

---

---

**EL PODER PREDICTIVO DE LAS EVALUACIONES**

*Ignacio Redondo Bellón y Carlos Lamas Alonso*

---

---

**13º Seminario AEDEMO sobre Audiencia de Televisión**

La Coruña, Febrero de 1997

**Título:** El poder predictivo de las evaluaciones.

**Autores:**

- Carlos Lamas Alonso. Director Adjunto de AIMC.
- Ignacio Redondo Bellón. Profesor Asociado del Dpto. de Investigación Comercial de la Universidad Autónoma de Madrid.

**Resumen**

La evaluación de planes y la reconstrucción de campañas a partir de la audimetría está sujeta a errores que dependen de varios factores: aquí han sido analizados con un doble enfoque, teórico y práctico, atendiendo tanto a las hipótesis de partida como a la precisión de las herramientas disponibles en el mercado español. Un estudio previo compara los *ratings* del minuto publicitario y del cuarto de hora correspondiente, durante los 66.083 anuncios emitidos en abril de 1995: así se han cuantificado tanto las sobrestimaciones derivadas de la medición en cuartos de hora, como la permanencia ante la publicidad, en función de las franjas horarias y de las cadenas. El segundo estudio mide la estabilidad de los soportes semana a semana, a partir de 6.188 *ratings* de cuartos de hora, entre febrero y abril de 1995: se han determinado entonces las franjas y canales que, con *ratings* más estables, se adaptaron mejor a las hipótesis implícitas en las evaluaciones. En la postrera y principal investigación hemos cotejado, sobre 112 campañas emitidas en abril de 1995, las exposiciones recontadas por Telescope y las evaluaciones de Marksel, estimadas sobre distintos periodos de referencia, para diversos *targets* y con ficheros fusionados o concatenados: así se ha estudiado la influencia de estos factores y del número de inserciones en la precisión de los *GRP's*, la cobertura, la frecuencia media y la cobertura efectiva.

**Sumario**

1. Introducción
  2. La evaluación de planes con Marksel
  3. La reconstrucción de campañas con Telescope
  4. Investigaciones realizadas
    - a) ¿Cómo difieren los ratings del minuto publicitario y del cuarto de hora?
    - b) ¿Cómo evoluciona la audiencia de los cuartos semana a semana?
    - c) ¿Cómo se ajustan las evaluaciones de las 112 campañas a la realidad?
      - Selección del periodo de referencia
      - Segmentación en targets
      - Fusión de ficheros
      - Número de inserciones
  5. Conclusiones
- Apéndice  
Anexo

## **1. Introducción**

En España, la inversión publicitaria<sup>1</sup> en TV rondaba los 220.000 millones de pesetas durante 1995. Para planificar cuándo y cómo distribuir ese importante presupuesto, las agencias y centrales usan generalmente los programas Marksel y Telescope, sobre la información proporcionada por los audímetros de la empresa Sofres A.M.

El *software* de Marksel permite evaluar y comparar planes alternativos antes de la emisión de la campaña. Telescope, por su parte, hace posible reconstruir la campaña *a posteriori*, recontando los contactos alcanzados. El primero se utiliza para predecir las exposiciones de una campaña mientras que el segundo, mediante el módulo de "Reconstrucción de campañas", se usa para verificar las exposiciones emitidas.

El objetivo básico de esta ponencia es medir la precisión de las predicciones de Marksel. Su cuantificación y análisis tiene un interés tanto empresarial como académico, pero es en el sector publicitario donde existen más partes afectadas: anunciantes, que demandan la eficacia de sus inversiones; planificadores, que deben seleccionar canales y franjas horarias según objetivos y restricciones; además debe añadirse el interés general de contar con herramientas suficientemente precisas, para dar mayor transparencia al sector y para recuperar el razonable protagonismo de la planificación.

Esta investigación se ha realizado conjuntamente desde el sector publicitario y la universidad. Los coautores pertenecen a la Asociación para la Investigación de Medios de Comunicación (AIMC) y al Departamento de Investigación Comercial de la Universidad Autónoma de Madrid. Esas entidades colaboran, según expresa el convenio firmado en marzo de 1995, para fomentar la investigación de los medios y de los programas de planificación, y su difusión a todos los niveles.

## **2. La evaluación de planes con Marksel**

Los programas Marksel tienen básicamente tres utilidades: [1] obtener *rankings* por canales y franjas horarias, según la cobertura, la rentabilidad y la afinidad, en diferentes segmentos; [2] evaluar la cobertura, los *GRP*'s y los demás indicadores de cualquier plan, que luego pueda constituir una campaña publicitaria; y [3] aproximarse a una combinación óptima de inserciones en función de la eficacia, del coste y de las restricciones impuestas por el planificador. Aquí sólo vamos a tratar del segundo punto, el cual constituye la pieza clave del *software*.

El fichero base de Marksel es semanal y contiene, para cada individuo, la probabilidad de exposición a cada soporte (que se define por una cadena, un día de la semana y un cuarto de hora) así como el factor de elevación correspondiente. Sólo incluye los panelistas que formaron parte de la muestra útil durante todos los días de la semana<sup>2</sup>. Esta "muestra semanal común" o "muestra intersección" ha pasado la rutina

---

<sup>1</sup> INFOADEX. "Estudio INFOADEX de la inversión publicitaria en España 1996". Madrid. 1996.

<sup>2</sup> MARKSEL. "Manual del usuario". Octubre de 1993.

especial de equilibrio que, usando las mismas condiciones del equilibrio estándar diario, calcula congruentemente los factores de elevación usados por Marksel. La probabilidad de exposición al soporte se halla, para cada individuo, mediante el cociente entre los minutos realmente vistos y los 15 minutos que podía haber visto.

Cuando el usuario toma varias semanas de referencia, debe optar por fusionar o concatenar los ficheros, dilema que tiene distintas implicaciones:

**1. Fusionar los ficheros** produce tres efectos: [a] se agregan todos los individuos, aunque no sean muestra útil de alguna semana; [b] la ponderación individual resulta de promediar las ponderaciones semanales de ese individuo (que valen 0 si no estuvo alguna semana); y [c] la probabilidad individual resulta también de promediar las probabilidades semanales (tomando sólo las semanas donde ha participado en la muestra útil). Por ejemplo, supongamos dos individuos que, en las últimas cuatro semanas y de 9.00 a 9.15 p.m., han visto un canal tantos minutos como aparecen en la **Tabla 1**; como se observa en la última columna, las probabilidades resultantes serían iguales.

**Tabla 1**

	minutos vistos de 9.00 a 9.15 p.m. en ...				Probabilidad de exposición
	1ª semana	2ª semana	3ª semana	4ª semana	
Individuo 1	0	0	12	0	$\frac{0+0+12+0}{15*4} = 0'2$
Individuo 2	3	3	3	3	$\frac{3+3+3+3}{15*4} = 0'2$

La probabilidad de exposición resultante es un coeficiente promedio que está asimilando dos conductas televisivas muy distintas: el primer espectador del soporte es esporádico y perseverante, mientras que el segundo es sistemático y cambiante.

**2. Concatenar los ficheros** supone adicionarlos sin ningún tipo de consolidación para un mismo individuo del panel. No se excluyen aquellos individuos que no han sido muestra útil durante alguna semana y la muestra no es de individuos sino que la forman las observaciones individuo/semana donde, como es obvio, un mismo panelista puede aportar más de un dato semanal. El tamaño de la muestra aumenta geométricamente con el número de semanas de referencia. El factor de elevación asociado a cada observación se halla simplemente dividiendo, el factor de elevación del propio individuo en el fichero base semanal, entre el número de semanas a concatenar.

El programa Marksel se clasifica dentro de los modelos de individuación que, a diferencia de los modelos de agregación, estima las exposiciones individuo a individuo<sup>3</sup>.

<sup>3</sup> REDONDO BELLÓN, Ignacio. "La cobertura en Revistas y Suplementos: cómo lograr mayor cobertura entre los planificadores". Ponencia presentada en el VIII Encuentro de Profesores Universitarios de Marketing. Zaragoza, 26-27 de septiembre de 1996. Libro de ponencias publicado por ESIC, págs. 417-427.

