

Análisis del cacao en Colombia para el año 2015 (Abril de 2024)

Aristizabal S. Juan, Cardona T. Santiago y Tobon M. Juan

Resumen

Para la producción agrícola en Colombia, se realizó un análisis estadístico del cultivo de cacao en los departamentos del Huila y Tolima, principales exponentes en la región. Se evaluaron áreas sembradas, áreas cosechadas y rendimientos. En 2015, ambos departamentos presentaron medias de área sembrada similares, pero el Huila mostró un mayor porcentaje de cosecha y rendimiento. Con un nivel de confianza del 94%, se determinó que la media del área cosechada en el Huila es comparable a la del Tolima, indicando un desarrollo similar. Además, con un nivel de significancia del 6%, se concluye que la media del área cosechada en el Huila es como máximo de 165 hectáreas, subrayando la necesidad de una gestión eficiente para un crecimiento sostenible. El análisis de regresión confirma una fuerte relación lineal positiva entre las variables estudiadas, validando la solidez del modelo para entender las dinámicas de producción y rendimiento en la cadena de cacao.

I. INTRODUCCIÓN

En este informe, realizamos un análisis descriptivo comparativo enfocado en la evolución de la cadena de cacao en Colombia, centrándonos en las variables críticas de área sembrada, producción y rendimiento durante los años 2015 y 2020. Nuestro objetivo es identificar tendencias, cambios significativos. Al comparar el área sembrada, observamos que en 2015 se registraron 165.006 hectáreas, mientras que en 2020, la cifra aumentó a 188.370 hectáreas, reflejando un crecimiento sostenido. En términos de producción, el año 2015 cerró con 54.798 toneladas, y en 2020, se alcanzó un nuevo hito con 63.416 toneladas. Este incremento es testimonio del fortalecimiento de la cadena de cacao en el país. Finalmente, el rendimiento del cacao, que es un indicador clave de eficiencia, pasó de 0,43 ton/ha en 2015 a 0,45 ton/ha en 2020, lo que sugiere mejoras en las prácticas de cultivo y posiblemente en la calidad de las variedades de cacao utilizadas. Se realizó un análisis estadístico del cultivo de cacao en Huila y Tolima, evaluando áreas sembradas, cosechadas y rendimientos. Se plantearon dos hipótesis: la media del área cosechada en Huila es igual a la del Tolima, y la media del área cosechada en Huila es como máximo de 165,000 hectáreas y se procedió a analizar la relación lineal entre sus variables y la robustez del cacao para una mayor comprensión de su producción y rendimiento.

II. MÉTODOS

Para realizar el análisis de los datos, se emplearon técnicas descriptivas e inferenciales; tales como el análisis descriptivo de las variables involucradas, el análisis bivariado, el juzgamiento de supuestos e hipótesis, inclusive el ajuste de datos con la media recortada al 10%; además, para ello fue clave utilizar las distintas herramientas que nos proporciona el área de la estadística, funciones que vienen implementadas por el Software R-STUDIO, más específicamente métodos como:

- mean()
- median()
- var()
- sd()
- shapiro.test()
- var.test()
- wilcox.test()
- lm()
- residuals()
- bptest()

Entre otros, tales funciones fueron de uso vital para realizar este análisis, siendo necesario tener conocimientos previos de su uso (**ver Anexo A**). Y es que estas nos permiten tener una visualización de la relación entre las variables, su varianza entre datos, y facilita la comparación entre departamentos y sus valores atípicos. Se analizaron tomando como variables independientes los departamentos en base a quienes se harían sus posteriores análisis, girando con sus variables de carácter continuo (Área Sembrada, Área Cosechada, Rendimiento) en torno a los departamentos en cuestión.

III. RESULTADOS

Análisis de las variables área sembrada y producción del cultivo de cacao en 7 departamentos.

Al analizar las variables área sembrada y producción del cultivo de cacao, se obtiene los siguientes resultados, los cuales fueron plasmados en la siguiente tabla:

Dpto.	ÁREA SEMBRADA			PRODUCCIÓN		
	Media	Varianza	Coef. De Variación	Media	Varianza	Coef. De Variación
ANTIOQUÍA	198.45	47163.28	1.09	147.63	40961.89	1.37
ARAUCA	153.22	132416.79	2.37	169.07	77165.76	1.64

CESAR	335.26	126124.98	1.06	138.58	25268.04	1.15
HUILA	201.19	38528.10	0.98	105.22	11791.03	1.03
NARIÑO	197.82	64097.64	1.28	68.31	7807.50	1.29
SANTAN DER	169.99	39914.87	1.18	132.35	30647.87	1.32
TOLIMA	237.70	51372.01	0.95	126.61	19036.79	1.09

Tabla 1. Análisis del área sembrada y producción de cacao en los departamentos Santander, Antioquia, Arauca, Huila, Cesar, Tolima y Nariño.

