

# BIENVENIDOS



# “PLAN B” – ANÁLISIS ESTADÍSTICO MULTIVARIANTE



Lic. Juan Pablo Suárez S.  
Dr. Jorge Flores A.

# INTRODUCCIÓN



# INTRODUCCIÓN

- Las técnicas multivariantes son técnicas poco difundidas en nuestro medio que no se encuentran en los programas de licenciatura.
- Las técnicas multivariantes suponen un tremendo poder analítico en manos del investigador, también crean una gran carga para este.

# OBJETIVOS

## OBJETIVO GENERAL

Realizar un análisis estadístico multivariante que determine en que medida la categoría de los equipos, condición de localía y altura inciden en los resultados del fútbol boliviano.

# OBJETIVOS

## OBJETIVOS ESPECÍFICOS

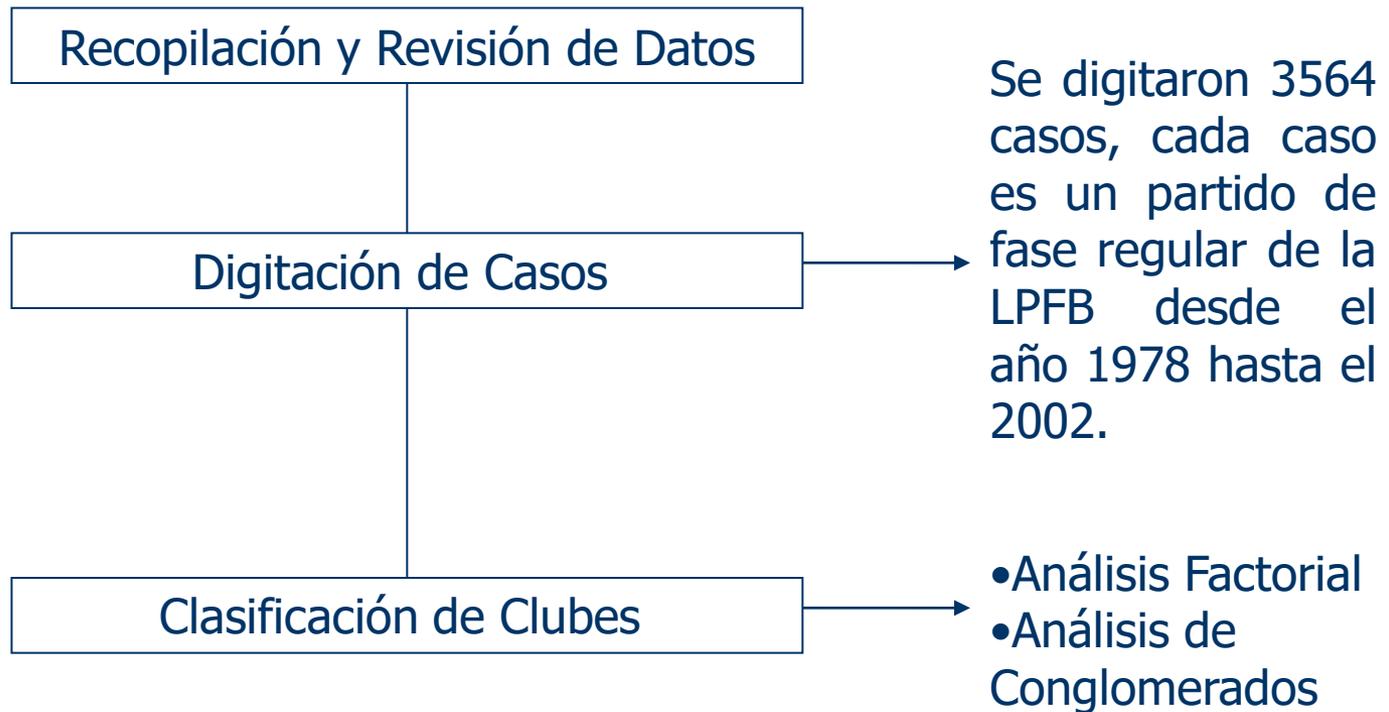
- Determinar las categorías de los equipos del Fútbol Boliviano.
- Determinar en que medida incide la categoría de los equipos en los resultados del Fútbol Boliviano.
- Determinar en que medida incide la condición de localía en los resultados del Fútbol Boliviano.

# OBJETIVOS

## OBJETIVOS ESPECÍFICOS

- Determinar en que medida incide la altura en los resultados del Fútbol Boliviano.
- Identificar las implicaciones económicas a causa de los efectos de la altura.

