

**“ANÁLISIS ESTADÍSTICO MULTIVARIANTE
APLICADO AL FÚTBOL BOLIVIANO”**

B1BO

1. Introducción	2
2. Planteamiento del problema	7
3. Objetivos	8
3.1. Objetivo General	8
3.2. Objetivos Específicos	8
4. Delimitación	8
4.1. Limite Espacial	8
4.2. Limite Sustantivo	9
1. Categoría de los Equipos	10
1.2. Análisis Factorial	11
1.2.1. Variables a Tomar en Cuenta Para el Análisis Factorial.....	11
1.2.1.1. Variables Numéricas	12
1.2.1.2. Variables Categóricas.....	13
1.3. Análisis de Componentes Principales	14
1.3.1. La matriz de Cargas, factorial o de Componentes	15
1.3.1.1. Comunalidades.....	16
1.4. Análisis Factorial Confirmatorio	18
1.4.1. Gráfico de Sedimentación	20
1.5. Análisis de conglomerados	21
1.5.1. Análisis de Conglomerados de k-medias en dos fases.....	21
2. Análisis de la Varianza	24
2.1. Bases del Análisis de Varianza	25
2.2. Modelo Lineal General Multivariante (MANOVA).....	25
2.3. Modelo Final del Análisis de la varianza (ANOVA).....	30
2.3.1. Estimación de los parámetros	35
2.3.2. Comparaciones de Medias Marginales	37
3. Implicaciones Económicas por el Efecto de la altura	38
4. Conclusiones	43

INTRODUCCION

1. Introducción

El campeonato de la Liga Profesional de Fútbol Boliviano es uno de los mas atípicos gracias a las caprichosa geografía de Bolivia, este presenta equipos en sus tres principales zonas geográficas, los llanos, los valles y el altiplano. Gracias a la diferencia de altura de estas regiones se han generado discusiones en torno a la ventaja *física y/o sicológica* que esta genera para los equipos locales.

Los dirigentes de los equipos del llano arguyen que la altura es una gran ventaja para los equipos del altiplano e inclusive para los equipos de los valles y que muchos de los resultados desfavorables para ellos se deben no a causas deportivas sino a este fenómeno. Esto, si es cierto, equivale a pérdidas económicas para los equipos de llano.

En el otro bando están algunos dirigentes que no aceptan la tesis de que la altura es una ventaja para los equipos del altiplano ya que ellos también sufren la diferencia de altura al bajar a jugar al llano, por lo tanto la organización del campeonato es totalmente justa y sus logros futbolísticos y económicos son ganados deportivamente.

Este trabajo académico tiene en cierto modo el motivo de develar la polémica en torno al tema mediante la utilización de técnicas estadísticas que a continuación se mencionan.

Las técnicas de estadística multivariante son técnicas poco difundidas en el medio y no se encuentran en los programas de estudio de licenciatura de las universidades de este país. Internacionalmente hablando, éstas están siendo difundidas en mayor proporción en los países adelantados, si bien su existencia es antigua su uso es relativamente reciente gracias a la aparición de potentes softwares estadísticos que facilitan grandemente el análisis de gran cantidad de datos y variables.

Estas técnicas se aplican a casi cualquier ámbito de estudio, en lo que respecta a el presente trabajo académico, el tema analizado es un muy buen medio para el uso y difusión de estas técnicas.

Las técnicas multivariantes suponen un tremendo poder analítico en manos del investigador, también crean una gran *carga* para éste, ya que tiene que asegurarse de que se mantengan los cimientos teóricos y estadísticos sobre las que se basan. Examinando con anterioridad los datos de la aplicación de una técnica multivariante, el investigador gana una comprensión básica de los datos y las relaciones entre las variables. En primer lugar, el investigador obtiene un conocimiento básico de los datos y las relaciones entre las variables. Las técnicas multivariantes plantean grandes demandas al analista en la comprensión, interpretación y articulación de resultados basados en relaciones cuya complejidad está en continuo aumento. El conocimiento de las interrelaciones de variables puede ayudar enormemente en la especificación y refinamiento del modelo multivariante. Así como proporcionar una perspectiva razonable para la interpretación de los resultados. En segundo lugar, las técnicas multivariantes demandan mucho más de los datos que se van a analizar. La potencia

estadística de las técnicas multivariantes requiere grandes conjuntos de datos y supuestos más complejos que los que encontramos en los análisis univariantes. La sofisticación analítica necesaria para asegurar que se consiguen los requerimientos estadísticos al aplicar la técnica multivariante elegida, obliga al analista a usar una serie de técnicas de examen de los datos que en muchas ocasiones rivaliza en complejidad con la propia técnica multivariante. Además, los efectos de los datos ausentes, los cuales por definición no se representan directamente en los resultados. Pueden ser sustanciales por el impacto que tienen sobre la naturaleza y carácter de los resultados. El propósito de este trabajo académico es proporcionar una visión general de las técnicas de examen de los datos, que van desde el simple proceso de inspección visual de los gráficos al proceso estadístico multivariante que requiere el análisis de datos y a comprobación de los supuestos subyacentes en todos los métodos multivariantes¹.

En cuanto a la explicación médica que se toma en cuenta para que jugadores, dirigentes deportivos y entendidos en el fútbol arguyan que la altura otorga una ventaja a los equipos del altiplano y los valles y por consiguiente efectos económicos; se puede resumir en los párrafos siguientes:

En la composición de la atmósfera terrestre, el oxígeno representa el 20%, el Nitrógeno el 79% y el resto lo conforman los gases raros y el anhídrido carbónico, por tanto la presión parcial de oxígeno en el aire respirado por el hombre y otros animales a nivel del mar será también el 20% de la presión total del aire atmosférico, salvo las modificaciones provocadas por el tenor de humedad ambiente y la

¹ www.unal.edu.co-estadística para la investigación social

saturación progresiva de vapor de agua que va sufriendo el aire inspirado en su recorrido por las vías aéreas superiores, el que llega al 100% en los alvéolos pulmonares que disminuyen la presión total del aire atmosférico. Además el anhídrido carbónico producto del metabolismo reemplaza al oxígeno consumido en la composición del aire alveolar, reduciendo el 20% de aire inspirado a solamente un 15% aproximadamente.

Esta presión parcial es más que suficiente para que el gas puesto en contacto con la hemoglobina de la sangre que llega desde los tejidos y pasa por el pulmón, la sature completamente.

En la altura la composición porcentual sigue siendo siempre la misma pero lo que ha cambiado es la presión atmosférica, es decir la presión total, y por ende la presión parcial a que dichos gases están sometidos.

Entonces como consecuencia de la altura y por ende de la menor presión atmosférica, cada molécula de gas ocupa mayor volumen a causa de la expansión, siendo sumamente importante este hecho para comprender algunas de las adaptaciones cardiorrespiratorias a la altura.

A medida que se asciende en la atmósfera disminuye la presión parcial de oxígeno, generando en las personas que viven a nivel del mar y se trasladan a zonas por encima de 2000 metros un estado de hipoxia, o se le producen una serie de cambios fisiológicos, que tienen profundos efectos sobre el rendimiento físico y que en numerosas ocasiones se transforman en cuadros clínicos, y que se incrementará a medida que la altura sea mayor. Y que conforman el mal de altura.

