UNIVERSIDAD NACIONAL DE JAÉN

UNIVERSIDAD NACIONAL DE JAÉN

DEPARTAMENTO ACADÉMICO DE CIENCIAS BÁSICAS Y APLICADAS

INGENERÍA CIVIL

MANUAL

ANÁLISIS DE LA VARIANZA CON UN FACTOR (ANOVA)

Autores:

Dra. Rosario Yaqueliny Llauce Santamaria

Dra. Marcela Yvone Saldaña Miranda

Mg. Mario Félix Olivera Aldana

Jaén – Perú, febrero 2025

Índice

1. Introducción

El ANOVA de un solo factor es una herramienta utilizada por los científicos para determinar si existen diferencias significativas entre tres o más grupos. En áreas como la ingeniería, economía, salud, psicología y educación, este método se aplica para comparar tratamientos, condiciones o enfoques con el propósito de evaluar su efectividad y analizar el impacto de ciertas variables sobre una variable específica de interés.

Este análisis distingue dos tipos de variabilidad: la intragrupal, que surge debido a factores aleatorios o no controlados, y la intergrupal, atribuida al factor en estudio. Para determinar si las diferencias entre las medias son estadísticamente significativas, se emplea el estadístico F, permitiendo así extraer conclusiones fundamentadas.

En consecuencia, el ANOVA facilita la identificación de patrones y discrepancias relevantes en grandes volúmenes de datos, optimizando la toma de decisiones. Su aplicación contribuye a mejorar la precisión y eficiencia en la investigación y el análisis estadístico, proporcionando a los investigadores información clave para perfeccionar estrategias y metodologías en sus respectivos campos.

En el contexto del análisis de datos, se generaron representaciones gráficas, como los diagramas de caja (box plots), mediante el uso del lenguaje de programación Python. Estos gráficos fueron seleccionados por su capacidad para resumir visualmente la distribución de los datos, detectar valores atípicos y representar con precisión la variabilidad dentro de cada grupo o conjunto de datos.

2. Objetivos

2.1. Objetivo principal

Ejecute un ANOVA de un factor para detectar diferencias potenciales entre las medias de múltiples grupos, analizado por separado. Además, está diseñado para medir la influencia que un cierto factor puede tener en una variable de interés.

2.2. Objetivo particular

- Examinar el valor F y el valor p con el fin de determinar con precisión la relevancia estadística de los datos y la influencia del factor en cuestión.
- Proporcionar sugerencias al utilizar datos analíticos para fundamentadas para la investigación científica, respecto a cómo aplicar ANOVA de un solo factor

destacando la notabilidad de su uso en la toma de decisiones en áreas clave como la medicina, psicología, economía y educación.

3. Marco teórico

El ANOVA es una herramienta estadística empleada para valorar la presencia de las diferencias entre tres o más conjuntos de datos autónomos (Montgomery, 2013, p. 34). Este método examina el impacto de una variable categórica individual en una variable cuantitativa. Además, ofrece claridad acerca de cómo los diferentes niveles del componente influyen en la variable que está siendo evaluada. El análisis de varianza (ANOVA) es un método de gran magnitud en las consideraciones de decisiones en diferentes contextos, ya que permite desglosar la variabilidad total en distintos componentes explicativos.

Desarrollo cronológico y progreso del ANOVA

El método de ANOVA fue consolidado por Ronald A. Fisher en 1920 y se emplea para evaluar investigaciones en el ámbito agrícola. La técnica en cuestión posibilita a los investigadores la comparación de las disparidades existentes entre los diversos tratamientos empleados en la agricultura. La estrategia desarrollada por Fisher tuvo como objetivo superar las limitaciones de la prueba en estudios multigrupo, lo que posibilitaría la realización de evaluaciones estadísticas más precisas en diversos ámbitos. (Fisher, 1925, p. 12).

Variantes del ANOVA

Este artículo aborda principalmente el ANOVA de un solo factor; sin embargo, es fundamental reconocer las distintas versiones.

- ANOVA de dos vías: evalúa la influencia de dos variables categóricas y sus interacciones potenciales.
- ANOVA de evaluaciones repetidas: se utiliza cuando voluntarios idénticos se someten a muchas pruebas en entornos variados.
- El MANOVA (análisis de varianza multivariante) se utiliza cuando hay muchas variables dependientes presentes.

Fundamentos del ANOVA

La variabilidad total de las muestras se descompone en dos componentes básicos: variabilidad intergrupal y variabilidad intragrupal.

El término "variabilidad intergrupal" se refiere a los cambios que se producen por el efecto del componente que se está investigando. El componente que se está investigando es responsable de la irregularidad que se aprecia en las medias del grupo si las medias del grupo muestran diferencias significativas.

Cuando hablamos de variabilidad intragrupal, nos referimos a las diferencias individuales que se observan dentro de cada grupo como resultado de eventos aleatorios o incontrolables.

Determinar si las diferencias que se han observado son estadísticamente significativas o no utilizando la estadística F, por lo tanto, se realiza un análisis de la irregularidad que existe entre los grupos en balance con la irregularidad que existe dentro de los grupos. (Montgomery,2013,p.45).

Concepto e interpretación del estadístico F

El estadístico F se utiliza para analizar el vínculo entre la varianza explicada por el factor y la varianza residual en un estudio estadístico. Asimismo,la distribución F presenta asimetría, cuya configuración se define por los grados de libertad dentro y fuera de los mismos. La negación de la hipótesis nula se produce debido a un alto valor de la estadística F y un valor p por debajo del nivel de significancia (por ejemplo, $\alpha = 0{,}05$), señala que al menos un grupo presenta una diferencia significativa. (Tabachnick & Fidell, 2013, p. 78).

$$F = \frac{\frac{\mathrm{SSB}}{\mathrm{dfB}}}{\frac{\mathrm{SSW}}{\mathrm{dfW}}} \quad O \quad F = \frac{\mathrm{MSB}}{\mathrm{MSW}}$$

- **La sumatoria de los cuadrados entre grupos (SSB)**: mide las irregularidades de las medias de los grupos en relación con la media.
- **El número de grados de libertad entre grupos (dfB)**: evalúa g-1, donde g representa el número de grupos en el estudio.
- **Sumatoria de los cuadrados dentro de los grupos (SSW)**: evalúa la irregularidad de los datos dentro de cada grupo en comparación con la media del grupo.

- **El número de grados de libertad dentro de los grupos (dfW)**: evalúa la diferencia entre el total de muestras y el número de grupos (N - g), donde N es el total de muestras y g es el número de grupos.
- $\boldsymbol{MSB} = \frac{SSB}{dfB}$ forma abreviada de definir promedio de la variabilidad presente entre de los grupos.
- $\boldsymbol{MSW} = \frac{SSW}{\mathrm{dfW}}$ forma abreviada de definir promedio de la variabilidad presente dentro de los grupos.

