

TRABAJO FINAL

KEVIN STEVEN GARCÍA^a, JOSE ALEJANDRO VARGAS^b

1. Problema 1

1.
 - a.
 - b.
 - c.
2.
 - a.
 - b.
 - c.
 - d.
 - e.
 - f.
 - g.
 - h.
 - i.
 - j.

^aCódigo: 1533173. E-mail: kevin.chica@correounivalle.edu.co

^bCódigo: 1525953. E-mail: jose.alejandro.vargas@correounivalle.edu.co

2. Problema 2

Un ingeniero de alimentos evalúa el rendimiento del almidón agrio de yuca (en kg), el cual ha sido sometido a tres métodos de secado: 1) al sol, 2) con aire caliente, 3) 50 % al sol y 50 % aire caliente. El experimento se lleva a cabo en dos rallanderías diferentes, utilizando el mismo lote de almidón agrio (obtenido en Julio de 2008) y por cada rallandería tres trabajadores llevan a cabo la experimentación. Los resultados se muestran a continuación:

Método de Secado	Rallandería I			Rallandería II		
	Trabajador			Trabajador		
	1	2	3	1	2	3
1	34.5	48.0	40.8	33.6	44.0	36.5
	40.7	45.2	43.0	35.2	44.4	40.4
	42.0	49.6	41.7	36.8	43.9	38.8
2	41.4	52.2	39.4	41.9	47.0	40.8
	43.8	51.5	45.3	42.6	46.5	40.9
	42.6	51.7	48.6	43.7	47.5	44.8
3	58.0	62.1	51.4	52.3	58.8	44.9
	54.8	62.5	49.9	48.5	62.0	45.9
	56.7	60.3	48.5	55.8	63.8	49.0

a. El objetivo del estudio es evaluar el rendimiento del almidón agrio de yuca (en Kg) con respecto a tres métodos de secado diferentes, dos rallanderías y tres trabajadores dentro de cada rallandería, el objetivo final sería obtener una conclusión acerca de qué método de secado, qué rallandería y cuáles trabajadores están generando un mayor rendimiento del almidón, lo cual es lo que interesa finalmente.

b.

- Factores:

- ◊ Factor 1: Método de secado.
- ◊ Factor 2: Rallandería.
- ◊ Factor anidado: Trabajador.

- Niveles:

- ◊ Factor 1: Sol(S), aire caliente(A) y 50 % al sol y 50 % aire caliente(SA).
- ◊ Factor 2: Rallandería 1(R1) y Rallandería 2(R2).
- ◊ Factor anidado: Trabajador 1(T1), trabajador 2(T2) y trabajador 3(T3).

- Tratamientos: Son 18 tratamientos en total. S-R1-T1, S-R1-T2, S-R1-T3, S-R2-T1, S-R2-T2, S-R2-T3, A-R1-T1, A-R1-T2, A-R1-T3, A-R2-T1, A-R2-T2, A-R2-T3, SA-R1-T1, SA-R1-T2, SA-R1-T3, SA-R2-T1, SA-R2-T2, SA-R2-T3.

c. La unidad experimental es el almidón agrio de yuca.

d. Dado que el almidón agrio utilizado para todo el experimento fue del mismo lote, simplemente se debe asignar al azar una parte de este a cada método de secado, a cada lavandería y a cada trabajador, es decir, se debe dividir el lote de almidón agrio de yuca en 18 partes, enumerar cada uno y seleccionar de forma aleatoria un número entre 1 y 18, asignándole a cada tratamiento el lote de almidón que le corresponda.

e. Para el análisis exploratorio de datos, dado que nos interesa encontrar diferencias entre método de secado, entre rallanderías y entre trabajadores dentro de las rallanderías, se obtendrán descriptivas para los factores principales y para el factor anidado, además, también se quiere evaluar si existen diferencias entre la interacción método de secado y rallandería, por ello también se sacaran estadísticas descriptivas de este cruce.