Cuando la altura es menor de 3000 metros la oxigenación arterial esta moderadamente disminuida en un sujeto en reposo, por tanto no suele exigir una mayor actividad del corazón, pero en cambio por encima de esa altura considerada crítica como promedio, la baja presión parcial de oxígeno en el alveolo produce una deficiente oxigenación arterial o hipoxemia y obligando al organismo a echar mano a recursos de emergencia para conseguir una compensación.

En general puede aceptarse que en alturas por debajo de los 3000 metros, para los sujetos en reposo no representa problema alguno, pero si el ejercicio produce una serie de cambios que son interesantes de analizar.

El aire en altura se halla sometido a una menor presión y por lo tanto las moléculas de los gases que lo constituyen tienden a expandirse. Por ello para un mismo ejercicio que provoca el mismo consumo de oxígeno tanto en el llano como en la altura, en esta ultima se necesita un mayor volumen del aire respirado para aportar la misma cantidad que a nivel del mar.

Entonces el deportista que estaba psíquicamente ubicado y aceptaba la hiperventilación para un determinado ejercicio a nivel del mar como habitual, cuando realiza el mismo en la altura se ve sorprendido por esta hiperventilación que le hace poner en tela de juicio sus aptitudes para soportarla. Por consiguiente la sorpresa psicológica es posiblemente tan importante como el desequilibrio somático que trae

aparejado el ejercicio en la altura, al extremo de que con cierta frecuencia se transforma en temor e inhibición para desarrollar sus habituales performances.

Además es aceptable pensar que cuando existe temor se produce un aumento de las catecolaminas circulantes, verificándose un franco incremento del consumo de oxígeno básico, con mecanismos netamente antieconómicos, lo cual provoca un agotamiento del sistema simpático adrenal, generándose entonces un círculo vicioso negativo que dificulta la buena adaptación del deportista a la altura, con la correspondiente performance deficitaria.

El principal efecto de la altura sobre el rendimiento deportivo se ve durante el máximo esfuerzo o ejercicio. A partir de los 2000 metros el consumo máximo de oxígeno se reduce en un 10% por cada 1000 metros, y el gasto cardíaco máximo se produce a una intensidad de ejercicio inferior al nivel del mar o submáximo.

Por tanto el deportista puede lograr el mismo rendimiento submáximo, pero requiere relativamente más trabajo fisiológico para conseguirlo.

Como pauta general se necesitan 2 semanas para adaptarse a 2300 metros y por cada 610 metros hay que sumar otra semana para que se produzca una adaptación completa.

2. Planteamiento del problema

Tomando en cuenta que la altura conlleva un efecto fisiológico y psicológico en el rendimiento de los deportistas que no están aclimatados a esta a causa de la menor

presión atmosférica y que en el fútbol este es un tema de permanente debate, se hace necesario hacer un análisis estadístico de los efectos de la altura en los resultados del fútbol boliviano, ya que este es un ámbito natural de observación por las características de ubicación geográfica de los equipos participantes de la Liga Boliviana.

3. Objetivos

3.1. Objetivo General

Realizar un análisis estadístico multivariante que determine en que medida la altura, la calidad de los equipos y la condición de localía inciden en los resultados del fútbol boliviano e identificar las implicaciones económicas a causa de esto.

3.2. Objetivos Específicos

Determinar en que medida incide la altura en los resultados del fútbol boliviano.

Determinar la categoría histórica de los equipos de la L.P.F.B.

Determinar el grado de incidencia de la categoría histórica de los equipos en los resultados del fútbol boliviano.

Determinar el grado de incidencia de la condición de localía de los equipos en el fútbol profesional boliviano.

Identificar implicaciones económicas a causa de los efectos de la altura.

4. Delimitación

4.1. Limite Espacial

El escenario futbolístico de La Liga del Fútbol Profesional Boliviano.

4.2. Limite Sustantivo

El trabajo se apoya en la teoría y técnicas estadísticas.

CAPITULO I

CATEGORÍA DE LOS EQUIPOS

1. Categoría de los Equipos

La definición de la categoría de los equipos es un tópico muy delicado en cuanto al análisis a realizarse, ya que dependiendo de la definición de la misma los resultados esperados del objetivo general del trabajo académico pueden variar; también se debe tener en cuenta que la definición de la categoría de los equipos puede conllevar susceptibilidades y estar inmersa en juicios de valor, todo esto por motivos obvios de la pasión que el fútbol inserta en la sociedad.

El presente capítulo está íntegramente desarrollado para llegar a categorizar los equipos de la liga, para ello se decidió realizar un **análisis factorial y un análisis de conglomerados**.

Previamente a la escritura de este trabajo académico se digitó al software estadístico SPSS una muestra de 3564 casos durante casi un mes, esto es necesario para realizar los cálculos en dicho software sin el cual fuera imposible hacer un trabajo de este tipo. Los 3564 casos representan casi todos los partidos oficiales de la Liga Profesional de Fútbol Boliviano desde el año 1979 hasta el 2002. Estos datos fueron proporcionados por la L.P.F.B. mediante el Libro “Una Historia en Números” del Lic. Walter Castedo².

Cada caso representa un partido, junto al cual se introdujo o digitó la siguiente información³:

- Goles a favor del equipo local

² Ver anexos

³ Ver anexos.

- Goles a favor del equipo visitante
- Diferencia de goles en el partido
- Altura de la ciudad donde se juega
- Posición Histórica del equipo (Se explica en profundidad mas adelante).
- Posición final en la tabla de posiciones del año anterior.
- Resultado del partido

Estos datos se deben completar mas adelante con resultados a obtener en el presente capítulo.

1.2. 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.

Tomando en cuenta el principio de **parsimonia e interpretabilidad** el objetivo del análisis factorial es explicar los fenómenos con el menor número de elementos posibles pero además estos factores deben poder ser interpretados mediante la teoría sustantiva.

1.2.1. Variables a Tomar en Cuenta Para el Análisis Factorial

Una vez hecha la codificación de los clubes se determina las variables a tomar en cuenta para llevar a cabo el análisis factorial, las cuales se detallan en la siguiente sección.

1.2.1.1. Variables Numéricas

- Goles a Favor del Equipo Local
- Goles en Contra del Equipo Local
- Diferencia de Goles en el partido.- Si la diferencia es positiva el resultado es a favor del equipo local, si es negativa el resultado es a favor del equipo visitante y si es “0” es un empate.
- Altura de la Ciudad donde se Juega .- Se toma la altura de la ciudad donde se juega el partido, se toma como medida los Metros Sobre el Nivel del Mar, a continuación se detalla la altura por ciudad.

Trinidad: 230 Mts. S. N. M.; Santa Cruz: 390 Mts. S. N. M.; Tarija: 1850 Mts. S. N. M.; Cochabamba: 2540 Mts. S. N. M.; Sucre: 2790 Mts. S. N. M.; La Paz: 3650 Mts. S. N. M.; Oruro: 3710 Mts. S. N. M.; Potosí: 4090 Mts. S. N. M.