4. Tabla de Anova

El análisis de varianza (ANOVA) se presenta de manera organizada en una estructura conocida, donde se detallan los cálculos principales y los resultados de la prueba. Esta tabla incluye las siguientes fórmulas:

Variación	Sumatoria de cuadrados	Grados de libertad	Promedio de los cuadrados	F
Factor	$SS_B = \sum_{k=1}^{g} (n_k * (X_k - \bar{X})^2)$	g - 1	$MS_B = \frac{SS_B}{(g-1)}$	$F = \frac{MS_B}{MS_W}$
Error	$SSE = \sum_{k=1}^{g} \sum_{i=1}^{nk} (\mathrm{X_i k} - \bar{\mathrm{X}}_\mathrm{k})^2$	N - g	$MS_W = \frac{SS_W}{(N-g)}$	
Total	$SST = \sum_{i=1}^{N} (\mathrm{X_i} - \bar{\mathrm{X}})^2$	N - 1		

Donde:

g: Número de grupos

n_k:Tamaño del grupo

X_k: Promedio aritmético de grupo

X: Promedio aritmético general

4.1.Concepto e Interpretación del Valor p

El valor de p se define como la posibilidad de conseguir un estadístico de prueba similar de extremo o dentro del margen de lo observado en los datos de muestra, bajo la premisa de que la hipótesis nula (H_0) es verdadera.

Por consiguiente, es interpretado en propósito del nivel de significancia (α) y sirve como criterio para concluir si se aprueba o niega la hipótesis nula en un estudio. En investigaciones se requieren un mayor nivel de precisión estadística, es factible observar niveles de significancia inferiores al estándar comúnmente aceptado de 0.05.

Si $(\boldsymbol{p} \leq \boldsymbol{\alpha})$

- Rechazamos (H_0).
- Se dispone de pruebas suficientes para poder negar la hipótesis nula.

Si $(\boldsymbol{p} > \boldsymbol{\alpha})$

- No rechazamos (H_0).
- No se dispone de pruebas suficientes para admitir la existencia de una disparidad o impacto de relevancia.

Fórmula para el valor p:

$$p = P(F \geq Fobservado \mid dfB, dfW)$$

Donde:

- F es una variable que se distribuye de acuerdo a grados de libertad específicos para cada grupo de datos. Esta distribución es ampliamente empleada en el análisis de varianza con el fin de contrastar la variabilidad existente entre dos o más muestras.
- F observado es el valor calculado del estadístico F para los datos observados.
- dfB (grados de libertad entre los grupos): es el número de grupos menos se representa como: g−1, donde g es el número de grupos.
- dfW (grados de libertad dentro de los grupos): es la diferencia entre el total de muestras y el total de grupos se representa como N−g, donde N denota el total de muestras y g el total de grupos.

4.2. Distribución de F de FICHER

Esta tabla contiene valores críticos para la distribución F definida por $P(F \geq F_{\alpha,x1,x2})$

$^{x1}/_{x2}$	1	2	3	4	5	6	7	8	9	10	15	20
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	245.95	248.01
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66

4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87
7	5.99	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.47	2.32	2.18	2.10
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.13	2.05
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.46	2.30	2.25	2.11	2.03
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.09	2.01
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.87	1.78
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.75	1.66
∞	3.85	3.00	2.61	2.38	2.22	2.10	2.01	1.94	1.88	1.84	1.67	1.58

La tabla muestra una pequeña parte de la tabla de distribución F de Fisher si necesita más valores puede visitar a distintos autores que proporcionen información más detallada.

5. Aplicaciones del ANOVA de un solo factor

A continuación, se enumeran algunos de los ámbitos que hacen un uso intensivo de este método (Field, 2013, p. 120):

- En el ámbito de la psicología, durante la realización de estudios para valorar la eficiencia de diversas terapias o intervenciones.
- En el ámbito de la medicina, se realizaría para evaluar la eficacia de diversas terapias o productos farmacéuticos.
- En el ámbito de la economía, el proceso de analizar los efectos de las políticas económicas en una serie de lugares o sectores diferentes.

5.1.Proceso del ANOVA de un solo factor

- La hipótesis nula (H_0) indica que no existe diferencias entre los grupos.

- La hipótesis alternativa (H_1) indica que existe como mínimo una diferencia entre muestras analizadas.

5.2.Análisis post hoc

Esta prueba es práctica para identificar discrepancias entre las medias. La prueba de Bonferroni es recomendable en situaciones en las que se pretende minimizar la posibilidad de producir un error de tipo I (Navarro, 2019, p. 145).

5.3.La prueba de Tukey

Es utilizado para reconocer cual grupos presenta discrepancias en sus medias.

También conocida como HSD, por sus siglas en inglés: Honest Significant Difference, se calcula siguiendo esta fórmula:

$$HSD = q \times \sqrt{\frac{MSE}{n}}$$

donde:

- q : Valor crítico de la distribución studentizada
- MSE : Cuadrado medio dentro de los grupos (del ANOVA).
- n : tamaño de los grupos.

5.4.Análisis de gráficos pertinentes para ANOVA de un solo factor

Para interpretar los resultados de un ANOVA de un solo factor, resulta fundamental realizar un análisis visual utilizando gráficos apropiados. Los diagramas de caja (box plots) se utilizan con frecuencia por su efectividad para visualizar la distribución de los datos y las diferencias entre los grupos analizados

6. Gráficos de diagnóstico

Para respaldar las suposiciones del ANOVA, se recomienda el uso de los siguientes gráficos:

- Diagrama de caja: Esta herramienta es ideal para detectar valores atípicos que podrían influir en los resultados del análisis.

La generación de estos gráficos puede realizarse mediante bibliotecas de software como Python (utilizando matplotlib y seaborn) o SPSS, lo que facilita la evaluación de los supuestos del ANOVA de manera más efectiva. (Howell, 2012).

7. Restricciones del ANOVA de un solo factor
    - Homogeneidad: Las varianzas entre grupos deben ser similares.
    - Las muestras deben ser autónomas entre ellas.

Las conclusiones pueden verse comprometidas si no se cumplen los criterios, lo que res ulta en resultados poco confiables La prueba Kruskal walis es una prueba no paramétrica que se utiliza en otros casos sin la necesidad de la normalidad d datos o la homogeneidad de las variaciones (Navarro, 2019).

8. Ejemplos

Ejemplo 1

Comparación de la Resistencia de Dos Materiales

Un ingeniero desea analizar la resistencia a la compresión de tres variedades de concreto para establecer si existen diferencias significativas entre ellas. Para ello, se recopilan datos de resistencia a partir de múltiples muestras de cada tipo de concreto.