Método de secado:

Método de secado	\bar{X}	S	$CV(\%)$	Min	Max
Sol	41.06	4.530049	11.03	33.6	49.6
Aire caliente	45.12	3.965521	8.78	39.4	52.2
Sol y Aire caliente	54.73	6.128525	11.19	44.9	63.8

Rallanderia:

Rallanderia	\bar{X}	S	$CV(\%)$	Min	Max
Rallanderia 1	48.38	7.358581	15.21	34.5	62.5
Rallanderia 2	45.57	7.633227	16.75	33.6	63.8

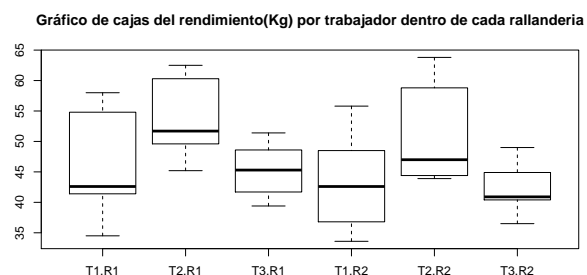
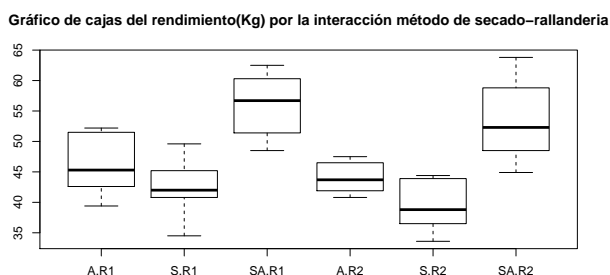
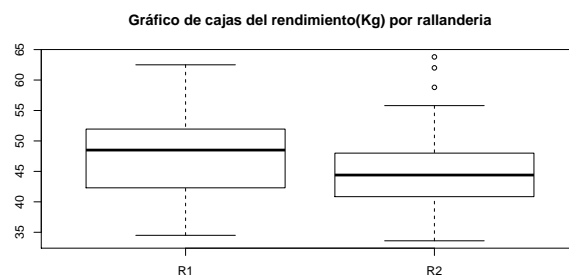
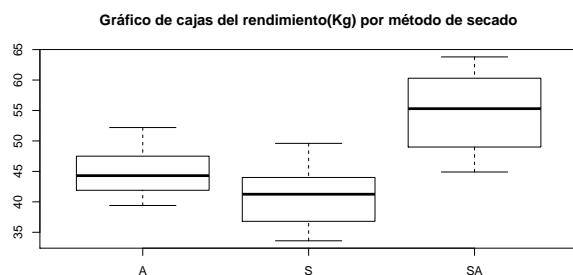
Trabajador:

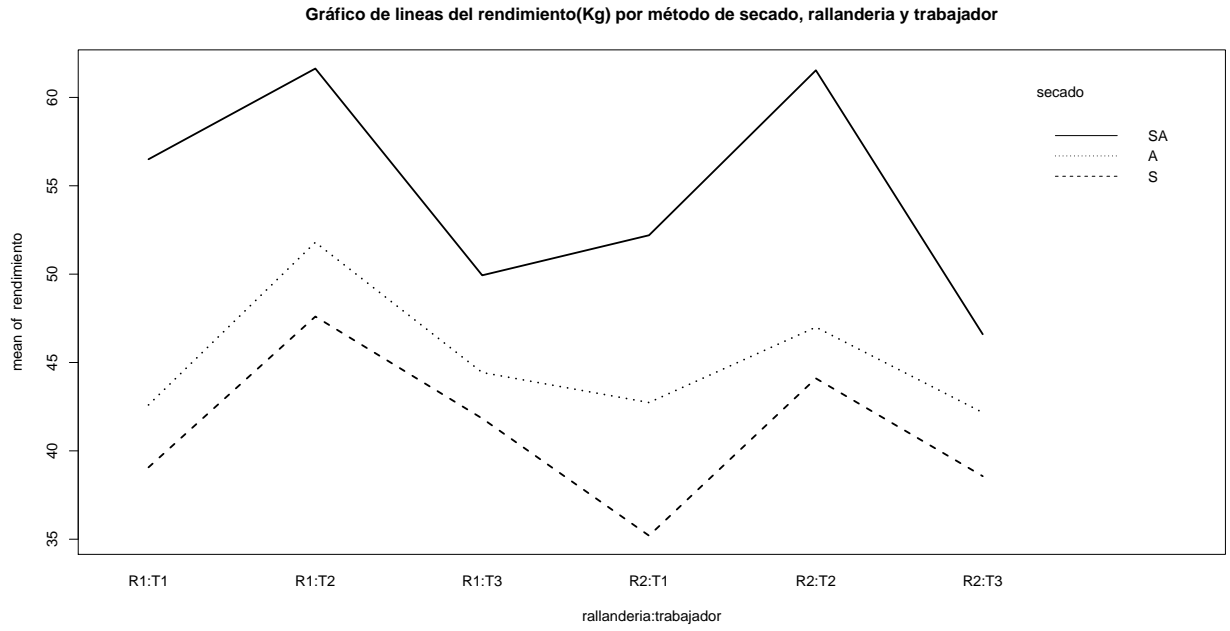
Trabajador	\bar{X}	S	$CV(\%)$	Min	Max
Rallanderia 1					
Trabajador 1	46.06	8.290071	18	34.5	58
Trabajador 2	53.68	6.365881	11.86	45.2	62.5
Trabajador 3	45.4	4.36921	9.62	39.4	51.4
Rallanderia 2					
Trabajador 1	43.38	7.655027	17.64	33.6	55.8
Trabajador 2	50.88	8.193103	16.10	43.9	63.8
Trabajador 3	42.44	3.943067	9.29	36.5	49

Interacción método de secado - rallanderia:

Método de secado - rallanderia	\bar{X}	S	$CV(\%)$	Min	Max
Rallanderia 1					
Sol	42.83	4.44719	10.38	34.5	49.6
Aire caliente	46.28	4.858698	10.5	39.4	52.2
Sol y aire caliente	56.02	5.226083	9.32	48.5	62.5
Rallanderia 2					
Sol	39.29	4.094339	10.42	33.6	44.4
Aire caliente	43.97	2.608639	5.93	40.8	47.5
Sol y aire caliente	53.44	6.983035	13.06	44.9	63.8

Para ver de forma gráfica posibles diferencias entre los niveles de los factores principales(método de secado y rallanderia) y entre los niveles del factor anidado(trabajador), se realizaron los siguientes gráficos de cajas. No se hizo el de la interacción triple, porque se perdía claridad en la interpretación, para esta interacción se tiene un gráfico de líneas posterior.





f.

El modelo estadístico es:

$$y_{ijkl} = \mu + \tau_i + \alpha_j + \gamma_{k(j)} + (\tau\alpha)_{ij} + (\tau\gamma)_{ik(j)} + \varepsilon_{(ijk)l}$$

$$i = 1, 2, 3 \quad j = 1, 2 \quad k = 1, 2, 3 \quad l = 1, 2, 3$$

Supuestos:

$$\begin{aligned} \varepsilon_{ijkl} &\simeq N(0, \sigma^2) \\ E[\varepsilon_{ijkl}] &= 0 \quad \forall_{ijkl} \\ V[\varepsilon_{ijkl}] &= \sigma^2 \quad \forall_{ijkl} \\ Cov[\varepsilon_{ijkl}, \varepsilon_{i'j'k'l'}] &= 0 \quad \forall_{i \neq i' \wedge j \neq j' \wedge k \neq k' \wedge l \neq l'} \end{aligned}$$

Donde:

y_{ijkl} es el rendimiento del almidón agrio de yuca en Kg bajo el método de secado i, en la rallandería j y bajo el trabajador k en la l-ésima réplica.

μ es la media general del rendimiento del almidón agrio de yuca en Kg sin tener en cuenta el método de secado, la rallandería ni el trabajador.

τ_i es el efecto del i-ésimo método de secado sobre el rendimiento del almidón agrio de yuca en Kg.

α_j es el efecto de la j-ésima rallandería sobre el rendimiento del almidón agrio de yuca en Kg.