- Posición Histórica del Equipo (Local y visitante).- Se determinó una posición histórica de los equipos, promediando la suma de posiciones de los campeonatos en que participaron. Esta tabla va desde el 1 al 14 y quedó de la siguiente manera

TABLA II.1:
POSICIÓN HISTÓRICA

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
14	Iberoamericana			
14	Universitario			
14	Real Beni			
14	Universitario del Beni			
14	Pompeya			
14	Wilster Coop			
14	Litoral			
14	Always Ready			
14	Mcal Braun			
14	Universitario de Pot	EQUIPOS DE PARTICIPACION CORTA E		
14	Metalsan	IRRELEVANTE EN LOS CAMPEONATOS		
14	Orcobol			
14	Petrolero			
14	San Pedro			
14	Municipal			
14	Magisterio Rural			
14	20 de Agosto			
14	Ferroviano			
14	Ind Unificada			
14	Chaco Petrolero			
14	1° de Mayo			
14	Magisterio Rural			
14	Bata			

La tabla anterior muestra la suma de posiciones de los equipos a lo largo de los años estudiados, es decir si un equipo quedó quinto en la tabla general del año 1995 se agrega cinco a la suma de posiciones, la suma total que se muestra en la tercera columna se divide entre el número de campeonatos jugados por el equipo en cuestión y el producto es la posición promedio mostrada en la última columna.

1.2.1.2. Variables Categóricas

Las variables categóricas son ellas que como su nombre indica asignan categorías y si bien se utilizan números para clasificarlas estos son simplemente códigos de clasificación.

- Condición de Localía.- Si hay algún equipo local se lo codifica como “1”, si los dos clubes son de la misma ciudad se codifica “0”.
- Resultado.- Se la toma como tres variables dicotómicas, Gana el Local, empate o pierde el local, para las tres los códigos son “1” ó “0”.
- Posición en el año Anterior (equipo local y visitante).- Se la toma como cuatro variables dicotómicas, si estuvo en el primera, segunda, tercera o cuarta parte de la tabla de posiciones de años anterior, para las cuatro los códigos son “1” ó “0”, si el equipo es uno recién ascendido se lo toma como uno de la cuarta parte de la tabla.

Una vez realizada la codificación se exportan los datos al software estadístico SPSS en el cual se realiza el primer análisis factorial.

1.3. Análisis de Componentes Principales

El tipo de análisis factorial a utilizar es el análisis de componentes principales que extrae factores mediante combinaciones lineales independientes de las variables utilizadas. El primer componente explica la varianza máxima y los sucesivos explican proporciones menores de las varianzas, los componentes no están correlacionados entre sí.

El análisis explora toda la varianza de cada variable: la común al resto, la específica y la debida a errores de observación⁴.

⁴ Mahía, Ramón, Guía de Curso Redes de Información y Análisis de Datos, Pág. 10

Se escogió el método de análisis de componentes principales debido a que el uso de este se recomienda para la reducción de datos. Antes de hacer el análisis final se debe interpretar la primera matriz de componentes principales que se muestra a continuación, esto con el fin de determinar las variables que deben seguir en la “segunda ronda” del análisis, ya que el objetivo del mismo es reducir el número de variables y componentes.

- El ACP es una técnica que transforma ciertas variables en otras incorrelacionadas, de media cero, que pueden escribirse como combinaciones lineales de las primeras y que se llaman factores o componentes principales, las cuales pueden ordenarse por la magnitud de su varianza la cual está dada por un valor propio de la matriz Σ (en la práctica de \mathbf{S})
- Las primeras r componentes principales bastan para describir en alto porcentaje la variabilidad total de las variables originales. Con frecuencia r vale 2 o 3, siendo el primero de ellos el caso más deseable.
- Cuando el porcentaje de variabilidad explicado por dos componentes principales es alto (70%) se puede realizar una representación gráfica de las variables originales y de los individuos de la muestra (mapas perceptuales) que muestran algunas relaciones de correlación o semejanza entre ellos.
- Aunque todas las variables originales entran en la composición de cada componente principal, algunas son más importantes que otras. Estas, las más importantes, determinan la naturaleza de cada componente⁵

1.3.1. La matriz de Cargas, factorial o de Componentes

⁵ J. A. Clavijo, Monografía sobre ACP, Pág. 6

Esta matriz relaciona factores y variables para aproximarnos a su significado. Matriz de estructura: Matriz que contiene los coeficientes de correlación entre factores y variables originales⁶.

TABLA II.2: MATRIZ 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

Método de extracción: Análisis de componentes principales.
a 6 componentes extraídos

Elaboración: Propia

En la anterior tabla se pueden apreciar los resultados obtenidos del software estadístico S.P.S.S. del Análisis Factorial en el que se extraen 6 componentes, los cuales muestran las distribuciones de varianza total del caso en cada variable.

Si se observa la tabla se aprecia que las variables aportan de distintas maneras a los componentes, las correlaciones o “cargas” relevantes están resaltadas en la matriz de componentes para determinar las variables que deben mantenerse en análisis factorial se usará el criterio de comunalidad.

1.3.1.1. Comunalidades

⁶ Op. Cit., pág. 11

Se denomina "comunalidad" a la proporción de la varianza explicada por los factores comunes en una variable. La comunalidad (h^2) es la suma de los pesos factoriales al cuadrado en cada una de las filas⁷.

TABLA II.3: 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

Método de extracción: Análisis de Componentes principales.

La tabla II.3 muestra las comunalidades de las variables utilizadas en el análisis, en esta se puede apreciar que solo la variable año no supera el 0.5 recomendado para continuar a la variable en el análisis.

Se determinó que las variables a continuar en el análisis son las siguientes:

- Posición Histórica Club Local
- Posición Histórica Club Visitante
- Goles Club Local
- Goles Club visitante
- Diferencia de Goles
- Gana Local
- Empate

⁷ Op. Cit., Pág. 31

- Gana Visitante
- Localía
- 1° Cuarto de la Tabla de Posiciones año anterior
- 2° Cuarto de la Tabla de Posiciones año anterior
- 3° Cuarto de la Tabla de Posiciones año anterior
- 4° Cuarto de la Tabla de Posiciones año anterior

Se retiró del análisis la variable año ya que esta tampoco presenta una distribución general de “cargas” en los componentes extraídos, esto denota que el resultado de los partidos a la larga no tiene que ver con años excepcionales y mas bien indica que el comportamiento de los equipos es relativamente constante en un tiempo largo. Recurriendo a criterios estadísticos se determinó retirar del análisis la variable altura ya que el objetivo final de este capítulo es determinar una categorización de los equipos de la L.P.F.B. que debe estar determinada por la el “comportamiento” o sucesión de resultados de los equipos, el incluir esta variable implicaría una tendencia a clasificar los equipos de acuerdo a la altura en donde estos juegan como local.

1.4. Análisis Factorial Confirmatorio

Con las variables ya definidas se procede al siguiente análisis factorial, el cual muestra los siguientes resultados:

TABLA II.4: Varianza total explicada

Com ponente	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			

Elaboración: Propia

En la tabla II.4 se aprecia que el número total de componentes es de 12. Para determinar los componentes “útiles” al análisis se utiliza el criterio de la “suma de saturaciones al cuadrado” mayor a 1, es decir que la suma de los coeficientes al cuadrado aportada por cada variable a cada componente debe ser mayor o igual a “1” para ser tomado en cuenta como un componente principal. Los componentes que cumplen el criterio son cinco, los cuales se muestran junto a las “cargas” asignadas de cada variable en la siguiente matriz de componentes. Estos cinco componentes explican el 78% de la varianza lo cual está por encima del 70% recomendado en este tipo de análisis.