Emplea un algoritmo<sup>4</sup> cuya hipótesis fundamental es la homogeneidad de soportes<sup>5</sup>, promediando las probabilidades de exposición a varios soportes, como si se redujera a un teórico soporte medio. Siendo  $p_{ij}$  la probabilidad de que el individuo  $i$  contacte con el soporte  $j$ , y  $r_j$  el nº de inserciones del soporte  $j$ , la probabilidad media  $\bar{p}_i$  se calcula:

$$\bar{p}_i = \frac{\sum_{j=1}^m r'_j * p_{ij}}{\sum_{j=1}^m r'_j} \quad \begin{array}{l} r'_j = r_j \quad \text{si } p_{ij} \neq 0 \\ r'_j = 0 \quad \text{si } p_{ij} = 0 \end{array}$$

Se está suponiendo que la exposición al soporte medio en un momento es independiente de su exposición en otro momento. Los contactos se distribuyen según una distribución binomial; siendo  $N$  el número total de inserciones de la campaña, la probabilidad de que el individuo  $i$  esté expuesto exactamente a  $k$  inserciones vale:

$$P_i(x=k/ N, \bar{p}_i) = \binom{N}{x} \bar{p}_i^x * (1-\bar{p}_i)^{N-x}$$

Entonces, tal individuo tiene esta probabilidad de contactar al menos con una de las inserciones (probabilidad de ser alcanzado por la campaña):

$$1 - P_i(x=0/ N, \bar{p}_i) = 1 - (1-\bar{p}_i)^N$$

Tras determinar la distribución de contactos del individuo muestral (por asimilación, de la parte del universo que representa) el programa agrega los datos de todos los individuos muestrales que componen el *target* seleccionado. Entonces el *target* recibe unos contactos que se distribuyen a partir de la media de las probabilidades individuales  $P_i$ , ponderadas por el factor de elevación individual  $W_i$ :

$$\frac{\sum_i W_i * P_i(x=k/ N, \bar{p}_i)}{\sum_i W_i}$$

Observamos que este proceso contiene básicamente dos deficiencias teóricas:

1) Supone que los soportes son homogéneos, luego las probabilidades de exposición a los soportes se asimilan a la probabilidad de exposición a un teórico soporte medio. Los contactos siguen una distribución binomial a partir de esa probabilidad media: dicha distribución es menos rigurosa que la multinomial, aunque algunas comprobaciones empíricas<sup>6</sup> muestran que esta simplificación no produce errores significativos.

<sup>4</sup> LAMAS ALONSO, Carlos. "Eso de las evaluaciones no lo veo claro". *Investigación y Marketing*. Asociación Española de Estudios de Mercado, Marketing y Opinión (AEDEMO). Nº 42. Julio 1993. Págs. 58-62.

<sup>5</sup> El soporte es la unidad de referencia de la audiencia, definida en relación a un canal determinado y a un cuarto de hora de la semana.

<sup>6</sup> LAMAS ALONSO, Carlos. "Eso de las evaluaciones no lo veo claro". Ponencia presentada en el 9º Seminario sobre Audiencia de Televisión (62º Seminario AEDEMO). Lisboa. Febrero de 1993.

2) Se presume la independencia entre las exposiciones a diferentes emisiones. Los efectos de esta presunción, sin estar en absoluto justificada, probablemente estén atenuados porque las sobrestimaciones de unos individuos se pueden compensar con las subestimaciones de otros.

### **3. La reconstrucción de campañas con Telescope**

Los programas Telescope también explotan las audiencias de TV individuo a individuo. Tiene múltiples utilidades<sup>7</sup>: analizar las audiencias por minuto y por cuartos de hora, estudiar la acumulación y la duplicación de la audiencias, observar las migraciones o flujos de audiencia, realizar *rankings* y perfiles, simular campañas según un plan de inserciones y analizar bloques publicitarios. La aportación más genuina, a los efectos que nos ocupa, es la reconstrucción de planes, para recontar exactamente las exposiciones obtenidas en una campaña. Para ello debe utilizar la técnica de muestra única en el periodo, que es la muestra del periodo central considerado.

Se parte de un fichero donde está registrada unívocamente la cadena contactada día a día, individuo a individuo y minuto a minuto, así como el factor de elevación correspondiente a cada individuo y día. Puede entonces medir las audiencias por minuto y seleccionar cualquier *target* a partir de los criterios disponibles. La audiencia de un anuncio, con independencia de su duración, corresponde a la del minuto de su inicio. El criterio es simple y razonable, exceptuando los casos de publirreportajes u otros mensajes publicitarios de gran duración, cuyos *ratings* quizá podrían medirse bien con el promedio de los minutos correspondientes.

La reconstrucción de campañas no se reduce sólo al recuento de las exposiciones recibidas por cada individuo. Telescope también se acomoda a la inestabilidad de la muestra durante la campaña, que se debe a: [1] la rotación entre los panelistas, ya que algunos se dan de baja y otros de alta durante ese periodo; y [2] la merma muestral diaria, consecuencia de averías, bajas temporales, fallos de comunicación, etc. Luego la muestra útil (sobre la que se estiman realmente las audiencias) varía diariamente y, como consecuencia, varía también el coeficiente de elevación de cada panelista.

La merma muestral diaria en el panel de Sofrés AM viene a ser de un 6% (en relación al panel instalado). Y para medir de forma global la estabilidad del un panel durante un conjunto consecutivo de días, AIMC sugiere<sup>8</sup> el siguiente índice:

$$I = 100 \frac{\sum_{i=1}^m (n_i - 1)}{m(d - 1)} \quad \text{siendo}$$

*m*: La muestra unión en el intervalo de días considerado (individuos que han sido útiles al menos un día).

<sup>7</sup> SOFRES, A.M. "Telescope. Fichas de análisis". Septiembre de 1993.

<sup>8</sup> AIMC. Informe Anual sobre el Control de la Audimetría, 1995.

$n_i$ : Número de días que el individuo muestral “ $i$ ” ha sido útil en el período analizado.

$d$ : Longitud en días del período considerado.

El índice I así definido toma valores comprendidos entre cero y cien. El valor cero corresponde a la hipotética situación de que la muestra del panel fuera completamente distinta cada día (ausencia total de panelización muestral) y el de cien se obtendría como consecuencia de tener una muestra útil totalmente constante durante todos los días del período (máxima estabilidad). Empíricamente se ha visto que, en el panel de audimetría español, los valores promedio que el índice de estabilidad I toma son de 95 para una semana, 92 para un mes, 88 para un trimestre, 82 para un semestre y 72 para un año.

Actualmente Telescope toma la muestra del día medio de la campaña efectuar las reconstrucciones y aplica a los panelistas los factores de elevación correspondientes a ese día. Esta simplificación implica, entre otras consecuencias, que:

- Los individuos que no fueron muestra útil el día medio de la campaña, no se consideran en la reconstrucción, aunque sí hayan sido muestra útil durante algún/os día/s de la campaña. Así quedan fuera, entre otros, todos los panelistas dados de alta después del día medio de la campaña.
- Cuando un panelista de esa muestra útil del día medio no es parte de la muestra útil algún otro día, se le supone elemento activo de la muestra pero inactivo para estar expuesto a la TV. En otras palabras, se le tiene en cuenta en el análisis, pero como si no hubiera visto la TV ese día, sin ser contactado por ningún mensaje publicitario.

La aplicación estricta de este método implica que, para un mismo plan, la cobertura y distribución de contactos del módulo "Reconstrucción de Campañas" no coincidan con los datos oficiales estimados a partir de la muestra útil diaria. Estas diferencias no son en absoluto neutras, sino que muestran una infravaloración sistemática de la cobertura y GRP's reportados por el módulo de Reconstrucción.

El método de la *muestra del día medio* se usa en muchos otros países y sus desviaciones sistemáticas se intentan corregir, en general, ajustando el desfase entre la cifra oficial de los GRP's y la generada por la reconstrucción. A estos efectos, en el Reino Unido se supone que la distribución de contactos sigue una “binomial negativa”<sup>9</sup> y en España se aplica un algoritmo de deformación que traslada la curva manteniendo básicamente la misma forma<sup>10</sup>.

La Reconstrucción de Campañas mide la cobertura, los GRP's, etc. en un grupo de individuos, pero obviamente esos indicadores no son reales porque los datos no corresponden a todos los individuos sino a una pequeña muestra. Debe controlarse que su tamaño resulte suficientemente representativo y, por tanto, que los errores de muestreo se mantengan en niveles admisibles. Para este fin, AIMC está distribuyendo la

<sup>9</sup> WILCOX, Steven. “The UK Reach and Frequency Model”. Worldwide Broadcast Audience Research Symposium. Toronto, 1992.

<sup>10</sup> PEÑA, Rafael. "Deformación de la función de distribución de contactos". Comité de Usuarios de Sofrés.

---

versión 2.0 del programa ESPA, que incluye un módulo que calcula los errores muestrales implícitos en los indicadores medidos.

#### **4. Investigaciones realizadas**

Las diferencias entre las evaluaciones de Marksel y las reconstrucciones de Telescope se deben, junto a las imperfecciones teóricas ya señaladas, a cuatro factores intrínsecos (en contraposición a otros extrínsecos -aquí no estudiados- como la discrepancia entre los pases planificados y los realmente emitidos):

**1. El pasado se usa para predecir el futuro:** para predecir los indicadores de una campaña futura, el planificador toma como referencia el comportamiento pasado de la audiencia. Cuando se emiten los anuncios, casi siempre comprobamos que las predicciones han sido defectuosas porque las audiencias no han permanecido estacionarias entre la planificación y la emisión: ni los espectadores mantienen los mismos hábitos de exposición ni las cadenas retransmiten la misma programación. Esta incertidumbre, habitual en la mayoría de los índices socioeconómicos, se puede reducir con una adecuada elección del periodo de referencia: la decisión afecta a cuántas y qué semanas se deben seleccionar.