Tipo de concreto	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
Convencional	25	27	24	28	26
Estructural	30	31	29	32	30
Permeable	35	36	34	37	36

Paso 1: Calcular el promedio para cada grupo.

Media de A: $\frac{25+27+24+28+26}{5} = 26$

Media de B: $\frac{30+31+29+32+30}{5} = 30.4$

Media de C: $\frac{35+36+34+37+36}{5} = 35.6$

Media General (MG):

$$MG = \frac{25+27+24+28+26+30+31+29+32+30+35+36+34+37+36}{15}$$

$$MG = 30.67$$

Paso 2: Calcular la sumatoria de los cuadrados entre los grupos (SSB)

Utilizamos la fórmula:

$$SSB = n \times \sum (Mediadelgrupo - Mediageneral)^2$$

Para cada grupo:

Tipo A: $(26.0 - 30.67)^2 = 21.79$

Tipo B: $(30.4 - 30.67)^2 = 0.0729$

Tipo C: $(35.6 - 30.67)^2 = 24.30$

Calculamos SSB sin redondeos:

$$SSB = 5 \times (21.79 + 0.0729 + 24.30) = 5 \times 46.163 = 230.815$$

Paso 3: Calcular la sumatoria de los cuadrados dentro de los grupos (SSW)

Usamos la fórmula:

$$SSW = \sum (valorindividual - Mediadelgrupo)^2$$

Para el Tipo A:

$$(25 - 26)^2 + (27 - 26)^2 + (24 - 26)^2 + (28 - 26)^2 + (26 - 26)^2 = 10$$

Para el Tipo B:

$$(30 - 30.4)^2 + (31 - 30.4)^2 + (29 - 30.4)^2 + (32 - 30.4)^2 + (30 - 30.4)^2 = 4.8$$

Para el Tipo C:

$$(35 - 35.6)^2 + (36 - 35.6)^2 + (34 - 35.6)^2 + (37 - 35.6)^2 + (36 - 35.6)^2 = 5.6$$

Calculamos SSW sin redondeos:

$$SSW = 10 + 4.8 + 5.6 = 20.4$$

Paso 4: Calcular los Grados de Libertad (df)

Entre los grupos (dfB): g−1=3−1=2

Dentro de los grupos (dfW): N−g=15−3=12

Paso 5: Calcular MSB y MSW (Media de los Cuadrados)

MSB (Media de los cuadrados entre grupos):

$$MSB = \frac{SSB}{dfB} = \frac{230.815}{2} = 115.4075$$

MSW (Media de los cuadrados dentro de los grupos):

$$MSW = \frac{SSW}{dfW} = \frac{20.4}{12} = 1.7$$

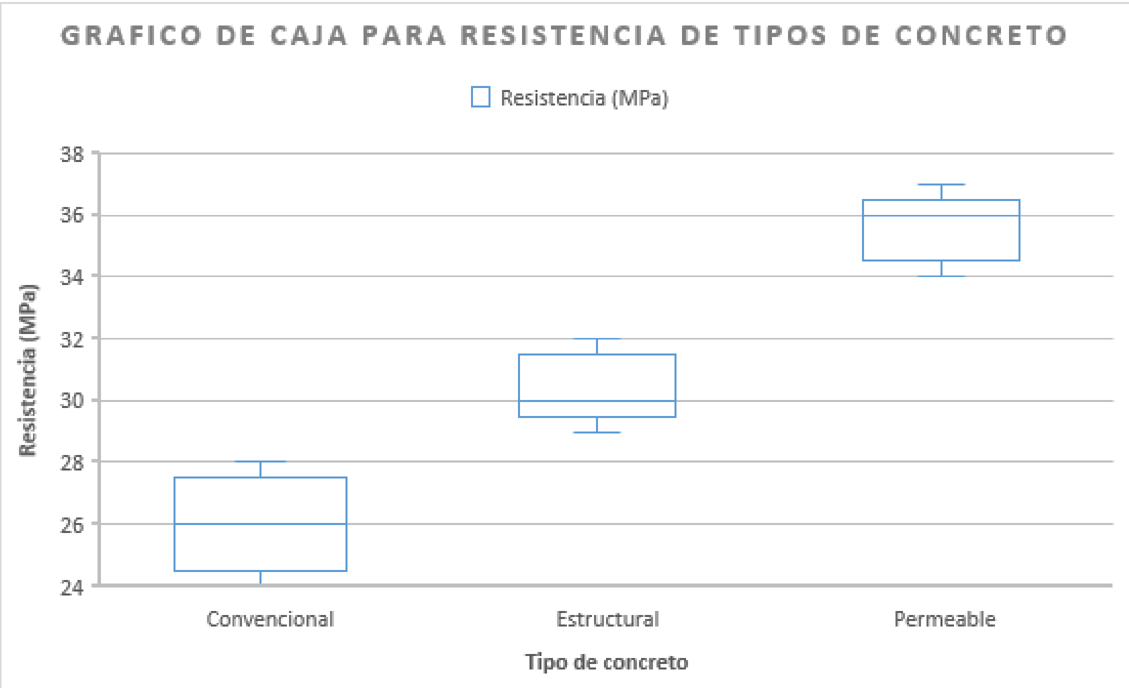
Paso 6: Calcular el Valor F

$$F = \frac{MSB}{MSW} = \frac{115.4075}{1.7} = 67.887$$

$$Valor\ F: 67.887$$

Interpretación: El valor F calculado es 67.887 es muy alto, indica que existe al menos una diferencia en la resistencia a la compresión entre los tipos de concreto. Asimismo, se confirma con un p-valor que será mucho menor que 0.05, obteniendo suficiente evidencia para negar la hipótesis nula.

Gráfico N°1: Box plot de resistencia de tipo de concreto



Fuente: Elaboración Propia.

Interpretación de la gráfica:

- Tipo C tiene la mayor resistencia, seguido por Tipo B y luego Tipo A.
- Tipo A presenta menos variabilidad en la resistencia, mientras que Tipos B y C tienen más dispersión.
- En general, Tipo C es más adecuado para proyectos que requieren alta resistencia, mientras que Tipo A podría usarse en aplicaciones menos exigentes.

Ejemplo 2:

Se quiere evaluar el efecto de tres tipos de cemento Pacasmayo (A), Inca (B), Mochica (C) en la resistencia a compresión de muestras de concreto. Los datos recopilados son:

Paso 1: Hipótesis

- **Nula (H_0)**: No hay diferencia en la resistencia a compresión entre los tipos de cemento Pacasmayo (A), Inca (B), Mochica (C).
- **Alternativa (H_1)**: Al menos un tipo de cemento tiene una resistencia significativamente diferente.