TABLA II.5: Matriz de Componentes

	Componente				
	1	2	3	4	5
Posición Histórica Local	-,472	,683	,009	,076	,024
Goles Club Local	,748	,111	,321	,034	-,008
Goles Club Visita	-,535	-,239	,563	,038	-,079
DIF Goles	,919	,224	-,039	,008	,036
Gana Local	,866	,253	,202	,076	-,021
Empate	-,422	-,109	-,778	-,147	,071
Pierde Local	-,660	-,209	,612	,066	-,052
Localía o no	,014	,371	,187	,131	,090
1° cuarto Tabla Local	,420	-,657	,089	-,446	,193
2° cuarto Tabla Local	,063	-,106	-,184	,571	-,783
3° cuarto Tabla Local	-,163	,076	-,034	,624	,735
4° cuarto Tabla Local	-,316	,655	,101	-,608	-,131

Método de extracción: Análisis de componentes principales.
a 5 componentes extraídos
Elaboración: Propia

Esta matriz muestra las cargas factoriales asignadas a cada componente por cada variable.

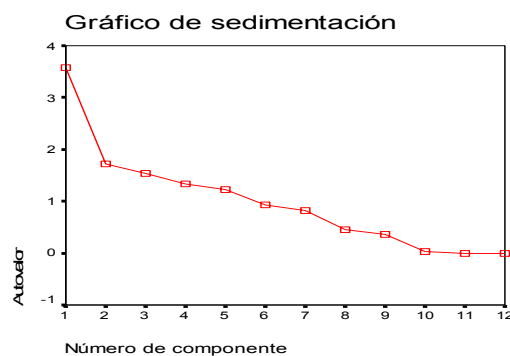
1.4.1. Gráfico de Sedimentación

El gráfico de sedimentación que a continuación se muestra (GRAFICO II.1) es resultado de análisis factorial realizado anteriormente y que dio como resultado cinco componentes útiles al análisis.

Para determinar el número de conglomerados es necesario visualizar el gráfico de sedimentación en el segmento del componente “1” al componente “5” esto por ser los cinco componentes extraídos por cumplir con el criterio de saturaciones mayores que “1”, en ese segmento se aprecia que hay como máximo tres “quebres” lo que podría interpretar como la repartición de varianza.

Estos tres “quebres” se identifican desde el componente “1” al “2”, del componente “2” al “4” y del componente “4” al “5”. Por lo tanto el número de conglomerados o categorías a determinar es de 3.

GRAFICO II.1



Elaboración: Propia

1.5. Análisis de conglomerados

Una vez llevado a cabo el análisis factorial que determina los componentes principales se precede a realizar el análisis de conglomerados que determina la clasificación que se le otorga a cada equipo.

Este procedimiento que también se realiza con la ayuda del software S.P.S.S. intenta identificar grupos de casos relativamente homogéneos basándose en las características seleccionadas y utilizando un algoritmo que puede gestionar un gran número de casos.

1.5.1. Análisis de Conglomerados de k-medias en dos fases

Por su mejor acoplamiento al manejo de distinto tipo de variables y gran número de casos el tipo de análisis de conglomerados a utilizar es el de k-medias en dos fases.

El método del análisis de conglomerados de k-medias tiene el siguiente proceso:

Seleccionar “G” puntos como centros de los procesos, esto puede hacerse de la siguiente manera:

- Asignando aleatoriamente los objetos a los grupos y tomando los centros de los grupos así formados;
- Tomando como centro los “G” puntos mas alejados entre sí;
- Seleccionando los centros a “priori” con criterios empíricos;

El criterio a utilizar en el siguiente análisis es el de “tomar como centro los puntos G mas alejados entre sí”; esto por ser un método totalmente estadístico en donde no caben juicios de valor y se evita la mano del investigador.

Una vez realizado lo anteriormente indicado se calculan las distancias euclídeas de cada elemento al centro “G” de los grupos y se asigna cada elemento al grupo mas próximo. La asignación se realiza secuencialmente y al introducir un nuevo elemento en un grupo se recalculan las coordenadas en una nueva media de grupo⁸.

El criterio de cálculo es el de la suma de cuadrados dentro de los grupos, el cual tiene como modelo matemático la siguiente forma:

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

Partiendo de “h” grupos y “p” variables, en donde:

SCDG: Suma de cuadrados dentro del grupo

$\sum_{j=1}^p$: Suma de desviaciones en todas las variables (p) para todos los sujetos (nj)

dentro del grupo “g”

x_{ijg} : Valor de la variable “j” para cada sujeto “i” perteneciente al grupo “g”.

\bar{x}_{jg} : Media de la variable “j” en el grupo “g”

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

El criterio persigue minimizar la varianza intra-grupos o dentro de los grupos.

⁸ Fondo Social Europeo, Informática Aplicada al Análisis Económico, Pág. 12

Una vez realizado el procedimiento necesario para la programación del análisis en el software S.P.S.S. se obtuvo el siguiente resultado.

TABLA II.6
CONGLOMERADOS O CATEGORIAS DEFINIDAS

CATEGORIA 1	CATEGORIA 2	CATEGORIA 3
Bolivar	Wilstermann	Aurora
Oriente Petrolero	Real Bamin	Ind Unificada
The Strongest	Guabirá	Real Beni
Blooming	Stormers	Orcobol
	Ind Petrolero	Petrolero
	Destroyers	San Pedro
	Unión Central	Always Ready
	Litoral	Mcal Braun
	Metalsan	Universitario de Pot
	Municipal	Universitario del Beni
		Pompeya
		Wilster Coop
		San José
		Real Santa Cruz
		Ciclón

Al distribuirse los equipos en torno a grupos con varianza mínima el resultado indica la clasificación de los equipos obtenida mediante el método de análisis de conglomerados.

Cada conglomerado es una categoría de equipos, en cada una de ellas se aglutinan los equipos con un comportamiento histórico similar, es decir que tienen resultados “parecidos” a lo largo de los años hablando en términos de triunfos, goles a favor o goles en contra.

El análisis realizado acaba de definir una de las variables a las cuales según los objetivos de este trabajo académico se les debe determinar su efecto sobre los resultados del fútbol; esta variable que como ya se dijo se denomina categoría de los equipos será analizada a profundidad en siguientes capítulos.

CAPITULO II

ANALISIS DE LA VARIANZA

2. Análisis de la Varianza

Previo al análisis de varianza se realizaron las pruebas de homogeneidad de varianzas, independencia y normalidad, ya que estos son supuestos del análisis, todas las pruebas resultaron ser positivas en torno al cumplimiento de los supuestos; se debe tener en cuenta que el gran tamaño de la muestra (3565 partidos) facilita el cumplimiento de los mismos.

En el capítulo presente se lleva a cabo el análisis de varianza múltiple de efectos fijos, se dice múltiple ya que toma tres variables como “resultados” o variables dependientes y de efectos fijos pues las variables independientes están decididas de antemano por quien realiza el análisis. En el análisis de varianza a realizarse en este capítulo se ha “incluido” la variable obtenida en el capítulo anterior, la categoría de los equipos.