# PLAN DE TRABAJO



# PLAN DE TRABAJO





# ANÁLISIS FACTORIAL

- El análisis factorial intenta identificar variables subyacentes o factores, que expliquen la configuración de las correlaciones dentro de un conjunto de variables observadas.

# ANÁLISIS FACTORIAL

## VARIABLES A TOMAR EN CUENTA EN EL ANÁLISIS

### Variables Categóricas:

- Condición de localía
- Resultado
- Posición del equipo en la tabla del año anterior

# ANÁLISIS FACTORIAL

## VARIABLES A TOMAR EN CUENTA EN EL ANÁLISIS

### Variables Numéricas:

- Goles a favor del equipo local
- Goles a favor del equipo visitante
- Diferencia de goles en el partido
- Altura de la ciudad donde se juega
- **Posición histórica del equipo local**

# ANÁLISIS FACTORIAL

## DETERMINACION DE LA POSICION HISTORICA DE LOS EQUIPOS

POS GRAL.	EQUIPO	SUMA DE POSICIONES	CAMPEONATOS JUGADOS	POSICION PROMEDIO
1	Bolivar	60	26	2,3
2	The Strongest	79	26	3,0
3	Oriente Petrolero	100	26	3,8
4	Blooming	123	25	4,9
5	Wilstermann	142	26	5,5
6	Real Bamin	40	6	6,7
7	San José	181	25	7,2
8	Guabirá	132	17	7,8
9	Destroyers	102	13	7,8
10	Ind Petrolero/Stormers	132	16	8,3
11	Real Sta Cruz	217	24	9,0
12	Ciclón/Unión Central	134	14	9,6
13	Aurora	146	13	11,2

Nota: El resto de los equipos tuvieron una participación corta y/o irrelevante en los campeonatos.

# ANÁLISIS FACTORIAL

## ANÁLISIS DE COMPONENTES PRINCIPALES

- Extrae factores mediante combinaciones lineales independientes de media cero e incorrelacionadas de las variables utilizadas.
- Este tipo de análisis factorial se recomienda para la reducción de datos y en conjuntos de varianza común elevada.

# ANÁLISIS FACTORIAL

## MATRÍZ DE COMPONENTES

	Componente					
	1	2	3	4	5	6
Posición Histórica Local	-,471	,680	-,069	,104	,012	-,091
Goles Club Local	,748	,143	,312	,042	-,017	,051
Goles Club Visita	-,536	-,183	,584	,036	-,095	,065
DIF Goles	,919	,220	-,058	,016	,038	,007
Gana Local	,866	,249	,160	,111	-,025	-,179
Empate	-,421	-,146	-,728	-,202	,097	,305
Pierde Local	-,660	-,162	,610	,082	-,076	-,110
Localía o no	,014	,408	,203	,127	,111	,505
1° cuarto Tabla Local	,420	-,606	,174	-,488	,192	,131
2° cuarto Tabla Local	,063	-,130	-,193	,518	-,797	,167
3° cuarto Tabla Local	-,163	,036	-,057	,646	,704	-,018
4° cuarto Tabla Local	-,315	,663	,041	-,545	-,097	-,268
Altura	-,009	,209	,258	-,110	,055	,714
Año 1	-,013	-,219	,033	,170	,129	-,277

# ANÁLISIS FACTORIAL

## COMUNALIDADES

	Inicial	Extracción
Posición Histórica Local	1,000	,708
Goles Club Local	1,000	,682
Goles Club Visita	1,000	,675
DIF Goles	1,000	,898
Gana Local	1,000	,882
Empate	1,000	,872
Pierde Local	1,000	,859
Localía o no	1,000	,591
1° cuarto Tabla Local	1,000	,866
2° cuarto Tabla Local	1,000	,990
3° cuarto Tabla Local	1,000	,943
4° cuarto Tabla Local	1,000	,919
Altura	1,000	,635
Año	1,000	,172

Se retira la variable  
“año” (comunalidad <6)  
para el siguiente  
Análisis Factorial  
Confirmatorio

Método de extracción: Análisis de Componentes principales.

