ARTÍCULO ESPECIAL / *SPECIAL ARTICLE*

Fundamentos para la aplicación de Bioestadística en Odontología (Parte 2)

Fundamentals for the Application of Biostatistics in Dentistry (Part 2)

Clarisse Virginia Díaz Reissener[1] Gustavo Ignacio Rivas Martínez[2]

RESUMEN

El conocimiento de la estadística descriptiva es esencial en toda disciplina científica. El ámbito odontológico no escapa a esta realidad. Sin embargo, se cuenta con estudios que demuestran su baja comprensión y uso por parte de odontólogos. Dada la utilidad de la estadística, en este artículo como continuación de la primera parte, se muestran procedimientos referentes al análisis de variables cuantitativas, como ser las medidas de centralización: media, mediana, moda;de dispersión: varianza, la desviación típica,rango;de posición: percentiles, cuartiles, deciles y de forma: coeficiente de simetría y curtosis. Así mismo, se presentan gráficos como ser: el histograma, polígono de frecuencia y el diagrama de cajas y bigotes. Se busca brindar herramientas simples y útiles para saber elegir la medida de resumen que representa de la mejor forma a un conjunto de datos a fin de evitar sesgos en la interpretación que deriven en conclusiones o ideas erróneas.

Palabras clave: Educación Continua en Odontología–Estadística–Técnicas

ABSTRACT

The knowledge of descriptive statistics is essential in any scientific discipline. The dental field does not escape this reality. However, there are studies that demonstrate their understanding and low use by dentists. Given the usefulness of the statistics in this article as a continuation of the first part, shown procedures relating to the analysis of quantitative variables, such as centralization measures: mean, median, mode; dispersion: variance, standard deviation, range; position: percentiles, quartiles, deciles and shape coefficient of symmetry and kurtosis. Also, are presented graphics such as Histogram, frequency polygon and box and whisker diagram. It seeks to provide simple and useful tools to know how to choose the summary measure that represents the best way to a set of data in order to avoid bias in interpretation that result in conclusions or misconceptions.

Keywords:*Education Dental Continuing– Statistics– Techniques*

INTRODUCCION

La estadística descriptiva permiteestudiar variablesaleatoriasa través de medidas de resumen para un conjunto, normalmente grande, de datos. En este artículo se muestran procedimientos referentes al análisis de variables cuantitativas, como ser las medidas de centralización, dispersión, posición y forma. Es de vital importancia saber elegir la medida de resumen que representa de la mejor forma al conjunto de datos,a fin de evitar sesgos en la interpretación que deriven en conclusiones o ideas erróneas. En muchos casos, se tiene la creencia de que el promedio del conjunto de datos es una buena medida de resumen del conjunto, pero no siempre representa de manera adecuada ydebe ser reportado acompañadode otras medidas como ser la desviación típica. Por estas razones es fundamental comprendercorrectamente los conceptos ycálculos relacionados, tanto para su utilización como para su interpretación.

La desviación típica nos indica que tan cercao lejosse encuentran los valores individuales con respecto al promedio. Por ejemplo, si sabemos que el promedio de caries en grupo de niños de 12 años es de 6,¿Se podría esperar que la mayoría tenga 6 dientes cariados?¿Se podría encontrar un niño con

Recibido el 22 de junio de 2015, aceptado para publicación el 17 de agosto de 2015

[1] Dirección de Editorial y Estadística. Facultad de Odontología. Universidad Nacional de Asunción.
[2] Departamento de Ciencias Básicas. Facultad Politécnica. Universidad Nacional de Asunción.

15 caries en ese grupo?.Para responder a estas preguntas es importante conocer la desviación típica y el modo en el cual se distribuyen los datos. Esto nos indicará como se encuentran los datos alrededor de este valor de centralización, pues podría darse el caso de queel número de caries sea de 5, 6 o 7 en todos los niños que conforman la población de estudio, entonces la desviación típica tendrá un valor pequeño existiendo así poca variabilidad en cuanto al nivel de caries en los niños de la población estudiada. Mientras que si los valores se encuentran entre 0 y 20 caries existe una mayor variabilidad, siendo este un valor más elevado. Esto nos revela la importancia de comprender y presentar la desviación típica junto con el promedio.

