



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA QUÍMICA Y BIOPROCESOS

Tarea 1 - IIQ3402

Diseño estadístico, optimización y análisis multivariado

Grupo 1

Integrantes:

William Aarland, Daniel Navarro V.

Fecha de entrega: 16 de Abril de 2025

Índice

1. Análisis Teórico de los Datos	2
1.1. Muestreo y recolección de datos	2
1.2. Tipo de estudio	2
1.3. Implicancia del método de recolección de datos sobre el alcance la inferencia estadística	2
2. Preguntas de Investigación, Hipótesis y Análisis Exploratorio de Datos	3
2.1. ¿Los olores frutales son más fáciles de identificar para una persona que padece COVID - 19?	3
2.2. ¿Son las personas con comorbilidades son menos propensas a contraer COVID - 19?	3
2.3. ¿Existe una relación entre las dificultades respiratorias y la función olfativa en individuos?	4
3. Análisis Exploratorio de Datos Cualitativos	6
A. Descargo de responsabilidad	8
B. Código Python	8

1. Análisis Teórico de los Datos

1.1. Muestreo y recolección de datos

El estudio de Eyheramendy et al. [1] se llevó a cabo con la participación de voluntarios no remunerados que firmaron un consentimiento informado y proporcionaron información anónima. Los participantes fueron reclutados en 5 centros de la red UC-Christus en Santiago de Chile. En la fase inicial del estudio, los voluntarios completaron un cuestionario que evaluó la presencia de 6 síntomas asociados al COVID-19, así como información demográfica y clínica, incluyendo edad, comorbilidades (alergia, resfriado, diabetes, hipertensión, Parkinson, rinitis y Alzheimer) y hábitos de fumar. También se registró la presencia de síntomas específicos como anosmia/hiposmia, ageusia, dolor de cabeza, diarrea, fatiga y dolor de estómago y pecho. Se excluyeron del estudio a los individuos que reportaron disfunción olfativa basal debido a trauma previo o problemas de salud agudos o crónicos, como sinusitis crónica o rinitis alérgica. Posteriormente, se seleccionaron sujetos que presentaban síntomas de COVID-19 o habían tenido contacto estrecho con un caso confirmado. A estos participantes se les administró el test KOR y una prueba de PCR para determinar la presencia del virus.

1.2. Tipo de estudio

El estudio contiene datos numéricos binarios, donde 1 representa un resultado positivo o acierto en la variable, y 0 representa un resultado negativo o desacierto en la variable. En cuanto al diseño del estudio, se trata de un estudio observacional prospectivo, ya que los datos fueron recopilados durante el transcurso del estudio.

1.3. Implicancia del método de recolección de datos sobre el alcance la inferencia estadística

Dado que se trata de un estudio observacional sin aleatorización, no es posible establecer una relación causal entre las variables. Además, los autores mismos destacan que la muestra no es representativa de la población, lo que limita aún más las conclusiones que se pueden extraer. En este sentido, y como se ha discutido en clase, los estudios observacionales pueden aspirar, como máximo, a identificar asociaciones entre las variables, pero no a establecer causalidad

2. Preguntas de Investigación, Hipótesis y Análisis Exploratorio de Datos

En esta sección desarrollamos las preguntas 2 y 3 del enunciado de la tarea, es decir, las preguntas de investigación, hipótesis y análisis exploratorio de las hipótesis propuestas con la misma base de datos entregadas al inicio de la tarea.

2.1. ¿Los olores frutales son más fáciles de identificar para una persona que padece COVID - 19?

Considerando que la capacidad de distinguir olores de alimentos podría ser una ventaja competitiva dada la alta conservación de los genes asociados a esta [2, 3], nos preguntamos si los olores frutales son más persistentes en individuos con COVID-19. Varios estudios han demostrado una relación entre la pérdida del olfato y el padecimiento de COVID-19 [4, 5, 6, 7, 8, 9, 10]. Sin embargo, pocos estudios han evaluado la anosmia mediante pruebas olfativas con olores control conocidos. La presencia de olores frutales en algunos estudios [11, 12, 13] sugiere que estos olores pueden ser percibidos con mayor facilidad.

Por lo tanto, planteamos la hipótesis de que los olores frutales son percibidos un mayor número de veces en comparación con los olores no frutales.

