**Ejercicio Propuesto 1  (Resuelto)**

**Se realiza un estudio con el objetivo de conocer si hay colores que sean más atractivos que otros para los insectos. Para ello se diseñaron trampas con los siguientes colores:  azul, verde, blanco y amarillo. Se cuantificó el número de insectos que quedaban atrapados:**

****

Tabla 1: Datos del Ejercicio Propuesto 1

1. **¿Existen diferencias significativas entre el número de insectos atrapados según el color de las trampas? Utilice un nivel de significación del 5%.**
2. **En caso afirmativo en el apartado anterior, ¿entre qué colores? Utilice un nivel de significación del 5%.**
3. **Estudiar las hipótesis del modelo: Homocedasticidad, independencia y normalidad.**

**Solución**

1. **¿Existen diferencias significativas entre el número de insectos atrapados según el color de las trampas? Utilice un nivel de significación del 5%.**

Nos interesa saber si el color de las trampas influye significativamente en el número de insectos atrapados

* El número de insectos atrapados es una **variable numérica**
* El color de las trampas es una **variable categórica** que divide a los individuos en grupos

Nuestro contraste de hipótesis es:



Es decir,

El número de insectos atrapados es el mismo para los cuatro colores de trampas

El número de insectos atrapados es diferente según el color de las trampas

En primer lugar, cargamos el paquete BrailleR

> library(“BrailleR”)

almacenamos los datos de las dos variables en dos vectores:

> color =factor(c(rep("Azul",6), rep("Verde",6), rep("Blanco",6), rep("Amarillo",6)), levels = c("Azul", "Verde", "Blanco", "Amarillo"))

> atrapados=c(16, 11, 20, 21, 14, 7, 37, 32, 15, 25, 39, 41, 21, 12, 14, 17, 13, 17, 45, 59, 48, 46, 38, 47)

Agrupamos las 2 variables en un data frame, al que llamamos *Datos*:

> Datos = data.frame (color, atrapados)

Realizamos el ANOVA pedido mediante la función **OneFactor** sin que nos muestre el Test post-hoc.

> OneFactor("atrapados", "color", Datos, HSD = FALSE)

Muestra una salida en html:

# **[Analysis of the Datos data, using Atra­­pados as the response variable and Color as the single grouping factor.](C:\\Users\\Usuario\\Desktop\\datos\\Atrapados.Color-OneFactor.html)**

#### **[Prepared by BrailleR](C:\\Users\\Usuario\\Desktop\\datos\\Atrapados.Color-OneFactor.html)**

A continuación, se muestra y se detalla cada resultado:

1. **Una tabla resumen que incluye para cada grupo la media, la desviación típica, el número de datos y el error estándar**

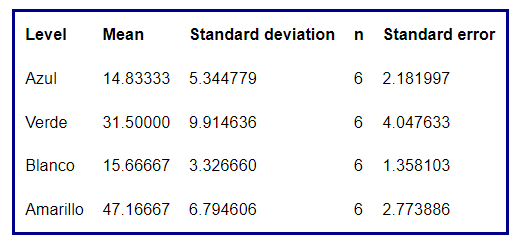
****

Tabla 2: Resumen para cada color

1. **Diagramas de caja y bigotes comparativos para cada grupo**

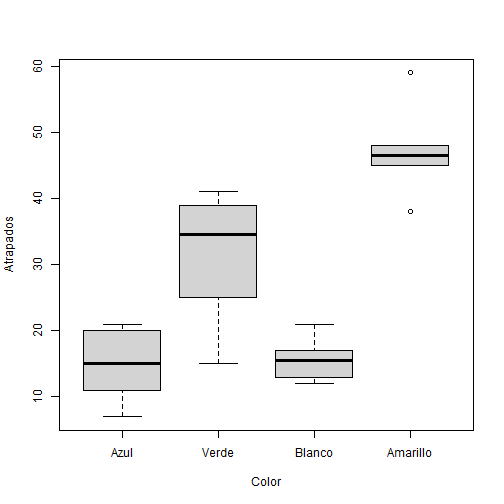
****

Figura 1: Diagrama de caja y bigotes para cada color

* + Los grupos Azul, Blanco y Amarillo parecen seguir una distribución simétrica, mientras que en el grupo Verde no se ve claramente tal simetría.
  + En el color Amarillo se detectan dos valores atípicos que habrá que estudiar con detalle por si fuese necesario eliminarlos.
  + El tamaño de las cajas parece diferente para los distintos niveles, por lo que podría incumplirse la hipótesis de homocedasticidad. Lo confirmaremos posteriormente gráfica y analíticamente.
  + Se puede intuir que existirá diferencia de medias entre los grupos Azul, Blanco y Verde con respecto al Amarillo; y Azul y Blanco con respecto al Verde.