# ANÁLISIS FACTORIAL

## VARIANZA TOTAL EXPLICADA

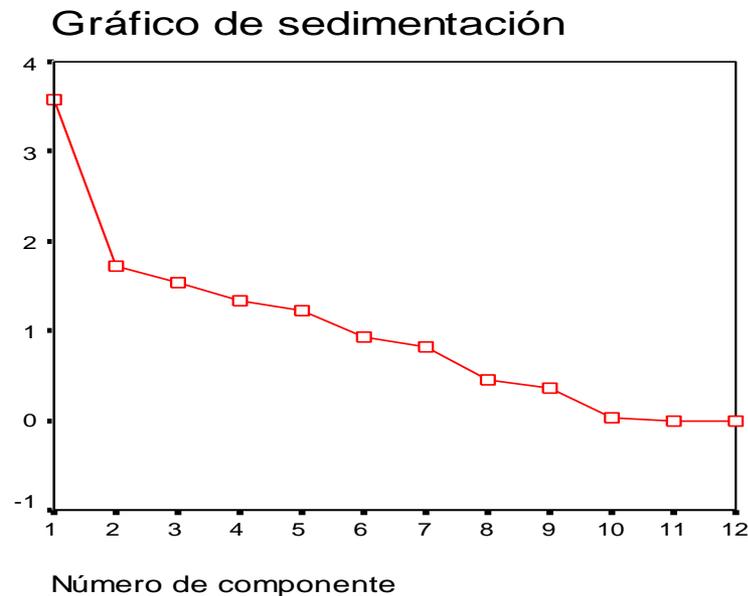
Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	3,582	29,849	29,849	3,582	29,849	29,849
2	1,722	14,348	44,197	1,722	14,348	44,197
3	1,530	12,751	56,949	1,530	12,751	56,949
4	1,341	11,171	68,120	1,341	11,171	68,120
5	1,231	10,258	78,378	1,231	10,258	78,378
6	,926	7,719	86,097			
7	,824	6,863	92,959			
8	,452	3,763	96,722			
9	,355	2,962	99,683			
10	,038	,317	100,000			
11	6,320E-16	5,267E-15	100,000			
12	-1,407E-15	-1,172E-14	100,000			

Método de extracción: Análisis de Componentes principales.

Se extraen los factores con sumas de saturaciones mayores a 1.

# ANÁLISIS FACTORIAL

## GRÁFICO DE SEDIMENTACIÓN



Hasta el componente “5” se aprecian tres “quiebres”:  
del “1” al “2”, del “2” al “4” y del “4” al “5”.

# ANÁLISIS DE CONGLOMERADOS

- El propósito del mismo es el de agrupar las observaciones de forma que los datos sean homogéneos dentro de los grupos (mínima varianza) y que los grupos sean lo mas heterogéneos entre ellos (máxima varianza)..

# ANÁLISIS CONGLOMERADOS

## ANÁLISIS DE CONGLOMERADOS DE K - MEDIAS

- Tiene un mejor acoplamiento al manejo de distintos tipos de variables Y gran número de casos.

# ANÁLISIS CONGLOMERADOS

## ANÁLISIS DE CONGLOMERADOS DE K - MEDIAS

### Criterios

- Asignando aleatoriamente los objetos a los grupos y tomando los centros de los grupos así formados.
- Tomando como centros los "G" puntos mas alejados entre si.
- **Seleccionando los centros a "priori" con criterios empíricos.**

# ANÁLISIS DE CONGLOMERADOS

ANÁLISIS DE CONGLOMERADOS DE K - MEDIAS

## CATEGORÍA 1

**Oriente Petrolero**

**Bolívar**

**Blooming**

**The Strongest**

# ANÁLISIS DE COMGLOMERADOS

## ANÁLISIS DE CONGLOMERADOS DE K - MEDIAS

### CATEGORÍA 2

**Wilstermann**  
**Real Potosí**  
**Guabirá**  
**Stormers**  
**Ind. Petrolero**

**Destroyers**  
**Unión Central**  
**Litoral**  
**Aurora**

# ANÁLISIS DE CONGLOMERADOS

## ANÁLISIS DE CONGLOMERADOS DE K - MEDIAS

### CATEGORÍA 3

**Ind. Unificada**  
**Orcobol**  
**Petrolero**  
**San Pedro**  
**Mcal. Braun**

**U. del Beni**  
**Pompeya**  
**Wilster. Coop.**  
**Real Santa Cruz**  
**Ind. Unificada**

# ANÁLISIS DE VARIANZA



# SUPUESTOS DEL ANÁLISIS

## INDEPENDENCIA

Los errores experimentales deben ser aleatorios, se consigue esta condición si los elementos de los diversos grupos han sido escogidos aleatoriamente o se utilizan muestras grandes.

# SUPUESTOS DEL ANÁLISIS

## HOMOGENEIDAD DE VARIANZAS

La varianza de los subgrupos debe ser homogénea ya que esta se debe al error. Se puede comprobar el supuesto mediante la prueba de Levene, Q de Crochan o Barlett-Box.

La heterocedasticidad influye en los contrastes si existe una diferencia de dos a uno en los tamaños muestrales de cada grupo.