Paso 2: Datos

Los datos de resistencia a compresión para cada tipo de cemento están organizados en la tabla siguiente:

Tipo de Cemento	**Muestra 1**	**Muestra 2**	**Muestra 3**	**Muestra 4**	**Muestra 5**
Pacasmayo (A)	25	26	24	25	27
Inca (B)	30	31	29	32	30
Mochica (C)	35	36	34	35	36

Paso 3: Cálculo de la media general

$$\overline{X} = \frac{(25+26+24+25+27)+(30+31+29+32+30)+(35+36+34+35+36)}{15}$$

$$\overline{X} = 30.33$$

Paso 4: Cálculo de la suma de cuadrados

Suma de cuadrados total (SST)

$$SST = \sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2$$

$$SST = (25-30.33)^2 + (26-30.33)^2 + (24-30.33)^2 + \cdots + (36-30.33)^2$$

Después de realizar los cálculos para cada observación, obtenemos:

$$SST = 253.33$$

Suma de cuadrados entre grupos (SStr)

$$SStr = \sum_{j=1}^{k} n_j \left(\overline{X}_j - \overline{X}\right)^2$$

Donde:

- n_j número de muestras en cada grupo.
- $\overline{X}_j$ media de cada grupo.
- $\overline{X}$ media general.

Para cada tipo de cemento, tenemos las medias:

- **Media de A** = 25.4
- **Media de B** = 30.4
- **Media de C** = 35.2

Aplicamos la fórmula:

$$SStr = 5 \times (25.4-30.33)^2 + 5 \times (30.4-30.33)^2 + 5 \times (35.2-30.33)^2$$

Realizando estos cálculos:

$$SSStr = 240.13$$

Sumatoria de cuadrados dentro de grupos (SSE)

Se obtiene mediante el siguiente cálculo:

$$SSE = SST - SStr$$

Sustituyendo los valores:

$$SSE = 253.33 - 240.13 = 13.20$$

Paso 5: Cálculo de los Grados de Libertad

- Grados de libertad entre grupos (df_b) = $g - 1 = 3 - 1 = 2$
- Grados de libertad dentro de los grupos (df_w) = $N - g = 15 - 3 = 12$
- Grados de libertad total (df_t) = $N - 1 = 15 - 1 = 14$

Paso 6: Cálculo de las medias cuadradas y el estadístico F

Media cuadrada entre grupos (MStr)

Se calcula como:

$$MStr = \frac{SStr}{df_b}$$

Sustituyendo los valores:

$$MStr = \frac{240.13}{2} = 120.07$$

Media Cuadrada dentro de los grupos (MSE)

Se calcula como:

$$MSE = \frac{SSE}{df_w}$$

Sustituyendo los valores:

$$MSE = \frac{13.20}{12} = 1.10$$

Estadístico F

Se obtiene como:

$$MSEF = \frac{MStr}{MSE}$$

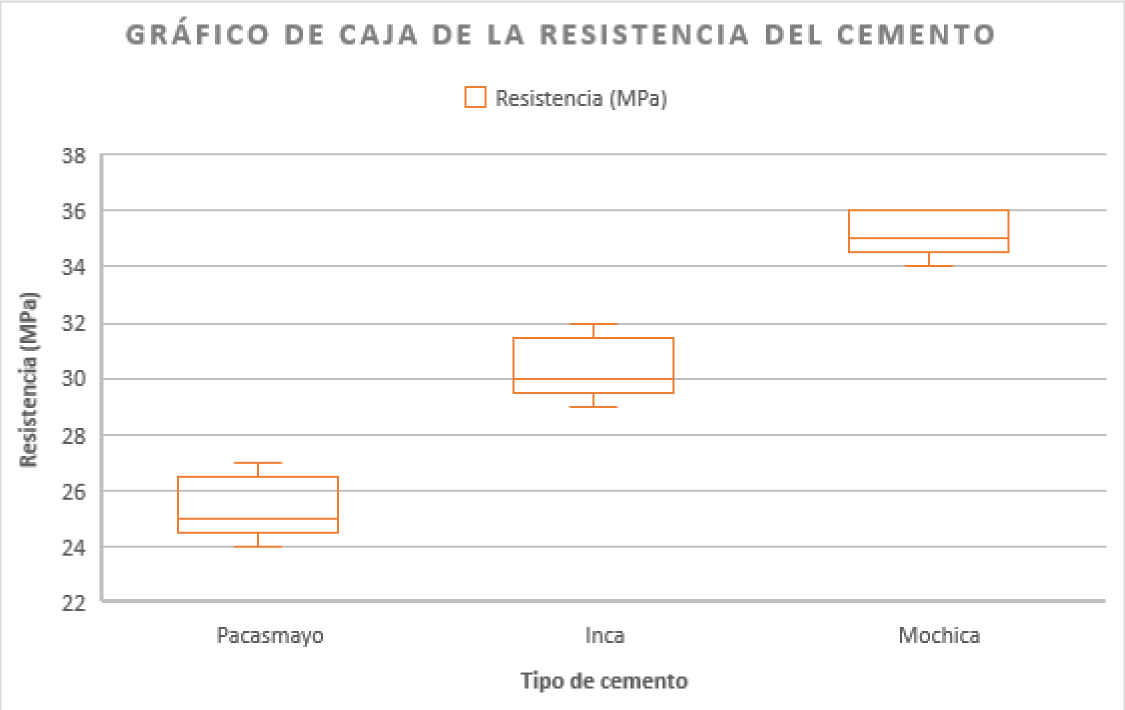
Sustituyendo los valores:

$$F = \frac{120.07}{1.10} = 109.15$$

Paso 7: Interpretación

Con un valor de 109.15 para la prueba F, se observa que este valor es significativamente elevado. El valor observado de F supera el valor crítico para el nivel de significancia (α=0.05), por lo cual se tiene evidencia suficiente para negar la hipótesis nula, concluyendo que existe por lo menos una diferencia en la resistencia a compresión entre los tipos de cemento (Pacasmayo, Inca, Mochica),

Gráfico N°2: Gráfico de caja de la resistencia del cemento



Fuente: Elaboración Propia.

Ejemplo 3:

Analizar los métodos de compactación: Placa Vibratoria (X), Modificado Proctor (Y) y sismógrafo (Z) en la densidad del suelo.

Paso 1: Hipótesis

Nula (H_0): No hay diferencia en la densidad del suelo entre los métodos de compactación (Anillo de densidad, Densidad aparente y Proctor).

Alternativa (H_1): Al menos un método de compactación produce una densidad significativamente diferente.