**2. Minuto publicitario versus cuarto de hora:** la audiencia, igual que todo fenómeno dinámico, se mide en relación al tiempo: por ejemplo, durante un intervalo de tiempo alguien ha podido ver una o varias cadenas o dejar de verla/s. Cuanto menor sea ese intervalo temporal, menos podrá evolucionar la audiencia. Los cambios en el comportamiento del espectador se pueden suceder en espacios de tiempo muy pequeños. La audimetría permite medir esas variaciones cada segundo, pero la fracción mínima disponible en *software* es de un minuto: ello tiene unas implicaciones que ya han sido certeramente estudiadas<sup>11</sup> y que, por consiguiente, no son objeto de esta investigación. En cambio hemos estudiado cómo influye que Marksel use el cuarto de hora como fracción mínima de tiempo, en lugar del minuto utilizado por Telescope. Con ello Marksel mantiene la hipótesis de estacionariedad en el cuarto ya que la probabilidad individual de exposición es constante durante los 15 minutos: ésta y otras hipótesis restrictivas se utilizan comunmente para simplificar los fenómenos reales en términos probabilísticos. En la primera investigación hemos analizado la *estabilidad intra cuartos*, comparando los *ratings* del minuto publicitario y del cuarto de hora correspondiente, durante los 66.083 anuncios emitidos en abril de 1995.

**3. Cuarto concreto versus promedio de cuartos:** Cuando se toman varias semanas de referencia, la probabilidad de exposición a un soporte resulta de promediar los minutos vistos durante ese cuarto en las semanas correspondientes. Eso supone mantener implícitamente la hipótesis irrealista de que la audiencia es estacionaria entre los cuartos de hora de distintas semanas. El segundo estudio se refiere a la *estabilidad entre cuartos*, y mide la variación de 6.188 *ratings* (cadena-cuarto de hora) en los mismos días de distintas semanas, entre febrero y abril de 1995.

---

<sup>11</sup> QUEIROS, Luis. "Medición de la audiencia de televisión a nivel del segundo". Ponencia presentada en el Encuentro Internacional AIMC sobre investigación de medios. Madrid, 30 de noviembre y 1 de diciembre de 1995. Libro de ponencias, págs. 45-64.



**4. Proceso de cálculo:** Como decíamos en el segundo epígrafe, Marksel estima los contactos recibidos, individuo a individuo, distribuyendo binomialmente la probabilidad media de exposición a varios soportes. Aunando los datos individuales, se estiman la cobertura y la distribución de contactos: esos resultados difieren de los realmente producidos, que sí son recontados en la "Reconstrucción de Campañas" de Telescope.

A continuación exponemos las tres investigaciones empíricas realizadas<sup>12</sup>. Primero, y con carácter previo, las que se refieren a los factores segundo y tercero. Después, aunando la influencia de todos los factores, exponemos la investigación principal de esta ponencia, que compara las evaluaciones de Marksel con las observaciones de Telescope, en 112 campañas que se emitieron en abril de 1995.

#### a) ¿Cómo difieren los *ratings* del minuto publicitario y del cuarto de hora?

Las evaluaciones de Marksel se ven afectadas por la hipótesis de que todos los minutos del cuarto tienen la misma audiencia. La realidad es bien distinta pero lo que nos interesa, tanto a nivel teórico como práctico, es valorar las implicaciones de esta hipótesis simplificadora. Para analizar la llamada *estabilidad intra cuartos*, vamos a medir la permanencia de la audiencia durante el cuarto de hora. Obviamente sólo nos interesan los minutos donde se emite publicidad, que previsiblemente tendrán una ligera caída en la audiencia debida, entre otras razones, al *zapping*.

La *estabilidad intra cuartos* se ha observado en todos los momentos donde se emitieron anuncios convencionales (sin sobreimpresiones, patrocinios, etc.) durante abril de 1995. Los datos se tomaron del módulo "Audiencia de Spots" de Telescope<sup>13</sup>. La pérdida general de audiencia en los minutos de publicidad ha sido calculada dividiendo la suma de *ratings* de minutos entre la suma de *ratings* de cuartos: para los 66.083 anuncios controlados<sup>14</sup> ese cociente valía 0'96, es decir, en general la audiencia del cuarto de hora se redujo un 4% durante el minuto publicitario.

Esta pérdida de audiencia no fue constante en las franjas horarias definidas por SOFRES (véase **Tabla 2**). Al contrario, la permanencia de los espectadores durante los minutos publicitarios se fue reduciendo progresivamente según avanzaba el día.

---

<sup>12</sup> Los autores agradecen la colaboración de la empresa Sofrés, A.M. que ha facilitado la información proveniente de la audimetría y los programas Marksel y Telescope. Asimismo agradecen a Raquel Ortega, Técnica de Estudios de AIMC, su valiosa ayuda en el tratamiento de la información.

<sup>13</sup> Un espectador que cambia de canal en un minuto queda registrado como audiencia del canal donde vio más segundos o, en caso de empate, del primer canal conectado.

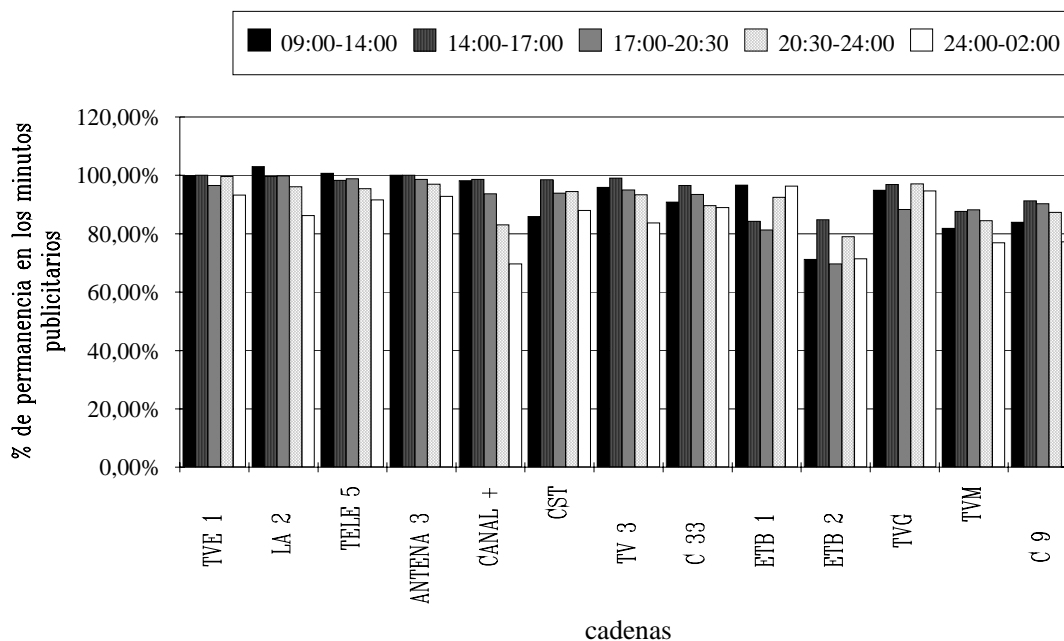
<sup>14</sup> No se han tenido en cuenta los anuncios de la franja limitada entre las 2.00 a.m. y las 9.00 a.m, debido a la escasez de elementos muestrales y al consiguiente carácter atípico de la información recogida.

Tabla 2

Franjas horarias	Permanencia en los minutos publicitarios	Número de anuncios
09:00-14:00	99'4%	10.214
14:00-17:00	98'7%	16.257
17:00-20:30	97'2%	17.589
20:30-24:00	95'4%	13.753
24:00-02:00	91'1%	8.270
total	96'2%	66.083

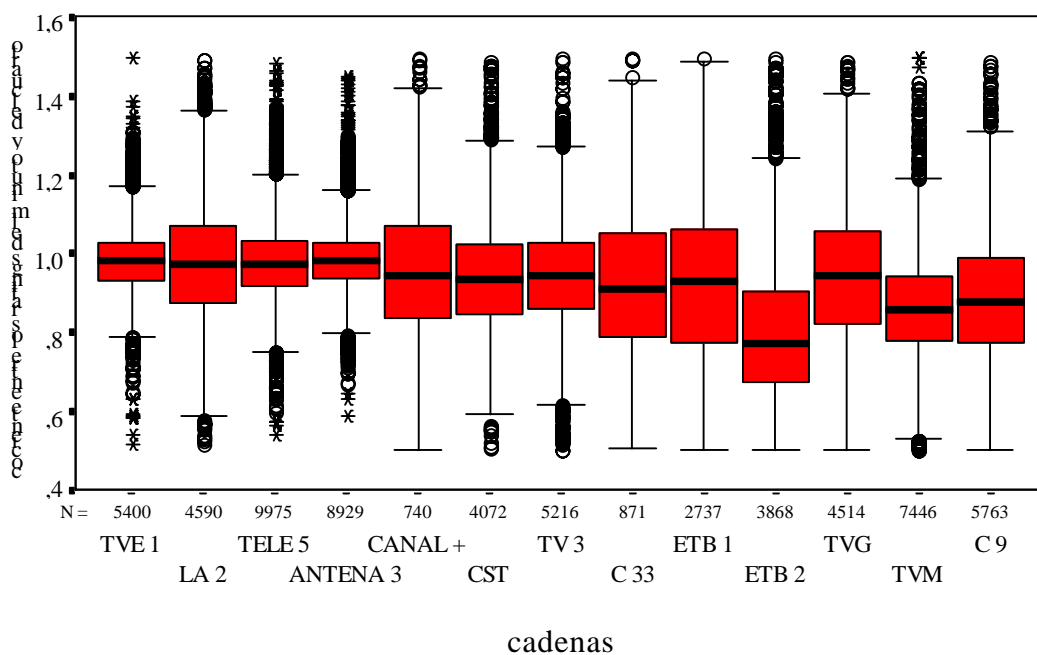
La estabilidad intra cuartos tampoco era constante para las diversas cadenas. En líneas generales parece que la cobertura nacional y la amplitud de *share* implicaban menor pérdida proporcional de espectadores durante los minutos de publicidad: TVE 1 (97'8%), ANTENA 3 (97'7%), LA 2 (97'6%), TELE 5 (96'4%), TVG (95'5%), TV 3 (95'2%), CST (94'7%), C 33 (90'9%), CANAL + (90'6%), ETB 1 (89'8%), C 9 (87'4%), TVM (84'9%) y ETB 2 (76'3%). Las cadenas nacionales de mayor audiencia son también las que menos variaciones presentaban en las distintas franjas horarias (cfr. **Figura 1**).

