Taller 1

Para cada uno de los siguientes problema organice la sintaxis en $\mathbf{R}$ que le permita obtener las respectivas salidas para responder las preguntas formuladas. Apóyese de la sintaxis en $\mathbf{R}$ presentada en el Ejemplo que se está analizando.

1. Con los datos del Caso de Estudio 1, verifique usando el software R, el cumplimiento de los supuestos del modelo de un factor.
2. Retomando el Ejemplo 1, ver diapositiva 8. ¿Hay alguna diferencia entre los cuatro fertilizantes en la producción promedio de kilos de papas?. En todas las pruebas use un $\alpha = 0.05$.
   - *a*) Identifique la variable respuesta, la variable que define el factor y los tratamientos. Realice un análisis de varianza para determinar si existe diferencia significativa entre los fertilizantes, escriba las respectivas hipótesis e interprete sus términos. Escriba la expresión para el valor P de la prueba.
   - *b*) Estime e interprete (signo) los efectos de cada uno de los tratamientos.
   - *c*) Verifique el supuesto de homogeneidad de varianza entre los tratamientos. Explicar en qué consiste el procedimiento utilizado. Haga en forma explícita la prueba de Bartlett, realice cada uno de los cálculos.
   - *d*) Verifique el supuesto de normalidad de los errores. Plantee las hipótesis y concluya.
   - *e*) Compare los tratamientos y recomiende el mejor, para ello tenga en cuenta las comparaciones de Tukey y Duncan, además del análisis descriptivo usando los boxplot?s. Con este diseño puede concluirse ¿cuál de los fertilizantes proporciona la mejor producción?
   - *f*) Suponiendo que fertilizante 4 es fertilizante estándar, control, determine si conjuntamente los resultados de los otros fertilizantes, difieren del estándar, use el método de comparaciones de Dunnet.
   - *g*) ¿Se podrá afirmar que el fertilizante 2 ofrece en promedio los mismos resultados de los otros tres fertilizantes? Plantee un contraste y realice la respectiva prueba de hipótesis.

TAREA 2. DISEÑO DE EXPERIMENTOS

Fecha entrega: Miércoles 7 de septiembre de 2022, enviar vilopez@unal.edu.co en formato pdf, con asunto: Tarea 2 Diseño de Experimentos, y nombre del archivo: T2_Apellido1_Apellido2_Apellido3.pdf. EQUIPOS Máximo de tres personas

Primer Problema a resolver por todos:

Problema 1: Se desea investigar el efecto del pH en el crecimiento de cierto microorganismo en un medio específico. Para ello se realiza un experimento, teniendo como punto de partida la misma cantidad de microorganismos. Se hacen cuatro repeticiones y se obtienen los siguientes resultados. ¿Estos datos son evidencia suficiente para afirmar que los niveles de pH donde se logra menor y mayor crecimiento son el 3 y e1 2, respectivamente? Explique y liste las consideraciones que se deben tener en cuenta para que las conclusiones obtenidas sean válidas.

Nivel de pH	Crecimiento promedio (en $\%$)
1	80
2	105
3	75

Comparación entre medias

Tarea 2 Continuación

NOTA Para escoger el segundo problema a resolver tenga en cuenta el siguiente procedimiento, para escoger uno de los cinco problemas cuyo enunciado aparece posteriormente.

A. Sumar el último dígito de cada uno de los integrantes del equipo
B. Divida el número obtenido por CINCO.
C. Si el residuo es CERO, escoja PROBLEMA 1. Si el residuo es UNO, escoja PROBLEMA 2. Si el residuo es DOS, escoja PROBLEMA 3. Si el residuo es TRES, escoja PROBLEMA 4. Si el residuo es CUATRO, escoja PROBLEMA 5.Por ejemplo: La suma del último dígito de tres documentos dio: 14, al dividirlo por CINCO, el residuo es 4, en este caso corresponde el PROBLEMA CINCO.

Con el problema que les tocó, hacer lo siguiente:

I. Identificar y describir los siguientes elementos: **Unidad experimental, Factor, niveles del factor, variable respuesta**.
II. Escribir el modelo estadístico asociado al respectivo diseño, indicando cuál es la variable respuesta y los supuestos del modelo.
III. Verificar los supuestos del modelo: Normalidad, homogeneidad de varianza (Prueba de Bartlett, y la prueba de Levene Modificada) e independencia.
IV. Construir la tabla ANOVA y probar la significancia de los efectos de los tratamientos. Use α=0.05. A partir de las expresiones para SSTrat, SSE, etc, verifique los resultados de la Tabla ANOVA reportada por el R.

