

## Práctica gráfica en tesis de Actuaría (ITAM, 1990-2007)

12 de junio de 2013

Felipe González y grupo de Estadística Aplicada I, Otoño 2007

Los métodos gráficos son una herramienta poderosa en el análisis de datos. En contraste con otros métodos, sólo hasta recientemente han aparecido guías sólidas (aunque todavía parciales) de lo que constituye la buena práctica gráfica, y estas guías son relativamente poco conocidas.

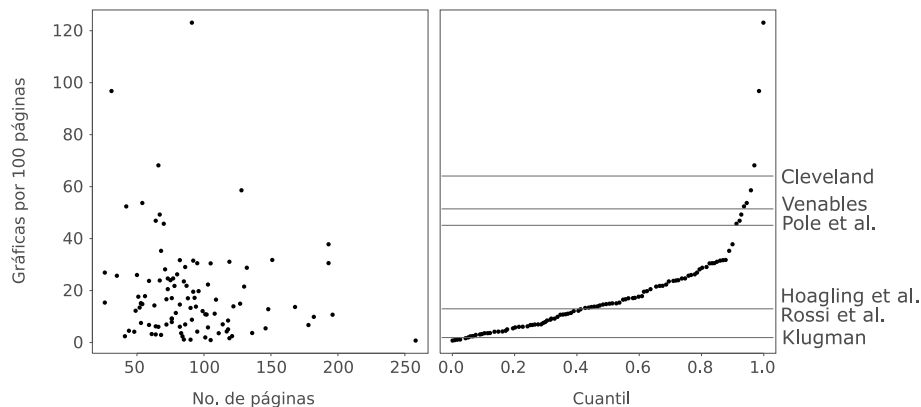
Un punto de partida para la difusión de estas guías es el diagnóstico de la práctica actual, que varía mucho según el contexto (periódicos, industria, publicaciones científicas). En este trabajo buscamos producir estimaciones de varios indicadores básicos para las tesis de Actuaría en el periodo 1990-2007: qué tanto y qué tipos de gráficas se usan, y algunos indicadores de la calidad de la práctica gráfica.

Mostramos que las gráficas son utilizadas comunmente en las tesis de actuaría (en particular las de dispersión y las de barras), y que podrían mejorar en varios aspectos: en primer lugar, reduciendo la basura gráfica (rejillas, tramas, efectos en 3D), en segundo lugar, ampliando el repertorio de tipos de gráficas (gráficas de caja y brazos, gráficas de paneles múltiples como matrices de gráficas dispersión), y en tercer lugar, más concretamente, sustituyendo las gráficas de barras por mejores alternativas: tablas, diagramas de caja y brazos, o diagramas de puntos.

## Resultados principales y Recomendaciones

### Densidad gráfica

Una buena parte de las tesis de actuaría contiene gráficas de datos ( $84\% \pm 6\%$ ). Excluyendo páginas de bibliografía, las tesis con gráficas tienen un promedio de  $19 \pm 3$  gráficas por cada 100 páginas, y una mediana de  $14.5 \pm 3$  gráficas por cada 100 páginas. Alrededor del 90% de las tesis con gráficas en la muestra de distribuyen entre 0 y 30 gráficas por 100 páginas. Por arriba de esto, tenemos una cola de tesis que son mucho más intensivas en su uso de métodos gráficos. Abajo mostramos la distribución muestral con algunas referencias más o menos conocidas.<sup>1</sup>



<sup>1</sup>Cleveland, William S., *The Elements of Graphing Data*, Hobart Press, 1994.

Venables, W.N. y Ripley, B.D., *Modern Applied Statistics with S-PLUS*, Springer, 2002.

Pole A., West M. y Harrison, J., *Applied Bayesian Forecasting and Time Series Analysis*, Chapman & Hall, 1994.

Hoaglin, D.C., Mosteller, F. y Tukey, J.W., *Understanding Robust and Exploratory Data Analysis*, Wiley, 1983.

Rossi, P.E., Allenby, G. y McCulloch, R., *Bayesian Statistics and Marketing*, Wiley, 2005.

Klugman, S. A., *Loss Models: From Data to Decisions*, Wiley-Interscience, 2004.

Donde vemos también que no hay una relación simple entre el número de páginas y la densidad gráfica, aunque la gráfica de la izquierda sugiere que las tesis con densidades atípicamente altas tienden a ser cortas.

Los tres casos que tienen densidad gráfica mayor al libro de Cleveland son interesantes: el mayor, con 123 gráficas por 100 páginas (IAC-2005-EAR, *Análisis comparativo del nivel de religiosidad en estudiantes católicos de dos universidades...*) contiene algunas gráficas de barras y muchas gráficas de pay decorativas acompañadas de tablas de datos (que es el estilo usual de la graficación de datos en la mercadotecnia). El segundo (IAC 2003.RRI) es un dato atípico por varias razones: se trata de una tesis de 31 páginas (sin contar bibliografía).

Finalmente tenemos IAC-1999-FAN, *Sistema Bonus-Malus 200 para la conservación de la cartera*, que es atípica por el uso extensivo de histogramas (para representar distribuciones de pérdida).

## Tipos de gráficas

Las gráficas más comunes son las dispersión (que no son series de tiempo). En la tabla de la derecha mostramos nuestras estimaciones para varios tipos de gráficas que fueron medidos para la población de tesis que tienen al menos una gráfica de datos.

Tipo	% de tesis	
Dispersión	68	9
Barras	51	10
Serie de tiempo	51	10
Histograma	20	7
Pay	17	7
Caja y brazos	9	5
Mapas	4	4
Otras	2	3

Estimaciones para la población de tesis con al menos una gráfica de datos, basadas en 102 tesis.

## Indicadores básicos de calidad gráfica

En nuestro estudio tenemos dos indicadores básicos de la calidad de cada gráfica:

- Contenido de basura gráfica (*chartjunk* de Tufte, como efectos 3d, *grid* o rejillas que ocultan o dificultan la percepción de los puntos de datos, marcos superfluos, etc.).
- Si la explicación de la gráfica está en la misma página de la gráfica o no. En el trabajo gráfico serio texto, tablas, diagramas y gráficas se integran para mostrar la evidencia (\*\*Tufte\*\*).

**\*\*Basura\*\*** Series de tiempo ok - el caso grave es el de las gráficas de barras. Por tesis, mostrar ejemplos - collage!

Hablar también de barras de 3+ variables - este es un problema

**\*\*Explicación en la misma página - agregado**

## Otros resultados

En el tiempo, tablas, discutir gráficas analíticas, número de páginas por tesis, también en el tiempo.

## Metodología

Describir marco muestral

Muestreo aleatorio por conglomerados - parte bietápica para otra ocasión.

Describir cuestionario y problemas que se encontraron **\*\*Nota: revisar la cuestión de bietápico con UPMs chicas\*\***

**\*\*Multipanel - poco usado**

Describir funcionamiento de grupo.

## Antecedentes

Explicación de basura gráfica, por qué es importante esto (oportunidades perdidas, claridad, etc.)