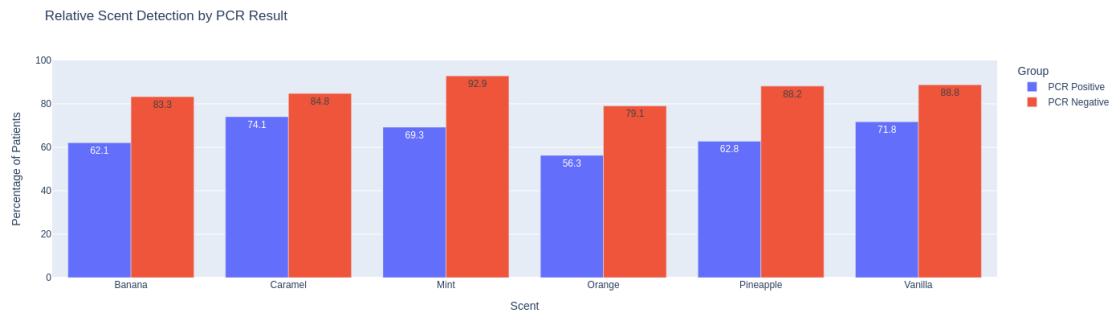


Figura 1: Comparación de la capacidad de detección de olores entre individuos que dieron positivo y negativo en las pruebas de COVID - 19 mediante PCR. El color azul fue asignado a las variables con resultado positivo para PCR, el color rojo fue asignado a las variables con resultado negativo.

En este estudio, consideramos como olores frutales al plátano, la naranja y la piña, mientras que el caramelo, la menta y vainilla como olores no frutales. Los resultados muestran que los individuos positivos a COVID-19 identifican un mayor número de veces los olores no frutales, lo que sugiere que nuestra hipótesis podría ser incorrecta. Sin embargo, es necesario aplicar tests estadísticos para determinar la validez de la hipótesis.

2.2. ¿Son las personas con comorbilidades son menos propensas a contraer COVID - 19?

Los mamíferos contamos un sistema inmune que nos defiende de las infecciones, este sistema se activa cuando detecta algún componente foráneo [14]. Basándonos en este hecho, nos preguntamos si las personas con comorbilidades tienen un umbral más alto para infectarse con el virus SARS-CoV-2. Proponemos la hipótesis de que los individuos que reportan enfermedades asociadas a cuadros inflamatorios crónicos tienen menor tasa de PCR positivos para COVID - 19, siendo estas enfermedades el dolor muscular y la dificultad respiratoria. Como medida de comparación, usaremos los datos reportados por las comorbilidades restantes.

Como se muestra en la Figura 2, en general se ve que los pacientes que reportan comorbilidades

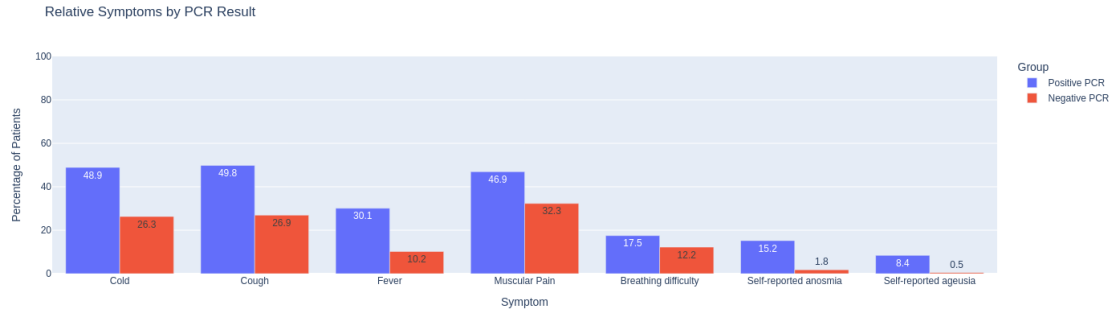


Figura 2: Porcentaje de pacientes que reportan comorbilidades y su resultado al test PCR de COVID - 19. En la figura se muestra en azul los pacientes con resultado positivo para COVID - 19 y en rojo los pacientes con resultado negativo para COVID - 19. Las categorías representan los síntomas reportados por los pacientes. El porcentaje representa los pacientes que reportaron la comorbilidad del total de pacientes con resultado positivo o negativo para el PCR de COVID - 19.

tienen mayor porcentaje de resultados positivos en el test PCR y no se aprecia un menor porcentaje de PCR's positivos en los individuos que reportaron dolor muscular y dificultad respiratoria en relación al resto de comorbilidades, por lo que nuevamente erramos la hipótesis.