Comparación entre medias

Tarea 2 Continuación

V. Según sea el caso, realice agrupaciones entre medias de los tratamientos. Usar el método de Duncan. Plantee un contraste para comparar la media del primer tratamiento con las otras medias y haga la prueba de significancia de dicho contraste, use $\alpha = 0.05$.

NOTA: Se debe adjuntar la sintaxis realizada en R, e incoporar en el informe las tablas y gráficos bien editadas.

Comparación entre medias

Tarea 2 Continuación

Considere el enunciado de los siguientes cinco problemas:

1. Se hace un estudio sobre la efectividad de tres marcas de spray para matar moscas. Para ello, cada spray se aplica a un grupo de 100 moscas, y se cuenta el número de moscas muertas, expresado en porcentajes. Se hacen seis réplicas y los resultados obtenidos se muestran enseguida:

Marca de Spray	Nro de moscas muertas
1	72 65 67 75 62 73
2	55 59 68 70 53 50
3	64 74 61 58 51 69

2. Para estudiar la confiabilidad de ciertos tableros electrónicos para carros, se someten a un envejecimiento acelerado durante 100 horas a determinada temperatura, midiéndose como variable de interés la intensidad de corriente que circula entre dos puntos, cuyos valores aumentan con el deterioro. Se probaron 20 módulos repartidos equitativamente en cinco temperaturas y los resultados obtenidos son:

20 C	40 C	60C	80 C	100 C
15	17	23	28	45
18	21	19	32	51
13	11	25	34	57
12	16	22	31	48

3. Un experimento para comparar el tiempo de respuesta en segundos de tres grupos de animales a tres diferentes estímulos. Se opta por asignar 14 animales a cada uno de los estímulos.

Estímulo	Tiempos de Respuesta
I	16 14 14 13 13 12 12 17 17 17 19 14 15 20
II	6 7 7 8 4 8 9 6 8 6 4 9 5 5
III	8 10 9 10 6 7 10 9 11 11 9 10 9 5

Comparación entre medias

Tarea 2 Continuación

4. Un experimento se condujo para comparar tres métodos de empaque para cierto alimento. El criterio fue el contenido de ácido ascórbico (mg/100g) después de un periodo de tiempo. Se decide asignar n unidades a cada método.

Método de Empaque	Contenido ácido ascórbico (mg/100g)
A	14.29 19.10 19.09 16.25 15.09 16.61 19.63
B	20.06 20.64 18.00 19.56 19.47 19.07 18.38
C	20.04 26.23 22.74 24.04 23.37 25.02 23.27

5. Una compañía farmaceútica desea evaluar el efecto que tiene la cantidad de almidón en la dureza de las tabletas. Se decidió producir lotes en una cantidad determinada de almidón. Se probó con $2\,\%$, $5\,\%$, $10\,\%$ de cantidad de almidón y se midió la dureza promedio de 20 tabletas de cada lote.

% Almidón	Dureza
2	4.3 5.2 4.8 4.5
5	6.5 7.3 6.9 6.1
10	9.0 7.8 8.5 8.1

Comparación entre medias

Anotación con relación a la Prueba de Durbin-Watson

Para probar que los Errores están incorrelacionados

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

El estadístico de Prueba es:

$$d = \frac{\sum_{t=2}^{N} (e_t - e_{t-1})^2}{\sum_{t=1}^{N} e_t^2}$$

Para un nivel de significancia dado (α), si el $\mathbf{valor-p}$ de la prueba es mayor que dicho nivel ($\mathbf{Valor-p} > \alpha$) entonces se puede concluir que los errores tienen correlación cero. En $\mathbf{R}$, con la función **durbinWatsonTest()** de la librería: **car**, se ha observado que el **Valor p** cambia cada vez que lo corre para el mismo conjunto de datos. Lo anterior se debe a que el progrma internamente hace un remuestreo en los residuales para hallar el valor P del Estadístico. Sin embargo, la conclusión no debe cambiar de forma drástica.

Modelo de un Factor: Efectos Aleatorios

Modelo Factorial de Efectos Aleatorios

Un diseño de un factor es de efectos aleatorios si los niveles del factor son una muestra de una población de niveles, que tiene una media cero y una varianza dada por σ_τ^2, esta última conocida como componente de varianza del factor.