Figura 1



Además de lo expuesto, los *ratings* del minuto variaban menos respecto a los *ratings* del cuarto, en aquellas cadenas de cobertura nacional y amplio *share*. Así se concluye tras calcular el cociente entre los *ratings* del minuto publicitario y del cuarto de hora correspondiente, en los 66.083 casos. Al representarlos por cadenas en la **Figura 2**, se observa que efectivamente ANTENA 3, TELE 5 y TVE 1 tienen las cajas más achatadas, es decir, los valores de los cuartiles centrales están más concentrados que en el resto de cadenas. En el **Apéndice** explicamos brevemente qué significa cada símbolo de estos *boxplots* o diagramas de cajas.

Figura 2



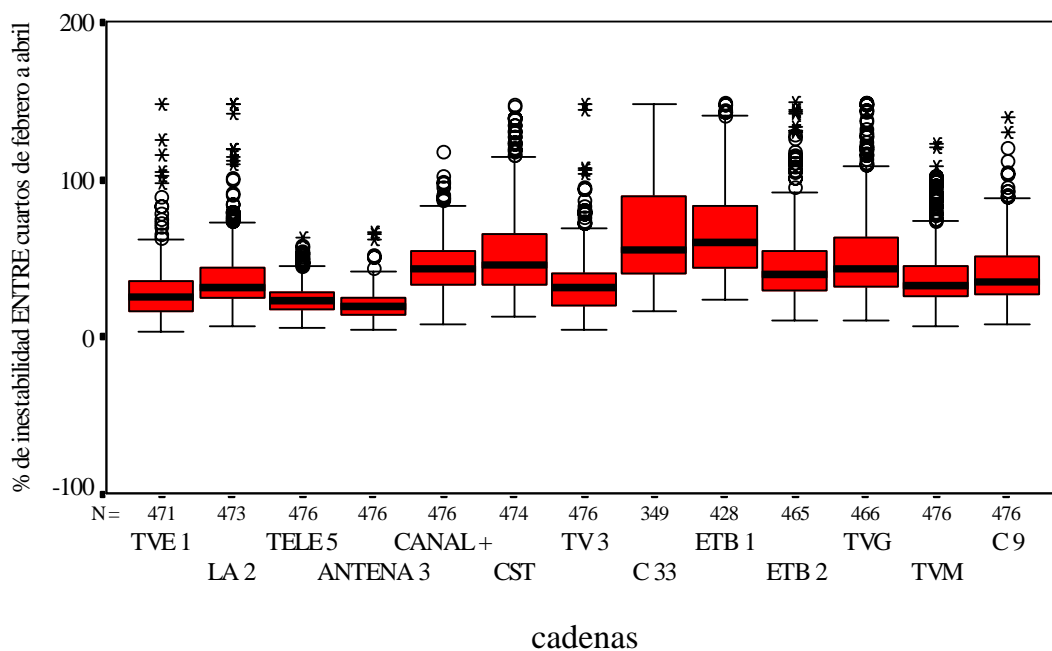
**b) ¿Cómo evoluciona la audiencia de los cuartos semana a semana?**

Tal como hemos expuesto, la audiencia de los soportes (cuarto de hora/cadena/día de la semana) se supone estacionaria durante el periodo de referencia. Esta hipótesis es también irrealista y, por tanto, fuente de errores para las evaluaciones. Su efecto se ha estudiado con la evolución de los soportes desde febrero a abril de 1995, excluyendo el atípico horario de 2.00 a 9.00 a.m. (en total 6.188 soportes, como resulta de multiplicar 4 cuartos por 17 horas, por 7 días y por 13 canales).

La *estabilidad entre cuartos* se ha medido con el coeficiente de variación en esas doce semanas: dicho *ratio* se calcula dividiendo la desviación típica de los 12 *ratings* del mismo soporte entre la media de tales *ratings*. La mediana de los 6.188 coeficientes de variación valía , pero existen importantes diferencias entre las medianas para cada cadena: TVE 1 (25'3%), LA 2 (31'5%), TELE 5 (22'6%), ANTENA 3 (19'2%), CANAL+ (43'6%), CST (45'4%), TV 3 (30'9%), C 33 (54'9%), ETB 1 (60'4%), ETB 2 (39'5%), TVG (43'4%), TVM (36'2%) y C 9 (34'6%).

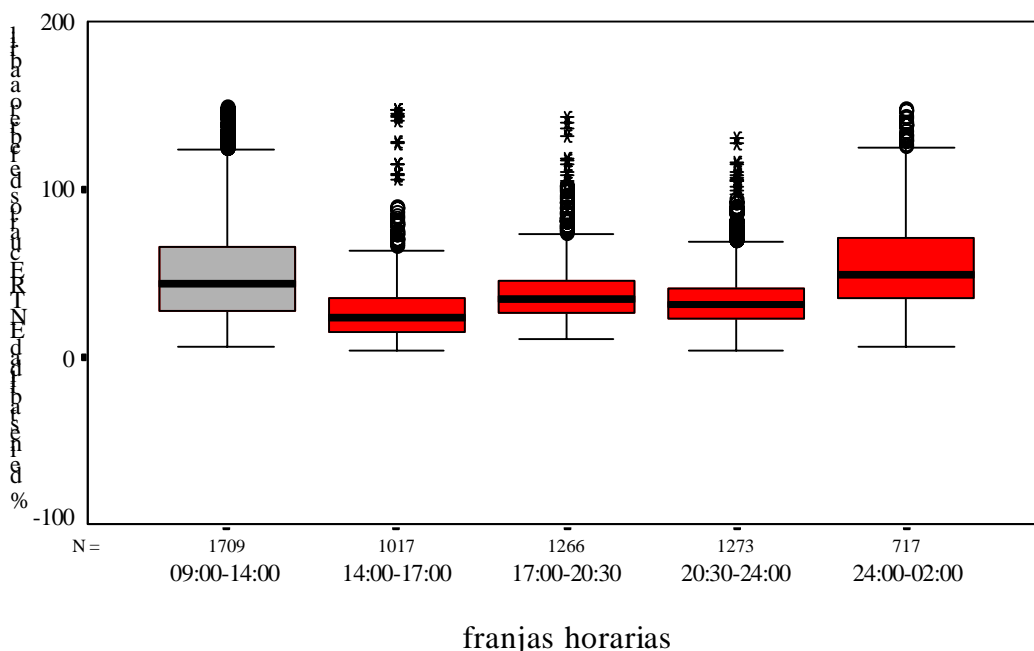
La **Figura 3** muestra que las cadenas nacionales, especialmente ANTENA 3 y TELE 5, son las que tienen menores coeficientes de variación, es decir, sus cuartos mantienen unos *ratings* más estables. Los canales autonómicos tienen variaciones superiores, en particular C 33 y ETB 1. Sin otras valoraciones, que incluso podrían aconsejar la incertidumbre, la capacidad predictiva se mejora cuando se utilizan cadenas con poca inestabilidad entre cuartos de hora.

Figura 3



Los coeficientes de variación también difieren según las franjas horarias, como indican sus respectivas medianas: 09:00-14:00 (43'3%), 14:00-17:00 (23'1%), 17:00-20:30 (34'8%), 20:30-24:00 (31'1%) y 24:00-02:00 (49'4%). Al cruzar los coeficientes de variación y las franjas horarias, las cajas parecen formar una W aplanada (ver **Figura 4**): los extremos tienen los valores más altos y están flanqueados interiormente por los más bajos que, a su vez, envuelven unos valores centrales relativamente altos. Los coeficientes de variación y la audiencia de las franjas están entonces inversamente relacionados, puesto que los coeficientes inferiores corresponden a las horas de mayor audiencia y los superiores viceversa. Luego las franjas con mayores *ratings* tienen más estabilidad entre cuartos y, por tanto, menos incertidumbre en la planificación.