Paso N°2: Datos

Los datos de densidad del suelo por cada tipo métodos de compactación están organizados en la tabla siguiente:

Método	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
Placa Vibratoria (X)	1.85	1.87	1.86	1.84	1.88
Modificado Proctor (Y)	1.90	1.92	1.89	1.91	1.90
Proctor (Z)	1.95	1.97	1.96	1.94	1.95

Paso 3: Cálculo de la media general

$$\overline{X} = \frac{(1.90 + 1.92 + 1.89 + 1.91 + 1.90) + (1.95 + 1.97 + 1.96 + 1.94 + 1.95)}{15} + \frac{(1.85 + 1.87 + 1.86 + 1.84 + 1.88)}{15}$$

$$\overline{X} = 1.906$$

Paso 4: Cálculo de las sumas de cuadrados

Suma de cuadrados total (SST)

Se estima mediante el siguiente cálculo:

$$SST = \sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2$$

$$SST = (1.85 - 1.906)^2 + (1.87 - 1.906)^2 + (1.86 - 1.906)^2 + \cdots + (1.95 - 1.906)^2$$

Luego de realizar estos cálculos obtenemos:

$$SST \ = \ 0.02416$$

Suma de cuadrados entre grupos (SStr)

El cálculo se realiza de la siguiente manera:

$$SStr = \sum_{j=1}^{k} n_j \left(\overline{X}_j - \overline{X}\right)^2$$

Donde:

- n_j número de muestras de cada grupo.
- $\overline{X}_j$ media de cada grupo.
- $\overline{X}$ media general.

Calculamos la media de cada método:

- **Media de X** = 1.86
- **Media de Y** = 1.90
- **Media de Z** = 1.95

Luego, aplicamos la fórmula:

$$SStr = 5 \times (1.86 - 1.906)^2 + 5 \times (1.90 - 1.906)^2 + 5 \times (1.95 - 1.906)^2$$

Después de realizar estos cálculos, obtenemos:

$$SStr \ = \ 0.02212$$

Suma de cuadrados dentro de los grupos (SSE)

Se obtiene mediante el siguiente cálculo:

$$SSE = SST - SStr$$

Sustituyendo los valores que calculamos anteriormente:

$$SSE = 0.02416 - 0.02212 = 0.00204$$

Paso 5: Cálculo de los grados de libertad

- Grados de libertad Entre Tratamientos $(\boldsymbol{df_b})$ = $g - 1 = 3 - 1 = 2$
- Grados de libertad del Error $(\boldsymbol{df_w})$ = $N - g = 15 - 3 = 12$
- Grados de libertad Total $(\boldsymbol{df_t})$ = $N - 1 = 15 - 1 = 14$

Paso 6: Cálculo de las medias cuadradas y el estadístico F

Media cuadrada entre gruposs (MStr)

Se calcula como:

$$MStr = \frac{SStr}{df_b}$$

Sustituyendo los valores:

$$MStr = \frac{0.02212}{2} = 0.01106$$

Media cuadrada dentro del grupo (MSE)

Se calcula como:

$$MSE = \frac{SSE}{df_{within}}$$

Sustituyendo los valores:

$$MSE = \frac{0.00204}{12} = 0.00017$$

Estadístico F

Se calcula como:

$$F = \frac{MStr}{MSE}$$

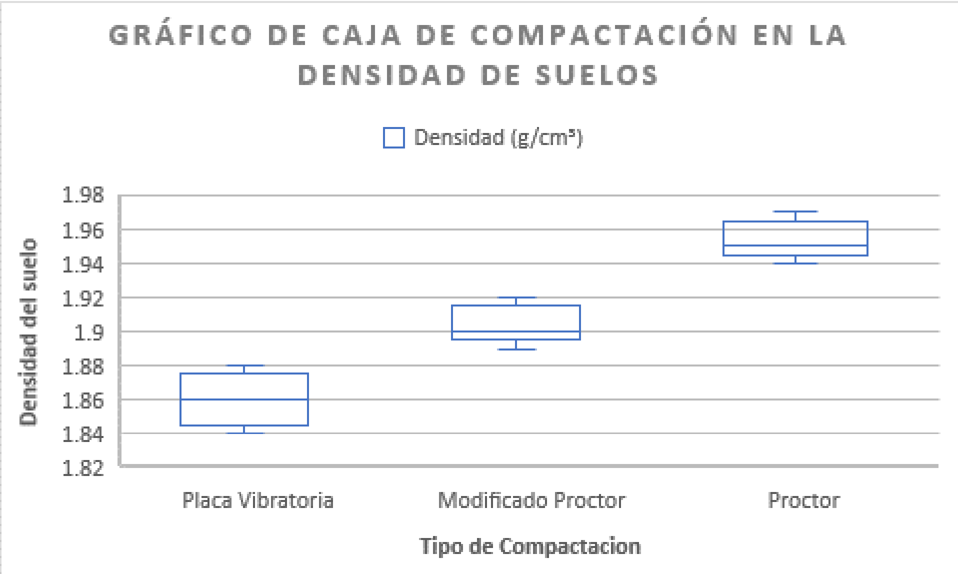
Sustituyendo los valores:

$$F = \frac{0.01106}{0.00017} = 65.06$$

Paso 7: Interpretación

Con un valor de 65.06 para la prueba F, se observa que este valor es significativamente elevado. El valor observado de la estadística F supera el valor crítico para un nivel de significancia (α=0.05), por lo cual se tiene evidencia suficiente para negar la hipótesis nula, lo que se concluye que existe al menos una diferencia entre densidades del suelo entre los distintos métodos de compactación (X, Y, Z).

Gráfico N°3: Gráfico de caja de Compactación en la densidad de suelos



Fuente: Elaboración Propia.

Ejemplo 4:

Evaluar el efecto de tres tipos de agregados :Arena de trituración (D), Arena de sílice (E), Polvo de piedra (F) en la durabilidad del concreto.

Hipótesis:

H_0: No hay diferencia en la durabilidad del concreto entre los tipos de agregados.

H_1: Al menos un tipo de agregado produce una durabilidad significativamente diferente.

Paso 2: Datos

Los datos de la durabilidad del concreto para cada tipo de agregados del concreto están organizados en la tabla siguiente:

Agregados	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
Arena de trituración (D)	100	102	98	101	99
Arena de sílice (E)	110	112	109	111	110
Polvo de piedra (F)	120	121	119	122	120

Paso 3: Cálculo de la media general

$$\overline{X} = \frac{(110 + 112 + 109 + 111 + 110) + (120 + 121 + 119 + 122 + 120)}{15} + \frac{(100 + 102 + 98 + 101 + 99)}{15}$$

$$\overline{X} = 110.27$$

Paso 4: Cálculo de las sumas de cuadrados

Suma de cuadrados total (SST)

La fórmula es la siguiente:

$$SST = \sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2$$

Sustituyendo los valores:

$$SST = (100 - 110.27)^2 + (102 - 110.27)^2 + (98 - 110.27)^2 + \cdots + (120 - 110.27)^2$$

Realizando los cálculos, obtenemos:

$$SST = 1060.93$$

Suma de cuadrados entre grupos (SStr)

La fórmula utilizada para su cálculo es la siguiente:

$$SStr = \sum_{j=1}^{k} n_j \left(\overline{X}_j - \overline{X}\right)^2$$

Donde:

- jn_j número de muestras en cada grupo.
- $\overline{X}_j$ media de cada grupo.
- $\overline{X}$ media general.