Nota: En este tipo de diseños la variabilidad asociada a la variable respuesta (e.d. la variabilidad del modelo) depende de la **varianza del término de error** ($V(\varepsilon) = \sigma^2$) y de la **varianza** asociada al **efecto aleatorio** ($V(\tau) = \sigma_\tau{}^2$)

Modelo de un Factor: Efectos Aleatorios

Caso de Estudio 2:

Una fábrica de **cuchillas de afeitar** utiliza una gran cantidad de máquinas en la producción.

Se desea que las máquinas sean homogéneas para producir objetos de la misma calidad.

Para investigar si existen **variaciones significativas** entre las máquinas, se **seleccionan cuatro (4) máquinas al azar** y se mide el porcentaje de un cierto componente de la hoja.

Variable Respuesta: Y: Porcentaje de un cierto componente de la hoja de afeitar.

Factor: Tipo de Máquina.

Niveles del Factor: Maq1, Maq2, Maq3, Maq4.

Tipo de diseño: Diseño de un factor con Efectos Aleatorios.

Un Ejemplo más

En una fábrica se han observado anomalías en la calidad de las piezas producidas por un **tipo de máquinas**: a partir de una revisión reciente de las máquinas se sospecha que los defectos es debido a los trabajadores. Para contrastar esta hipótesis se toma una muestra aleatoria de trabajadores y se controla la calidad de las distintas piezas que cada uno obtiene.

- ¿Cuál es la variable respuesta?
- ¿Qué factor se está estudiando?
- Los niveles del factor son una muestra aleatoria de una población mayor de tratamientos.

Modelo de un Factor: Efectos Aleatorios

Modelo de efectos aleatorios

El modelo de efectos aleatorios se representa como:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \;\; i = 1,\, 2,\, \ldots,\, a \text{ y } j = 1,\, 2,\, \ldots,\, n$$

con τ_i y ε_{ij} variables aleatorias con los siguientes supuestos:

Supuestos:

- $\tau_i \sim NID(0, \sigma_\tau^2)$
- $\varepsilon_{ij} \sim NID(0, \sigma^2)$
- τ_i y ε_{ij} son v. a. independientes luego:

$$Var(Y_{ij}) = \sigma_\tau^2 + \sigma^2.$$

Componentes de Varianza: σ_τ^2 y σ^2.

Modelo de un Factor: Efectos Aleatorios

...

En este tipo de diseño, se sigue cumpliendo la descomposición de la Suma de cuadrados totales (SS_T) en: Sumas de Cuadrados de los Tratamientos (SS_{Trat}) y en Suma de cuadrados de los Errores, (SS_E), es decir,

$$\boxed{SS_T = SS_{Trat} + SS_E}$$

Pruebas de Hipótesis de Interés

En el caso del modelo de **Efectos Aleatorios**, la realización de las pruebas de hipótesis coinciden con la realizada para el Modelo de Efectos Fijos, cambia la formulación de las Hipótesis y las conclusiones obtenidas.
La tabla **ANOVA**, se construye al descomponer las sumas de cuadrados, de acuerdo al diseño experimental usado.
Se obtienen las medias cuadráticas (MS_{Trat} y MS_E) para cada una de las fuentes de variación.
Luego se obtienen las Esperanzas de las Medias Cuadráticas, i. e. $E[MS_{Trat}]$ y $E[MS_E]$).
Y se hallan los estimadores de cada **Media Cuadrática**, MS.

Modelo de un Factor: Efectos Aleatorios

Formulación de las Hipótesis en el modelo de Efectos Aleatorios

La Hipótesis Nula en este caso es: **igualar la varianza del efecto aleatorio** a CERO, equivalentemente a probar que las esperanzas de las medias cuadráticas son iguales, i. e.:

$$\begin{cases} H_0 : E[MS_{Trat}] = E[MS_E] \\ H_a : E[MS_{Trat}] > E[MS_E] \end{cases} \iff \begin{cases} H_0 : \sigma^2 + n\sigma^2_\tau = \sigma^2 \\ H_a : \sigma^2 + n\sigma^2_\tau > \sigma^2 \end{cases}$$

$$\iff \begin{cases} H_0 : \sigma^2_\tau = 0 \\ H_a : \sigma^2_\tau > 0 \end{cases}$$

...