Otro medida de centralización muy conocida es la mediana.En ocasiones, podrá resultar mejor representante del conjunto de datos que la media, como se muestra más adelante. En otros casos, la medidas de posición como los percentiles aportan también resultados útiles, pues saber que el 25% de los niños tiene 3 dientes cariados o menoso que un solo un 5% tenía más de 13 dientes cariados (siendo este no típico en la muestra estudiada) puede ser relevante según el objetivo del estudio.

Aunque el conocimiento de la estadística descriptiva se considera básico e indispensable actualmente, por suutilidad para determinar necesidades de la población, diseñar y evaluar programas de salud entre otras cosas.En un estudio realizado en Dinamarca por Scheutz et al, que contenía 9 preguntas sobre conceptos básicos de estadística en odontólogos y estudiantes, obtuvieron un bajo promedio de respuestas correctas, específicamente de 2,2 y 3,4 respectivamente.Esto nos da una idea de lo mucho que falta por hacer con respecto al conocimiento cabal de esta disciplina tan importante y muchas veces subvalorada.

A continuación se desarrollará la segunda parte de estadística descriptiva que consiste en presentar las medidas y los gráficos comúnmente utilizados, para variables cuantitativas para datos simples y agrupados, aunque para este último no se desarrolla el procedimiento en su totalidad debido quelas fórmulas son muy similares en ambos, entonces nos centraremos en la interpretación.

1. MEDIDAS DE CENTRALIZACIÓN

Para conocer la distribución de las variables cuantitativas se emplean principalmente las medidas de tendencia central y dispersión. Las medidas de tendencia central comúnmente utilizadas son la media aritmética, mediana y moda; las medidas de dispersión: la varianza, desviación típica, el coeficiente de variación y el rango. A continuación se explica cómo se calcula e interpreta cada una de ellas. Se debe tener en cuenta que estas medidas se calculan para la población (parámetros) y para la muestra (estadísticos), aquí presentamos fórmulas para muestras de tamaño *n*.

1.1 Media

Generalmente, a la media aritmética se la denomina media sin calificativos o promedio, pero se debe tener en cuenta que también hay otros tipo de medias como la geométrica, armónica y cuadrática, entre otras, pero escapan al objetivo de este artículo por lo que no serán tratadas. Se emplea el símbolo para representar la media aritmética poblacional y para representar la media aritmética muestral. Esta última se calcula promediando todos los datos de un conjunto de datos, se expresa mediante la siguiente fórmula:

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n},$$

donde x_i indica *i – ésimo* el dato observado y es el tamaño muestral, siendo por tanto, la sumatoria de cada dato divido el total de la muestra.

1.2 Mediana

La mediana divide a un conjunto de datos ordenados en dos partes iguales, si *n* es impar:

$$M = \frac{X_{(n+1)}}{2},$$

y si *n* es par la mediana es la media aritmética de los dos valores centrales:

$$M = \frac{X_{n/2}+X_{n/2+1}}{2}.$$

1.3 Moda

La moda (M_o) es el valor que más de repite en la serie.Si existe un solo valor que es el que más se repite se dice que el conjunto es unimodal, en caso que haya dos valores es bimodal y así sucesivamente. En esos casos tomar la mayor de todas las modas podría ser de utilidad para describir un conjunto de datos o el promedio de las modas.

2. MEDIDAS DE DISPERSIÓN

2.1 Varianza

La varianza poblacional se representa por o^2 y la varianza muestral por S^2, siendo la expresión matemática de esta última:

$$S^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1},$$

donde x es la media muestral, x_i es la i – *ésima* observación y n es el tamaño muestral.

2.2 Desviación típica

La desviación típica muestral es la raíz cuadrada de la varianza muestral

$$S = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}}.$$

Habitualmente, para la comparar dispersión en variables en unidades de medida diferentes,o muestras diferentes, se utiliza una combinación entre la desviación típica y la media, el coeficiente de variación: $\mathrm{CV} = \frac{S}{\bar{x}} \cdot 100,$ el cual mide el grado de dispersión en ambas muestras respecto al promedio y es adimensional lo cual facilita la comparabilidad. Valores mayores al 30% indican datos dispersos.

2.3 Rango

Otra medida de dispersión es el rango, que es la diferencia entre el valor máximo y el mínimo en un conjunto de datos.

$$R = x_{máx} - x_{min},$$

donde, $x_{máx}$ es el mayor dato observado y x_{min} es el mínimo. Es el intervalo entre el valor máximo y el valor mínimo. Permite obtener una idea de la dispersión de los datos, cuanto mayor es el rango, más dispersos están los datos de un conjunto.

