

# PRINCIPIOS DE DISEÑOS EXPERIMENTALES

## PRACTICA No.2

### Refrescos

Una empresa embotelladora de refrescos está interesada en obtener alturas de llenado más uniformes en las botellas que se fabrican en su proceso de manufactura. Teóricamente, la máquina de llenado llena cada botella a la altura objetivo correcta, pero en la práctica, existe variación en torno a este objetivo, y a la embotelladora le gustaría entender mejor las fuentes de esta variabilidad y, en última instancia, reducirla. El ingeniero del proceso puede controlar tres variables durante el proceso de llenado:

- el porcentaje de carbonatación,
- la presión de operación en el llenador y
- las botellas producidas por minuto o rapidez de línea.

Es sencillo controlar la presión y la rapidez, pero el porcentaje de carbonatación es más difícil de controlar durante la manufactura real debido a que varía con la temperatura. Por lo tanto los factores de diseño son la presión y la rapidez. Para los fines de un experimento, el ingeniero puede controlar la carbonatación en tres niveles: 10, 12 y 14%. Elige dos niveles para la presión (25 y 30psi) y dos niveles para la rapidez de línea (200 y 250bpm). El ingeniero decide correr dos réplicas de un diseño factorial con estos tres factores, haciendo las 24 corridas de manera aleatoria.

La variable respuesta observada es la desviación promedio absoluta de la altura del llenado objetivo que se observa en una corrida de producción de botellas con cada conjunto de condiciones. Las desviaciones positivas son alturas de llenado promedio en la corrida arriba del objetivo, mientras que las negativas son alturas de llenado promedio abajo del objetivo. Sin embargo, interesa solo la desviación absoluta ( $d_2$ ).

### Ejercicios

1. Usando como respuesta la desviación absoluta ( $d_2$ ):
  - Analice las interacciones de primer orden (dobles) para cada par de factores. Haga un análisis gráfico. Es importante que decida bien cuál factor se pone en el eje X.
  - Debe decir si espera que exista interacción en cada caso y por qué.
  - A partir de los tres gráficos obtenidos anteriormente obtenga conclusiones preliminares sobre los efectos de los factores de diseño.
2. Tome solamente las variables carbonatación y presión. Escriba un modelo con aov que considere la interacción entre los factores.
  - Obtenga las estimaciones de los efectos simples y de interacción usando `model.tables`.
  - A partir de los promedios adecuados, estime manualmente los efectos simples de presión y carbonatación. Compárelos con la tabla que da R.

- Estime manualmente los efectos de interacción. Compárelos con la tabla de R.
- Haga el cálculo de la Suma de Cuadrados y el Cuadrado Medio de Presión.
- Haga el cálculo de la Suma de Cuadrados y el Cuadrado Medio de Carbonatación.
- Qué representa cada uno de estos números?
- Usando las interacciones calculadas previamente, haga el cálculo de la Suma de Cuadrados y el Cuadrado Medio de interacción.

- Calcule el Cuadrado Medio Residual a partir de las variancias de los tratamientos. Interprete este valor.
- Compare este valor con el que da la tabla del análisis de variancia.
- Escriba la hipótesis nula referente a la interacción con símbolos y con palabras.
- Haga la prueba de esta hipótesis.
- Dé una estimación de los promedios de cada tratamiento.
- Construya los contrastes adecuados y haga las pruebas de hipótesis para estos contrastes.
- Construya los intervalos de confianza adecuados.
- Dé una conclusión del efecto del factor de diseño en este caso.
- Escriba el modelo usando lm y obtenga las estimaciones de los coeficientes del summary. Dé una interpretación de todos los coeficientes.

3. Tome solamente las variables carbonatación y rapidez. Escriba un modelo con aov que considere la interacción entre los factores.

- Obtenga las estimaciones de los efectos simples y de interacción usando model.tables.
- A partir de los promedios adecuados, estime manualmente los efectos simples de rapidez y carbonatación. Compárelos con la tabla que da R. Compare el efecto de carbonatación con el obtenido en el punto 2.
- Estime manualmente los efectos de interacción. Compárelos con la tabla de R.
- Usando los promedios adecuados, haga el cálculo de la Suma de Cuadrados y el Cuadrado Medio de Rapidez.
- Usando las interacciones calculadas previamente, haga el cálculo de la Suma de Cuadrados y el Cuadrado Medio de interacción.

- Calcule el Cuadrado Medio Residual a partir de las variancias de los tratamientos. Interprete este valor.
- Compare este valor con el que da la tabla del análisis de variancia.
- Escriba la hipótesis nula referente a la interacción con símbolos y con palabras.
- Haga la prueba de esta hipótesis.
- Pase al modelo adecuado.
- Dé una estimación de los promedios de cada tratamiento.