Cabe mencionar que esta pregunta fue planteada ya que en el *paper* de donde se obtuvieron los datos [1] no se encontró información respecto a si las comorbilidades reportadas eran crónicas o recientes, abriendo paso a que el dato se pudiese tomar como otra enfermedad de base y no como un síntoma del padecimiento de COVID - 19.

2.3. ¿Existe una relación entre las dificultades respiratorias y la función olfativa en individuos?

Considerando que el sentido del olfato está estrechamente relacionado con la respiración [15, 16], hipotetizamos que el padecimiento de dificultades respiratorias está asociado a dificultades en la identificación de olores.

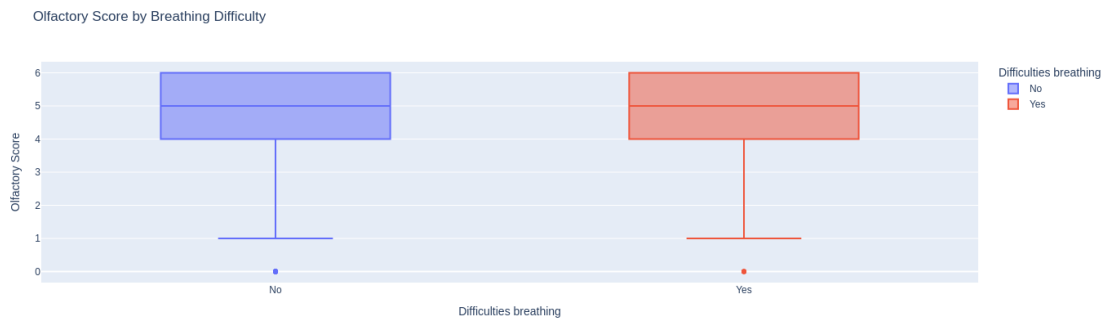


Figura 3: Puntaje en test olfativo según presencia o ausencia de dificultad olfatoria. En azul representado por 0 se muestran el puntaje de los pacientes que no reportaron dificultades respiratorias, mientras que en rojo representado por 1 se muestran los pacientes que reportaron dificultades respiratorias

El análisis del *boxplot* en la Figura 3 revela que ambas muestras presentan medias, rangos intercuartiles y dispersión de datos similares. Esta similitud se confirma al examinar la Figura 4, donde se observa que la distribución de los puntajes del test olfatorio es comparable entre los pacientes con y sin dificultad respiratoria. En ambos grupos, se aprecia un aumento en la cantidad de pacientes a medida que aumenta el puntaje del test olfatorio.

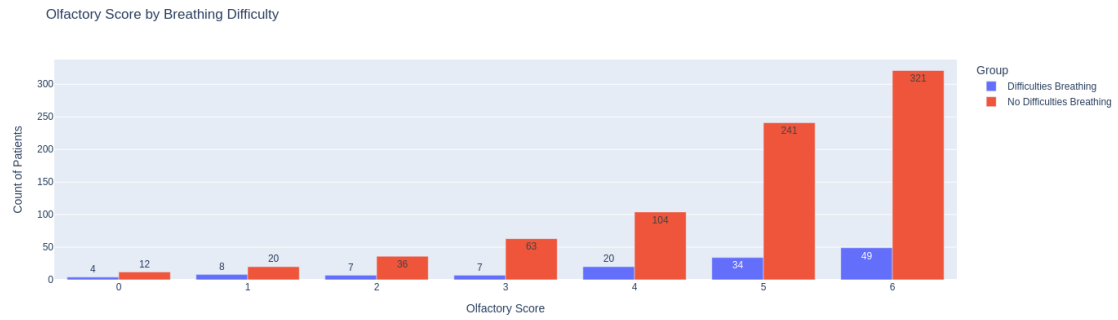


Figura 4: Conteo de pacientes según su puntaje en el test olfatorio discriminado si reportan o no dificultad olfatoria. En rojo se muestra los pacientes que no tienen dificultad respiratoria, mientras que en azul los que si poseen dificultad respiratoria. Las categorías del 1 al 6 representan los puntajes obtenidos en el test olfatorio, En el eje Y se muestra la cantidad de pacientes

En resumen, nuestros resultados sugieren que no hay diferencias significativas entre las dos muestras, lo que nos lleva a rechazar nuestra hipótesis inicial de que existe una relación entre la dificultad respiratoria y la función olfativa.