# SUPUESTOS DEL ANÁLISIS

## PRUEBA DE LEVENE -HOMOGENEIDAD DE VARIANZAS

	Estadístico de Levene	gl1	gl2	Sig.
<b>Conglomerado</b>	,108	1	3618	<b>,743</b>
Localía	14,606	1	3618	,000
<b>Altura</b>	1,882	1	3618	<b>,170</b>

- La variable conglomerado y altura presentan heterocedasticidad.
- El análisis es robusto a la heterocedasticidad si los tamaños muestrales no son distintos en una proporción de “2” a “1”.
- Conglomerado: mayor – 1266 casos; menor 1140 casos.
- Altura: mayor – 1377 casos; menor 845 casos.

# SUPUESTOS DEL ANÁLISIS

## PRUEBA DE KOLMOROV SMIRNOV - NORMALIDAD

- La prueba de normalidad de Kolmorov-Smirnov indica si se cumple el supuesto de distribución normal de la variable dependiente, es decir que los errores experimentales se distribuyen normalmente.

		Conglomerado	Localía	Altura
N		3620	3620	3620
Parámetros normales(a,b)	Media	2,01	,85	2,00
	Desviación típica	,804	,361	,874
Diferencias más extremas	Absoluta	,220	,511	,257
	Positiva	,213	,335	,254
	Negativa	<b>-,220</b>	<b>-,511</b>	<b>-,257</b>
Z de Kolmogorov-Smirnov		13,255	30,758	15,434
Sig. asintót. (bilateral)		<b>,000</b>	<b>,000</b>	<b>,000</b>

a La distribución de contraste es la Normal.

b Se han calculado a partir de los datos.

# MODELO LINEAL GENERAL

VARIACION	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	MEDIA	F
Debido a la Regresión	$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	k-1	$S_E^2 = \frac{1}{k-1} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	$F = \frac{S_E^2}{S_R^2}$
Debido a los residuos	$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	n-k	$S_R^2 = \frac{1}{n-k} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	
Total	$\sum_{i=1}^n (Y_i - \bar{Y})^2$	n-1		

# MODELO LINEAL GENERAL

Para el análisis se decidió utilizar como variable dependiente la diferencia de goles ( $GFL - GCL$ ) ya que en el modelo preliminar da el mayor coeficiente de correlación (0.88).

Con los goles a favor del equipo local, el modelo tiene un coeficiente de correlación de 0.80 y con goles en contra 0.23

# MODELO LINEAL GENERAL

## CONTRASTES MULTIVARIADOS

El estadístico Lambda de Wilks en el caso de dos factores permite contrastar tres hipótesis relativas al factor A, al factor B y a la interacción entre ambos (cuyas matrices son  $\Sigma_A$  y  $\Sigma_{AB}$  respectivamente). El estadístico tiene la siguiente expresión.

$$\Lambda = \frac{|SC_{entre}|}{|F_H + SC_{intra}|}$$

FH = Suma de cuadrados y productos cruzados residual

# MODELO LINEAL GENERAL

## CONSTRASTES MULTIVARIADOS

Efecto		Valor	F	Gl de la hipótesis	Gl del error	Significación
Intercept	Traza de Pillai	,412	1252,120(a)	2,000	3574,000	,000
	Lambda de Wilks	,588	1252,120(a)	2,000	3574,000	,000
CONGLO	Traza de Pillai	,032	29,066	4,000	7150,000	,000
	Lambda de Wilks	,968	29,295(a)	4,000	7148,000	,000
LOCALÍA	Traza de Pillai	,007	11,892(a)	2,000	3574,000	,000
	Lambda de Wilks	,993	11,892(a)	2,000	3574,000	,000
ALTURA	Traza de Pillai	,005	4,302	4,000	7150,000	,002
	Lambda de Wilks	,995	4,304(a)	4,000	7148,000	,002
CONGLO * LOCALÍA	Traza de Pillai	,001	,552	4,000	7150,000	,698
	Lambda de Wilks	,999	,552(a)	4,000	7148,000	,698
CONGLO * ALTURA	Traza de Pillai	,002	1,001	6,000	7150,000	,422
	Lambda de Wilks	,998	1,001(a)	6,000	7148,000	,422
LOCALÍA * ALTURA	Traza de Pillai	,005	4,487	4,000	7150,000	,001
	Lambda de Wilks	,995	4,492(a)	4,000	7148,000	,001
CONGLO * LOCALÍA * ALTURA	Traza de Pillai	,002	1,280	6,000	7150,000	,263
	Lambda de Wilks	,998	1,280(a)	6,000	7148,000	,262

a Estadístico exacto

b El estadístico es un límite superior para la F el cual ofrece un límite inferior para el nivel de significación.