Se determinó de acuerdo a los resultados del análisis factorial que las variables a analizar serán las siguientes:

Variables Dependientes

- Diferencia de Goles (positiva si gana el local, negativa el visitante)
- Goles del Equipo Local
- Goles del Equipo Visitante

Variables Independientes

- Conglomerado o Categoría del Equipo
- Condición de Localía
- Altura del Estadio donde se juega

Cabe aclarar que la altura que en anterior capítulo se la definía como numérica en el presente capítulo es categórica y se le asigna códigos del “1” al “3”, esto para facilitar una lectura mas clara de los resultados.

2.1. Bases del Análisis de Varianza

Supónganse k muestras aleatorias independientes, de tamaño n , extraídas de una única población normal. A partir de ellas existen dos maneras independientes σ de estimar la varianza de la población.

1) Una llamada varianza intra-grupos (ya que sólo contribuye a ella la varianza dentro de las muestras), o varianza de error, o cuadrados medios del error, que se calcula como la media de las k varianzas muestrales. La varianza del error es un cociente: al numerador se le llama suma de cuadrados del error y se representa por SSE y al denominador grados de libertad por ser los términos independientes de la suma de cuadrados.

2) Otra llamada varianza entre grupos (sólo contribuye a ella la varianza entre las distintas muestras), o varianza de los tratamientos, o cuadrados medios de los tratamientos). Se calcula a partir de la varianza de las medias muestrales y es también un cociente; al numerador se le llama suma de cuadrados de los tratamientos y al denominador $(k-1)$ grados de libertad⁹.

2.2. Modelo Lineal General Multivariante (MANOVA)

Para el cálculo estadístico adecuado para el análisis de varianza en el Modelo Lineal General Multivariante se recurre a las tablas de análisis de la varianza.

⁹ Levine, D.M., Statistical Methods and Applications, Pag. 240

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		

La suma de cuadrados del análisis de varianza puede definirse del siguiente modo:

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum (Y_i - \hat{Y}_i)^2$$

Es decir, la suma del cuadrado del total de desviaciones, que mide la variabilidad de Y_i , es igual a la suma del cuadrado de desviaciones de la regresión, que mide la variabilidad de Y_i que es eliminada al estimarse Y_i , más la suma del cuadrado de los residuos, que mide la variabilidad de Y_i que no se reduce al estimarse Y_i .

\bar{Y} en $(Y_i - \bar{Y})^2$ denota la media de los valores observados de Y_i , mientras $\bar{\hat{Y}}$ en $(\hat{Y}_i - \bar{\hat{Y}})^2$ denota la media de los valores Y estimados. Dado que $U_i = Y_i - \hat{Y}_i$, ello implica que $U_i = Y_i - \hat{a} - cX_i$. Los grados de libertad están asociados a la suma de cuadrados, pero serán diferentes dependiendo del tipo de regresión.

La tabla de análisis de varianza con “q” factores presenta la siguiente forma:

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	$MC_t = \frac{SC_t}{gl_t}$	

El primer modelo da como resultado tres posibles cuantificaciones de los efectos sobre el resultado, la diferencia de goles, los goles del equipo local y los goles del equipo visitante. Los análisis preliminares indican que los modelos que tienen como resultado la diferencia de goles y los goles del equipo local se ajustan mejor, ya que presentan Coeficientes de determinación de .89 y .80 respectivamente.

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.

2.2.1. Descomposición de la Varianza y Estadístico Lambda de Wilks

El análisis multivariante de la varianza con “q” factores se basa en que la variabilidad total de la muestra puede descomponerse en la variabilidad debida a las diferencias entre grupos y la debida a la diferencia dentro de los grupos.

$$SC_{total} = SC_{entre} + SC_{intra}$$

Donde:

SC_{total} : Mide las desviaciones de cada observación al centro total sin distinguir grupos.

SC_{entre} : Mide las desviaciones entre los centros de cada grupo al centro total

distinguiendo los grupos.

SC_{intra} : Mide las desviaciones de cada observación al centro del grupo

correspondiente.

A partir de esta descomposición, para determinar que parte de la variabilidad total es debida a cada uno de los términos, bastaría con calcular el cociente de cada uno de ellos y la variabilidad total. En este sentido, el estadístico Lambda de Wilks compara las desviaciones dentro de cada grupo con las desviaciones totales sin distinguir grupos.

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

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 F_A, F_B y F_{AB} respectivamente). El estadístico tiene la siguiente expresión.

$$\Lambda = \frac{|SC_{entre}|}{|F_H + SC_{intra}|} \quad H=A, B, AB$$

En donde:

F_H = Suma de cuadrados y productos cruzados residual

TABLA II.2
Contrastes multivariados(c)

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

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

Elaboración: Propia

El anterior cuadro muestra la significación del estadístico Lambda de Wilks por variable del modelo:

$$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3 + \beta_7 X_1 X_2 X_3 + e$$

En donde:

α : 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

$\beta_4 X_1 X_2$: Efecto de la interacción categoría-localía

$\beta_5 X_1 X_3$: Efecto de la interacción categoría-altura

$\beta_6 X_2 X_3$: Efecto de la interacción localía-altura

$\beta_7 X_1 X_2 X_3$: Efecto de la interacción categoría-localía-altura

e: Error en el modelo

y: Resultado de un partido

Se deduce que las variables, localía, conglomerado o categoría de los equipos y altura son significantes en el modelo ya que la significación otorgada por el estadístico Lambda de Wilks es de .000, .000 y .002 respectivamente, también se aprecia que la interacción localía-altura es significativa, es decir, su variación inter-grupos es significativa.

Como conclusión de lo descrito anteriormente se puede decir que las tres variables introducidas de antemano en el modelo son significantes por lo cual se deduce que tienen un efecto significativo en los resultados del fútbol boliviano y que la interacción de la localía con la altura también lo es.

Una vez que se sabe las variables e interacciones significativas del modelo es posible determinar cuales de ellas se deben “extraer” en el modelo lineal general final, estas son todas las que se definieron significativas en la tabla III.2; estas son:

- Conglomerado o categoría de los equipos; Condición de localía; Altura; Interacción Localía - Altura

2.3. Modelo Final del Análisis de la varianza (ANOVA)

El modelo final de Análisis de varianza se debe realizar con las variables e interacciones descritas anteriormente.

La tabla de análisis de varianza tiene la forma mostrada anteriormente en el presente capítulo y la cual es el primer producto del análisis de varianza realizado en el software S.P.S.S. El resultado se muestra a continuación:

Tabla II.3

ANOVA

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
Conglomerado	Inter-grupos	149,396	18	8,300	13,635	,000
	Intra-grupos	2191,913	3601	,609		
	Total	2341,309	3619			
Localía	Inter-grupos	6,113	18	,340	2,629	,000
	Intra-grupos	465,183	3601	,129		
	Total	471,296	3619			
NTILES of ALTURA	Inter-grupos	28,722	18	1,596	2,101	,004
	Intra-grupos	2735,224	3601	,760		
	Total	2763,946	3619			

Elaboración: Propia

La tabla muestra la suma de cuadrados para las tres variables a analizar de la varianza intra-grupos e inter-grupos, se aprecia que la significancia de las tres variables es menor al .05 por lo tanto se define que las varianzas inter-grupos e intra-grupos son significativamente distintas, por lo tanto se deduce que las variables tienen efectos significativos en sus distintos niveles.