3. Análisis Exploratorio de Datos Cualitativos

Para evaluar la función olfativa como predictor del estado de COVID - 19, se creó una variable continua y , definida como una suma ponderada de variables binarias del reconocimiento de aromas, según se muestra en la Ecuación 1.

$$y = \sum_{i=1}^6 w_i x_i \quad (1)$$

El diagrama de *boxplot* Figura 5 compara la distribución del puntaje ponderado de reconocimiento olfativo (y) entre participantes que dieron positivo y negativo en la prueba PCR para COVID - 19.

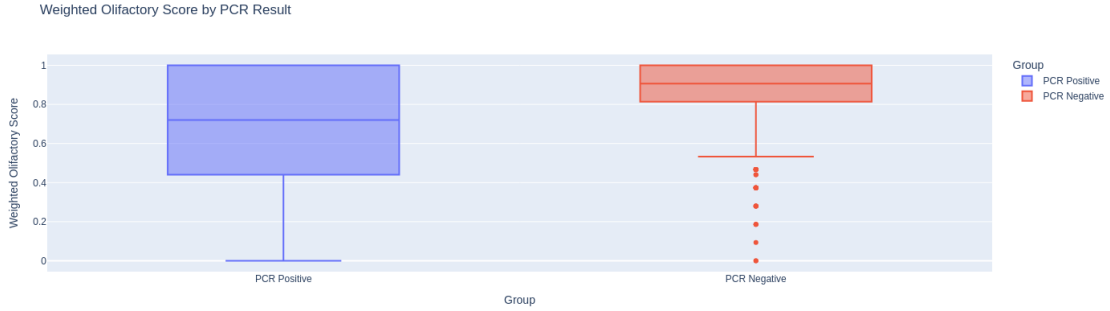


Figura 5: Puntaje ponderado de reconocimiento olfativo según resultado PCR

Del análisis del gráfico se desprenden varios patrones clave. Los participantes que dieron positivo para COVID-19 tienden a tener puntajes olfativos considerablemente más bajos, con una mediana de aproximadamente 0,72, en comparación con 0,90 en el grupo PCR negativo (estos valores fueron obtenidos del gráfico interactivo, accesible al ejecutar el código en el Anexo B).

Además, el grupo PCR positivo presenta una mayor variabilidad en los puntajes olfativos, incluyendo valores cercanos a 0, lo que podría indicar una mayor incidencia de anosmia o hiposmia entre los individuos infectados. En contraste, el grupo PCR negativo muestra una distribución más concentrada, con la mayoría de los participantes obteniendo puntajes cercanos al valor máximo, lo que sugiere una función olfativa generalmente preservada.

Es importante destacar que algunos individuos PCR negativos también presentan puntajes olfativos bajos, lo cual podría atribuirse a trastornos olfativos preexistentes u otras condiciones no relacionadas.

En general, la clara separación entre ambos grupos en términos de mediana y dispersión proporciona evidencia visual y estadística de que el puntaje ponderado de reconocimiento olfativo es un **predictor** prometedor del estado de infección por COVID - 19.

