

Escritura académica en español 2: analizando datos

Fernanda Barrientos Contreras

18 de julio de 2018

(Materiales: ejerciciosExcel.zip - bajar [aquí](#))

Ahora que ya sabemos utilizar Praat para obtener los datos que necesitamos, tenemos que analizar dichos datos. En este tutorial utilizaremos Excel para lograr los siguientes objetivos:

1. Obtener estadígrafos descriptivos sobre un determinado conjunto de datos.
2. Obtener estadísticas inferenciales de datos obtenidos.
 - Relación entre dos variables numéricas (continuas)
 - Comparación entre los promedios de dos grupos (variable continua vs. variable categórica)
 - Comparación entre los promedios de tres grupos (variable continua vs. variable categórica)
 - Comparación entre cuentas obtenidas por dos o más grupos (variable categórica vs. variable categórica)
3. Generar gráficos adecuados para el tipo de datos:
 - Gráfico de dispersión (scatterplot)
 - Carta de formantes
 - Gráfico de barras
 - Gráfico de cajas

1. Los estadígrafos descriptivos

Con el fin de explicar nuestros resultados, necesitamos describirlos y resumirlos. Consideraremos tres tipos esenciales:

- Estadígrafos de concentración: Muestran la tendencia en una muestra. La más conocida de estas medidas es la **media aritmética** (también conocida como **promedio**). De cualquier modo, el promedio por sí solo es un descriptor bastante pobre de los datos: no mostrará cuán dispersos están los datos, o la probabilidad de que un dato cualquiera tenga un valor X .
- Estadígrafos de posición: Son puntos de referencia contra el cual un punto específico puede ser comparado. La **mediana** es el descriptor más utilizado.

- Estadígrafos de dispersión: ¿Qué tan dispersos o concentrados están los datos en una muestra? ¿Qué tan lejos está un dato X del promedio? Esto puede cuantificarse calculando la **desviación estándar**.

1.1. Obtener estadígrafos descriptivos

Hay una manera muy simple de obtener todos los estadígrafos juntos. Para esto, abriremos `vot.csv`. En Excel 2016, vamos a *Daten* > *Datenanalyse* > *Populationskenngrößen* (Fig. 1).

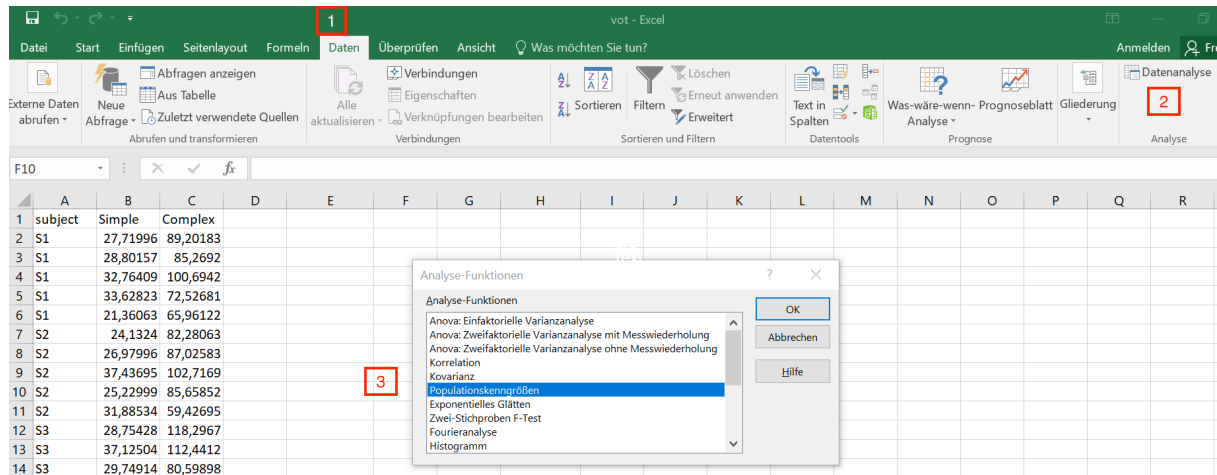


Figura 1: Calculando estadígrafos descriptivos.

Luego, una segunda ventana (Fig. 2) nos pedirá seleccionar el rango de datos (*Eingabebereich*) para lo cual hacemos clic en el recuadro (1): esto nos lleva de vuelta a la hoja de cálculo y ahí seleccionamos con el mouse todas las celdas. Luego seleccionamos dónde queremos tener la información: podemos crear una hoja nueva para eso (*Ausgabe* > *Neues Tabellenblatt*) (2). Le damos OK y en la hoja tendremos todos los datos deseados, junto a otros más.

Si no tenemos Excel 2016, tenemos las siguientes alternativas:

- Cargar el add-on Analysis Toolpak (*Analyse-Funktionen*). Esto es igual al botón *Datenanalyse* que aparece en Excel 2016.
- También podemos calcular la media, la mediana y la desviación estándar haciendo clic en una celda vacía, luego clic en el ícono de funciones (ver Fig. 3a) y buscamos: MITTELWERT para promedio, MEDIAN para la mediana, y STABW.S para la desviación estándar (Fig. 3b). Luego hacemos clic en los recuadros pequeños (que nos llevan de vuelta a la hoja de datos) y seleccionamos los datos. El valor correspondiente aparecerá en la celda vacía donde hicimos clic al inicio.