- Represente gráficamente estos promedios y comente lo que observa. Comente la relación que hay entre estos promedios estimados y los observados.
  - Observe el CMResidual en este nuevo modelo y explique por qué no coincide con la media de las variancias.
  - Escriba la hipótesis nula referente al efecto del factor de diseño con símbolos y con palabras.
  - Pruebe la hipótesis.
  - Haga el intervalo de confianza adecuado.
  - Dé una conclusión del efecto del factor de diseño en este caso.
4. Haga un modelo con todas las interacciones dobles.
- Observe los grados de libertad de cada fuente de variación y justifique cada número.
  - Obtenga la media de las variancias de todos los tratamientos.
  - Observe el CMResidual y vea que no coincide con la media de las variancias de los tratamientos. ¿Por qué?
  - Obtenga las estimaciones de los efectos principales e interacciones usando `model.table` y compárelos con los obtenidos en los puntos anteriores. ¿Cambian?
5. Evaluación de interacciones de primer orden:
- ¿Cuál es la primera interacción que puede asumirse como nula?
  - Elimine esa interacción y continúe el proceso para decidir si puede asumir que otras interacciones son nulas.
  - En el modelo final sólo se pueden hacer supuestos sobre las interacciones pero nunca se pueden eliminar los factores principales.
  - Escriba el modelo final.
6. Usando el modelo final haga las pruebas e intervalos de confianza pertinentes:
- ¿De qué forma se pueden analizar los efectos de Rapidez?
  - ¿De qué forma se pueden analizar los efectos de Presión?
  - ¿Interesa analizar los efectos de Carbonatación?
7. Obtenga los promedios estimados con el modelo reducido usando los coeficientes estimados y la matriz de estructura. Verifique sus resultados con la función “predict”.
- En este experimento se tienen 12 tratamientos, pero se obtienen 24 estimaciones de los promedios. Esto significa que cada promedio está repetido 2 veces porque hay 2 réplicas por tratamiento. Obtenga las estimaciones de los promedios para los 12 tratamientos sin repetir.
  - Compare sus resultados con los promedios observados.
  - ¿Por qué los observados y los estimados no coinciden?

## Salsas

Una diseñadora de productos gourmet tiene como objetivo determinar el nivel de picante (suave, medio y fuerte) que tiene mayor satisfacción en promedio. Sus pruebas las realiza con dos tipos de tomate como base (cherry y pera).

El interés del experimento es decidir cuál nivel incrementa la satisfacción de los consumidores potenciales, para esto hace un estudio con personas que gustan de productos alimenticios especializados.

La variable respuesta es un índice de satisfacción calculado a partir de 8 preguntas en una escala de 0 al 10. El índice es simplemente el promedio de los puntajes de las 8 preguntas para cada persona.

En total se trabajó con 8 personas (hombres y mujeres entre 20 y 40 años) y a cada una se le dio a probar todos los tipos de salsa que consistía en la combinación de picante y tipo de tomate.

La diseñadora espera que una diferencia de medio punto entre dos medias sea suficiente para demostrar que una salsa es más satisfactoria que otra.

Los datos se encuentran en el archivo "salsa3.Rdata".

## Ejercicios

1. Si el objetivo del estudio es analizar solamente el efecto del nivel de picante y su interacción con el tipo de tomate, ¿sería necesario incluir el sexo y la edad de cada persona en el modelo? ¿Por qué?
  - Si además se quiere saber si el efecto del tipo de picante es diferente según sexo, ¿qué tipo de diseño se tendría que hacer?
  - Si se asume que no hay interacción entre tipo de picante y sexo, pero se quiere saber si la salsa es de más agrado a hombres o a mujeres, ¿qué tipo de diseño se tendría que hacer? ¿Qué cambia en este caso con respecto al inciso anterior?
2. Represente los datos mediante un gráfico de líneas para analizar si se puede esperar que haya un efecto del nivel de picante.
  - A partir del gráfico, ¿considera que las personas son diferentes entre sí en cuanto al puntaje que dan a las salsa?
3. Represente los datos mediante un boxplot para analizar si se puede esperar que haya un efecto del nivel de picante, sin olvidar que también se cuenta con el otro factor (tipo de tomate). ¿Por qué es importante hacer el gráfico considerando también el tipo de tomate?
  - Compare el gráfico correcto (con datos centrados) con el gráfico incorrecto (puntajes originales).
4. Obtenga la Suma de Cuadrados Total directamente de la variable respuesta.
  - Haga un modelo con picante y tomate solamente (con interacción) y verifique que la suma de cuadrados total coincide con la obtenida anteriormente.
  - Obtenga manualmente los efectos simples de picante y tomate. Verifique sus resultados comparando con los coeficientes del modelo anterior (suma nula)
  - Obtenga manualmente la suma de cuadrados de cada factor.
  - Haga un modelo donde incluya la persona como bloque y además la interacción entre picante y tomate. verifique que la suma de cuadrados total se mantiene.
  - Obtenga manualmente la suma de cuadrados de bloques