Se puede demostrar que:

$$E[MS_{Trat}] = \sigma^2 + n\sigma_\tau^2 \quad \text{y que} \quad E[MS_E] = \sigma^2,$$

de donde:

$$\hat{\sigma}^2 = MS_E \quad \text{y} \quad \hat{\sigma}_\tau^2 = \frac{MS_{Trat} - MS_E}{n}$$

y si el diseño es desbalanceado:

$$n_0 = \frac{1}{a-1}\left[\sum_{i=1}^{a} n_i - \frac{\sum n_i^2}{\sum n_i}\right].$$

El Estadístico de Prueba, está dado por:

$$F = \frac{SS_{Trat}/(a-1)}{SS_E/(N-a)} = \frac{MS_{Trat}}{MS_E} \sim F_{a-1,N-a}$$

Se rechaza H_0 si $F_{cal} > F_{\alpha;a-1,N-a} = qf(1-\alpha,\ a-1,\ N-a)$.

Modelo de un Factor: Efectos Aleatorios

...

En resumen:

Para el caso de un solo factor la prueba de hipótesis para un modelo de efectos aleatorios (o modelo de componentes de varianza), es idéntica a la prueba PH del modelo de efectos fijos, sólo cambian las conclusiones.

Todos los tratamientos serán iguales si $\sigma_\tau^2 = 0$. Sin embargo, si $\sigma_\tau^2 > 0$ existe variabilidad entre los tratamientos.

El procedimiento de cálculo es igual que en el modelo de efectos fijos, aunque las conclusiones se aplican a toda la población de tratamientos.

Modelo de un Factor: Efectos Aleatorios

¿Componentes de Varianza Negativos?

Las estimaciones de los componentes de varianza (i. e. $\hat{\sigma}^2$ y $\hat{\sigma}_\tau^2$) por el método de momentos (ANOVA) puede dar estimaciones negativas.

Algunas sugerencia a seguir en este caso son:

1. La estimación es evidencia de que esta etimación debe ser **cero** y por tanto se usa el **cero** como la estimación, aunque sea **sesgada**.
2. Interpretar la estimación de la componente negativa como una indicación de un **modelo estadístico incorrecto**.
3. Utilizar un método diferente de estimación.
4. Reunir más datos y analizarlos por separado o junto con los existentes y esperar que el aumento de información conduzca a estimaciones positivas.

Modelo de un Factor: Efectos Aleatorios

Retomando el Caso de Estudio 2:

Una fábrica de cuchillas de afeitar utiliza una gran cantidad de máquinas en la producción. Se desea que las máquinas sean homogéneas para producir objetos de la misma calidad.

Para investigar si existen variaciones significativas entre las máquinas, se seleccionan cuatro (**4**) máquinas al azar y se mide el porcentaje de un cierto componente de la hoja.

El experimento se realiza en orden aleatorio. Los resultados obtenidos son:

Máquina	Porcentaje del Componente			
Máquina-1	98	97	99	96
Máquina-2	91	90	93	92
Maquina-3	96	95	97	95
Maquina-4	95	96	99	98

Modelo de un Factor: Efectos Aleatorios

...

Para de terminar la significancia del factor: **Tipo de máquina**, se contrastan las siguientes hipótesis:

$$\begin{cases} H_0 : \sigma_\tau^2 = 0 \\ H_a : \sigma_\tau^2 > 0 \end{cases}$$

Suponga que el nivel de significancia utilizado es $\alpha = 0.05$.
Lueggo de usar la misma función **aov** del $\mathbf{R}$ se obtiene la siguiente tabla ANOVA:

Fuente de Variación	Df	Sum Sq	Mean Sq	F value
Maquinas	3	89.19	29.73	15.68
Residuales	12	22.75	1.90	
Total	15	111.94		

Como

$$F_{cal} = 15.68 > 3.49 = F_{\alpha,a-1,N-a} = F_{0.05,3,12} = qf(0.95, 3, 12)$$

$\therefore$ Se rechaza $H_0 : \sigma_\tau^2 = 0$, i. e. que **existe una variabilidad entre los tratamientos**, i.e. existe **variabilidad entre las máquinas** a un nivel de significancia de 0.05.

Componentes de Varianza-Estimación

Estimación de las componentes de varianza:

$$\hat{\sigma}_\tau^2 = \frac{MS_{Trat} - MS_E}{n} = \frac{29.73 - 1.90}{4} = 6.96$$

y

$$\hat{\sigma}^2 = MS_E = 1.90.$$

La varianza de cualquier observación de la muestra es:

$$\widehat{Var}(Y_{ij}) = \hat{\sigma}_\tau^2 + \hat{\sigma}^2 = 6.96 + 1.90 = 8.86,$$

Se observa que la mayor parte de la variabilidad se atribuye a diferencias entre las máquinas.

NOTA: Todas las demás pruebas para verificar los supuestos del modelo, en particular las que tienen que ver con el Término de Error, se hacen de manera similar al modelo de Efectos Fijos.