Figura 4



**c) ¿Cómo se ajustan las evaluaciones de las 112 campañas a la realidad?**

Hemos elegido aleatoriamente 112 marcas que emitieron pases publicitarios en alguna de las 13 cadenas españolas, durante el mes de abril de 1995. No hemos considerado pues las inserciones que tales marcas emitieron antes o después de ese mes. La muestra seleccionada (véase **Anexo**) contiene una amplia variedad de inserciones, canales, presupuestos, anunciantes y sectores de actividad. La única restricción es que ninguna contiene más de 300 inserciones, porque Marksel no permite entonces evaluar los planes.

Con el programa Telescope hemos reconstruido las 112 campañas, obteniendo los *GRP's*, la cobertura y la distribución de contactos. Esta reconstrucción se ha repetido para tres *targets* (el total individuos, las amas de casa y los individuos entre 25 y 44 años) para analizar la influencia de la segmentación en públicos objetivo. Para esta experimentación, se ha invertido el proceso normal que sigue un planificador: el plan que se evalúa se ha tomado de una realidad ya conocida y se ha cuantificado el rendimiento del mismo que el Marksel hubiera pronosticado.

Las observaciones exactas de Telescope se comparan con las estimaciones de Marksel<sup>15</sup>, a través de un sencillo coeficiente:

$$\% \text{ de error en la estimación del indicador} = \frac{\text{indicador según Marksel} - \text{indicador según Telescope}}{\text{indicador según Telescope}}$$

Los indicadores contrastados son los cinco que parecen más útiles para la planificación: *GRP's*, cobertura, frecuencia (o número medio de contactos) y cobertura

<sup>15</sup> Agradecemos que la empresa MARKTEST proporcionase para este proyecto un sistema de lectura automática de inserciones, que ha permitido hacer un total de 1.120 evaluaciones.

efectiva que, como se entiende de manera diversa<sup>16</sup>, hemos definido de dos formas, "3 ó más contactos" y "5 ó más contactos".

Estos indicadores se podrían considerar variables dependientes de varios factores que parecen influir en la precisión del Marksel: selección del periodo de referencia, segmentación en *targets*, fusión de los ficheros y número de inserciones. Como hay múltiples variables dependientes y factores, se puede contrastar la existencia de diferencias estadísticamente significativas mediante el análisis multivariable de la varianza<sup>17</sup> o MANOVA (Multivariate ANalysis Of VAriance). Esta técnica nos permite verificar si cada factor afecta a los indicadores, en su conjunto y uno por uno. A continuación exponemos la influencia de los factores por separado.

#### • Selección del periodo de referencia

Hemos evaluado los 112 planes con distintas alternativas en cuanto al número de semanas (1, 4 y 12) y a su ubicación (desde diciembre de 1994 hasta junio de 1995). Se resumen ahora los 6 periodos utilizados como referencia, nominados por el/los mes/es que ocupan:

- diciembre-enero-febrero: de la semana 49<sup>a</sup> a la 52<sup>a</sup> de 1994, y de la semana 1<sup>a</sup> a la 8<sup>a</sup> de 1995. Lo notaremos DIC-ENE-FEB.
- 9<sup>a</sup> semana de 1995.
- febrero: de la 6<sup>a</sup> a la 9<sup>a</sup> semana de 1995.
- marzo: de la 10<sup>a</sup> a la 13<sup>a</sup> semana de 1995.
- abril: de la 14<sup>a</sup> a la 17<sup>a</sup> semana de 1995. Las evaluaciones suelen hacerse antes de que se emitan los anuncios pero una planificación simultánea con la emisión, siendo inviable en la práctica, permite aislar el efecto *periodo de referencia* y determinar la influencia de los demás.
- junio: de la 23<sup>a</sup> a la 26<sup>a</sup> semana de 1995. Igualmente, sería ilógico evaluar una campaña *a posteriori*, pues el Telescope ya proporcionaría las exposiciones realmente alcanzadas. Pero ello también permite estudiar qué semanas deben incluirse en el periodo de referencia.

El factor *periodo de referencia* afectaba significativamente al conjunto de los indicadores, porque el valor de la F aproximada al lambda de Wilks es muy superior al valor en tablas, con los grados de libertad correspondientes y para cualquier intervalo de confianza (véase **Tabla 3**). Ese factor influía también significativamente en los *GRP's*, la cobertura, la frecuencia y las coberturas efectivas, porque, en cada caso, el valor de la F es superior al de tablas, con los grados de libertad específicos y en un intervalo de confianza del 99%.

---

<sup>16</sup> KRESHEL, Peggy J.; LANCASTER, Kent M.; and TOOMEY, Margaret A. "How Leading Advertising Agencies Perceive Effective Reach and Frequency". *Journal of Advertising*. Vol. 14. Nº 3. 1985. Págs. 32-8.

<sup>17</sup> BISQUERRA ALCINA, Rafael. "Introducción conceptual al análisis multivariable. Un enfoque informático con los paquetes SPSS-X, BMDP, LISREL y SPAD". Ed. Promociones y Publicaciones Universitarias. Barcelona. 1989. Págs. 147-195.

**Tabla 3:** Contraste de la influencia de los factores según la distribución F

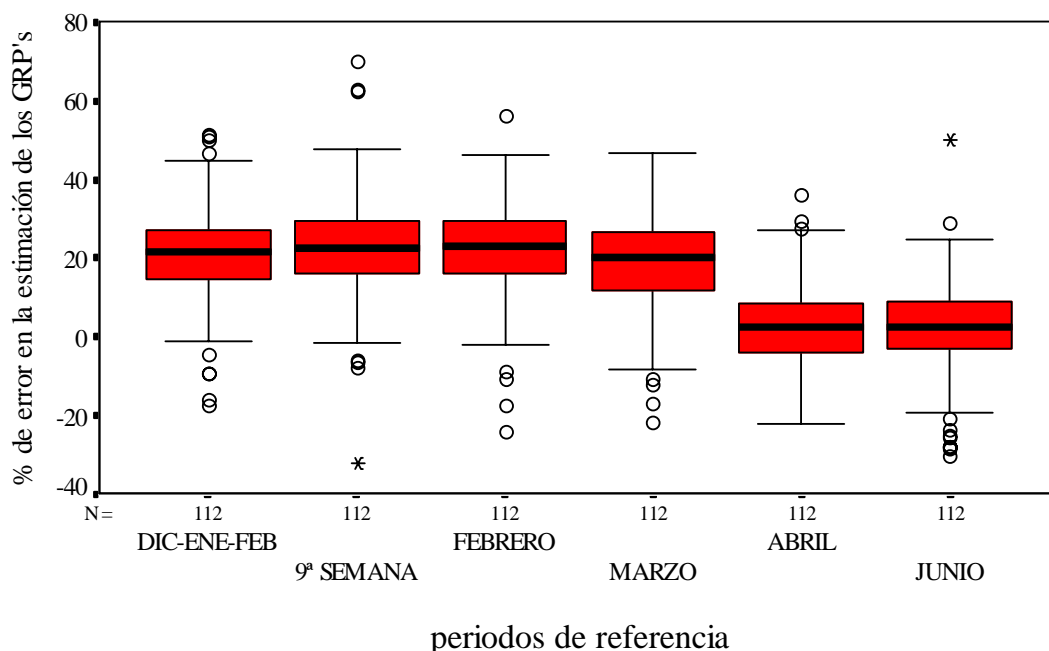
	periodos de referencia	segmentación en <i>targets</i>	fusión de ficheros	número de inserciones
Efecto conjunto (F aproximada al $\lambda$ de Wilks)	20'83**	0'36	16'00**	36'78**
Efecto en los <i>GRP's</i>	87'98**	0'39	0'04	4'47**
Efecto en la cobertura	13'46**	0'01	55'72**	13'04**
Efecto en la frecuencia	59'18**	0'25	59'54**	3'72*
Efecto en la cobertura efectiva ( $\geq 3$ contactos)	24'87**	0'01	0'49	86'15**
Efecto en la cobertura efectiva ( $\geq 5$ contactos)	29'07**	0'38	2'40	10'59**

\* diferencias significativas con un intervalo de confianza del 95%

\*\* diferencias significativas con un intervalo de confianza del 99%

La **Figura 5** muestra cómo evolucionan los *GRP's* estimados, según los periodos de referencia. No tenía relevancia el número de semanas, porque los tres primeros periodos no presentan diferencias y se componen respectivamente de 12, 1 y 4 semanas. Sin embargo, es clara la trascendencia de las semanas seleccionadas pues, comparando los meses de febrero, marzo, abril y junio, se observa una ligera y progresiva reducción de los errores, con una importante caída entre marzo y abril: este fenómeno se debe, al menos en su mayor parte, a que la audiencia de la TV se va reduciendo en ese tiempo y provoca que las predicciones estén sobrestimadas; la evolución de la audiencia durante esos meses puede verse en la **Tabla 4**.