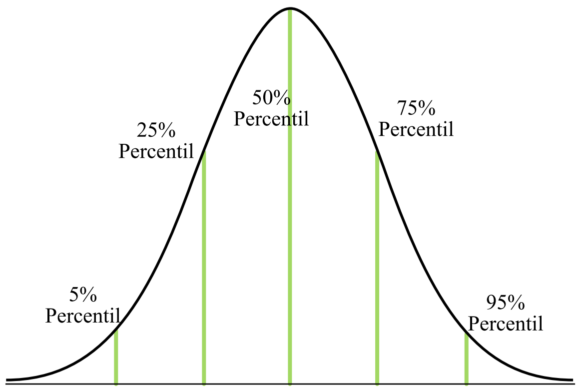
3. MEDIDAS DE POSICIÓN

Son útiles para conocer el porcentaje acumulado para una serie ordenada ascendente, es decir, divide la distribución en partes iguales como ser: percentiles (100 partes iguales), deciles (10 partes iguales), quintiles (5 partes iguales) y cuartiles (4 partes iguales).

3.1 Percentil

El percentil indica, una vez ordenados los datos de menor a mayor, el valor de la variable por debajo del cual se encuentra un porcentaje dado de observaciones. Por ejemplo, el percentil 25 es el valor debajo del cual se encuentran el 25% de las observaciones.Se representan con la letra . Para el percentil , donde la toma valores del 1 al 99. El de la muestra son valores menores que él y el restante son mayores. P_{50}, por ejemplo, coincide con la mediana, tal como se observa en la Figura 1.

Figura 1. Percentiles en una distribución normal.



3.2 Deciles

En los deciles la mediana representa P_{50}, ya que contiene al 50% de los datos, mientras que al 20% se le denomina P_{20} y al que contiene el 70% de los datos P_{70}, esto es divido en 10 partes iguales.

3.3 Cuartiles

En algunos casos son más conocidos y utilizados los cuartiles. Estos son los tres valores que dividen al conjunto de datos ordenados en cuatro partes porcentualmente iguales, entonces el Q_1 es el que contiene al 25% de los datos, el Q_2 se corresponde a la mediana conteniendo así al 50% de los datos y el Q_3 al 75% de los datos.Se representa gráficamente como la anchura de las cajas en los llamados diagramas de cajas y lo cual se muestra más adelante. La diferencia entre el tercer cuartil y el primero se conoce como rango intercuartílico (RI). ElRI se obtiene de la diferencia entre el tercer cuartil y el primer cuartil, se recomienda utilizarlo para acompañar a la mediana, ya que elimina los valores extremos, tanto superiores como inferiores.

Los valores extremos y atípicos son aquellos que se diferencian de los demás valores por encontrarse muy alejados con respecto a la mediana, motivo por el cual debería verificarse inicialmente si es un valor genuino y luego decidir si representan realmente al conjunto de datos que se está estudiando, pues estos podrán influir en los resultados.

4.1 Histograma

Los tres gráficos que más se utilizan para variables cuantitativas, en estudios descriptivos, son el histograma de frecuencias, el polígono de frecuencias y el diagrama de cajas y bigotes; constituyendo este último el grafico que aporta mayor información para dicho tipo de variables. Sin embargo, en la sección 3.2 del estudio de López-Soto et alse presenta un análisis muy interesante mediante el uso del histograma.

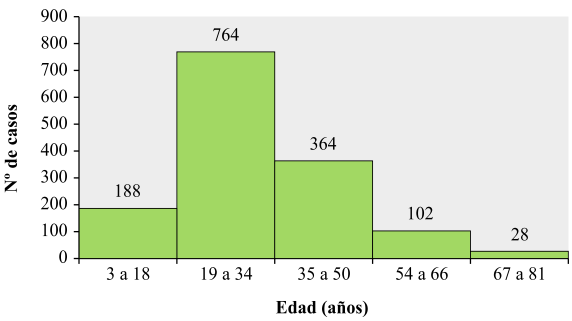


Figura 2. Distribución de pacientes con traumatismo maxilofacial por agresión según grupo etario.

El histograma se puede graficar con las frecuencias absolutas o relativas, aunque es poco frecuente esta última. Si bien está conformada también por rectángulos que representan la altura conforme la frecuencia, se diferencia del grafico de barras simples en que las barras se presentan sin espacio entre ellas para dar un aspecto de continuidad y la base de la barra es proporcional a los intervalos de clase.