Usando las medias de cada tipo de agregado:

- **Media de D** = 100.0
- **Media de E** = 110.2
- **Media de F** = 120.6

Realizando los cálculos:

$$SStr \ = \ 1040.53$$

Suma de cuadrados dentro de grupos (SSE)

Se obtiene mediante el siguiente cálculo:

$$SSE = SST - SStr$$

Sustituyendo los valores:

$$SSE = 1060.93 - 1040.53 = 20.40$$

Paso 5: Cálculo de los grados de libertad

- Grados de libertad Entre Tratamientos ($\boldsymbol{df_b}$) = g−1=3−1=2
- Grados de libertad del Error ($\boldsymbol{df_w}$) = N−g=15−3=12
- Grados de libertad Total **($\boldsymbol{df_t}$)** = N−1=15−1=14

Paso 6: Cálculo de las medias cuadradas y el estadístico F

Media cuadrada entre grupos (MStr)

Se calcula como:

$$MStr = \frac{SStr}{\boldsymbol{df_b}}$$

Aplicamos en la fórmula:

$$SStr = 5 \times (100.0 - 110.27)^2 + 5 \times (110.2 - 110.27)^2 + 5 \times (120.6 - 110.27)^2$$

$$MStr = \frac{SStr}{\boldsymbol{df_b}}$$

Sustituyendo los valores:

$$MStr = \frac{1040.53}{2} = 520.27$$

Media Cuadrada dentro de los grupos (MSE)

Se calcula como:

$$MSE = \frac{SSE}{df_w}$$

Sustituyendo los valores:

$$MSE = \frac{20.40}{12} = 1.70$$

Estadístico F

Se obtiene como:

$$F = \frac{MStr}{MSE}$$

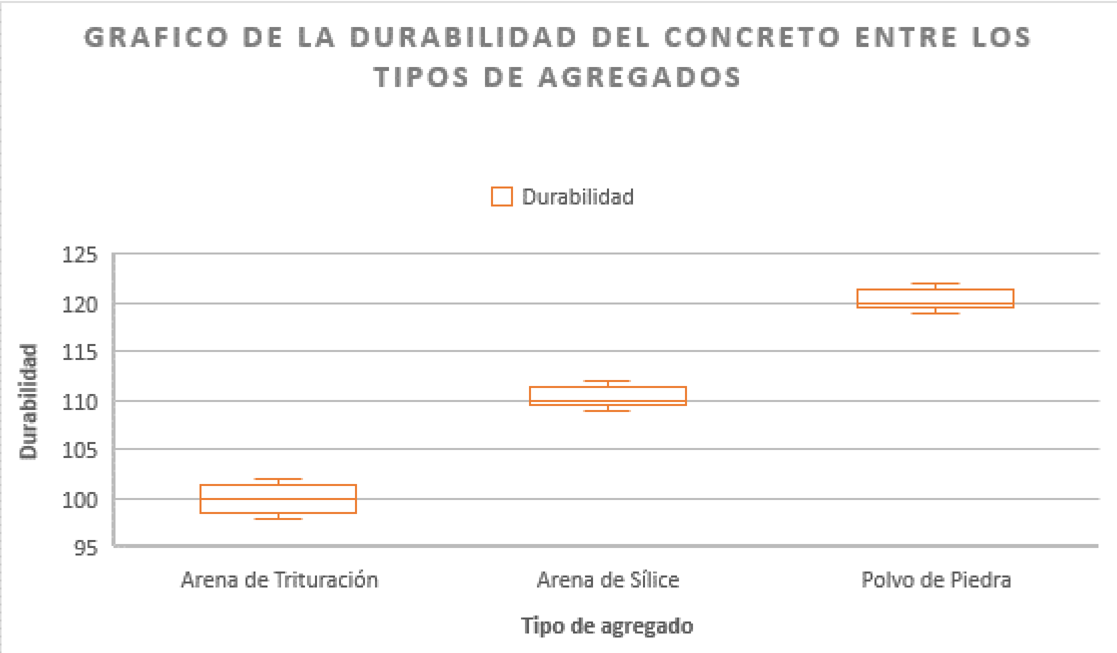
Sustituyendo los valores:

$$F = \frac{520.27}{1.70} = 306.04$$

Paso 7: Interpretación

El valor de F, que asciende a 306.04, se considera significativamente elevado. El valor observado de la estadística F supera el valor crítico para un nivel de significancia (α = 0.05), por lo cual e tiene suficiente evidencia para rechazar la hipótesis nula, concluyendo que existe al menos una diferencia en la resistencia del concreto según el tipo de agregados utilizados (D, E, F).

Gráfico N°4: Gráfico de caja de la durabilidad del concreto entre los tipos de agregados



Fuente: Elaboración Propia

Ejemplo 5:

Analizar el efecto de tres condiciones climáticas Frío y Húmedo (A), Cálido y Húmedo (B), Costero (C) en el tiempo de fraguado del concreto.

Hipótesis:

H₀: No hay diferencia en el tiempo de fraguado del concreto entre las condiciones climáticas: Frío y Húmedo (A), Cálido y Húmedo (B), Costero (C) .

H₁: Al menos una de las condiciones climáticas produce un tiempo de fraguado significativamente diferente.

Paso 2: Datos

Los datos del tiempo de fraguado del concreto para cada tipo de condiciones climáticas están organizados en la tabla siguiente:

Clima	Muestra 1	Muestra 2	Muestra 3	Muestra 4	Muestra 5
Frío y Húmedo (A),	180	182	178	181	179
Cálido y Húmedo (B),	150	152	148	151	150
Costero (C)	120	122	119	121	120

Paso 3: Cálculo de la media general

$$\overline{X} = \frac{+(150+152+148+151+150)+(120+122+119+121+120)}{15}$$
$$+\frac{(180+182+178+181+179)}{15}$$

$$\overline{X} = 150.2$$

Paso 4: Cálculo de las sumas de cuadrados

Suma de cuadrados total (SST)

Se calcula como:

$$SST = \sum_{i=1}^{N} \left(X_i - \overline{X}\right)^2$$

$$SST = (180-150.2)^2 + (182-150.2)^2 + (178-150.2)^2 + \cdots + (120-150.2)^2$$

Después de realizar los cálculos para cada observación, obtenemos:

$$SST = 8904.4SST \; = \; 8904.4SST = 8904.4$$

Suma de cuadrados entre grupos (SStr)

$$SStr = \sum_{j=1}^{k} n_j \left(\overline{X}_j - \overline{X}\right)^2$$

Donde:

- n_j número de muestras en cada grupo.
- $\overline{X}_j$ media de cada grupo.
- $\overline{X}$ media general.