1. **Gráfico de puntos comparativos para cada grupo**

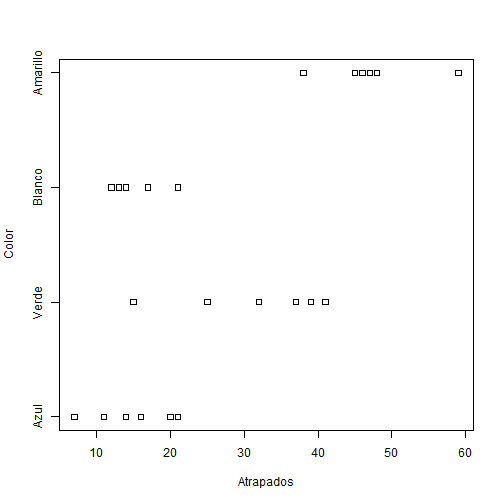
****

Figura 2: Gráfico de puntos para cada color

Al igual que en el diagrama de caja y bigotes, se puede tener una idea intuitiva de si las medias difieren unas de otras o no. Obteniendo las mismas conclusiones que con el gráfico anterior.

1. **Análisis de la varianza de un factor**

Df Sum Sq Mean Sq F value Pr(>F)

color 3 4218 1406 30.55 1.15e-07 \*\*\*

Residuals 20 920 46

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Esta es la tabla ANOVA que nos proporciona un p-valor=1.15e-07 por lo que se rechaza la hipótesis nula, esto quiere decir que existen diferencias significativas entre el número de insectos atrapados según el color de las trampas (al menos en dos de ellos).

1. **En caso afirmativo en el apartado anterior, ¿entre qué colores? Utilice un nivel de significación del 5%.**

Se realiza un análisis post-hoc, mediante el método Tukey HSD. Añadimos a la función el test post-hoc de la siguiente forma:

> OneFactor("atrapados", "color", Datos)

[Muestra una salida en html](file:///C:\Users\Usuario\Desktop\datos\Atrapados.Color-OneFactor1.html) que nos devuelve todo lo anterior junto con lo que se especifica a continuación:

# **[Analysis of the Datos data, using Atrapados as the response variable and Color as the single grouping factor.](C:\\Users\\Usuario\\Desktop\\datos\\Atrapados.Color-OneFactor1.html)**

#### **[Prepared by BrailleR](C:\\Users\\Usuario\\Desktop\\datos\\Atrapados.Color-OneFactor1.html)**

1. **Test post-hoc mediante el método de Tukey Honestly Significant Difference**

Tukey multiple comparisons of means

95% family-wise confidence level

factor levels have been ordered

Fit: aov(formula = atrapados ~ color, data = Datos)

$color

diff lwr upr p adj

Blanco-Azul 0.8333333 -10.129663 11.79633 0.9964823

Verde-Azul 16.6666667 5.703670 27.62966 0.0020222

Amarillo-Azul 32.3333333 21.370337 43.29633 0.0000004

Verde-Blanco 15.8333333 4.870337 26.79633 0.0032835

Amarillo-Blanco 31.5000000 20.537004 42.46300 0.0000006

Amarillo-Verde 15.6666667 4.703670 26.62966 0.0036170

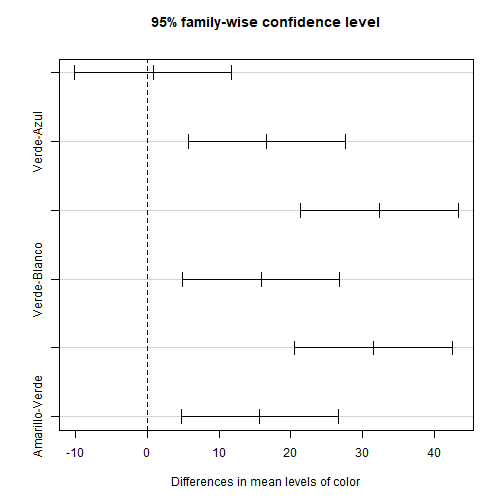


Figura 3: Intervalos de confianza de la diferencia de medias para cada par de colores

Tanto numéricamente mediante el p-valor como gráficamente se tiene que:

* Blanco y Azul: Puesto que p-valor=0.9964823, no se puede rechazar H0, es decir, no hay diferencias significativas en el número de insectos atrapados según si el color es blanco o azul.
* Verde y Azul: Puesto que p-valor=0.0020222, se rechaza H0, es decir, existen diferencias significativas en el número de insectos atrapados entre los colores verde y azul.
* Amarillo y Azul: Puesto que p-valor=0.0000004, se rechaza H0, es decir, existen diferencias significativas en el número de insectos atrapados entre los colores amarillo y azul.
* Verde y Blanco: Puesto que p-valor=0.0032835, se rechaza H0, es decir, existen diferencias significativas en el número de insectos atrapados entre los colores verde y blanco.
* Amarillo y Blanco: Puesto que p-valor=0.0000006, se rechaza H0, es decir, existen diferencias significativas en el número de insectos atrapados entre los colores amarillo y blanco.
* Amarillo y Verde: Puesto que p-valor=0.0036170, se rechaza H0, es decir, existen diferencias significativas en el número de insectos atrapados entre los colores amarillo y verde.

En definitiva, existen diferencias significativas en el número de insectos atrapados según todos los colores excepto entre el blanco y el azul.

Se pueden deducir las mismas conclusiones, observando que los intervalos de confianza, tanto analítica como gráficamente no contienen al cero.

1. **Estudiar las hipótesis del modelo: Homocedasticidad, independencia y normalidad.**

Para comprobar la idoneidad del modelo propuesto, la salida de la función **OneFactor** muestra los gráficos:

1. **Valores ajustados frente a residuos**
2. **Gráfico Q-Q de normalidad**
3. **Valores ajustados frente a raíz cuadrada de los residuos estandarizados (en valor absoluto)**
4. **Residuos estandarizados frente a leverages**

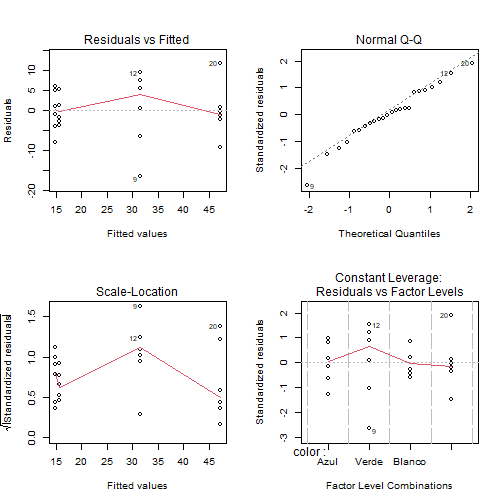
****

Figura 4: Gráficos de comprobación de las hipótesis

Los gráficos 1 y 3 se utilizan para contrastar gráficamente si la media de los residuos es cero y la homocedasticidad de los residuos. La idea intuitiva para que se cumplan estas hipótesis es que no se observe ningún patrón ni forma de embudo. Además, en el gráfico 3 se usa como regla general que cualquier punto por encima de 2 en el eje Y sugeriría heterogeneidad de varianza. Por tanto, se confirma que los residuos tienen media cero y son homocedásticos.

Observando el gráfico Q-Q, se observa que los residuos estandarizados están próximos a la diagonal punteada que aparece en el gráfico, por tanto, los residuos siguen una distribución normal.

El gráfico de Leverage para cada grupo frente a los residuos estandarizados se utiliza para detectar valores atípicos. En caso de detectarse alguna observación fuera del rango [-2,2] debe estudiarse este punto de forma aislada para detectar, por ejemplo, si la elevada importancia de esa observación se debe a un error. Solo se detecta un valor atípico.

La hipótesis de independencia es intrínseca al diseño del experimento, por tanto, se supone que las observaciones son independientes ya que se están considerando distintos colores.

1. **Dos tests de homogeneidad de varianzas**

Vamos a confirmar la homocedasticidad mediante dos métodos analíticos.

* **El test de homogeneidad de varianzas de Bartlett**

Bartlett test of homogeneity of variances

data: atrapados by color

Bartlett's K-squared = 5.2628, df = 3, p-value = 0.1535

Considerando como nivel de significación 0.05, y observando el p-valor=0.1535 no tenemos evidencia muestral para rechazar la hipótesis nula, por lo que la varianza permanece constante entre grupos.

* **El test de homogeneidad de varianzas de Fligner-Killeen**

Fligner-Killeen test of homogeneity of variances

data: atrapados by color

Fligner-Killeen:med chi-squared = 3.7498, df = 3, p-value = 0.2898

Considerando como nivel de significación 0.05, y observando el p-valor=0.2898 no tenemos evidencia muestral para rechazar la hipótesis nula, por lo que la varianza permanece constante entre grupos.