Referencias

- [1] Susana Eyheramendy et al. «Screening of COVID-19 cases through a Bayesian network symptoms model and psychophysical olfactory test». En: *iScience* 24.12 (dic. de 2021), pág. 103419. ISSN: 2589-0042. DOI: 10.1016/j.isci.2021.103419. URL: <https://www.sciencedirect.com/science/article/pii/S2589004221013900> (visitado 9 de abr. de 2025).
- [2] Jessica Cande, Benjamin Prud'homme y Nicolas Gompel. «Smells like evolution: the role of chemoreceptor evolution in behavioral change». en. En: *Curr. Opin. Neurobiol.* 23.1 (2013), págs. 152-158.
- [3] Kara C Hoover. «Smell with inspiration: the evolutionary significance of olfaction». en. En: *Am. J. Phys. Anthropol.* 143 Suppl 51.S51 (2010), págs. 63-74.
- [4] Akosua Adom Agyeman et al. «Smell and taste dysfunction in patients with COVID-19: A systematic review and meta-analysis». en. En: *Mayo Clin. Proc.* 95.8 (2020), págs. 1621-1631.
- [5] S B Gane, C Kelly y C Hopkins. «Isolated sudden onset anosmia in COVID-19 infection. A novel syndrome?» en. En: *Rhinology* 58.3 (2020), págs. 299-301.
- [6] Andrea Giacomelli et al. «Self-reported olfactory and taste disorders in patients with severe acute respiratory Coronavirus 2 infection: A cross-sectional study». en. En: *Clin. Infect. Dis.* 71.15 (2020), págs. 889-890.
- [7] C Hopkins, P Surda y N Kumar. «Presentation of new onset anosmia during the COVID-19 pandemic». en. En: *Rhinology* 58.3 (2020), págs. 295-298.
- [8] Ling Mao et al. «Neurologic manifestations of hospitalized patients with Coronavirus disease 2019 in Wuhan, China». en. En: *JAMA Neurol.* 77.6 (2020), págs. 683-690.
- [9] Cristina Menni et al. «Real-time tracking of self-reported symptoms to predict potential COVID-19». en. En: *Nature Medicine* 26.7 (jul. de 2020). Publisher: Nature Publishing Group, págs. 1037-1040. ISSN: 1546-170X. DOI: 10.1038/s41591-020-0916-2. URL: <https://www.nature.com/articles/s41591-020-0916-2> (visitado 9 de abr. de 2025).
- [10] Antje Welge-Luessen et al. «What is the correlation between ratings and measures of olfactory function in patients with olfactory loss?» en. En: *Am. J. Rhinol.* 19.6 (2005), págs. 567-571.
- [11] Shima T Moein et al. «Smell dysfunction: a biomarker for COVID-19». en. En: *Int. Forum Allergy Rhinol.* 10.8 (2020), págs. 944-950.
- [12] Luigi Angelo Vaira et al. «Validation of a self-administered olfactory and gustatory test for the remotely evaluation of COVID-19 patients in home quarantine». en. En: *Head Neck* 42.7 (2020), págs. 1570-1576.
- [13] Luigi Angelo Vaira et al. «Objective evaluation of anosmia and ageusia in COVID-19 patients: Single-center experience on 72 cases». en. En: *Head Neck* 42.6 (2020), págs. 1252-1258.
- [14] Abul K Abbas, Andrew H Lichtman y Shiv Pillai. *Inmunología Celular Y Molecular*. es. 10.^a ed. Elsevier, 2022.
- [15] Kerem Sami Kaya et al. «Olfactory function in patients with obstructive sleep apnea using positive airway pressure». en. En: *Ear Nose Throat J.* 99.4 (2020), págs. 239-244.
- [16] Ellen Bonde et al. «Dissociation of dysfunctional breathing and odour intolerance among adults in a general-population study: Dysfunctional breathing in a general population». en. En: *Clin. Respir. J.* 7.2 (2013), págs. 176-182.

A. Descargo de responsabilidad

Las AI's se utilizaron solo como traductor, motor de búsqueda y redactor. No se utilizaron como generadores de contenido y/o desarrolladores de esta tarea.

B. Código Python

IIQ3402 - Tarea 1

This task entails the Exploratory Data Analysis (EDA) of a dataset containing information whether individuals could smell certain smells, and comorbidities such as fever, coughing, muscle pain, and more. At last it contains a column with the result of a PCR test for that individual.

```
In [ ]: import polars as pl
import numpy as np
import plotly.express as px
from sklearn import linear_model

In [ ]: df = pl.read_csv("Enunciado_Tarea1/datos_tarea1.csv")
df = df.rename(
    {
        "1__Cold": "Cold",
        "2__Cough": "Cough",
        "3__Fever": "Fever",
        "4__Muscular_pain": "Muscular Pain",
        "5__Breathing_difficulty": "Breathing difficulty",
        "6__Self_reported_anosmia": "Self-reported anosmia",
        "7__Self_reported_ageusia": "Self-reported ageusia",
        "COVID19_PCR": "COVID19 PCR Result",
    }
)
df

In [ ]: df_neg = df.filter(pl.col("COVID19 PCR Result") == 0)
df_pos = df.filter(pl.col("COVID19 PCR Result") == 1)

neg_pcr = df.filter(pl.col("COVID19 PCR Result") == 0).shape[0]
pos_pcr = df.filter(pl.col("COVID19 PCR Result") == 1).shape[0]

print(f"Total number of positive PCR: {pos_pcr}")
print(f"Total number of negative PCR: {neg_pcr}")
```