c Diseño: Intercept+CONGLO+LOCALÍA+ALTURA+CONGLO \* LOCALÍA+CONGLO \* ALTURA+LOCALÍA \* ALTURA+CONGLO \* LOCALÍA \* ALTURA

# MODELO LINEAL GENERAL

## MODELO FINAL DE ANOVA

Variable dependiente: DIF de Goles

Fuente	Suma de cuadrados tipo III	gl	Media cuadrática	F	Significación	Eta al cuadrado parcial
Modelo corregido	1158,578(b)	7	165,511	46,753	,000	,88
Intersección	1041,600	1	1041,600	294,231	,000	,85
CONGLO	905,949	2	452,974	127,956	,000	,70
LOCALÍA	45,869	1	45,869	12,957	,000	,05
NALTURA	27,934	2	13,967	3,945	,007	,03
LOCALÍA * NALTURA	35,096	2	17,548	4,957	,008	,03
Error	12786,775	36	3,540			
Total	18297,000	12				
Total corregida	13945,353	36				
		20				
		36				
		19				

a Calculado con alfa = ,05

b R cuadrado = ,88 (R cuadrado corregida = ,86)

# MODELO LINEAL GENERAL

## ESTADÍSTICO ETA

FACTOR	ETA	PORCENTAJE
Conglomerado	0,70	86%
Localía	0,05	6%
Altura	0,03	4%
Altura*Localía	0,03	4%
<b>TOTAL</b>	<b>0,81</b>	<b>100%</b>

# MODELO LINEAL GENERAL

## PORCENTAJE DE EFECTO DEJANDO DE LADO CATEGORÍA DE EQUIPOS

FACTOR	ETA	PORCENTAJE
Localía	0,05	46%
Altura	0,03	27%
Altura*Localía	0,03	27%
<b>Total Altura+Interacción</b>	<b>0,08</b>	<b>54%</b>
<b>Total</b>	<b>0,11</b>	<b>100%</b>

# MODELO LINEAL GENERAL

## VENTAJA LOCAL SIN ALTURA vs. LOCAL CON ALTURA

- La ventaja obtenida para un equipo local sin el efecto de la altura es de un 46%.
- La ventaja obtenida para un equipo local de la altura "3" (altiplánica) es de un 70.8%.  $-((0.46*0.54)+0.46)$

# MODELO LINEAL GENERAL

## MODELO FINAL

$$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_2 X_3 + e$$

En donde:

$\alpha$ : Constante del modelo

$\beta_1 X_1$  : Efecto de la categoría de los equipos

$\beta_2 X_2$  : Efecto de la Condición de localía

$\beta_3 X_3$  : Efecto de la altura

$\beta_4 X_2 X_3$  : Efecto de la interacción localía-altura

$e$ : Error en el modelo

# MODELO LINEAL GENERAL

## ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

Variable dependiente: DIF de Goles

Parámetro	B	Error típ.	t	Significación	Intervalo de confianza al 95%.		Eta al cuadrado parcial
					Límite inferior	Límite superior	
Intersección	,823	,066	12,510	,000	,694	,953	,42
[CONGLO=1]	1,267	,079	16,036	,000	1,112	1,421	,67
[CONGLO=2]	,277	,076	3,640	,000	,128	,426	,04
[CONGLO=3]	0(a)	.	.	.	.	.	.
[LOCALÍA=0]	-,580	,147	-3,948	,000	-,868	-,292	,04
[LOCALÍA=1]	0(a)	.	.	.	.	.	.
[ALTURA=1]	-,095	,077	-1,221	,222	-,246	,057	,00
[ALTURA=2]	-,173	,091	-1,912	,056	-,351	,004	,01
[ALTURA=3]	0(a)	.	.	.	.	.	.
[LOCALÍA=0] *							
[ALTURA=1]	-,272	,185	-1,466	,143	-,635	,092	,01
[LOCALÍA=0] *							
[ALTURA=2]	,637	,315	2,020	,043	,019	1,254	,01
[LOCALÍA=0] *							
[ALTURA=3]	0(a)	.	.	.	.	.	.
[LOCALÍA=1] *							
[ALTURA=1]	0(a)	.	.	.	.	.	.
[LOCALÍA=1] *							
[ALTURA=2]	0(a)	.	.	.	.	.	.
[LOCALÍA=1] *							
[ALTURA=3]	0(a)	.	.	.	.	.	.