En la **tabla 1**, se puede observar que el departamento que en promedio tiene una mayor área de cacao sembrada es el Cesar y el que tiene la menor, es Arauca; sin embargo, el departamento que en promedio tiene una mayor producción de cacao es Arauca y el que tiene la menor, es Nariño. De esta información se puede inferir que no necesariamente si se siembra una mayor área de cacao, esto significa que habrá una mayor producción de éste cultivo.

Además, se realiza un diagrama de dispersión para poder realizar un mejor análisis de los datos:

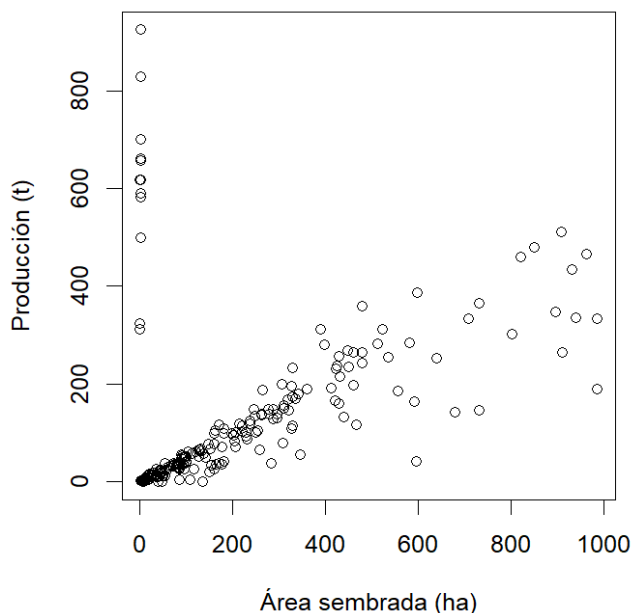


Gráfico 1. Gráfico de dispersión del área sembrada y la producción en los departamentos Santander, Antioquia, Arauca, Huila, Cesar, Tolima y Nariño.

Se observa una relación positiva entre el área sembrada y la producción. Esto indica que a medida que aumenta el área

sembrada, también lo hace la producción, lo cual es esperado en una cadena de producción agrícola. Sin embargo, se identifican datos atípicos en ambas variables, lo cual podría ser causado porque, según la **tabla 1**, las medias del área sembrada y de la producción, en los siete departamentos, son diferentes, esto conlleva que exista una dispersión entre los datos, además de que hay presencia de datos nulos; lo que podría ser objeto de análisis adicional para comprender las razones detrás de estos valores inusuales.

Análisis de la variable rendimiento del cultivo de cacao en los departamentos Huila y Tolima.

Al realizar la revisión de la variable rendimiento del cultivo de cacao para los departamentos Huila y Tolima, tomados de la base de datos de la producción agrícola en Colombia reportada para el año 2015, se obtuvo lo siguiente:

Dpto.	Media	Mediana	Varianza	Coefficiente de variación
HUILA	0.609	0.58	0.0225	0.246
TOLIMA	0.535	0.52	0.0078	0.165

Tabla 2. Rendimiento del cultivo de cacao en los departamentos del Huila y Tolima.

De acuerdo a la **tabla 2**, se puede observar que el rendimiento de cacao en el Huila y Tolima presentan medidas de tendencia central algo similares, también presentan varianzas que si bien no son iguales para ambos departamentos, no son muy diferentes. Esto indica, que el rendimiento del cacao no es igual en ambos departamentos, sin embargo, la diferencia de rendimientos no es demasiada.

Con la misma variable, se realizan los gráficos histogramas del rendimiento del cacao para ambos departamentos, con el propósito de tener una representación gráfica de dichos valores, y de esta forma poder analizarlos a detalle, así:

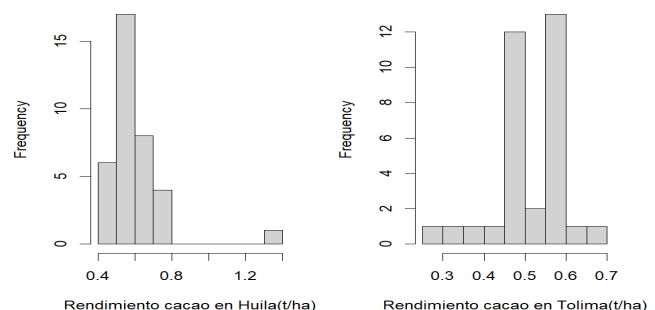


Gráfico 2. Histogramas del rendimiento del cultivo de cacao en los departamentos del Huila y Tolima.

De acuerdo a los histogramas, los departamentos Huila y Tolima presentan una distribución asimétrica y diferente en su rendimiento respecto al cultivo de cacao.