5. Obtenga la media de las variancias de los 6 tratamientos y compárela con el CMRes del modelo con interacción (sin bloques)
  - Obtenga el cuadrado medio residual a partir de la suma de cuadrados total, cuando se considera picante, tomate, interacción entre ambos y bloques.
  - Cuántos parámetros tiene el modelo con interacción y bloques?
  - Usando el número de parámetros del modelo, obtenga los grados de libertad residuales.
6. Verifique que no hay interacción. ¿Qué implicaciones tiene este resultado para el análisis posterior?
7. Si se considera que es mejor que una persona no pruebe varios tipos de salsa sino una sola, explique cómo se debe ejecutar el experimento en este caso.
  - ¿Cuántas personas se necesitarían para asegurar que se logre el mismo nivel de información que dio este experimento?
  - ¿Qué tendría que pasar para que el número de personas que se necesiten si se hace el experimento de esta forma sea igual a 48, es decir, el mismo número de datos que en el experimento actual? Justifique.

## Mariposas

Se quiere determinar el efecto de la altura (dosel o sotobosque) sobre la presencia de mariposas en dos especies (Colobura dirce y Dioristes tauropolis). Se tomaron 4 árboles separados donde se pusieron trampas.

En cada árbol se pusieron 2 trampas: una se situó en el dosel, y la otra en un pequeño espacio de luz inmediatamente bajo la primera (sotobosque). Al cabo de cierto tiempo se visitaron las trampas y se registró el número de individuos de cada especie en cada trampa. Se hicieron 4 visitas a cada trampa, una vez por semana. Se trata de un experimento con dos factores, cada uno con dos niveles y 16 réplicas para cada tratamiento. Se tienen en total  $2 \times 2 \times 16 = 64$  observaciones.

Los datos se encuentran en el archivo `Mariposas2.csv`.

## Ejercicios

1. Lea los datos y defina una nueva variable llamada `bloque` que consista en el árbol + la semana de visita. Para esto puede usar la función `paste` de la siguiente forma: `paste(variable1,variable2)`. Además verifique que la variable `bloque`, `tipo`, `altura` y `especie` sean factores.
2. Haga un gráfico de líneas para observar si hay diferencias entre bloques, ponga los bloques en el eje X y use la combinación de especie y altura en `groups` poniendo `groups=paste(altura,especie)`. También puede hacer un gráfico de puntos colocando las medias de la frecuencia por bloque.
3. Centre los datos de frecuencia y haga gráficos de cajas con la frecuencia original y centrada poniendo en el eje X `altura+especie`. Asegúrese de que ambos gráficos tiene la misma escala en el eje Y. Comente.
4. Haga el análisis de varianza para probar si hay interacción entre altura y especie, tomando en consideración el bloque.
  - Haga el análisis de varianza para probar si hay interacción entre altura y especie pero no tome en consideración el bloque. Compare el CMRes de este análisis con el anterior. Explique qué está pasando.
5. Ahora vamos a tomar en cuenta la variable llamada `tipo`. Esta variable clasifica los árboles según la ubicación en el bosque, que puede ser en la sombra o en la luz. Por lo tanto, ahora tiene un nuevo factor,

pero los árboles pueden estar solo en una de las dos ubicaciones, por lo que este diseño se convierte en parcelas divididas.

- ¿Cuál es el factor de parcela?
  - ¿Cuáles son los factores de subparcela?
  - Analice gráficamente si se puede esperar interacción entre altura y tipo, así como entre especie y tipo.
6. Escriba un modelo mixto con las interacciones entre altura y especie, tipo y altura, y finalmente tipo y especie. Recuerde que el bloque es el elemento aleatorio.
- Para obtener los grados de libertad para el error de parcela, correspondiente a tipo, debe multiplicar  $a(r-1)$ , donde  $a$  es el número de niveles de tipo, y  $r$  es el número de repeticiones dentro de cada tipo para cada tratamiento. Obtenga los grados de libertad de parcela.
  - Para obtener los grados de libertad residuales puede seguir el siguiente truco. Haga un modelo equivalente al anterior pero normal, sin bloques, tome los grados de libertad residuales de ese modelo y reste los grados de libertad del error de parcela.
  - Verifique las hipótesis relativas a las interacciones.