En la Figura 2 los intervalos de edades son iguales, de 15 años en cada grupo de edad. Se observa que el mayor número de pacientes atendidos con traumatismo maxilofacial lo constituyeron los adultos jóvenes del grupo etario comprendido entre 19 y 34 años.

4.1 Polígono de Frecuencias

El polígono de frecuencias se construye uniendo en una línea recta el punto medio (marca de clase) de las bases superiores del histograma de frecuencias. El polígono comienza con una hipotética clase anterior a la primera clase de igual longitud de intervalo y termina trazando una línea que va desde el punto medio de la base superior del rectángulo de la última clase con una hipotética clase posterior a la última de igual longitud de intervalo. Como se observa en la Figura3,inicia y finaliza el polígono en cero casos, con un pico a la edad de 19 a 34 años de edad. Este puede presentarse con frecuencia absoluta o porcentaje.

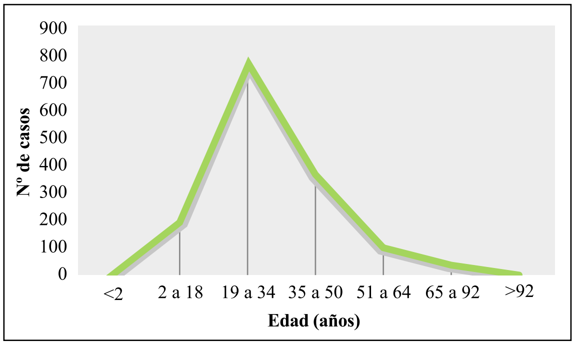


Figura 3. Pacientes con traumatismo maxilofacial por agresión según grupo etario.

4.3 Cajas y bigotes

Comprender los conceptos de cuartiles, nos permitirá entender el gráfico de cajas y bigotes de la Figura 4. Se puede ver que ambos lados son

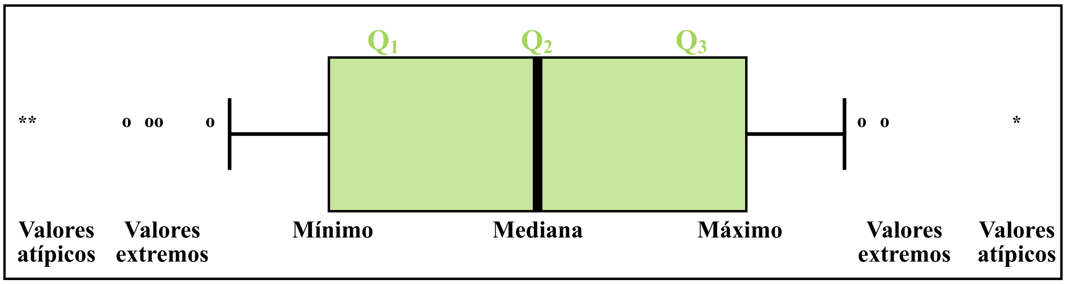


Figura 4. Gráfico de cajas y bigotes.

iguales, esto sería lo ideal para una distribución simétrica, por eso justamente es útil para observar cómo se distribuyen los datos. En algún que otro software, este diagrama se representa verticalmente.

MEDIDAS DE FORMA

5.1 Asimetría

En distribuciones unimodales, la asimetría es una medida que nos permitirá saber hacia dónde se agrupan los valores, pues lo que esperaríamos es que la mayor parte de los valores se encuentren en el centro y en los extremos se encuentre valores en menor cantidad, es por esto que si hay más valores agrupados a la izquierda se denomina asimetría positiva, mientras que si están agrupados a la derecha la asimetría será negativa, tal como se observa en la Figura 5.

Existen distintas formas para medir el nivel de asimetría, aquí presentamos el coeficiente de asimetría de Carl Pearson.

$$As = \frac{3(\bar{x}-M)}{S},$$

Donde x, es la media muestral, M es la mediana y S la desviación típica.El Coeficiente de Pearson varía entre -3 y 3.

Si As < 0, la distribución será asimétrica negativa.

Si As = 0, la distribución será simétrica.

Si As > 0, la distribución será asimétrica positiva.