Para cada clima, tenemos las medias:

- **Media de A** = 180.0
- **Media de B** = 150.2
- **Media de C** = 120.4

Aplicamos la fórmula:

$$SStr = 5 \times (180.0 - 150.2)^2 + 5 \times (150.2 - 150.2)^2 + 5 \times (120.4 - 150.2)^2$$

Realizando estos cálculos:

$$SStr = 8880.4 SStr \ = \ 8880.4 SStr = 8880.4$$

Suma de cuadrados del error (SSE)

Se obtiene mediante el siguiente cálculo:

$$SSE \ = \ SST \ - \ SStr$$

Sustituyendo los valores:

$$SSE = 8904.4 - 8880.4 = 24.0$$

Paso 5: Cálculo de los grados de libertad

- Grados de libertad Entre Tratamientos $(\boldsymbol{df_b}) = \ g - 1 = 3 - 1 = 2$
- Grados de libertad del Error $(\boldsymbol{df_w}) = \ N - g = 15 - 3 = 12$
- Grados de libertad Total $(\boldsymbol{df_t})$ = $N - 1 = 15 - 1 = 14$

Paso 6: Cálculo de las medias cuadradas y el estadístico F

Media cuadrada entre grupos (MStr)

$$MStr = \frac{SStr}{df_b}$$

Sustituyendo los valores:

$$MStr = \frac{8880.4}{2} = 4440.2$$

Media cuadrada dentro de los grupos (MSE)

Se obtiene como:

$$MSE = \frac{SSE}{df_w}$$

Sustituyendo los valores:

$$MSE = \frac{24.0}{12} = 2.0$$

Estadístico F

Se obtiene como:

$$MSEF = \frac{MStr}{MSE}$$

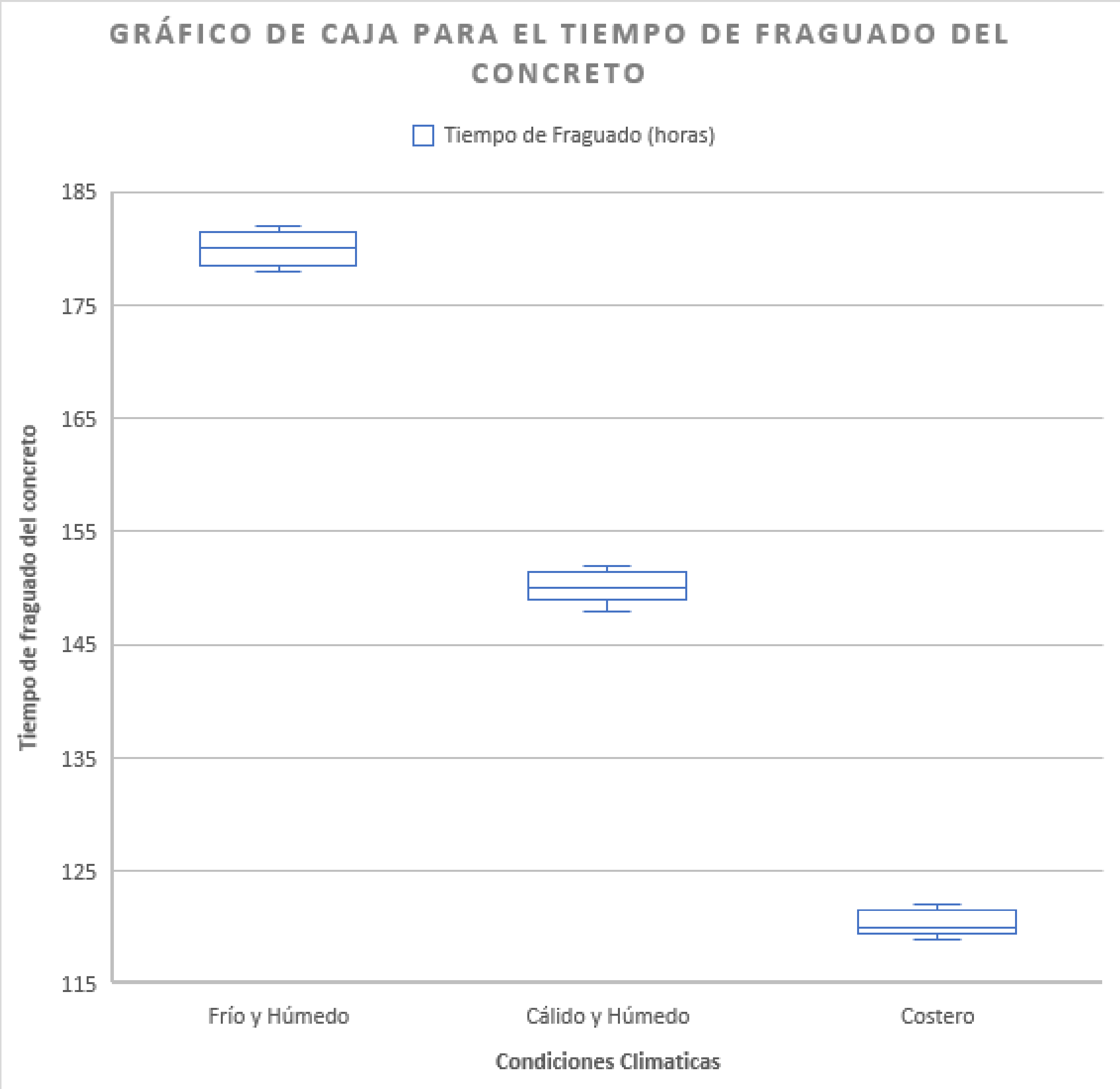
Sustituyendo los valores:

$$F = \frac{4440.2}{2.0} = 2220.1$$

Paso 7: Interpretación

Con un valor de 2220.1 para la prueba F, se puede afirmar que este valor es significativamente elevado. El valor observado de la estadística F supera el valor crítico correspondiente para un nivel de significancia ($\alpha = 0.05$), con ello se tiene suficiente prueba para negar la hipótesis nula concluyendo que existe en al menos un grupo diferencia en el tiempo de fraguado del concreto bajo distintas condiciones climáticas (A, B, C).