Hypothesis 1

Covid-positive individuals most frequently identify orange, pineapple and mint odors.

```
In [ ]: scent_cols = ["Banana", "Caramel", "Mint", "Orange", "Pineapple", "V

scent_data = {
    "Scent": [],
    "Group": [],
    "Count": [],
    "Percentage": [],
```

```

}

for i, scent in enumerate(scent_cols):
    scent_data["Scent"].append(scent)
    scent_data["Group"].append("PCR Positive")
    scent_data["Count"].append(df_pos.filter(pl.col(scent) == 1).shape[0])
    scent_data["Percentage"].append(round(
        100 * df_pos.filter(pl.col(scent) == 1).shape[0] / pos_pcr,
    ))

    scent_data["Scent"].append(scent)
    scent_data["Group"].append("PCR Negative")
    scent_data["Count"].append(df_neg.filter(pl.col(scent) == 1).shape[0])
    scent_data["Percentage"].append(round(
        100 * df_neg.filter(pl.col(scent) == 1).shape[0] / neg_pcr,
    ))

scent_df = pl.from_dict(scent_data)
scent_df

```

```

In [ ]: fig = px.bar(
    scent_df,
    x="Scent",
    y="Count",
    color="Group",
    barmode="group",
    labels={"Count": "Count of Patients"},
    title="Scent Detection by PCR Result",
    text="Count",
)
fig.show()

```

```

In [ ]: fig = px.bar(
    scent_df,
    x="Scent",
    y="Percentage",
    color="Group",
    barmode="group",
    labels={"Percentage": "Percentage of Patients"},
    title="Relative Scent Detection by PCR Result",
    text="Percentage",
    range_y=[0, 100],
)
fig.show()

```

Based on the plot, the hypothesis does not seem to be supported. Here the relative scent detection is the number of individuals who could smell the scent divided by the total number of individuals who tested positive, and negative for COVID-19.

Hypothesis 2

¿Son las personas con comorbilidades menos propensas a contraer

COVID - 19?

English: Are people with comorbidities less prone to contracting COVID - 19?

Meaning that we have to check if there is a correlation between the people with e.g. fever and the PCR test result.

```
In [ ]: symptoms = [
    "Cold",
    "Cough",
    "Fever",
    "Muscular Pain",
    "Breathing difficulty",
    "Self-reported anosmia",
    "Self-reported ageusia",
]

symptom_data = {
    "Symptom": [],
    "Group": [],
    "Count": [],
    "Percentage": [],
}

for symptom in symptoms:
    # Positive
    symptom_data["Symptom"].append(symptom)
    symptom_data["Group"].append("Positive PCR")
    symptom_data["Count"].append(df_pos.filter(pl.col(symptom) == 1))
    symptom_data["Percentage"].append(round(
        100 * df_pos.filter(pl.col(symptom) == 1).shape[0] / pos_pcr
    ))

    # Negative
    symptom_data["Symptom"].append(symptom)
    symptom_data["Group"].append("Negative PCR")
    symptom_data["Count"].append(df_neg.filter(pl.col(symptom) == 1))
    symptom_data["Percentage"].append(round(
        100 * df_neg.filter(pl.col(symptom) == 1).shape[0] / neg_pcr
    ))

symptom_df = pl.from_dict(symptom_data)
symptom_df

In [ ]: fig = px.bar(
    symptom_df,
    x="Symptom",
    y="Count",
    color="Group",
    barmode="group",
    labels={"Count": "Count of Patients"},
    title="Symptoms by PCR Result",
    text="Count",
)
fig.show()
```

```
In [ ]: fig = px.bar(
    symptom_df,
    x="Symptom",
    y="Percentage",
    color="Group",
    barmode="group",
    labels={"Percentage": "Percentage of Patients"},
    title="Relative Symptoms by PCR Result",
    text="Percentage",
    range_y=[0, 100],
)
fig.show()
```

Hypthesis 3

¿Existe una relación entre las dificultades respiratorias y la función olfativa en individuos?