De la misma manera, se realiza un diagrama de caja para el rendimiento del cacao en cada uno de los departamentos anteriormente mencionados, así:

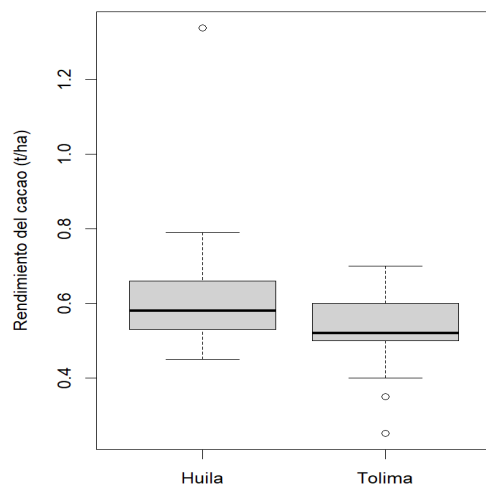


Gráfico 3. Diagramas de caja del rendimiento del cultivo de cacao en los departamentos del Huila y Tolima.

De acuerdo al diagrama de caja gráfico 3, los departamentos Huila y Tolima presentan una variación en su rendimiento respecto al cultivo de cacao, esto apoyado con los datos de varianza y coeficiente de variación obtenidos en la **tabla 2**. Para ambos departamentos, se encontró que hay presencia de datos atípicos.

Análisis de las variables área sembrada y área cosechada.

Al realizar la revisión de las variables área sembrada y área cosechada, tomadas de la base de datos de la producción agrícola en Colombia reportada para el año 2015, se obtuvo lo siguiente:

Covarianza	Coeficiente de Correlación	Coeficiente de variación área Sembrada	Coeficiente de variación área Cosechada
28132.98	0.8440	1.6788	1.7938

Tabla 3. Relación entre el área sembrada y el área cosechada.

Con las mismas dos variables se realizan los gráficos de dispersión y los diagramas de caja para cada variable, con el fin de tener una representación gráfica de dichos datos, y de esta forma poder analizarlos a detalle, así:

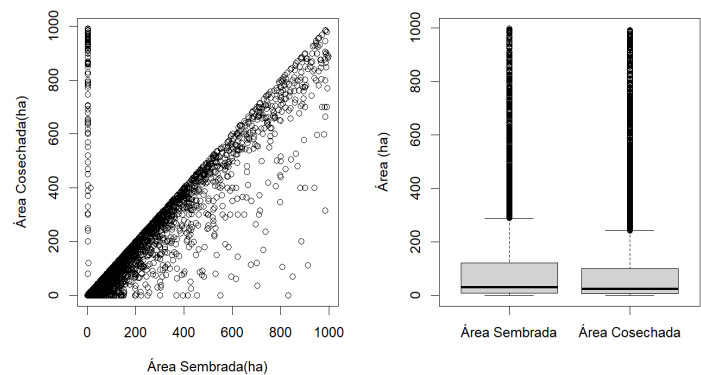


Gráfico 4. Gráfico de dispersión y diagramas de caja para las variables área sembrada y área cosechada.

De acuerdo al gráfico de dispersión, se puede observar que existe una relación positiva entre las variables área sembrada y área cosechada, esto apoyado con el valor dado por el coeficiente de correlación en la **tabla 3**. Además, se puede observar en el diagrama de caja, que entre las variables existe una gran variabilidad; esto puede ser debido a que en ambas variables hay presencia de una cantidad considerable de datos atípicos.

Análisis inferencial del cacao.

Se pretende analizar el área cosechada de cacao en los departamentos del Huila y el Tolima, para lo cual se recurre al juzgamiento de hipótesis que permitan determinar de qué manera influye el cacao en estos lugares; por medio del software estadístico R.

Para este análisis, primero se plantea la hipótesis de que la media del área cosechada de cacao en el Huila es igual a la media del área cosechada de cacao en el Tolima. Se procede a juzgar la hipótesis y verificar si es verdadera o no lo es.

Antes de juzgar dicha hipótesis, se evalúan los supuestos de normalidad y homocedasticidad en las poblaciones; al realizar la prueba “Shapiro-Wilk” se encuentra que ninguna de las dos presenta una distribución normal, sin embargo, al realizar la prueba de homocedasticidad, se encontró que si cumplen con este supuesto, es decir que tienen varianzas iguales. Como las poblaciones no se distribuyen de forma normal, para poder juzgar la hipótesis se recurre al uso de la prueba de “Wilcoxon”, la cual es la apropiada para este caso; al realizar la prueba, se concluye que con un nivel de confianza del 94%, existe evidencia estadística suficiente para afirmar que la hipótesis planteada es cierta.

Adicionalmente se pretende analizar la media del área cosechada de cacao, pero ahora solo del departamento del Huila.

Se plantea la hipótesis de que la media del área cosechada de cacao en el Huila es a lo sumo 165 ha. El supuesto de normalidad ya fue evaluado con anterioridad, y los resultados fueron que no se distribuye de forma normal. Por consiguiente, se procede a realizar la prueba de “Wilcoxon”, por medio de la cual, se concluye que con un nivel de significancia del 6%, existe evidencia estadística significativa para afirmar que la media del área cosechada de cacao es a lo sumo 165 ha. Además, verificando el intervalo de confianza, se concluye que la verdadera media del área cosechada de cacao se encuentra entre 93 y 202 ha.