$\gamma_{k(j)}$ es el efecto del k-ésimo trabajador anidado a la j-ésima rallandería sobre el rendimiento del almidón agrio de yuca en Kg.

$(\tau\alpha)_{ij}$ es el efecto de la interacción del método de secado i-ésimo y la rallandería j-ésima sobre el rendimiento del almidón agrio de yuca en Kg.

$(\tau\gamma)_{ik(j)}$ es el efecto de la interacción del método de secado i-ésimo y el trabajador k-ésimo anidado a la j-ésima rallandería sobre el rendimiento del almidón agrio de yuca en Kg.

$\varepsilon_{(ijk)l}$ es el error aleatorio debido al i-ésimo método de secado, a la j-ésima rallandería, al k-ésimo trabajador anidado a la j-ésima rallandería y a la l-ésima réplica.

g. Las hipótesis a evaluar son:

1).

$$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau$$

H_a : Al menos una de estas igualdades no se cumple

2).

$$H_0 : \alpha_1 = \alpha_2 = \alpha$$

H_a : Al menos una de estas igualdades no se cumple

3).

$$H_0 : \gamma_{1(1)} = \gamma_{2(1)} = \gamma_{3(1)} = \gamma_{1(2)} = \gamma_{2(2)} = \gamma_{3(2)} = \gamma$$

H_a : Al menos una de estas igualdades no se cumple

4).

$$H_0 : (\tau\alpha)_{ij} = 0$$

$$H_a : (\tau\alpha)_{ij} \neq 0$$

5).

$$H_0 : (\tau\gamma)_{ik(j)} = 0$$

$$H_a : (\tau\gamma)_{ik(j)} \neq 0$$

h.

Fuente de Variación	Grados de Libertad	Suma de Cuadrados	Cuadrados Medios	F	Pr>F	ECM
τ	2	1774.8	887.4	180.140	0	$\sigma^2 + 9 \sum_{i=1}^3 \tau_i^2 + 3\sigma_{\tau\gamma}^2$
α	1	106.7	106.7	21.656	0.0000431	$\sigma^2 + 27 \sum_{j=1}^2 \alpha_j^2 + 9\sigma_{\tau\gamma}^2$
$\tau\alpha$	2	3.8	1.9	0.385	0.683429	$\sigma^2 + \frac{9}{2} \sum_{i=1}^3 \sum_{j=1}^2 (\tau\alpha)_{ij} + 3\sigma_{\tau\gamma}^2$
$\gamma_{(j)}$	4	765.9	191.5	38.868	0	$\sigma^2 + 9\sigma_{\tau\gamma}^2$
$\tau\gamma_{(j)}$	8	201	25.1	5.101	0.000274	$\sigma^2 + 3\sigma_{\tau\gamma}^2$
Error	36	177.3	4.9			σ^2
Total	53	3029.5				

A un nivel de significancia $\alpha = 0.05$ se rechaza H_0 en las hipótesis (1),(2),(3) y (5) del literal anterior.

i.

Normalidad:

Para verificar el supuesto de normalidad en los errores, se hicieron pruebas gráficas(histograma y Q-Q plot) y se realizó la prueba de normalidad Anderson-Darling por ser más potente, ya que se tiene una cantidad de datos superior a 30.

En las gráficas no se observa un posible incumplimiento de la normalidad, se puede ver que las barras del histograma se ajustan muy bien a la curva teórica de la normal, y en el QQ-plot se observa que la gran mayoría de los puntos están sobre la recta de los cuantiles teóricos, cabe aclarar que la prueba gráfica no es muy confiable porque depende de la experiencia del investigador, por ello nos vamos a centrar en los resultados de la prueba estadística, ya que estas son más objetivas.