English: Is there a relationship between breathing difficulties and olfactory function in individuals

Thought process

Make an olfactory score, summing the number of olfactory tests that were positive. Then, check if there is a correlation between the olfactory score and the breathing difficulties.

```
In [ ]: scents = ["Banana", "Caramel", "Mint", "Orange", "Pineapple", "Vanilla"]
df = df.with_columns(
    [
        pl.sum_horizontal(pl.col(scent) for scent in scents).alias("Olfactory Score")
    ]
)
olfactory_df = df.select(scents + ["Olfactory Score", "Breathing difficulty"])
olfactory_df.columns
```

```
In [ ]: olfactory_df = olfactory_df.with_columns(
    pl.when(pl.col("Breathing difficulty") == 1)
        .then(pl.lit("Yes"))
        .otherwise(pl.lit("No"))
        .alias("Difficulties breathing")
)

fig = px.box(
    olfactory_df,
    x="Difficulties breathing",
    y="Olfactory Score",
    color="Difficulties breathing",
    title="Olfactory Score by Breathing Difficulty",
    # range_y=[0, 6],
)
fig.show()
```

```

In [ ]: olfactory_scores = [0, 1, 2, 3, 4, 5, 6]
        olfactory_data = {
            "Olfactory Score": [],
            "Group": [],
            "Count": []
        }

        for score in olfactory_scores:
            olfactory_data["Olfactory Score"].append(score)
            olfactory_data["Group"].append("Difficulties Breathing")
            olfactory_data["Count"].append(df.filter(pl.col("Breathing diffi

            olfactory_data["Olfactory Score"].append(score)
            olfactory_data["Group"].append("No Difficulties Breathing")
            olfactory_data["Count"].append(df.filter(pl.col("Breathing diffi

        olfactory_df = pl.from_dict(olfactory_data)
        olfactory_df

In [ ]: fig = px.bar(
        olfactory_df,
        x="Olfactory Score",
        y="Count",
        color="Group",
        barmode="group",
        labels={"Count": "Count of Patients"},
        title="Olfactory Score by Breathing Difficulty",
        text="Count",
    )
    fig.show()

```

Part 4

Weighted Olfactory Score

By giving certain weight to the olfactory tests, we can create a continuous variable that can be used to indicate COVID19 in patients. The score is defined as follows:

$$y = \sum_{i=1}^6 w_i x_i$$

where

- y is the olfactory score
- x_i is the result of the olfactory test i (1 if positive, 0 if negative)
- w_i is the weight of the olfactory test i .

The weights can be found in Table 1 in [Enunciado Tarea 1](#).

```

In [ ]: olfactory_weights = {
        "Banana": 0.09332,
        "Caramel": 0.09332,
        "Mint": 0.34668,
        "Orange": 0.09334,
        "Pineapple": 0.18667,
        "Vanilla": 0.18666,
    }

In [ ]: df = df.with_columns(
    [
        pl.sum_horizontal(
            pl.col(scent) * olfactory_weights[scent] for scent in scents
        ).alias("Weighted Olfactory Score"),
    ]
)

df = df.with_columns(
    pl.when(pl.col("COVID19 PCR Result") == 1)
    .then(pl.lit("PCR Positive"))
    .otherwise(pl.lit("PCR Negative"))
    .alias("Group"),
)
df

In [ ]: fig = px.box(
    df,
    x="Group",
    y="Weighted Olfactory Score",
    color="Group",
    title="Weighted Olfactory Score by PCR Result",
)
fig.show()

```

Based on these results we want to see if there is a relationship between the symptoms and the olfactory score, not just based on the positive/negative results of the PCR.

Logistic Regression

Using logistic regression on the weighted olfactory score

NOTE: Due to lack of time, a visual analysis of the data was not performed.

```

In [ ]: # Model training
X = df.select('Weighted Olfactory Score').to_numpy()
y = df.select('COVID19 PCR Result').to_numpy().ravel()

model = linear_model.LogisticRegression()
model.fit(X, y)

```

t1_code

file:///home/william/GitHub/IIQ3402-Statistical-Design-Optim...

```
In [ ]: x_vals = np.linspace(X.min(), X.max(), 1000).reshape(-1, 1)
        y_pred = model.predict_proba(x_vals)[:, 1]

        print(x_vals.flatten().shape)
        print(y_pred.shape)
        fig = px.scatter(
            x=x_vals.flatten(),
            y=y_pred,
            title="Logistic Regression Model",
            labels={"x": "Weighted Olfactory Score", "y": "Probability of PCR"}
        )

        y_jittered = y + np.random.normal(0, 0.03, size=y.shape) # Jittering
        fig.add_scatter(
            x=X.flatten(),
            y=y_jittered,
            mode="markers",
            name="PCR Results",
            marker=dict(color="gray", opacity=0.4),
        )
        fig.show()
```