El segundo resultado esperado es la tabla de efectos inter-sujetos solo para el resultado “diferencia de goles”.

Tabla II.4

Prueba de los Efectos Inter-sujetos

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
LOCALIA	45,869	1	45,869	12,957	,000	,05
NALTURA	27,934	2	13,967	3,945	,007	,03
LOCALIA * NALTURA	35,096	2	17,548	4,957	,008	,03
Error	12786,775	36	3,540			
Total	18297,000	36				
Total corregida	13945,353	20				
		36				
		19				

a Calculado con alfa = ,05

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

Elaboración: Propia

En el cuadro se aprecia que se confirma que las variables de altura, conglomerado y localía además de la interacción de la localía y la altura son significantes. Esto significa que la categoría de los equipos, la condición de localía y la altura por si solas tienen efectos altamente significativos en el resultado de los partidos, aún así la localía es más significativa que la altura ya que su valor crítico de significancia (0.000) es menor que el valor crítico de significancia de la altura (0.007); pero se puede apreciar claramente que la altura se refuerza con la condición de localía (0.008).

El estadístico eta parcial que mide la variabilidad del resultado atribuible a un factor muestra gran información para el análisis, tiene la siguiente forma:

$$\eta^2 = \frac{F_g * gl_g}{F_g * gl_g + gl_e} = 1 - \Lambda$$

En donde:

F_g : Estadístico F de un efecto.

gl_g : Grados de libertad de un efecto.

gl_e : Grados de libertad del error.

En cuanto a la categoría de los equipos se aprecia que se atribuye la variabilidad del resultado en 0.70 a la categoría o conglomerado de los equipos, es la variable más “potente” del modelo.

Para poder dar una cuantificación porcentual a las variables analizadas y sus respectivas interacciones se debe hacer una “transformación” del 0.81 que suman las mismas al 100%. Esto se realiza haciendo una simple división del estadístico eta

obtenido por cada variable entre la suma total de las tres variables más la interacción de la localía-altura. El resultado es el siguiente:

TABLA IV.5

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

Elaboración: Propia

El cuadro muestra los porcentajes de los efectos atribuibles a cada factor, se aprecia que la variable de categoría de los equipos o conglomerado tiene un efecto del 86% en el resultado, la localía del 6%, la altura más su interacción con la localía un 8%.

En cuanto a la condición de localía, esta presenta un porcentaje del 6% en los resultados del fútbol, esta es una ventaja deportiva que se toma en cuenta en la organización de los campeonatos ya que los campeonatos se realizan de “ida y de vuelta”, es decir que cada equipo juega igual número de partidos como local y como visitante.

La altura presenta un porcentaje del 4% en su efecto por sí sola pero se refuerza con la localía un 4% más, lo cual refleja una ventaja extra-deportiva del 8% para los equipos de la altura. Esta ventaja no se contempla en la organización del campeonato lo cual repercute en los resultados finales del mismo.

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; mas aún cuando el equipo que sube a jugar es uno de categoría “2” o “3”.

Para poder determinar la ventaja extradeportiva de la altura se procederá bajo el supuesto de que todos los equipos tienen la misma categoría. El cálculo muestra los siguientes resultados:

TABLA IV.6

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

Elaboración: Propia

La tabla anterior muestra los porcentajes que tienen de efecto sobre el resultado las variables, localía, altura e interacción altura-localía; como se dijo anteriormente estos resultados dejan de lado el efecto de la categoría de los equipos ya que se está tomando el supuesto de que todos los equipos tienen la misma categoría.

Para determinar la ventaja que tiene un equipo del llano jugando contra un equipo de la misma categoría de local, solo tenemos que ver el porcentaje de localía que es del 46%, ahora, si el equipo local es uno de la altura “3” contra un equipo del llano la probabilidad es de 46% más la multiplicación del mismo por el 54% de la altura más su interacción con la localía, lo cual da como resultado un 70.84%

Haciendo un análisis futbolístico a estos datos, se puede deducir que un equipo local del llano jugando contra un equipo visitante sea este de la altura o del llano tiene un 46% de ventaja sobre el equipo visitante por concepto de la localía, ya que la altura de “bajada” no tiene efectos (ver estimación de los parámetros). En cambio si un equipo de la altura del altiplano juega contra un equipo de similar categoría que sube desde el llano tiene un casi un 71% de ventaja sobre el visitante.

Traduciendo la ventaja que obtienen los equipos locales de las distintas regiones se puede deducir las probabilidades de cada resultado para un local del llano y para un local de la altura del altiplano

LOCAL LLANO	
Triunfo	64%
Empate	21%
Derrota	15%

LOCAL ALTURA "3" VS EQUIPO DEL LLANO	
Triunfo	80%
Empate	16%
Derrota	4%

2.3.1. Estimación de los parámetros

La tabla siguiente aparte de indicar los parámetros “B” calculados, los cuales son los coeficientes para cada nivel de cada variable y que se deben reemplazar en la fórmula del modelo para obtener medias marginales, indica la significación de cada parámetro y el estadístico “eta parcial al cuadrado” el cual indica la variabilidad de una variable en relación al nivel de la misma. *Cabe aclarar que el modelo a desarrollarse es un **modelo teórico** que sirve para hacer consideraciones de análisis estadístico y que no es bajo ningún criterio un instrumento de estimación de resultados futbolísticos.*

El modelo tiene la siguiente forma:

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

En donde:

α : 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

y : Resultado en un partido (diferencia de goles)

El resultado “extraído” por el software estadístico SPSS es el siguiente:

TABLA II.7
Estimaciones de los parámetros

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]	-,272	,185	1,466	,143	-,635	,092	,01
* [ALTURA=1]							
[LOCALÍA=0]	,637	,315	2,020	,043	,019	1,254	,01
* [ALTURA=2]							
[LOCALÍA=0]	0(a)
* [ALTURA=3]							
[LOCALÍA=1]	0(a)
* [ALTURA=1]							
[LOCALÍA=1]	0(a)
* [ALTURA=2]							
[LOCALÍA=1]	0(a)
* [ALTURA=3]							

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

Elaboración: Propia

La tabla mostrada proporciona valiosa información para el análisis como ser:

Los niveles de la variable conglomerado o categoría de los equipos son altamente significantes al modelo (.000), esto aparte de indicar que las tres categorías de los equipos tienen incidencia en el modelo “de paso” quiere decir que la variable estuvo muy bien definida en el capítulo anterior. El estadístico eta parcial al cuadrado que indica la variabilidad del resultado a causa del nivel de la variable, indica que la variabilidad en el resultado se da en gran medida a causa de la actuación de los

equipos del conglomerado o categoría “1” ya que este presenta un índice de .67 y la categoría o conglomerado “2” presenta un índice de .04.

En cuanto a la altura se aprecia que el nivel “1” de esta variable, es decir la altura de los llanos no es significativa a este nivel de confianza, lo cual se puede describir como que el bajar del altiplano o los valles no tiene efecto en el resultado del partido. Por lo tanto se infiere que el efecto de la altura en los resultados futbolísticos solo es de “subida”.