Finalmente, se pretende analizar las variables área cosechada y producción de cacao en el departamento del Huila, se recurre al uso de la regresión lineal, con el propósito de obtener un modelo con el cual se pueda predecir su comportamiento.

Las variables se presentan en un diagrama de puntos a continuación:

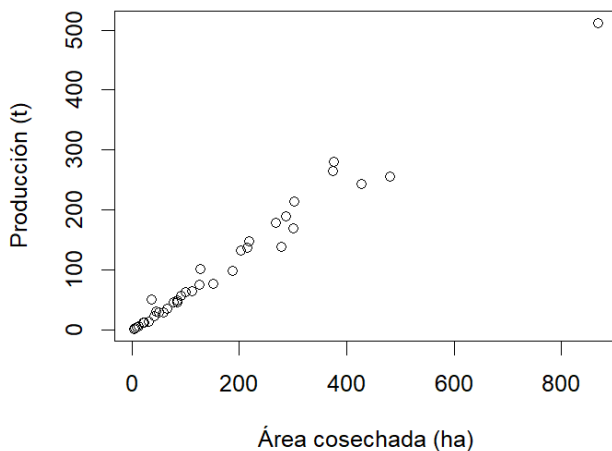


Gráfico 5. Gráfico de dispersión del área cosechada y la producción de cacao en el Huila.

En el **gráfico 5** se puede observar que las variables presentan una relación positiva, lo cual es lo que se espera. Sin embargo, existe presencia de valores atípicos por lo cual se hace necesario hacerle un tratamiento a los datos, en este caso, se le aplicó una media recortada al 10% a los datos, luego se realiza nuevamente el diagrama de dispersión como se muestra a continuación.

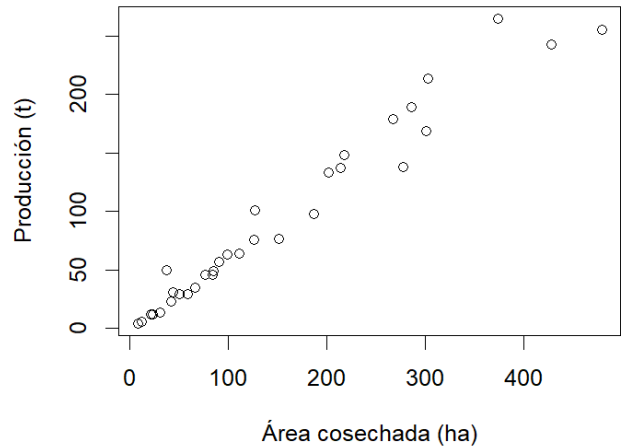


Gráfico 5. Gráfico de dispersión del área cosechada y la producción de cacao en el Huila, con la media recortada aplicada.

De esta forma, ahora se le determina el modelo de regresión apropiado, por medio de la función “lm” del software R. El cual calcula que el modelo de regresión lineal adecuado para este caso tiene la ecuación:

$$y = 4.335 + 0.576x$$

Donde (y) es la variable respuesta, producción, la cual depende de (x), el área cosechada.

Obteniendo una regresión significativa, con un r^2 de 0.96; adicionalmente, se le calculan los residuales de la regresión, y mediante la prueba de “Shapiro-Wilk”, la normalidad de estos; los cuales demostraron no tener una distribución normal y de igual forma, tampoco ser homocedásticos. Por tal motivo, se les realiza una corrección al modelo de regresión, por medio de un factor de raíz cuadrada a los datos de la variable respuesta, la cual es la producción. El modelo de regresión corregido resultó ser:

$$y = 4.276 + 0.029x$$

Igualmente, se determina que el modelo de regresión calculado es significativo, además, cumple con el supuesto de homocedasticidad, y sus residuales cumplen con el supuesto de normalidad. Por lo cual, se concluye que el modelo encontrado es óptimo.

IV. DISCUSIÓN

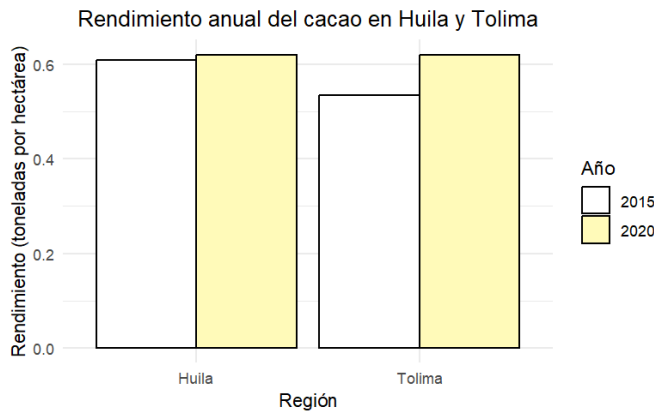


Gráfico 6. Comparativa del rendimiento anual del cacao en el Huila y el Tolima para los años 2015 y 2020.