5.2 Curtosis

Otra medida de forma es lacurtosis que nos indicará si la distribución tiende a ser uniforme, es decir, la frecuencia de valores es similar para cada categoría, entonces tiene una tendencia plana, denominándose por esto negativa; también se encuentra el otro caso en que es muy puntiaguda, siendo así positiva, acumulándose la mayor frecuencia en el centro de la distribución. Existen varias formas de calcular la curtosis, no obstante, está más extendida la siguiente definición del coeficiente de curtosis:

$$g_2 = \frac{\frac{1}{n} * \sum_{i=1}^{n}(x_i-\bar{x})^4}{\left(\frac{1}{n} * \sum_{i=1}^{n}(x_i-\bar{x})^2\right)^4} - 3,$$

donde el valor 3 es la curtosis de la Normal, con objeto de generar un coeficiente que valga 0 para la Normal y tome a ésta como referencia de apuntamiento. Tomando, pues, la distribución normal como referencia, una distribución puede ser, como se observa en la Figura 6:

- más apuntada y con colas menos anchas que la normal: leptocúrtica.

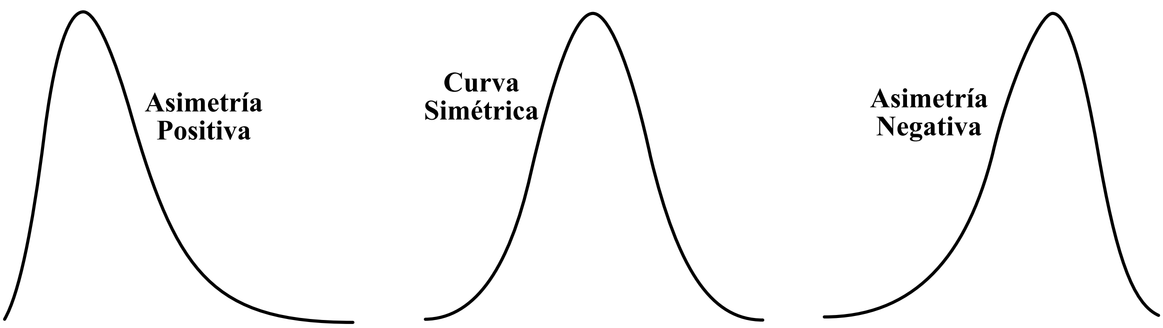


Figura 5. Tipos de asimetría.

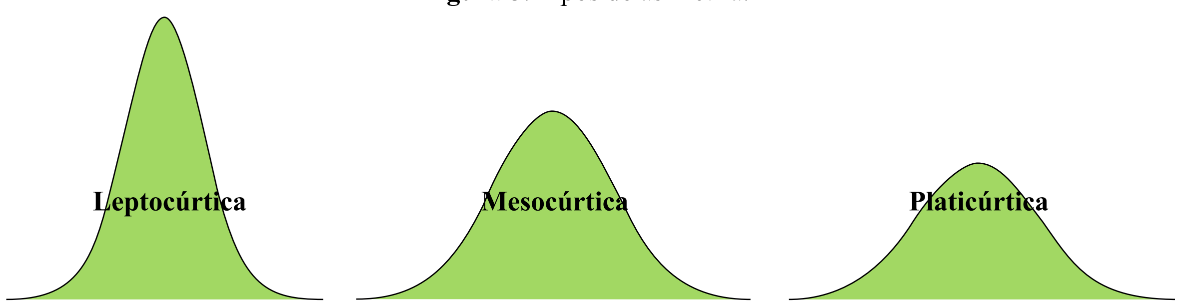


Figura 6. Tipo de curtosis.

- menos apuntada y con colas más anchas que la normal: platicúrtica.
- la distribución normal es:mesocúrtica.

CONSIDERACIONES FINALES

Las medidas de tendencia central nos informan sobre el valor central del conjunto de datos, mientras que las medidas de dispersión nos indican como varían los datos con respecto al valor central.En los estudio de Aranzaru-Moya et al y Ramírez-Puerta et al se aplican estos conceptos y sus interpretaciones.

Se debe tener en cuenta que si los datos no representan una muestra aleatoria probabilística, entonces constituyen una población en sí misma y los resultados deben expresarse con el valor de la media pero sin el error estándar de la media o la desviación típica, debido a que es un número sin ningún error aleatorioy las inferencias solo son válidas para el conjunto de datos analizado.Un trabajo sobre una muestra no probabilística puede verse en Carreño et al.

La media es única, su comprensión resulta fácil pero se debe considerar que los valores extremos pueden afectar su estimación, por lo que se recomienda utilizarla en distribuciones simétricas.