Figura 5



El tamaño de los errores no depende exclusivamente de las semanas elegidas, lo que implicaría que abril estimase perfectamente; sin embargo, se mantiene una ligera sobrestimación, como manifiesta el valor de la mediana: 27% (véase **Tabla 5**). Entonces hay factores exógenos que han motivado esta tendencia a sobrevalorar los *GRP's*: podemos afirmar que se debe (siendo posible la interrelación con otros factores) a que la audiencia de los minutos publicitarios se mide con la audiencia de los cuartos de hora (como antes exponíamos, ello supone una sobrestimación general de los *ratings* del 4%).

Tabla 4: Evolución de los minutos de visionado de TV

	TOTAL		PRIME TIME	
	Minutos promedio de visionado	ÍNDICE (Abril=100)	Minutos promedio de visionado	ÍNDICE (Abril=100)
DICIEMBRE 94	230	111,6	85	111,8
ENERO 95	238	115,5	91	119,7
FEBRERO 95	232	112,6	90	118,4
MARZO 95	225	109,2	87	114,5
ABRIL 95	206	100,0	76	100,0
JUNIO 95	202	98,1	71	93,4

Estos resultados sugieren que el planificador debe tomar el periodo de referencia más cercano, vigilando las previsibles variaciones de la audiencia debidas, como en este caso, a la llegada de la primavera. Si no resulta ya complejo, se añade que la estacionalidad de la audiencia no es igual todos los años: por ejemplo, en 1994 la principal caída del visionado se da en marzo, repuntando ligeramente en abril.



**Tabla 5:** Medidas del % de error en la estimación de los *GRP's*

	DIC- ENE-FEB	9 <sup>a</sup> SEMANA	FEBRE- RO	MARZO	ABRIL	JUNIO
máximo	51,6%	70,1%	85,7%	46,7%	36,2%	50,0%
percentil 75%	27,4	29,8	30,9	26,5	8,7	8,9
mediana	21,7	22,8	23,8	20,3	2,7	2,6
desviac. típica	12,4	14,4	14,4	12,6	10,1	12,4
percentil 25%	14,9	16,3	16,3	11,8	-3,8	-2,8
mínimo	-17,4	-32,0	-36,4	-21,7	-22,1	-30,4

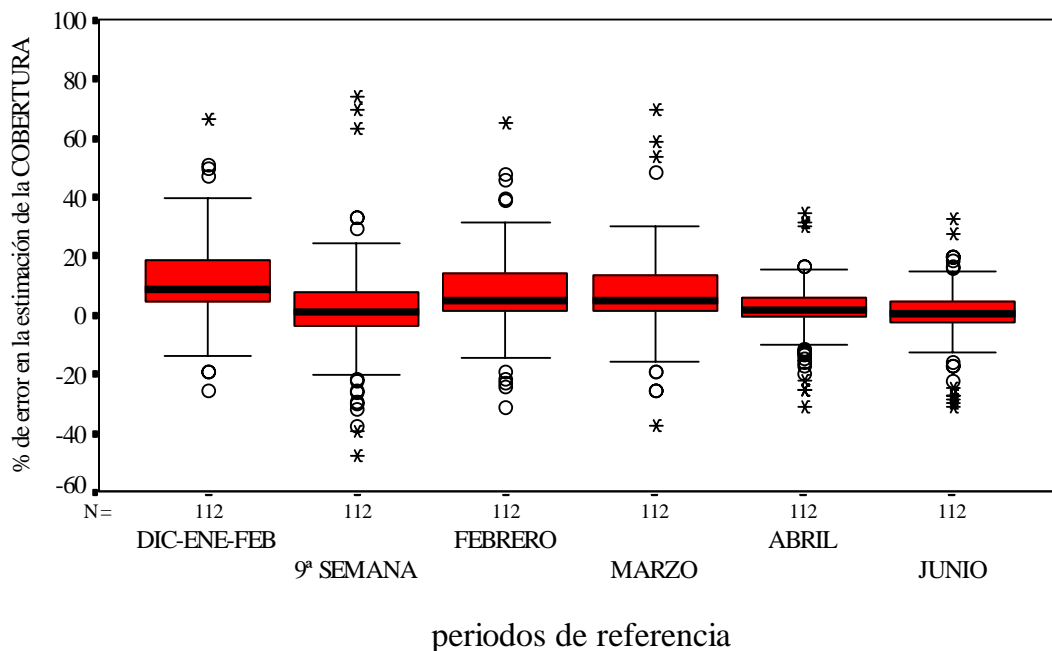
La **Tabla 6** muestra el valor central (mediana) de las diferencias obtenidas en la predicción de los tres indicadores de **cobertura** analizados.

**Tabla 6:** Medidas del % de error en la estimación de la *Cobertura*

	DIC- ENE-FEB	9 <sup>a</sup> SEMANA	FEBRE- RO	MARZO	ABRIL	JUNIO
Cobertura						
Simple (1+)	6,6%	29,8	6,6%	26,5	8,7	
Efectiva (3+)	21,7	22,8	19,8	20,3	2,7	2,6
Efectiva (5+)	12,4	14,4	14,4	12,6	10,1	12,4

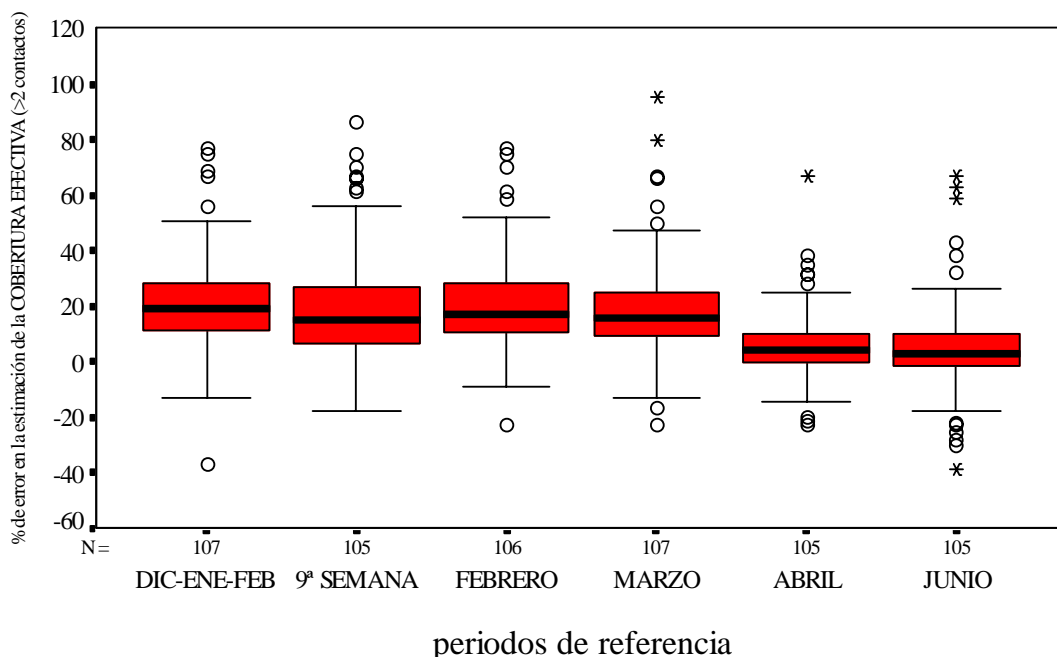
La **Figura 6** presenta de forma gráfica los errores obtenidos en la **cobertura simple**. La mitad izquierda muestra que la sobrestimación es superior cuando se ha seleccionado un mayor número de semanas: sucede así porque ampliar el periodo de referencia implica que más individuos tengan probabilidades de exposición no nulas. De nuevo la ubicación de las semanas es el fenómeno determinante: estimar sobre niveles de audiencia del invierno provoca sobrevaloraciones en las coberturas de primavera. En abril y junio predominan los errores de signo positivo, por el ya comentado efecto de la permanencia de la audiencia durante los minutos publicitarios.

Figura 6



La **cobertura efectiva**, definida como "5 ó más contactos", se estima mejor cuando se han incluido más semanas en el periodo de referencia (véanse las tres primeras cajas de la **Figura 7**) porque un nº amplio de contactos se simula mejor a partir de un periodo espaciado y no puntual, donde se recoja la repetición de exposiciones con mayor precisión. Se comprueba de nuevo que las estimaciones previas a abril son excesivas debido a la caída de la audiencia. También la medida temporal del *rating* esta influyendo en que los errores sean principalmente positivos, es decir, que las evaluaciones de Marksel sean superiores a las reconstrucciones de Telescope.