Al observar las gráficas es apreciable el incremento en cuanto al rendimiento anual que han tenido los departamentos Y permite vislumbrar cuál será su cambio o trayectoria a futuro y es que; El análisis del cacao en Colombia para el año 2015 revela un sector en crecimiento con perspectivas positivas para el futuro. El aumento del área sembrada, la producción y el rendimiento son indicadores del dinamismo del sector y del compromiso del país por fortalecer su posición como uno de los principales productores de cacao a nivel mundial. Las diferencias observadas en el rendimiento entre el Huila y el Tolima sugieren la necesidad de fortalecer las estrategias de transferencia de tecnología y asistencia técnica en el Huila, a fin de que los productores de este departamento puedan aprovechar al máximo el potencial del cultivo de cacao.

- Durante el período analizado (2015-2020), se observa un crecimiento continuo en el área sembrada, la producción y el rendimiento del cultivo de cacao en Colombia. Este incremento refleja una tendencia positiva y sostenida en el desarrollo de la cadena de cacao en el país [1].
- El rendimiento del cacao, definido como la cantidad de toneladas de cacao producidas por hectárea, presentó una tendencia positiva, con un aumento del 1.8% en el Huila y del 15.9% en el Tolima entre 2015 y 2020 [2].
- Se confirma una relación positiva entre el área sembrada y la producción de cacao, respaldada por un alto coeficiente de correlación. Este hallazgo indica que un aumento en el área sembrada se traduce en un incremento correspondiente en la producción, lo cual es coherente con los principios de la producción agrícola.

El cultivo de cacao en Colombia muestra una tendencia positiva y creciente en 2015, con diferencias significativas entre departamentos. El análisis destaca que:

Desempeño General de los Departamentos:

Tolima y Cesar son los departamentos con mayor área sembrada, con medias de 237.70 ha y 335.26 ha respectivamente, destacándose en la extensión de sus cultivos.

Arauca presenta la mayor variabilidad en el área sembrada (varianza de 132416.79), lo que indica fluctuaciones significativas en su cultivo de cacao.

Huila y Tolima son claves en la producción de cacao, con Huila mostrando un rendimiento medio superior (0.609 ton/ha) pero con mayor variabilidad, indicando áreas para mejorar en términos de estabilidad y eficiencia.

Existe una relación positiva entre el área sembrada y la producción en todos los departamentos, con un coeficiente de correlación de 0.8440. Esto confirma que un aumento en el área sembrada generalmente lleva a un incremento en la producción de cacao.

Se desarrolló un modelo de regresión lineal para Huila, ajustado para cumplir con los supuestos de normalidad y homocedasticidad. El modelo corregido tiene la ecuación: $y = 4.276 + 0.029x$, donde y es la producción y x es el área cosechada.

REFERENCES

- [1] CADENA DE CACAO Dirección de Cadenas Agrícolas y Forestales junio 2021. (2021, junio). <https://sioc.minagricultura.gov.co/Cacao/Documentos/2021-06-30%20Cifras%20Sectoriales.pdf>, pp 6.
- [2] Montes, S. (2023, 26 septiembre). Conozca a los nuevos 'cacaos' del país y su impacto en el exterior. Forbes Colombia. <https://forbes.co/2023/07/11/economia-y-finanzas/conozca-a-los-nuevos-cacaos-del-pais-y-su-impacto-en-el-exterior>
- [3] Cely Torres, L. A. (2017). Oferta productiva del cacao colombiano en el posconflicto. Estrategias para el aprovechamiento de oportunidades comerciales en el marco del acuerdo comercial Colombia-Unión Europea. *Equidad y Desarrollo*, 1(28), 167-195.