La mediana es única y los valores extremos no le afectan como a la media, pues divide a los datos en dos partes iguales dejando a cada lado 50%, es útil cuando no se tiene una distribución simétrica.

Se debe tener en cuenta que la modaes más recomendable utilizarla para variables cualitativas, aunque suele ser pocas veces reportada debido ala poca información que brinda dada su restringida interpretación.

Es muy importante conocer el tipo de variables con la que estamos trabajando, pues para las variables de intervalo es adecuado utilizar la media y desviación típica, mientras que para variables ordinales son la mediana y la distancia intercuartil, en tanto que, para variables nominales es la moda. Otra medida muy utilizada para variables nominales es el porcentaje, aplicaciones pueden verse en Builes et al y Jiménez Malagón et al, entre otros.

BIBLIOGRAFIA CONSULTADA

1. *Álvarez Cáceres R. Estadística aplicada a las ciencias de la salud. Díaz de Santos; 2007.*
2. *Aranzaru-Moya GC, Delgado-Jaimes RY, Gutiérrez MPP. Variaciones de riesgo en valores de tensión arterial en pacientes hipertensos durante procedimientos odontológicos. Rev Salud Uis. 2014;46(2):137-45.*
3. *Bravo Pérez M. Bioestadística. Principios, métodos y aplicaciones en odontología. En: Cuenca Sala E, Baca García P, editores. Odontología preventiva y comunitaria Principios, métodos y aplicaciones. 3ª ed. Barcelona: ElsevierMasson; 2005.*
4. *Builes AMV, Pino NM, Saldarriaga AFS, Galvis MM, Colorado KJC, Zuluaga YPM, et al. Caries dental y necesidades de tratamiento en el primer molar permanente en escolares de 12 años de las escuelas públicas del municipio de Rionegro (Antioquia, Colombia), 2010.UnivOdontol. 2012 Ene-Jun [citado 26 de noviembre de 2015];31(66):25-32. Disponible en: http://revistas.javeriana.edu.co/index.php/revUnivOdontologica/article/view/2708*
5. *Canavos G. Probabilidad y Estadística. Aplicaciones y métodos. México: McGraw-Hill; 1988.*
6. *Carreño MAA, Triana AB, Pabón MEC, Santamaría CEF, Ciódaro AR, Vivas MCH. Signos y síntomas de trastornos temporomandibulares en niños entre los 6 y los 13 años de edad. Serie de 50 casos.UnivOdontol. 2013;32(69):161-168*
7. *Celis de la Rosa A de J. Bioestadística. 2da ed. México: Manual Moderno; 2008.*
8. *Cobo E, Muñoz P, González JA. Bioestadística para no estadísticos. Base para interpretar artículos científicos. Barcelona: ElsevierMasson; 2007.*
9. *Hernández Sampieri, R., Fernández Collado, C., Baptista Lucio, P. Metodología de la Investigación. 5ta ed. México DF: McGraw-Hill; 2010.*
10. *López-Soto OP, Cerezo-Correa M del P, Paz-Delgado AL. Variables relacionadas con la satisfacción del paciente de los servicios odontológicos. Gerenc Políticas Salud. 2010 Ene-Jun [citado 26 de noviembre de 2015];9(18):124-36. Disponible en: http://revistas.javeriana.edu.co/index.php/gerepolsal/article/view/2639*
11. *Jara Ibarrola TS. Prevalencia de traumatismos maxilofaciales producidos por agresión, en pacientes que acudieron al Centro de Emergencias Médicas "Manuel Luis Giagni", en el periodo de enero a diciembre de 2011. Asunción: Universidad Autónoma del Paraguay "Pierre Fauchard"; 2013.*
12. *Jiménez Malagón M del C, Fang L, Diaz Caballero A. Perfil epidemiológico oral y necesidad de tratamiento odontológico de pacientes VIH/SIDA. Rev Clínica Med Fam. 2012;5(2):97-103.*
13. *Ramírez-Puerta BS, Agudelo-Suárez AA, Morales-Flórez JL, Sarrazola-Moncada ÁM. Dientes presentes en población de 25, 35, 45, 55 y 65 años, Antioquia (Colombia) 2011. 25(2):12-23.*
14. *Scheutz F, Andersen B, Wulff HR. What do dentists know about statistics?Eur J Oral Sci. 1988;96(4):281-7.*