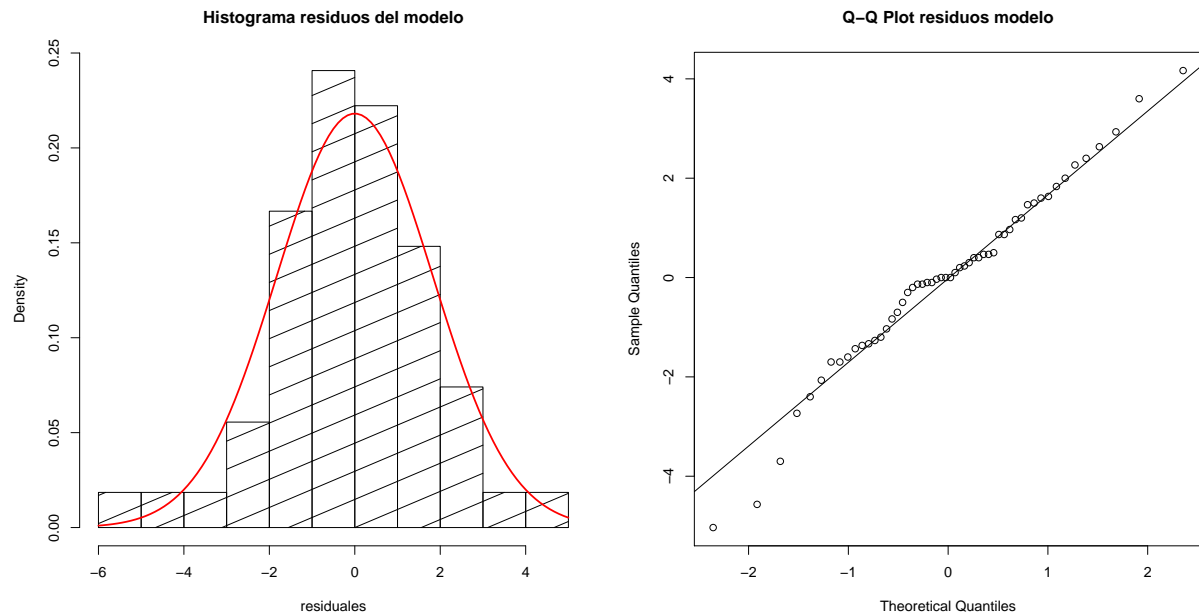
Los resultados para la prueba estadística realizada fueron:

H_0 : Los residuales provienen de una distribución normal

H_0 : Los residuales no provienen de una distribución normal

$$A = 0.43888, p - value = 0.2832$$

Como el p-valor=0.2832 es superior a nuestro nivel de significancia $\alpha = 0.05$, no se rechaza la hipótesis nula, y concluimos que no hay evidencia suficiente para decir que los residuales no provienen de una distribución normal. Lo cual nos indica que este supuesto de normalidad en los errores parece cumplirse.



Homocedasticidad:

Para analizar este supuesto se realizó el test de Bartlett, ya que bajo el cumplimiento de la normalidad este es más potente.

Los resultados la prueba realizada fueron:

H_0 : La varianza de los grupos no son diferentes (iguales)

H_a : La varianza de los grupos son diferentes

Bartlett's $K - squared = 25.586, df = 17, p - value = 0.08234$

Como el p valor=0.08234 es superior a nuestro nivel de significancia establecido $\alpha = 0.05$ no se rechaza la hipótesis nula, y concluimos que no hay evidencia suficiente para decir que la varianza de los grupos (18 grupos en este caso) son diferentes, por lo tanto, el supuesto de homocedasticidad en los errores parece cumplirse.

Independencia:

Para analizar el supuesto de independencia se realizó el test de Rachas, ya que este no se puede evaluar con un test Durbin- Watson o un correlograma porque estos requieren conocer el orden en el cuál fueron obtenidos los datos y esa información no se tiene.

