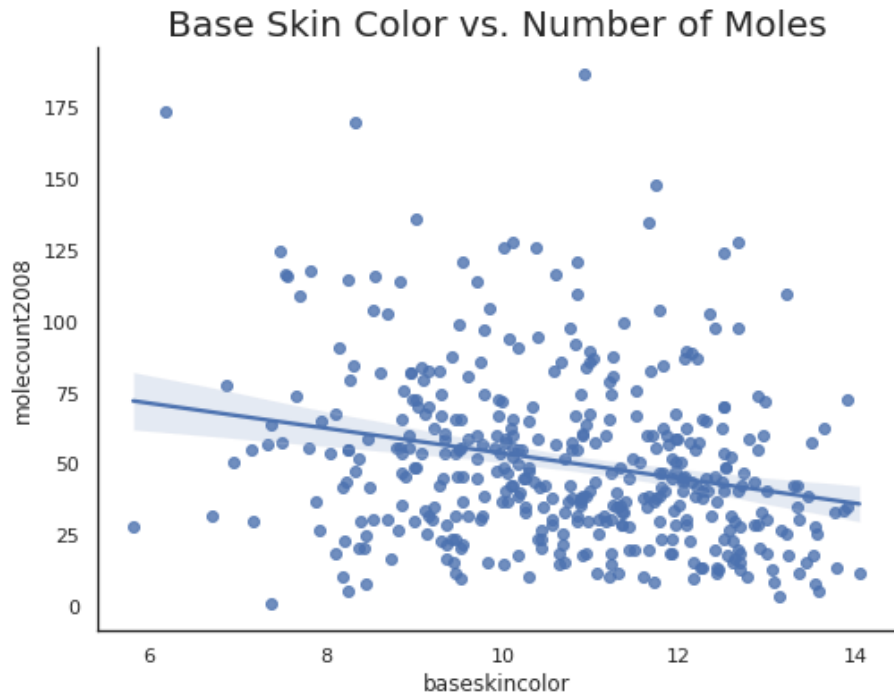
There **is** a statistically significant difference in the mean of the mole count between blue or green eye color and brown or hazel eye color groups



The Analyses - Hypothesis 6

H_0 : There is no correlation between the **base skin color** and the **mole count**

H_a : There is some correlation between the **base skin color** and the **mole count**



Since the scatter plot does not show an obvious correlation, Pearson's Correlation test is used to exam the hypothesis:

- $r = -0.23999390727705466$
- $p\text{-value} = 7.78102625749745e-07$
- Reject the null

There **is** a relationship between base skin color and mole count

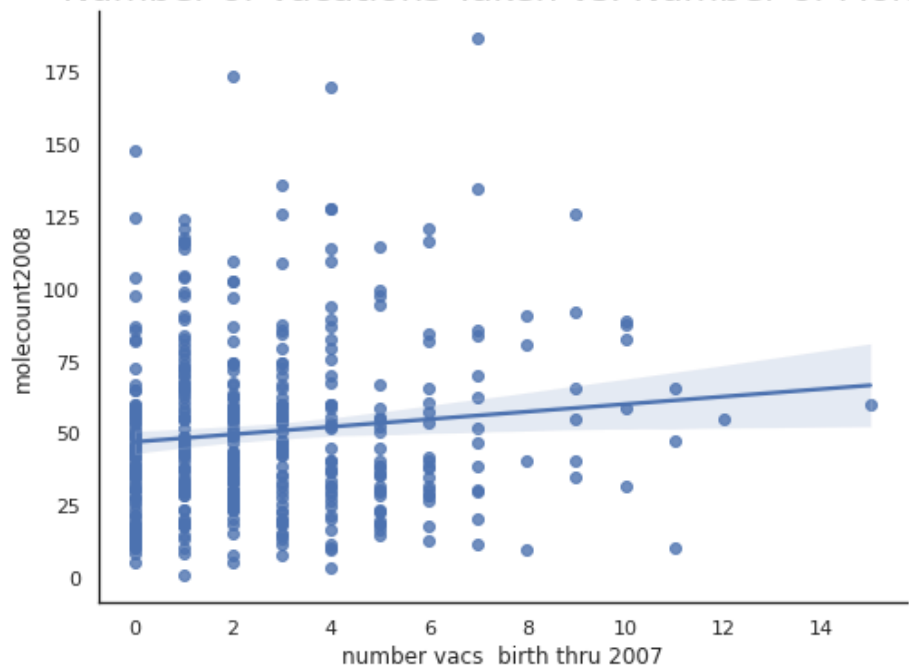
It is a negative linear relationship

The Analyses - Hypothesis 7

H_0 : There is no correlation between the **number of vacations taken** and the **mole count**

H_a : There is some correlation between the **number of vacations taken** and the **mole count**

Number of Vacations Taken vs. Number of Moles



Since the scatter plot does not show an obvious correlation, Pearson's Correlation test is used to exam the hypothesis:

- $r = 0.10859041411504917$
- $p\text{-value} = 0.02715068383576409$
- Reject the null

There **is** a relationship between number of vacations taken and mole count

It is a positive linear relationship

Summary

With our analyses on the mole count in Colorado children, we can make the following conclusions:

- There **is a** statistically significant difference in the mean of mole count between:
 - Hispanic and Non-Hispanic groups
 - Blue or Green Eye and Brown or Hazel Eye groups
 - Lighter Base Skin Color and Darker Base Skin Color groups
- There **is no** statistically significant difference in the mean of mole count between:
 - Male and Female
 - Light Hair and Dark Hair
- There **is a** relationship between mole count and:
 - Base Skin Color (negative)
 - Number of Vacations Taken (positive)

Next Step

- 01 Sun Exposure Study
- 02 Genotype Study
- 03 Longitudinal Study

Thank You!

Do you have any questions?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**

For reference:
The Google Colab Notebook is [here](#)