a Al parámetro se le ha asignado el valor cero porque es redundante

# MODELO LINEAL GENERAL

## COMPARACIONES DE MEDIAS MARGINALES (ALTURA)

Variable dependiente: DIF de Goles

(I) ALTURA	(J) ALTURA	Diferencia entre medias (I-J)	Error típ.	Significación(a)	Intervalo de confianza al 95 % para diferencia(a)	
					Límite inferior	Límite superior
1	2	-,376(*)	,154	,015	-,678	-,073
	3	-,230(*)	,093	,013	-,412	-,048
2	1	,376(*)	,154	,015	,073	,678
	3	,145	,162	,372	-,173	,464
3	1	,230(*)	,093	,013	,048	,412
	2	-,145	,162	,372	-,464	,173

Basadas en las medias marginales estimadas.

\* La diferencia de las medias es significativa al nivel ,05.

a Ajuste para comparaciones múltiples: Diferencia menos significativa (equivalente a la ausencia de ajuste).

# MODELO LINEAL GENERAL

## PROBABILIDADES DE RESULTADO

### LOCAL LLANO

Triunfo	64%
Empate	21%
Derrota	15%

46% DE VENTAJA

### LOCAL ALTURA "3" VS EQUIPO DEL LLANO

Triunfo	80%
Empate	16%
Derrota	4%

71% DE VENTAJA

# CONCLUSIONES



# CONCLUSIONES

- La cantidad de conglomerados o categorías de equipos (3) está muy bien definida y presenta un gran porcentaje de la distribución de la varianza
- La diferencia de goles al ser la combinación de los goles del equipo local y los goles del equipo visitante presenta un mayor ajuste al modelo
- las tres variables introducidas de antemano en el modelo son significantes.

# CONCLUSIONES

- La interacción de localía y categoría de los equipos, categoría y altura y la categoría-altura y localía no son significantes en el modelo, esto indica que los efectos de la altura son independientes a la categoría del equipo local.
- La variable de categoría de los equipos o conglomerado tiene un efecto del 86% en el resultado, la localía del 6%, la altura mas su interacción con la localía un 8%.

# CONCLUSIONES

- Haciendo un análisis sobre estos porcentajes se llega a la conclusión de que un equipo del llano y de categoría “1” puede compensar la desventaja cuando sube a la altura si es que juega contra un equipo de la categoría “2” o “3” pero si juega contra un equipo de la misma categoría es muy probable que no consiga buenos resultados
- Dejando de lado la ventaja deportiva de la variable categoría del equipo, se deduce que un equipo del llano jugando de local contra uno de la misma categoría tiene un 46% de ventaja.
- Un equipo de la altura jugando de local contra uno de la misma categoría del llano, tiene un 71% de ventaja.  $(0.54 + (0.54 * 0.46))$  ETA

# CONCLUSIONES

- La altura de los valles no presenta un efecto significativo al 95% de confianza en los efectos del fútbol Boliviano.

LOCAL LLANO

Triunfo	64%
Empate	21%
Derrota	15%

LOCAL ALTURA "3" VS EQUIPO DEL LLANO

Triunfo	80%
Empate	16%
Derrota	4%

# PROPUESTA DE COMPENSACIÓN – PLAN “B”



# PLAN "B"

## SUPUESTOS

- **Todos los jugadores (titulares y suplentes) tienen la misma capacidad técnica.**
- **Todos los cambios de jugadores (equipo local y visitante) se realizan al minuto 60 del partido.**

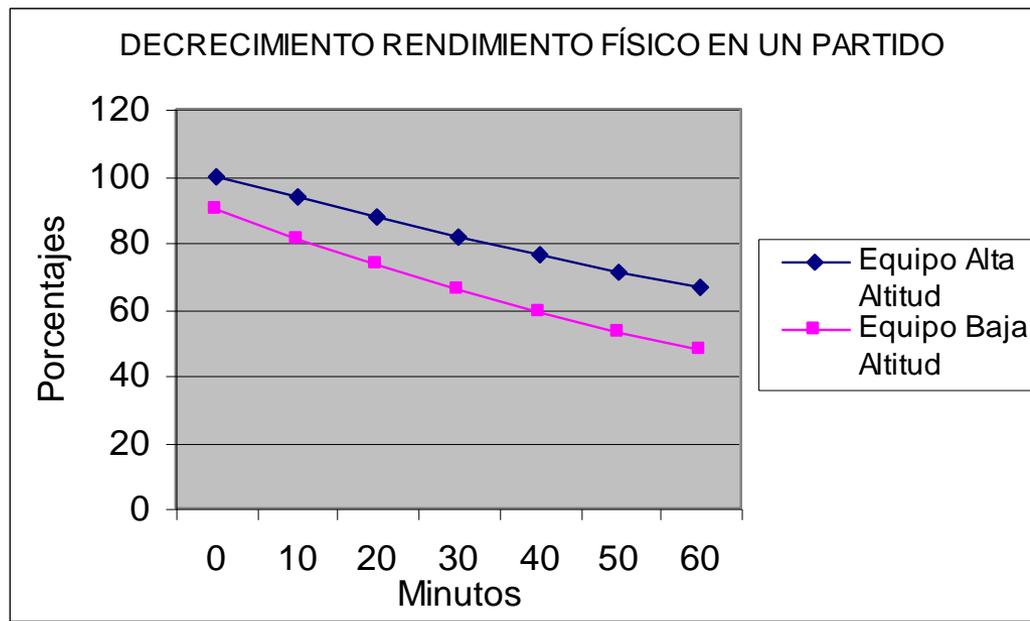
# PLAN "B"