ANEXOS

ANEXO A: CÓDIGO EN R UTILIZADO PARA EL ANÁLISIS DE DATOS.

```
library(readxl)
library(dplyr)
library(tidyverse)
# LIBRERIA PARA PASAR LAS TABLAS A ARCHIVOS WORD
library(officer) # (SE COMENTA PORQUE GENERA ERROR SI SE ACTIVA A LA
VEZ QUE LA LIBRERIA READXL)

DATOS_PRODUC_AGRIC <-
read_excel("C:/Users/JUAN/JARISTIZABAL/U/TRABAJOSU/ESTADIS
TICA/ DATOS PRODUC AGRIC.XLSX", sheet = "Hoja1")
view(DATOS_PRODUC_AGRIC)

DATA = DATOS_PRODUC_AGRIC
attach(DATA)

# -----
# PUNTO A) REALICE UN RESUMEN CON LAS VARIABLES, DE MANERA QUE
PLANTEE LAS TEMÁTICAS
# DADAS EN ESTADÍSTICA DESCRIPTIVA Y EL ANÁLISIS BIVARIADO (TABLAS,
GRAFICAS MTC, ML,
# MD Y MF) CADA GRUPO TIENE LA POSIBILIDAD DE PLANTEAR EL
ANÁLISIS.

TABTB = DATA %>% group_by(DATA$DEPARTAMENTO) %>%
filter(CULTIVO=='CACAO',DEPARTAMENTO=='SANTANDER'
| DEPARTAMENTO=='ANTIOQUIA' |
DEPARTAMENTO=='ARAUCA' | DEPARTAMENTO=='HUILA' |
DEPARTAMENTO=='CESAR' | DEPARTAMENTO=='TOLIMA' |
DEPARTAMENTO=='NARIÑO') %>%
summarise(MEDIA_A_SEMBRADA = round(mean(ÁREA
SEMBRADA(HA)),na.rm = T),2),
varianza_A_SEMBRADA = round(var(ÁREA
SEMBRADA(HA)),na.rm = T),2),
coeficiente_de_variacion_A_SEMBRADA = round(sd(ÁREA
SEMBRADA(HA))/mean(ÁREA SEMBRADA(HA)),2),
MEDIA_PRODUCCION = round(mean(PRODUCCIÓN(T)),na.rm =
T),2),
varianza_PRODUCCION = round(var(PRODUCCIÓN(T)),na.rm =
T),2),
coeficiente_de_variacion_PRODUCCION =
round(sd(PRODUCCIÓN(T))/mean(PRODUCCIÓN(T)),2))
view(TABTB)

DATAFILTRADA = DATA %>% group_by(DATA$DEPARTAMENTO)
%>%
filter(CULTIVO=='CACAO',DEPARTAMENTO=='SANTANDER'
| DEPARTAMENTO=='ANTIOQUIA' |
DEPARTAMENTO=='ARAUCA' | DEPARTAMENTO=='HUILA' |
DEPARTAMENTO=='CESAR' | DEPARTAMENTO=='TOLIMA' |
DEPARTAMENTO=='NARIÑO')
plot(DATAFILTRADA$ÁREA
SEMBRADA(HA),DATAFILTRADA$PRODUCCIÓN(T),ylab='PRODUCCIÓN
(T)',xlab='ÁREA SEMBRADA (HA)')

# CREA UN DOCUMENTO DE WORD
doc <- read_docx()
```

```

# AGREGA UNA TABLA AL DOCUMENTO
DOC <- DOC %>%
  BODY__ADD__TABLE(TABTB)

# GUARDAR EL DOCUMENTO DE WORD

PRINT(DOC, TARGET = "TABLA_EDITABLE.DOCX")

# -----
# PUNTO B) REALICE UN RESUMEN DESCRIPTIVO COMPARATIVO PARA LA
VARIABLE RENDIMIENTO,
# SEGÚN CORRESPONDA A LOS DEPARTAMENTOS PLANTEADOS DE ACUERDO
AL PRODUCTO DESIGNADO.
# TENGA EN CUENTA LA INFORMACIÓN PRESENTADA EN LA TABLA, EL
ANÁLISIS SOLAMENTE DEBE
# PLANTEARSE DE MANERA DESCRIPTIVA. NO DEBEN JUZGAR
HIPÓTESIS NI ESTIMAR INTERVALOS
# DE CONFIANZA.

TABR = DATA %>% GROUP_BY(DEPARTAMENTO) %>%
  FILTER(CULTIVO=='CACAO',DEPARTAMENTO=='HUILA' |
  DEPARTAMENTO=='TOLIMA') %>%
  SUMMARISE(MEDIA__RENDIMIENTO__CACAO =
  ROUND(MEAN('RENDIMIENTO(T/HA)',NA.RM = T),3),
    MEDIANA__RENDIMIENTO__CACAO =
  MEDIAN('RENDIMIENTO(T/HA)',NA.RM = T),
    VARIANZA__RENDIMIENTO__CACAO =
  ROUND(VAR('RENDIMIENTO(T/HA)',NA.RM = T),5),
    COEFICIENTE__VARIACION__RENDIMIENTO__CACAO =
  SD('RENDIMIENTO(T/HA)')/MEAN('RENDIMIENTO(T/HA)')
  )
VIEW(TABR)

# CREA UN DOCUMENTO DE WORD
DOC <- READ__DOCX()
# AGREGA UNA TABLA AL DOCUMENTO
DOC <- DOC %>%
  BODY__ADD__TABLE(TABR)
# GUARDAR EL DOCUMENTO DE WORD
PRINT(DOC, TARGET = "TABLA_EDITABLE.DOCX")

PAR(MFROW=C(1,2))

DATAFILTRADA1 = DATA %>% GROUP_BY(DEPARTAMENTO) %>%
  FILTER(CULTIVO=='CACAO',DEPARTAMENTO=='HUILA')
DATAFILTRADA2 = DATA %>% GROUP_BY(DEPARTAMENTO) %>%
  FILTER(CULTIVO=='CACAO',DEPARTAMENTO=="TOLIMA")
HIST(DATAFILTRADA1$`RENDIMIENTO(T/HA)`,MAIN="",XLAB='RENDIMIENTO
CACAO EN HUILA(T/HA)')
HIST(DATAFILTRADA2$`RENDIMIENTO(T/HA)`,MAIN="",XLAB='RENDIMIENTO
CACAO EN TOLIMA(T/HA)')

PAR(MFROW=C(1,1))
BOXPLOT(DATAFILTRADA1$`RENDIMIENTO(T/HA)`,DATAFILTRADA2$`RENDIM
IENTO(T/HA)`,NAMES=C('HUILA','TOLIMA'),YLAB='RENDIMIENTO DEL CACAO
(T/HA)')

# -----