Los resultados la prueba realizada fueron:

H_0 : La muestra es aleatoria (los residuales son independientes)

H_a : La muestra no es aleatoria (los residuales no son independientes)

Standard Normal = 1.8326, $p - value = 0.06685$

Como el p valor=0.06685 es mayor que nuestro nivel de significancia $\alpha = 0.05$, no rechazamos la hipótesis nula y concluimos que no hay evidencia suficiente para decir que los residuales no están distribuidos de manera aleatoria, en otras palabras, no podemos decir que no son independientes. Teniendo en cuenta esto, se puede concluir que este supuesto de independencia en los errores parece cumplirse.

j. INTERPRETAR

k. Para evaluar el control local del experimento nos centramos en la varianza estimada y en el coeficiente de variación

$$\sigma^2 = 4.9 \quad \bar{y}_{....} = 46.97222 \quad \rightarrow CV = \frac{\sqrt{\sigma^2}}{\bar{y}_{....}} = 0.047125 \cdot 100\% \simeq 4.7125\%$$

Dado que la desviación estándar $\sigma = 2.2136$ es pequeña y el coeficiente de variación también lo es, concluimos que hubo un buen control local en el experimento.

l.

Dado que en la anova se obtuvo que la interacción método de secado-rallandería no es significativa, se puede proceder a evaluar los factores individualmente, es decir, realizaremos contrastes para ver cuál o cuáles niveles del factor método de secado difieren en su efecto en el rendimiento del almidón agrio de yuca, también sería interesante saber qué trabajador difiere en cuanto a su efecto en el rendimiento del almidón. No se realizarán contrastes para el factor rallandería, ya que solo tiene dos niveles, y dado que en la anova ya se obtuvo que existen diferencias, resultaría redundante realizar contrastes para este factor.

Método de secado: En las estadísticas descriptivas y en el gráfico de cajas se vio que aparentemente, el método de secado Sol y aire caliente son parecidos, pero estos a su vez difieren del método 50 % sol y 50 % aire caliente. Esto se verificará teóricamente con los siguientes contrastes.

$$H_0 : \tau_1 - \tau_2 = 0$$

Bajo la hipótesis nula, tenemos que el estadístico de prueba es:

$$F = \frac{(\sum_{i=1}^t a_i \bar{y}_{i.})^2}{\hat{\sigma}^2 \sum_{i=1}^t \frac{a_i^2}{r_i}}$$

$$\rightarrow F = \frac{(-4.06)^2}{4.9 \cdot \frac{1}{9}} = 30.276$$

Como $F_{0.95(1,51)} = 4.030393 < F_{cal} = 30.276$, se rechaza H_0 y concluimos que hay evidencia estadística para decir que existen diferencias entre el efecto del método de secado al sol y el método de secado a aire caliente sobre el rendimiento (en Kg) del almidón agrio de yuca.

Ahora se verificará si el método de secado 50 % sol y 50 % aire caliente difiere de los otros dos métodos de secado (sol y aire caliente).

$$H_0 : \tau_1 + \tau_2 - \tau_3 = 0$$

$$\rightarrow F = \frac{(-23.28)^2}{4.9 \cdot \frac{1}{3}} = 331.81$$

Como $F_{0.95(1,51)} = 4.030393 < F_{cal} = 331.81$, se rechaza H_0 y concluimos que si hay diferencias significativas entre los efectos de los métodos de secado sol y aire caliente con el método de secado 50 % sol y 50 % aire caliente sobre el rendimiento del almidón agrio de yuca.

Con los resultados de las dos hipótesis anteriores, se puede concluir que los efectos de los 3 métodos de secado son distintos, complementando esto con las estadísticas descriptivas obtenidas, se podría decir que en general, el método de secado que mayor rendimiento (en Kg) produce sobre el almidón agrio de yuca es el de 50 % sol y 50 % aire caliente seguido por el método de secado aire caliente y finalmente, el método que aparentemente menor rendimiento produce es el de secado al sol.

Trabajador: Con respecto a la **rallandería 1**, se vio en las estadísticas descriptivas y en el gráfico de cajas que los trabajadores 1 y 3 parecen generar el mismo rendimiento del almidón agrio de yuca, y estos dos a su vez difieren del rendimiento del trabajador 2. Esto se verificará con los siguientes contrastes.

$$H_0 : \gamma_{1(1)} - \gamma_{3(1)} = 0$$

$$\rightarrow F = \frac{(0.66)^2}{4.9 \cdot \frac{2}{9}} = 0.4$$

Como $F_{0.95(1,51)} = 4.030393 > F_{cal} = 0.4$, no se rechaza H_0 y concluimos que no hay diferencias significativas entre los efectos del trabajador 1 y el trabajador 3 de la rallandería 1 sobre el rendimiento del almidón agrio de yuca.