## RENDIMIENTO FÍSICO POR EQUIPO

MINUTO	Decreciente 10% - EMPEZANDO CON 10% DESVENTAJA		VENTAJA
	Equipo Alta Altitud	Equipo Baja Altitud	
0	100	90	11,11%
10	93,40	81,00	15,31%
20	87,24	72,90	19,66%
30	81,48	65,61	24,19%
40	76,10	59,05	28,88%
50	71,08	53,14	33,75%
60	66,39	47,83	38,80%
<b>PROMEDIO</b>	<b>6,64</b>	<b>4,78</b>	24,53%
	POR JUGADOR	POR JUGADOR	POR RANGO DE TIEMPO

# PLAN "B"

## RENDIMIENTO FÍSICO POR EQUIPO



Fuente: Elaboración propia

# PLAN "B"

## RENDIMIENTO PROMEDIO POR JUGADOR

# de jugadores de campo	Equipo de Alta Altitud (Local)	Equipo de Baja Altitud (Visitante)	Equipo de Baja Altitud (con 4 cambios extras)
1	6,62	4,78	9,00
2	6,62	4,78	9,00
3	6,62	4,78	9,00
4	6,62	4,78	9,00
5	6,62	4,78	4,78
6	6,62	4,78	4,78
7	6,62	4,78	4,78
8	6,62	4,78	4,78
9	6,62	4,78	4,78
10	6,62	4,78	4,78
<b>Promedio Equipo</b>	<b>6,62</b>	<b>4,78</b>	<b>6,47</b>

Elaboración propia

# PLAN "B"

- **Efectuar 7 cambios de jugadores por partido solo para el equipo visitante, cambios que serán efectuados en tres entradas, los cambios se efectuarán de uno, dos, tres cuatro, cinco, seis o los siete jugadores por vez, pudiendo un jugador volver a entrar a jugar cuando le pase la fatiga o sofoco contando este retorno como un cambio, el equipo local adaptado solo tendrá derecho a tres cambios.**

# PLAN "B"

- **Efectuando 7 cambios el equipo no adaptado tendrá renovada el 50% de su Capacidad Anaeróbica( CAE), Capacidad Aeróbica(CA), Desequilibrio Acido Básico, Fuerza, Potencia y Resistencia, 50% de disminución del cansancio y ahogo o fatiga y por lo tanto menor riesgo potencial teórico para la salud del atleta y entre un 22% y 28% de VENTAJA COMPENSATORIA a la diferencia de altitud.**

# GRACIAS POR SU ATENCIÓN



# FORMULARIOS

## COVARIANZA

$$C_{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{X}\bar{Y}$$

## COEFICIENTE DE CORRELACIÓN

$$R = \frac{S_{xy}}{\sqrt{S_x S_y}}$$

## SUMA DE CUADRADOS DENTRO DE LOS GRUPOS

$$SCDG = \sum_{g=1}^G \sum_{j=1}^p \sum_{i=1}^{n_g} (x_{ijg} - \bar{x}_{jg})^2$$

$$\min SDCG = \min \sum_{g=1}^G \sum_{j=1}^p n_g s_{jg}^2$$

El criterio minimiza la varianza de los casos intra-grupos.

# FORMULARIOS

## DISTANCIA EUCLÍDEA

$$d_{ij} = \sqrt{\sum_{k=1}^t (X_{ik} - X_{jk})^2}$$

(Viene determinada por la escala de medidas de variables: binarias, ordinal, etc.)