Dando una mirada al parámetro estimado para la altura “2”, es decir, la altura de los valles se aprecia que esta no tiene un efecto significativo en los resultados futbolísticos ya que el índice de significancia es de .056, encima del .050 máximo para ser considerado significativa al 95% de confianza.

2.3.2. Comparaciones de Medias Marginales

Una vez realizado el procedimiento descrito anteriormente se llevan a cabo comparaciones de medias marginales con el fin de reforzar y complementar los resultados obtenidos, las mismas indican que la altura de subida ni tiene efecto en los resultados además de indicar que el subir de los llanos (altura “1”) a los valles (altura “2”) tampoco tiene un efecto significativo.

CAPÍTULO III

IMPLICACIONES ECONÓMICAS

3. Implicaciones Económicas por el Efecto de la altura

La realidad económica de los clubes bolivianos dice que el fútbol no solo es un deporte o un juego, en realidad en cada partido se juega la estabilidad económica de los clubes participantes en el campeonato. Ante esto, cualquier ventaja representa una ganancia económica para el equipo beneficiando y una pérdida para el equipo que sufre la desventaja.

Hay tres conceptos principales por los que un equipo pierde o gana económicamente por los resultados de los partidos:

- Recaudaciones por venta de boletos
- Traspaso de jugadores a otros equipos
- Auspicios de empresas privadas
- Clasificación a campeonatos internacionales

Dentro del primer concepto se debe considerar que las recaudaciones tienen una directa relación con los resultados de los partidos anteriores del equipo que juega como local, es decir que cuando un equipo viene de ganar los partidos anteriores recibe una mayor cantidad de público y viceversa.

En cuanto a los traspaso de los jugadores, también en cierto modo, se puede decir que un equipo con buenos resultados tiende a cotizar mejor sus jugadores que uno que está entre los últimos lugares de la tabla.

La clasificación a campeonatos internacionales se la otorga a los equipos que estuvieron posicionados en los primeros puestos del campeonato, hasta el año 2001 al campeón y sub-campeón, el principal campeonato internacional de clubes en Sudamérica es la “Copa Libertadores de América”.

En entrevistas a Directivos de la Liga y conocidos dirigentes de equipos de fútbol se obtuvieron valiosas opiniones en cuanto a los efectos deportivos que la altura deportivos de un equipo conllevan, por ejemplo el Lic. Roberto Paz Limpias dirigente del club Blooming de Santa Cruz durante mas de una década dice textualmente *“el año 1984 Blooming tenía el mejor equipo del país con nueve jugadores en la selección, ese año en la final le ganamos a Bolívar en Santa Cruz y cuando jugábamos el partido de vuelta en La Paz faltando veinte minutos estábamos ganando 3-0 empezaron a venir los periodistas a mi casa por que con eso Blooming ya era campeón en esos 20 minutos nos volcaron el partido 6-3 esos resultados se dan en la altura, también hace unos tres años Bolívar perdía 5-0 en La Paz contra un equipo Brasileiro y en el segundo tiempo empató a 5.”* El mismo dirigente comenta que el principal efecto económico de la altura es perder la participación a torneos internacionales los cuales en promedio generan una ganancia líquida de \$us 500.000, este monto es suficiente para tener la economía del club saneada durante el siguiente año y poder hacer inversiones en jugadores de categoría, además otro factor importante es el incremento de la cotización de los jugadores del club si estos realizan un buen papel, según la opinión experimentada del mencionado dirigente cuando un jugador es traspasado al exterior el club puede percibir diez veces lo que puede percibir en el mercado interno. También opina que respecto a las recaudaciones por taquilla el público no deja de ir al estadio después de una derrota en la altura pues está

consciente de la desventaja en la que se juega en la altura, en sus propias palabras dice *“no dejan de ir por que saben que va a perder, los dirigentes sabemos que hay un 95% de posibilidades que pierda”*, también observa que un equipo necesita 20 días para aclimatarse óptimamente a la altura y que con una semana *“algo se puede hacer”*, pero que es muy caro hacerlo y que el campeonato no brinda las facilidades de hacerlo ya que los partidos son seguidos.

En síntesis en Bolivia tenemos cuatro equipos en la categoría “a” que son Bolívar, The Strongest, Blooming y Oriente Petrolero pero los dos equipos de la altura duplican o triplican las participaciones en Copa Libertadores de América a los equipos del llano, esto en suma es un problema económico para los equipos que sufren la desventaja ya que cada año se perdería \$us 500000, esto suponiendo que en promedio un equipo por año de la región clasifique a la Copa Libertadores de América; si se hace un análisis a mas largo plazo se puede deducir que eso se convierte en un problema económico regional ya que la pérdida sistemática de \$us 500000 al año se traduce en millones de dólares de salida de la mediante prácticamente el único deporte en Bolivia que tiene capacidad de generación de empleos y un alto impacto social, por otro lado se aprecia un efecto económico positivo para las regiones de los equipos de altura ya que estas reciben ingresos en inversión de infraestructura, medicamentos, sueldos, etc. Además que si bien no solo se puede aducir a los efectos de la altura hay equipos de ciudades sin tradición futbolística que han tenido la posibilidad de tener un campeonato internacional y que inclusive obtuvieron resultados holgados en condición de local ante equipos tradicionales como Peñarol del Uruguay. Por lo expuesto anteriormente se puede concluir que la altura influye de manera negativa en la economía de los equipos de

llano y positiva en los de la altura ya que de lógica se esperaría que los equipos de la misma categoría tengan un número similar de participaciones en campeonatos internacionales lo cual se ve desbalanceado por los efectos de la altura.

Para hacer una estimación del monto de pérdidas económicas se puede hacer una comparación entre los equipos de categoría “1” del llano y los de categoría “1” del altiplano, las comparaciones deben ser por pares de los equipos con más participaciones y menos participaciones para cada nivel de altura.

Los pares a comparar son los siguientes:

TABLA III.1
PARTICIPACIÓN EN COPAS LIBERTADORES

CLUBES DEL ALTIPLANO	Participaciones en la Copa L.	CLUBES DEL LLANO	Participaciones en la Copa L.
Bolivar	16	Oriente Petrolero	11
The Strongest	10	Blooming	5
TOTAL	26	TOTAL	16

Elaboración: Propia

La comparación se la realiza de acuerdo a la tabla anterior, por un lado los equipos Categoría “1” de la altura, Bolivar y The Strongest con 16 y 10 participaciones en la Copa Libertadores de América respectivamente y por el otro los categoría “1” del llano, Oriente y Blooming con 11 y 5 participaciones respectivamente. Observando los datos se denota la casualidad que cada uno de los equipos del altiplano lleva una ventaja de cinco participaciones en campeonatos internacionales a cada equipo del llano o cruceño.

Después de haber realizado el análisis estadístico de los capítulos II, III y IV se puede suponer que la diferencia de participaciones en la Copa Libertadores se debe a los

efectos de la altura por lo tanto si el campeonato no tuviera tales efectos la distribución de participaciones quedaría de la siguiente manera:

TABLA III.2
PARTICIPACIÓN ESTIMADA POR REGIÓN

Equipos del Altiplano	Participaciones en la Copa L.	Equipos del llano	Participaciones en la Copa L.
Bolivar	xx	Oriente Petrolero	xx
The Strongest	xx	Blooming	xx
TOTAL	21	TOTAL	21

Elaboración: Propia

Se aprecia que el efecto de la altura sería la reducción de la participación de los equipos de la altura de 26 a 21 y el aumento de participaciones de los equipos del llano de 16 a 21.