Figura 7

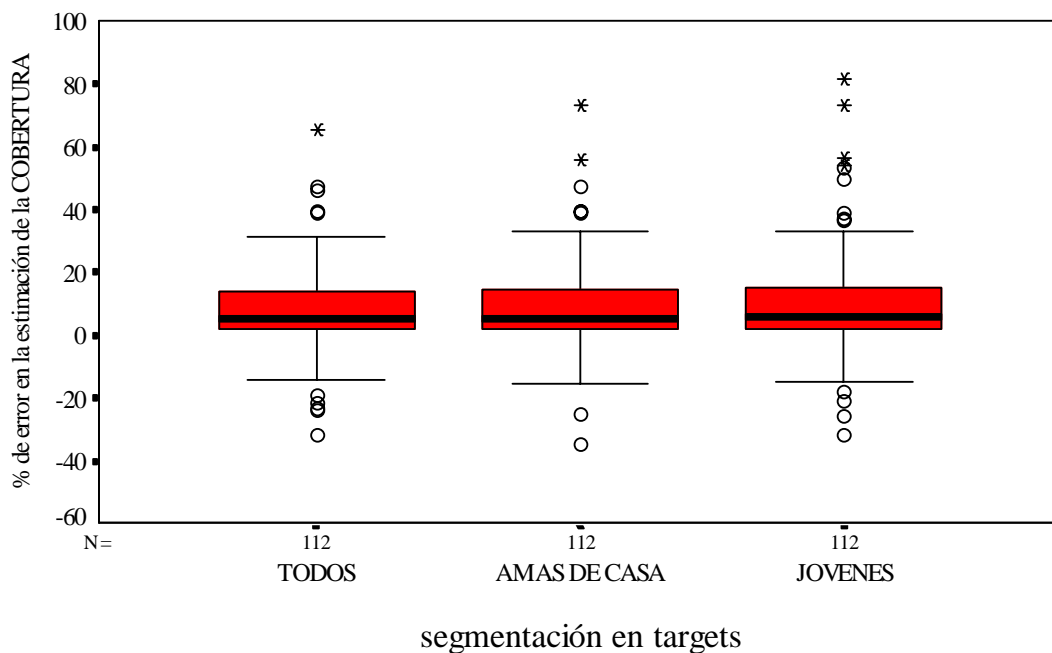


• Segmentación en targets

Se trataba de analizar si la especificación de *targets* influía en la precisión del Marksel, suponiendo siempre que la muestra sea lo suficientemente amplia para no sobrepasar el error máximo admitido. Tomando febrero como periodo de referencia, hemos evaluado y reconstruido los 112 planes para tres públicos objetivo: todos los individuos, las amas de casa y los individuos entre 25 y 44 años, grupo que denominaremos (adolesciendo de la precisión tan comentada) jóvenes.

El factor *segmentación* no ha afectado significativamente al conjunto de los indicadores revisados. Asimismo se observa en la **Tabla 3** que no existen diferencias significativas en ningún indicador tomado aisladamente. Como ejemplo, en la **Figura 8** reproducimos los errores de la **cobertura**: es diáfano que la tendencia a la sobrestimación no depende de los targets, sino que se debe explicar mediante otro/s factor/es.

**Figura 8**



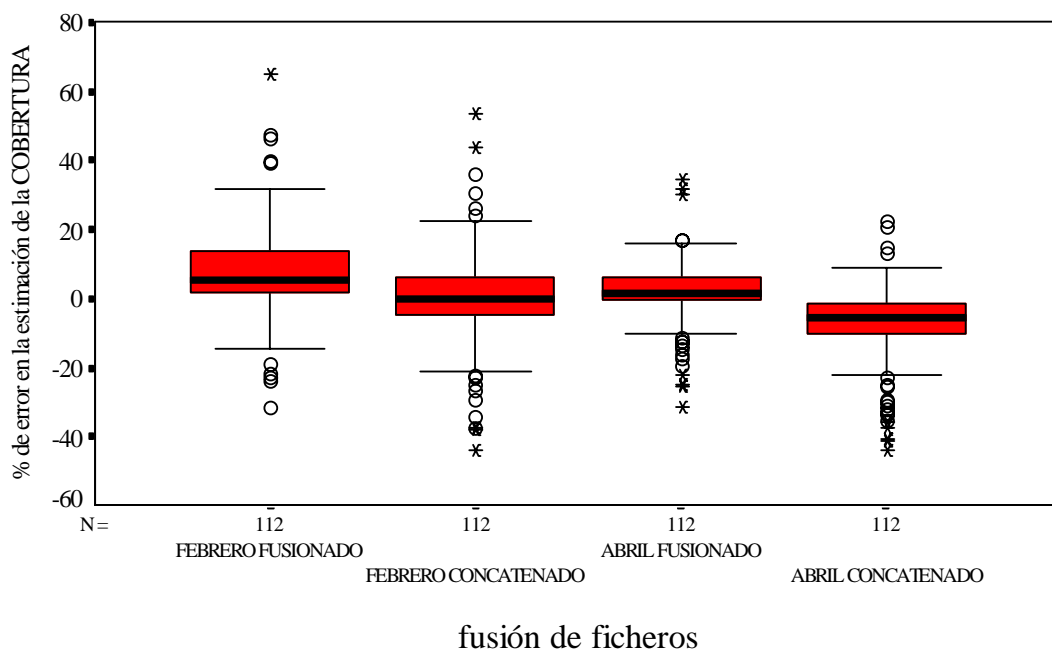
• Fusión de ficheros

El usuario de Marksel, antes de evaluar un plan concreto, debe decidir si fusiona o concatena los ficheros. La elección no resulta indiferente, pues afecta significativamente a la cobertura y a la frecuencia. Lo hemos comprobado analizando las diferencias que este factor produce en las 112 campañas, planificadas tanto en febrero como en abril.

El factor *fusión de ficheros* ha supuesto diferencias significativas en el conjunto de estimaciones, como se concluye al contrastar en tablas la F aproximada al  $\lambda$  de Wilks (cfr. **Tabla 3**). Esta influencia no se manifiesta en todos los indicadores, sino sólo en la cobertura y en la frecuencia, con intensas diferencias.

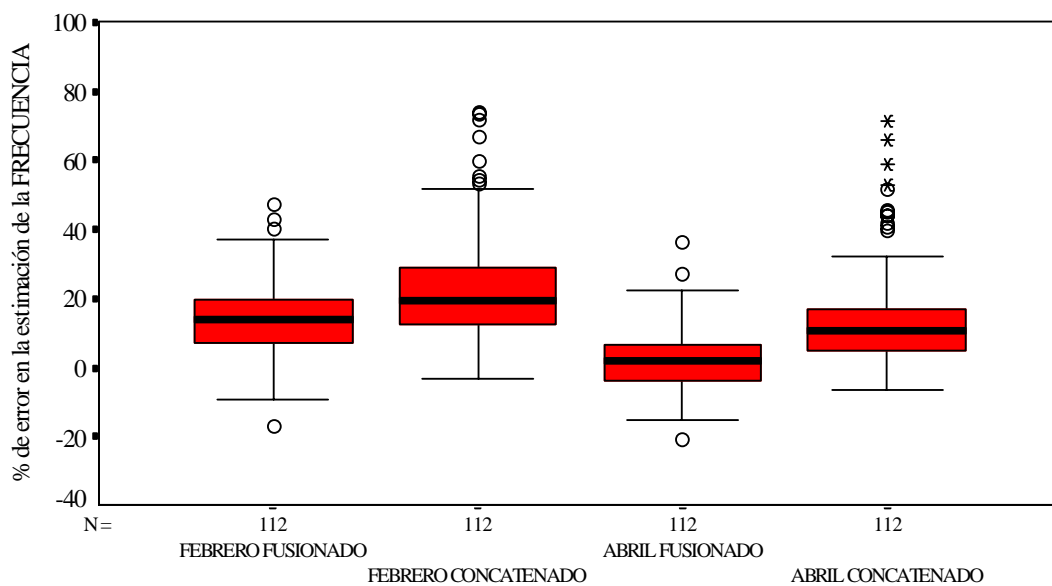
El dato de **cobertura** se ajusta más cuando los ficheros no se fusionan, tanto en febrero como en abril: en la **Figura 9** se puede apreciar que el escalón es descendente con independencia del mes de referencia. En principio, el sentido común -apoyado por el conocimiento de los procesos internos del Marksel- sugieren la conveniencia de utilizar el mismo criterio temporal en la selección del período base de la evaluación que la longitud de la campaña a evaluar. En otras palabras, se deben utilizar bases mensuales para estimar campañas mensuales y bases semanales para estimar campañas más cortas. Y, a estos efectos, una concatenación de ficheros semanales mantiene la base semanal (como ya dijimos, es una muestra de semanas). Cuando con base semanal, se trata de estimar una campaña mensual, es previsible esperar una infraestimación de la cobertura y viceversa. En el caso que nos ocupa, la infraestimación que produce la concatenación de semanas ha venido a corregir la sobrestimación ya detectada como consecuencia de las diferencias sistemáticas entre las audiencias del bloque publicitario y las del cuarto de hora.

**Figura 9**



Lo contrario sucede con la **frecuencia**: cuando se fusionan los ficheros se cometen menos errores, sea cual fuere el periodo de referencia: en la **Figura 10** ahora los escalones toman sentido ascendente. Aunque no recogemos los demás diagramas de cajas, los *GRP's* apenas presentan diferencias en función de este factor, hecho explicable en que los efectos contrarios se anulan al multiplicarse ( $GRP's = Cobertura \times Frecuencia$ ).