$$H_0 : \gamma_{1(1)} + \gamma_{3(1)} - 2\gamma_{2(1)} = 0$$

$$\rightarrow F = \frac{(-15.9)^2}{4.9 \cdot \frac{2}{3}} = 77.39$$

Como $F_{0.95(1,51)} = 4.030393 < F_{cal} = 77.39$, se rechaza H_0 y concluimos que hay diferencias significativas entre los efectos del trabajador 1 y 3 y el trabajador 2 de la rallandería 1 sobre el rendimiento del almidón agrio de yuca.

Con estos resultados se puede concluir que el trabajador 2 genera mayor rendimiento del almidón de yuca en la rallandería 1 sin importar el método de secado.

Con respecto a la **rallandería 2**, en las estadísticas descriptivas y en el gráfico de cajas se observó lo mismo que en la rallandería 1, que los trabajadores 1 y 3 parecen generar el mismo rendimiento del almidón agrio de yuca, y estos dos a su vez difieren del rendimiento del trabajador 2. Esto se verificará con los siguientes contrastes.

$$H_0 : \gamma_{1(2)} - \gamma_{3(2)} = 0$$

$$\rightarrow F = \frac{(0.94)^2}{4.9 \cdot \frac{2}{9}} = 0.8114$$

Como $F_{0.95(1,51)} = 4.030393 > F_{cal} = 0.8114$, no se rechaza H_0 y concluimos que no hay diferencias significativas entre los efectos del trabajador 1 y el trabajador 3 de la rallandería 2 sobre el rendimiento del almidón agrio de yuca.

$$H_0 : \gamma_{1(2)} + \gamma_{3(2)} - 2\gamma_{2(2)} = 0$$

$$\rightarrow F = \frac{(-15.94)^2}{4.9 \cdot \frac{2}{3}} = 77.78$$

Como $F_{0.95(1,51)} = 4.030393 < F_{cal} = 77.78$, se rechaza H_0 y concluimos que hay diferencias significativas entre los efectos del trabajador 1 y 3 y el trabajador 2 de la rallandería 2 sobre el rendimiento del almidón agrio de yuca.

Con estos resultados se puede concluir que el trabajador 2 genera mayor rendimiento del almidón de yuca en la rallandería 2 sin importar el método de secado.

m.

Para la primera hipótesis de literal g:

Las diferencias en valor absoluto entre los promedios son:

$$\tau_1 - \tau_2 = 41.06 - 45.12 = 4.06$$

$$\tau_1 - \tau_3 = 41.06 - 54.73 = 13.67$$

$$\tau_2 - \tau_3 = 45.12 - 54.73 = 9.61$$

Observando las diferencias, nos parece razonable asumir una diferencia mínima de 4 kg para decir que existen diferencias entre los tratamientos, entonces, se tienen dos formulas esenciales para calcular la potencia del experimento:

$$\hat{\lambda} = \frac{\sum_{i=1}^t (\tau_i - \bar{\tau})^2 r}{2\hat{\sigma}^2} \quad , \quad \hat{\Phi} = \sqrt{\frac{2\hat{\lambda}}{t}}$$

$$\rightarrow \hat{\lambda} = \frac{\sum_{i=1}^3 (4)^2 18}{2 \cdot 4.9} = \frac{3 \cdot 16 \cdot 18}{2(4.9)} = 88.1632$$

$$\rightarrow \hat{\Phi} = \sqrt{\frac{2(88.1632)}{3}} = \sqrt{58.7754} = 7.66$$

Utilizando la tabla de la beta no central con $f_1 = 2$ grados de libertad en el numerador y $f_2 = 51$ grados de libertad en el denominador, se obtiene:

$$P(\text{Error II}) = 0 \quad \rightarrow \quad \beta = \text{Potencia} = 1 - P(\text{Error II}) = 1$$

Para la segunda hipótesis de literal g:

Las diferencias en valor absoluto entre los promedios son:

n.