Si se toma el parámetro de \$us 500.000 de ganancia líquida promedio por participación en Copa Libertadores y se considera que la diferencia sería el aumento de cinco participaciones para la región del llano se puede deducir fácilmente mediante simple multiplicación que la región del altiplano tiene un efecto económico positivo de \$ 2.500.000 desde el año 1977 al 2002, es decir en 25 años de competencia.

Esta suma dividida entre los 25 años de competición analizados da como resultado \$us 100.000 por año, si se toma en cuenta el supuesto que esta debería ser dividida entre los dos clubes categoría “1” de la región se está hablando de \$us 50.000 de pérdida anual por club, esta suma a grandes rasgos alcanza para pagar gran parte de una planilla de jugadores de buen nivel.

CAPITULO IV

CONCLUSIONES

4. Conclusiones

El presente capítulo es un agregado de conclusiones obtenidas mediante los análisis realizados en los capítulos III, IV y V. Estas conclusiones son netamente estadísticas y no incluyen ningún tipo de juicio de valor, las mismas se describen a continuación:

- La cantidad de conglomerados o categorías de equipos (3) está muy bien definida y presenta un gran porcentaje de la distribución de la varianza, por lo tanto es lógico decir que los equipos de la Liga de Fútbol Profesional Boliviano se pueden clasificar en tres categorías.

- Como conclusión, se puede decir que las tres variables introducidas de antemano en el modelo son significantes por lo cual se deduce que tienen un efecto significativo en los resultados del fútbol boliviano y que la interacción de la localía con la altura también lo es, también se denota que 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 lo cual quiere decir que la altura afecta en el resultado del fútbol de la misma manera si el local es un equipo de la categoría “1” o de la categoría “3”. En cuanto a la no significación de la interacción de la categoría de los equipos y localía indica que el efectos de la localía sobre el resultado no se “refuerza” ni “debilita” con la categoría de los equipos, es decir que su efecto es el mismo sin

importar si el equipo es de la categoría “1” o la “3”. La interacción de las tres variables tampoco es significativa.

- En cuanto a la categoría de los equipos se aprecia que se atribuye la variabilidad del resultado en un 66% a la categoría o conglomerado de los equipos, es la variable más “potente” del modelo.
- La condición de localía determina la variabilidad del resultado en un 4%.
- La altura se atribuye un 2% de variabilidad al resultado pero esta se ve reforzada con la interacción localía-altura a la cual se atribuye un 3% de la variabilidad del resultado.
- Se aprecia que la suma del efecto de la altura por sí sola mas la interacción localía-altura es de alrededor 5%; esto hablando en términos futbolísticos significa que cuando un equipo del llano sube a la altura del altiplano a jugar un partido tiene la desventaja equivalente a jugar con un jugador menos durante 45 minutos, esto bajo el supuesto que todos los jugadores tienen la misma valía.
- Se aprecia claramente que la altura tiene un efecto mucho mayor, es decir mas del doble, que el efecto que tiene el ser local. Volviendo a hablar en términos

futbolísticos se podría decir que un equipo local de la altura del altiplano es “doble local” jugando contra un equipo del llano.

- La altura de los valles no representa efectos significativos sobre los resultados futbolísticos.
- En cuanto a la categoría de los equipos el resultado está en un 64% definido por la categoría “1” de los equipos y solo en un 4% para la categoría “2”, es decir que existe una mayor variabilidad en los resultados cuando un equipo históricamente “bueno” juega.
- La localía tiene un efecto de alrededor del 4% en la variabilidad de los resultados, si bien este porcentaje “cambia la historia” de un partido no es mas alto que el 5% que representa la altura mas la interacción localía-altura.
- Para poder dar una cuantificación porcentual a las variables analizadas y sus respectivas interacciones se debe hacer una “transformación” del 0.81 que suman las mismas al 100%. Esto se realiza haciendo una simple división del estadístico eta obtenido por cada variable entre la suma total de las tres variables mas la interacción de la localía-altura. El resultado es el siguiente:

- Se aprecia que la variable de categoría de los equipos o conglomerado tiene un efecto del 86% en el resultado, la localía del 6%, la altura más su interacción con la localía un 8%.
- Analizando estos resultados se denota que una variable con gran “fuerza” en el análisis es la categoría de los equipos, esta presenta un 88% de efecto en los resultados del fútbol lo cual es totalmente lógico hablando en términos futbolísticos ya que la principal ventaja que un equipo debe tener es su categoría.
- En cuanto a la condición de localía, esta presenta un porcentaje del 6% en los resultados del fútbol, esta es una ventaja deportiva que se toma en cuenta en la organización de los campeonatos ya que los campeonatos se realizan de “ida y de vuelta”, es decir que cada equipo juega igual número de partidos como local y como visitante.
- La altura presenta un porcentaje del 4% en su efecto por sí sola pero se refuerza con la localía un 4% más, lo cual refleja una ventaja extra-deportiva del 8% para los equipos de la altura. Esta ventaja no se contempla en la organización del campeonato lo cual repercute en los resultados finales del mismo.

- 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; mas aún cuando el equipo que sube a jugar es uno de categoría “2” o “3”.
- *Para poder determinar la ventaja extradeportiva de la altura se procedió bajo el supuesto de que todos los equipos tienen la misma categoría. El cálculo muestra los siguientes resultados:*
- Haciendo un análisis futbolístico a estos datos, se puede deducir que un equipo local del llano jugando contra un equipo visitante sea este de la altura o del llano tiene un 46% de ventaja sobre el equipo visitante por concepto de la localía, ya que la altura de “bajada” no tiene efectos (ver estimación de los parámetros). En cambio si un equipo de la altura del altiplano juega contra un equipo de similar categoría que sube desde el llano tiene un 70.8% de ventaja sobre el visitante.
- Traduciendo la ventaja que obtienen los equipos locales de las distintas regiones se puede deducir las probabilidades de cada resultado para un local del llano y para un local de la altura del altiplano.

LOCAL LLANO

Triunfo	64%
Empate	21%
Derrota	15%

LOCAL ALTURA "3" VS EQUIPO DEL LLANO

Triunfo	80%
Empate	16%
Derrota	4%

- Se aprecia en la estimación de los parámetros y las comparaciones por pares que el bajar de un equipo del altiplano o los valles no le presenta ninguna desventaja en cuanto a la altura, también mediante el eta parcial obtenido en la misma tabla se aprecia que el efecto negativo que tiene un equipo del llano al subir a la altura de los valles no es significativo en los resultados del fútbol.
- La altura tiene un efecto económico en los equipos del llano ya que la lógica indica que los equipos de la categoría “1” deberían tener similar número de participaciones en campeonatos internacionales, pero estas se ven disminuidas a favor de los equipos de la altura.
- Las pérdidas económicas para la región del llano en los 25 años de competencia analizados suman alrededor de \$us 2.500.000 los cuales repartidos entre los dos equipos categoría “1” de la misma y divididos en partes iguales durante los años de competición muestran una pérdida anual de \$us 50.000.