## MODELO LINEAL GENERAL MULTIVARIANTE

VARIACION	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	MEDIA	F
Debido a la Regresión	$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	k-1	$S_E^2 = \frac{1}{k-1} \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	$F = \frac{S_E^2}{S_R^2}$
Debido a los residuos	$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	n-k	$S_R^2 = \frac{1}{n-k} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	
Total	$\sum_{i=1}^n (Y_i - \bar{Y})^2$	n-1		

# FORMULARIOS

## LAMBDA DE WILKS

$$\Lambda = \frac{|SC_{entre}|}{|F_H + SC_{intra}|}$$

Razón entre el determinante de la matriz de variancias y covariancias dentro de grupos y el determinante de la matriz de variancias y covariancias total

## PRUEBA DE LEVENE

$$z_{ij} = \sum_{j=1}^i |y_{ij} - \bar{y}_i|$$

Y<sub>ij</sub>: Cada valor esperado de “y”  
Y: Promedio de “y”

## KOLMOROV - SMIRNOV

$$E_N = n(i)/N$$

$$D = \max_{1 \leq i \leq N} |F(Y_i) - \frac{i}{N}|$$

FUNCIÓN EMPÍRICA DE DISTRIBUCIÓN: Donde: n(i) es el n° de puntos menores que Y<sub>i</sub> y Y<sub>i</sub> está ordenado desde el mas pequeño al mas grande.

# FORMULARIOS

TABLA DE ANÁLISIS DE VARIANZA CON Q FACTORES

FUENTE DE VARIACIÓN	SUMA DE CUADRADOS	GRADOS DE LIBERTAD	MEDIA DE CUADRADOS	ESTADÍSTICO F
EFFECTOS PRINCIPALES	$SC_p$	$gl_p$	$MC_p = \frac{SC_p}{gl_p}$	$F_p = \frac{MC_p}{MC_r}$
INTERACCIÓN DE ORDEN 2	$SC_{i2}$	$gl_{i2}$	$MC_{i2} = \frac{SC_{i2}}{gl_{i2}}$	$F_p = \frac{MC_{i2}}{MC_r}$
INTERACCIÓN DE ORDEN Q	$SC_{iq}$	$gl_{iq}$	$MC_{iq} = \frac{SC_{iq}}{gl_{iq}}$	$F_p = \frac{MC_{iq}}{MC_r}$
EXPLICADA	$SC_e$	$gl_e$	$MC_e = \frac{SC_e}{gl_e}$	$F_e = \frac{MC_e}{MC_r}$
RESIDUAL	$SC_r$	$gl_r$	$MC_r = \frac{SC_r}{gl_r}$	
TOTAL	$SC_t$	$gl_t$	$MC_t = \frac{SC_t}{gl_t}$	

# FORMULARIOS

## MODELO ANÁLISIS FACTORIAL

- $X_{ij} = a_{i1} \cdot F_{1j} + a_{i2} \cdot F_{2j} + a_{i3} \cdot F_{3j} + \dots + d_i \cdot U_{ij}$
- En donde:
- **$X_{ij}$**  = Valor normalizado de la variable “i” para el sujeto “j”
- 
- **$F_{1j}$**  = Valor del Factor 1 para el sujeto “j”
- 
- **$a_{i1}$**  = Relación entre variable “i” y factor 1
- 
- **$F_{2j}$**  = Valor del Factor 2 para el sujeto “j”
- 
- **$a_{i2}$**  = Relación entre variable “i” y factor 2
- ...
- ...
- **$d_i \cdot U_{ij}$**  = Parte aleatoria independiente de los factores[1]
-

# FORMULARIOS

## MODELO ANÁLISIS FACTORIAL

- $X_{ij} = a_{i1} \cdot F_{1j} + a_{i2} \cdot F_{2j} + a_{i3} \cdot F_{3j} + \dots + d_i \cdot U_{ij}$
- En donde:
- **$X_{ij}$**  = Valor normalizado de la variable “i” para el sujeto “j”
- 
- **$F_{1j}$**  = Valor del Factor 1 para el sujeto “j”
- 
- **$a_{i1}$**  = Relación entre variable “i” y factor 1
- 
- **$F_{2j}$**  = Valor del Factor 2 para el sujeto “j”
- 
- **$a_{i2}$**  = Relación entre variable “i” y factor 2
- ...
- ...
- **$d_i \cdot U_{ij}$**  = Parte aleatoria independiente de los factores[1]
-

# FORMULARIOS

## TEOREMA RESULTADOS MUTUAMENTE EXCLUYENTES

- En el caso particular de sucesos independientes se verifica:  
$$P(A \cap B) = P(A) \cdot P(B)$$