```

```
# PUNTO C) EVALUÉ LA EXISTENCIA DE RELACIÓN ENTRE EL ÁREA
SEMBRADA Y COSECHADA,
# USANDO TÉCNICAS DESCRIPTIVAS
```

```
# COVARIANZA
```

```
COV(`ÁREA SEMBRADA(HA)`,`ÁREA COSECHADA(HA)`)
```

```
# COEFICIENTE DE CORRELACION
```

```
COR(`ÁREA SEMBRADA(HA)`,`ÁREA COSECHADA(HA)`)
```

```
TABLA = DATA %>% SUMMARISE(COVARIANZA = COV(`ÁREA
SEMBRADA(HA)`,`ÁREA COSECHADA(HA)`),
  COEFICIENTE_DE_CORRELACION = COR(`ÁREA
SEMBRADA(HA)`,`ÁREA COSECHADA(HA)`),
  COEFICIENTE_DE_VARIACION_A_SEMBRADA =
SD(`ÁREA SEMBRADA(HA)`)/MEAN(`ÁREA SEMBRADA(HA)`),
  COEFICIENTE_DE_VARIACION_A_COSECHADA =
SD(`ÁREA COSECHADA(HA)`)/MEAN(`ÁREA COSECHADA(HA)`))
VIEW(TABLA)
```

```
# CREA UN DOCUMENTO DE WORD
```

```
DOC <- READ_DOCX()
```

```
# AGREGA UNA TABLA AL DOCUMENTO
```

```
DOC <- DOC %>%
```

```
  BODY_ADD_TABLE(TABLA)
```

```
# GUARDAR EL DOCUMENTO DE WORD
```

```
PRINT(DOC, TARGET = "TABLA_EDITABLE.DOCX")
```

```
PAR(MFROW=c(1,2))
```

```
PLOT(`ÁREA SEMBRADA(HA)`,`ÁREA COSECHADA(HA)`)
```

```
BOXPLOT(`ÁREA SEMBRADA(HA)`,`ÁREA COSECHADA(HA)`,
```

```
  NAMES = c("ÁREA SEMBRADA", "ÁREA COSECHADA"),
```

```
  YLAB = "ÁREA (HA)")
```

```
# PUNTO D) REALICE UNA PRUEBA DE HIPÓTESIS QUE LE PERMITA EVALUAR
LAS HIPÓTESIS ENTRE LAS DOS
```

```
# SITUACIONES PLANTEADAS EN EL CUADRO, USE UN NIVEL DE
SIGNIFICANCIA DEL 6%, EVALÚE LOS
```

```
# SUPUESTOS NECESARIOS PARA EL ANÁLISIS. (TENER EN CUENTA LAS
TEMÁTICAS PLANTEADAS EN CLASE
```

```
# Ho: LA MEDIA DEL AREA COSECHADA DE CACAO EN EL HUILA ES IGUAL A
LA MEDIA AREA COSECHADA DE CACAO EN EL TOLIMA
```

```
# Ha: LA MEDIA DEL AREA COSECHADA DE CACAO EN EL HUILA ES DIFERENTE
A LA MEDIA AREA COSECHADA DE CACAO EN EL TOLIMA
```

```
# (NIVEL DE SIGNIFICANCIA DEL 6%) = 0,06
```

```
D_HUILA = DATA %>% FILTER(CULTIVO=="CACAO",
DEPARTAMENTO=="HUILA") %>%
```

```
  SUMMARISE(
    ÁREA_COSECHADA = `ÁREA COSECHADA(HA)`
  )
```

```
D_TOLIMA = DATA %>% FILTER(CULTIVO=="CACAO",
DEPARTAMENTO=="TOLIMA") %>%
```

```
  SUMMARISE(
    ÁREA_COSECHADA = `ÁREA COSECHADA(HA)`
  )
```



```
VIEW(D_HUILA) # N=36
VIEW(D_TOLIMA) # N=33
```

```
N_HUILA = D_HUILA$ÁREA_COSECHADA
N_TOLIMA = D_TOLIMA$ÁREA_COSECHADA
```

```
# VERIFICAR SUPUESTO DE NORMALIDAD
SHAPIRO.TEST(N_HUILA) # No SE DISTRIBUYE NORMAL
SHAPIRO.TEST(N_TOLIMA) # No SE DISTRIBUYE NORMAL
```

```
# VERIFICAR SUPUESTO DE HOMOCEDASTICIDAD
VAR.TEST(N_HUILA,N_TOLIMA,CONF.LEVEL = 0.94)