Gráfico N°5: Gráfico de caja para el Tiempo de Fraguado del Concreto



Fuente: Elaboración Propia.

9. Conclusiones

- Este estudio comprende plenamente los conceptos teóricos del ANOVA de un solo factor), con miras a la eficiencia de este enfoque estadístico para detectar variaciones notables entre grupos independientes. Demostrando el poderoso instrumento estadístico para examinar el efecto de una sola variable categórica en un resultado cuantitativo, los investigadores pueden obtener resultados más exactos tanto en estudios experimentales como observacionales.

10. Recomendaciones

- Los modelos mixtos, son recomendados para el análisis de datos con estructuras jerárquicas o mediciones repetidas. Al combinar ambos tipos de efectos, estos modelos ofrecen un enfoque más flexible y equilibrado.
- Cuando los datos no cumplen con los supuestos de normalidad y homogeneidad de varianzas requeridos para el ANOVA, se sugiere utilizar métodos alternativos como el MANOVA, que permite analizar múltiples variables independientes.

11. La creación de scripts en Python con el propósito de analizar ANOVA de un solo factor posee la capacidad de incrementar significativamente tanto la eficacia como la reproducibilidad de investigaciones subsiguientes. Códigos Utilizados en Python para el Análisis ANOVA

Para ejecutar códigos es necesario tener instalado Python y una serie de bibliotecas:

1. Instalar `numpy` para trabajar con arrays y funciones matemáticas:

   pip install numpy

2. Instalar `pandas` para la manipulación y análisis de datos:

   pip install pandas

3. Instalar `matplotlib` para la visualización de gráficos:

   pip install matplotlib

4. Instalar `statsmodels` para realizar el análisis ANOVA:

   pip install statsmodels

Código del programa

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import statsmodels.api as sm

from statsmodels.formula.api import ols
```

```
def ingresar_datos():
    """Función para que el usuario ingrese datos para varios grupos."""
    grupos = {}
    print("Ingrese los datos para cada grupo. Escriba 'fin' cuando haya terminado.")
    while True:
        nombre_grupo = input("\nNombre del grupo (o 'fin' para terminar): ")
        if nombre_grupo.lower() == 'fin':
            break
        datos = input(f"Ingrese los valores para {nombre_grupo} separados por comas: ")
        try:
            grupos[nombre_grupo] = list(map(float, datos.split(',')))
        except ValueError:
            print("Error: asegúrese de ingresar números separados por comas.")
    return grupos
def realizar_anova(grupos):
    """Función para calcular la ANOVA y mostrar los resultados junto al gráfico."""
    # Crear un DataFrame con los datos ingresados por el usuario
    data = pd.DataFrame({
        'Valor': [valor for grupo in grupos.values() for valor in grupo],
        'Grupo': [grupo for grupo, valores in grupos.items() for _ in valores]
    })
    # Realizar el análisis ANOVA
    model = ols('Valor ~ C(Grupo)', data=data).fit()
    anova_table = sm.stats.anova_lm(model, typ=2)
```

```
    try:
        f_value = anova_table['F'][0]
        p_value = anova_table['PR(>F)'][0]
        sum_sq_between = anova_table['sum_sq'][0]
        df_between = anova_table['df'][0]
        sum_sq_within = anova_table['sum_sq'][1]
        df_within = anova_table['df'][1]
    except KeyError:
        print("No se pudo calcular la estadística F y el valor p. Verifique los datos
ingresados.")
        return
    # Crear un gráfico con matplotlib
    fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(14, 6))
    # Gráfico de caja personalizado
    boxplot = ax1.boxplot([valores for valores in grupos.values()], labels=grupos.keys(),
patch_artist=True)
    # Personalizar el gráfico
    for patch in boxplot['boxes']:
        patch.set_facecolor('orange')  # Color de la caja (IQR)
    for median in boxplot['medians']:
        median.set_color('red')
        median.set_linewidth(2)
    ax1.set_title('Distribución de valores por grupo')
    ax1.set_xlabel('Grupo')
    ax1.set_ylabel('Valor')
```

```
    ax1.grid(True)

    # Mostrar los resultados como una lista en el lado derecho

    ax2.axis('off')

    ax2.set_title('Resultados del análisis ANOVA')

    # Crear la lista de resultados

    resultados_texto = (

        "Resultados del ANOVA:\n\n"

        f"- Suma de cuadrados entre grupos: {sum_sq_between:.4f}\n"

        f"- Grados de libertad entre grupos: {df_between}\n"

        f"- Suma de cuadrados dentro de grupos: {sum_sq_within:.4f}\n"

        f"- Grados de libertad dentro de grupos: {df_within}\n"

        f"- Estadística F: {f_value:.4f}\n"

        f"- Valor p: {p_value:.4f}\n"

        "\nInterpretación:\n"

        "- Un valor p bajo (< 0.05) indica diferencias significativas entre los grupos.\n"

        "- Un valor p alto (>= 0.05) sugiere que no hay diferencias significativas.\n"

    )

    # Mostrar el texto en el gráfico

    ax2.text(0.1, 0.5, resultados_texto, fontsize=12, va='center', wrap=True)

    plt.tight_layout()

    plt.show()

# Ingresar datos y realizar ANOVA

grupos = ingresar_datos()

if grupos:
```

```
  realizar_anova(grupos)
else:
  print("No se ingresaron datos.")
```

12. Referencias Bibliográficas

DataCamp. (2024). Prueba ANOVA: Guía detallada con ejemplos. https://www.datacamp.com/es/tutorial/anova-test

Field, A. (2013). Discovering statistics using IBM SPSS Statistics (4ª ed.). SAGE Publications.

Fisher, R. A. (1925). Statistical methods for research workers. Oliver & Boyd.

Howell, D. C. (2012). Statistical methods for psychology (8ª ed.). Wadsworth.

Montgomery, D. C. (2013). Design and analysis of experiments (8ª ed.). John Wiley & Sons.

Navarro, D. J. (2019). Aprendiendo estadística con R: Un tutorial para estudiantes de psicología y otros principiantes. LibreTexts Español. https://espanol.libretexts.org/Estadisticas/Estadistica_Aplicada/Libro_Aprendiendo_Estadistica_con_R

Probabilidad y Estadística. (2023). Análisis de la varianza (ANOVA). https://www.probabilidadyestadistica.net/analisis-de-la-varianza-anova/

Statologos. (2024). Cómo verificar los supuestos de ANOVA. https://statologos.com/supuestos-de-anova/

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics (6ª ed.). Pearson.

Universidad Loyola Andalucía. (s.f.). Pruebas ANOVA. Repositorio de la Universidad Loyola. https://repositorio.uloyola.es/bitstream/handle/20.500.12412/2166/Tema_10_Pruebas_ANOVA.pdf