Figura 10



fusión de ficheros

• Número de inserciones

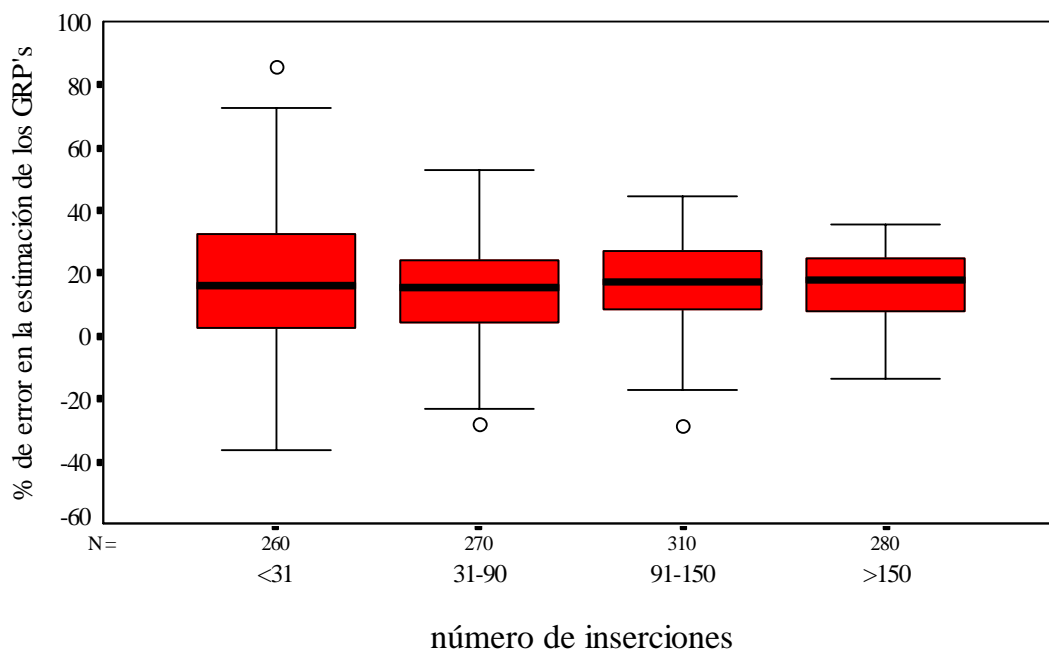
El nº de inserciones de los 112 planes es muy variado, pero nunca superior a 300 pases. Para estudiar como afecta este factor a las evaluaciones, hemos dividido el total en cuatro grupos: hasta 30 inserciones, de 31 a 90, de 91 a 150 y más de 150. Respectivamente cada grupo contiene 26, 27, 31, y 28 campañas.

La técnica MANOVA se ha aplicado sobre las 1.120 evaluaciones, es decir, para todas las alternativas de semanas de referencia, *targets* y fusión/concatenación de ficheros. Los resultados indican que el número de inserciones afecta significativamente al conjunto de indicadores, con intervalo de confianza del 99% (cfr. **Tabla 3**). Esto se debe al algoritmo de cálculo del Marksel, pero su estudio no es objeto de esta ponencia.

La tendencia general es que el grupo de menos inserciones ( $\leq 30$ ) es el que peor estima, mientras que el de más inserciones ( $> 150$ ) es evaluado con mayor precisión.

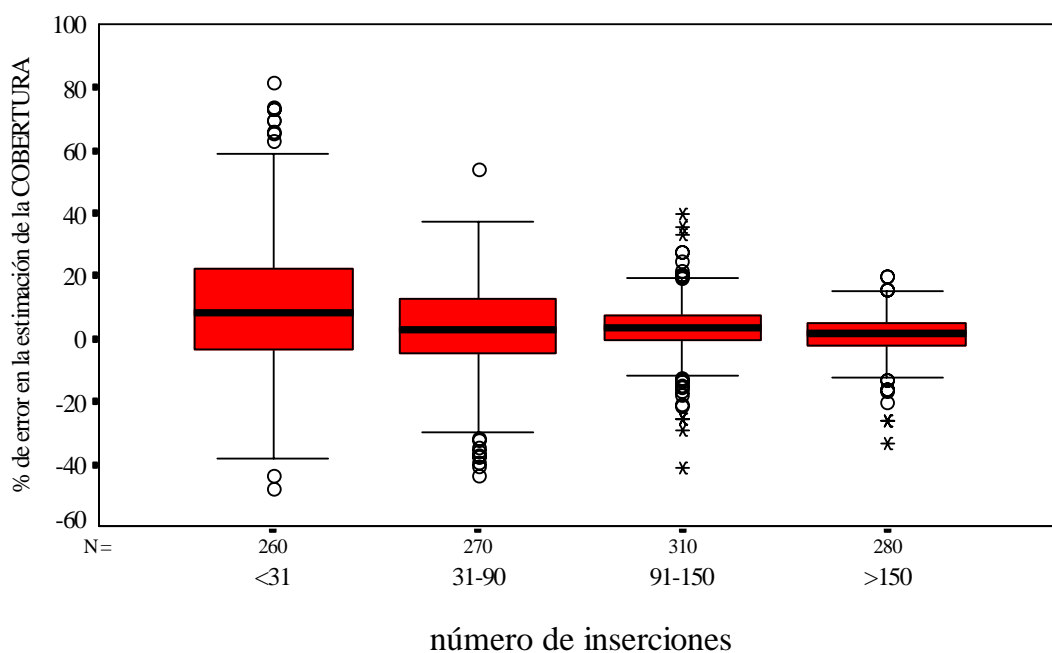
Los *GRP's* no presentan diferencias significativas con un intervalo de confianza del 99%. No obstante, en la **Figura 11** se observa que el recorrido (distancia entre el máximo y el mínimo) y los cuartiles centrales se van reduciendo ante el incremento de inserciones: entonces es menor la incertidumbre de la planificación.

**Figura 11**



La cantidad de países afecta a la **cobertura**, pues existen diferencias significativas entre los cuatro grupos (véase **Tabla 3**). Además la precisión y el número de inserciones están correlacionados negativamente: así la **Figura 12** revela que el incremento de países implica la disminución en el valor de la mediana, del recorrido y del tamaño de los cuartiles centrales. Ese mismo efecto ocurre con las coberturas efectivas aunque, al definirse como "5 ó más contactos", el fenómeno se atenúa. La frecuencia, que sólo presenta diferencias significativas al 95%, tiene mayor precisión en las campañas más sencillas, compensando la mayor imprecisión de la cobertura.

**Figura 12**



## 5. Conclusiones

Las diferencias entre las evaluaciones de Marksel y las reconstrucciones de Telescope se deben a cuatro factores: la capacidad predictiva, la medida temporal, el promedio entre cuartos y el algoritmo de cálculo.

La *capacidad predictiva* puede mejorarse seleccionando adecuadamente el periodo de referencia: lo más importante es que las semanas tengan niveles de audiencia similares, pero también debe evitarse elegir excesivas semanas porque supone inflar la cobertura total y las coberturas efectivas. Sin embargo, resulta indiferente el *target* especificado. Conviene fusionar los ficheros para no incurrir en subestimaciones de la cobertura ni en sobrestimaciones de la frecuencia. Además cuantas menos inserciones tengan la campaña, más incertidumbre tendrán los indicadores y, en particular, la cobertura.

La *medida temporal* no debe olvidarse porque planificar sobre cuartos, y no sobre minutos, implica una sobrestimación general en los *ratings* del 4%. A su vez, la permanencia de los espectadores durante la publicidad se va reduciendo según avanza el día; y asimismo difiere entre cadenas, pues las de cobertura nacional y mayor *share* presentan menos diferencias en los *ratings* del minuto y del cuarto.

El *promedio entre cuartos* también debe tenerse en cuenta, ya que en realidad los *ratings* varían semana a semana, pero esa inestabilidad es superior en las franjas horarias y en los canales de menor audiencia: su planificación, por tanto, está sujeta a mayor incertidumbre.

Posteriores investigaciones podrían analizar la influencia del *algoritmo de cálculo* a partir de datos descryptados, e incidir en una explicación causal de la precisión de las evaluaciones en función de estos factores.

## Apéndice

Los *boxplots* o diagramas de cajas permiten visualizar fácilmente la distribución de la variable. Una serie de símbolos representan convencionalmente algunos indicadores estadísticos muy utilizados<sup>18</sup>:

- La caja recoge el 50% de los casos centrales y, por tanto, está limitada por los percentiles 25° y 75°.
- La mediana se representa por una gruesa traza horizontal.
- De la caja salen, hacia arriba y hacia abajo, dos rectas que terminan en el mayor/menor valor que no sea *outlier*.
- *Outliers*: valores que se alejan más de 1'5 veces la altura de la caja, desde el percentil 75° hacia arriba o desde el percentil 25° hacia abajo.

---

<sup>18</sup> NORUSIS, Marija J. "SPSS for Windows. Base System User's Guide. Release 6.0". Chicago, 1996.

- *Extremes*: valores que se alejan más de 3 veces la altura de la caja, desde el percentil 75° hacia arriba o desde el percentil 25° hacia abajo.