```

```
# JUZGAR LA HIPÓTESIS E INTERVALO DE CONFIANZA
WILCOX.TEST(N_HUILA,N_TOLIMA, ALTERNATIVE =
"T",PAIRED=F,CONF.LEVEL = 0.94,MU=0,CONF.INT = T,NA.RM=T,EXACT =
F)
```

```
# PUNTO E) REALICE UN ANÁLISIS PARA PROBAR LA HIPÓTESIS PLANTEADA
EN “UNA CONDICIÓN” EN EL
# CUADRO SEGÚN EL GRUPO QUE LE CORRESPONDA, JUZGUE LA HIPÓTESIS
TENGA EN CUENTA LAS
# TEMÁTICAS DADAS EN CLASE.
```

```
# HIPOTESIS: LA MEDIA DEL ÁREA COSECHADA DEL HUILA ES A LO SUMO
165
```

```
# Ho: LA MEDIA DEL ÁREA DE CACAO COSECHADA DEL HUILA ES MENOR O
IGUAL A 165
# Ha: LA MEDIA DEL ÁREA DE CACAO COSECHADA DEL HUILA ES MAYOR A
165
```

```
# YA SE VERIFICO EL SUPUESTO DE NORMALIDAD Y SE ENCONTRO QUE NO
SE DISTRIBUYE NORMAL
```

```
# JUZGAR LA HIPOTESIS
WILCOX.TEST(N_HUILA, MU=165,ALTERNATIVE = "G",CONF.INT = T,EXACT
= F, CONF.LEVEL = 0.94)
```

```
# INTERVALOS DE CONFIANZA
WILCOX.TEST(N_HUILA, MU=165,ALTERNATIVE = "T",CONF.INT = T,EXACT
= F, CONF.LEVEL = 0.94)
```

```
# PUNTO F) SELECCIONE DOS VARIABLES DE ACUERDO AL PRODUCTO
ASIGNADO Y ESTABLEZCA ECUACIÓN DE
# REGRESIÓN ENTRE ÁREA COSECHADA Y PRODUCCIÓN
```

```
DAT$INCEROS <-
READ_EXCEL("C:/USERS/JUAN/JARISTIZABAL/U/TRABAJOSU/ESTADIS
TICA/DATOS PRODUC AGRIC.XLSX",
SHEET = "DAT$INCERO")
VIEW(DAT$INCEROS)
```

```
DAT$INC = DAT$INCEROS
ATTACH(DAT$INC)
```

```
# REGRESION LINEAL
D = LM(DAT$INC$PRODUCCIÓN(T)~DAT$INC$ÁREA COSECHADA(HA))
```

```
SUMMARY(D)
```

```
# ECUACION Y = 4.335 + 0.576x
```

```
# RESIDUALES
RESIDUALS(D)
# REVISAR NORMALIDAD DE LOS RESIDUALES CON SHAPIRO
SHAPIRO.TEST(RESIDUALS(D))
```

```
# HOMOCEDASTICIDAD DE LA REGRESION
#INSTALL.PACKAGES("LMTEST")
LIBRARY("LMTEST")
BPTEST(D)
```

```
# ARREGLO DE LA REGRESION
LIBRARY(MASS)
BC = BOXCOX(D)
```

```
(LAMBDA =BC $ x [WHICH.MAX (BC $ y)])
#TRANSFORMAR LA VARIABLE CON EL LAMBDA ENCONTADO
```

```
# CORRECCION DE LA REGRESION
REGRESION_ARREGLADA =
LM(DAT$INC$PRODUCCIÓN(T)^0.5~DAT$INC$ÁREA
COSECHADA(HA))
SUMMARY(REGRESION_ARREGLADA)
```

```
# NUEVA ECUACION: Y = 4.276 + 0.029x
```

```
SHAPIRO.TEST(RESIDUALS(REGRESION_ARREGLADA))
BPTEST(REGRESION_ARREGLADA)
```

```
PT = DATA %>%
FILTER(CULTIVO=="CACAO",DEPARTAMENTO=="HUILA")
PLOT(PT$ÁREA COSECHADA(HA),PT$PRODUCCIÓN(T), XLAB="ÁREA
COSECHADA (HA)", YLAB="PRODUCCIÓN (T)")
PLOT(DAT$INC$ÁREA COSECHADA(HA),DAT$INC$PRODUCCIÓN(T),
XLAB="ÁREA COSECHADA (HA)", YLAB="PRODUCCIÓN (T)")
```