2. Correlaciones entre dos variables continuas

En esta sección queremos saber si existe alguna relación entre la velocidad de lectura de una persona (medida en palabras por minuto o PPM) y el tiempo pasado frente a la TV (medido en horas por día). En el fondo, lo que queremos saber es si (a) existe una

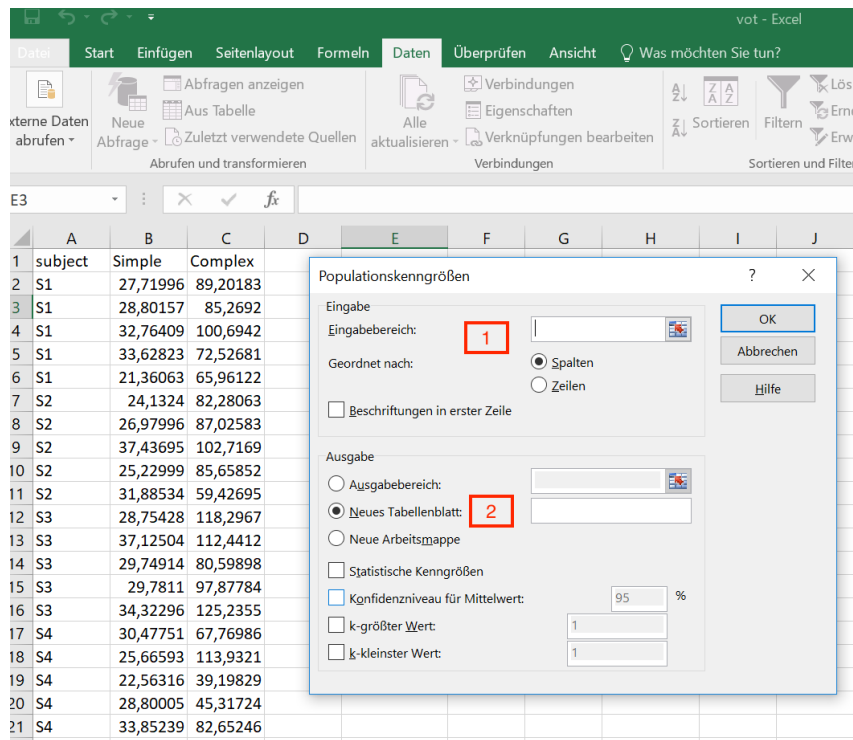


Figura 2: Estadígrafos descriptivos: selección de datos y del área para los resultados.

correlación; y (b) el sentido de la correlación (directa o inversa). En una **correlación directa**, mientras más alto sea el valor de una variable, más alto será el valor de la otra variable; en el caso de la **correlación inversa**, mientras más alto es el valor de una variable, más bajo será el valor de la otra variable.

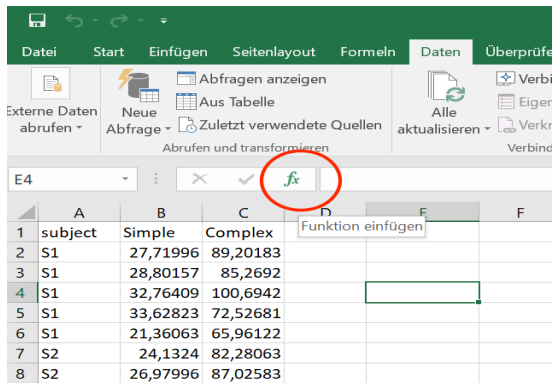
2.1. Cálculo de R de Pearson

El test que nos sirve para este caso es la **R de Pearson**, que da valores entre -1 y 1. Si el número obtenido está entre -1 y 0, entonces la correlación es inversa; si el valor obtenido está entre 0 y 1, entonces la correlación es directa. Además, si el número se encuentra entre -1 y -0.5; o bien entre 0.5 y 1, entonces la correlación es alta.

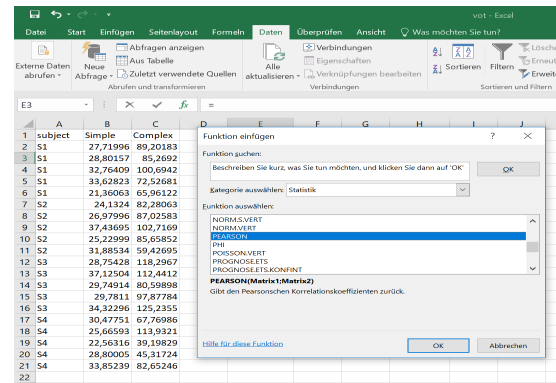
Luego, también tenemos que reportar el valor p de la R de Pearson. En estadística, las pruebas arrojan una probabilidad de que lo que estamos observando es significativo. A esto se le llama **valor p** (*p-Wert*). En lingüística lo normal es asumir que **un valor p menor a 0.05 es estadísticamente significativo**, por lo que si obtenemos este valor entonces podemos decir en este caso que **los resultados que obtuvimos son debido a una real correlación, con un 95 % de seguridad**.

El archivo *lectura.csv* tiene una pequeña muestra con valores de velocidad lectura y horas frente a la TV. Notemos que en este caso un solo dato consiste en dos mediciones: persona A tiene una velocidad de lectura X y una cantidad de horas Y frente a la televisión.

Para encontrar la correlación con Excel, hacemos clic en una celda vacía, vamos al botón de funciones (Fig.3a), y buscamos PEARSON. Hacemos clic en los botones que nos llevan de vuelta a la hoja de cálculo (Fig. 4) y seleccionamos las celdas de la variable 1 (cantidad de horas frente a la TV). Volvemos al cuadro de diálogo pulsando el botón nuevamente, y luego seleccionamos las celdas de la variable 2 (velocidad de lectura). El valor aparecerá en la celda donde hicimos clic: -0.56. (Es convención redondear a 2



(a) Botón de funciones



(b) Búsqueda de funciones

Figura 3: Calculando estadísticas descriptivas en versiones anteriores a Excel 2016.

decimales).

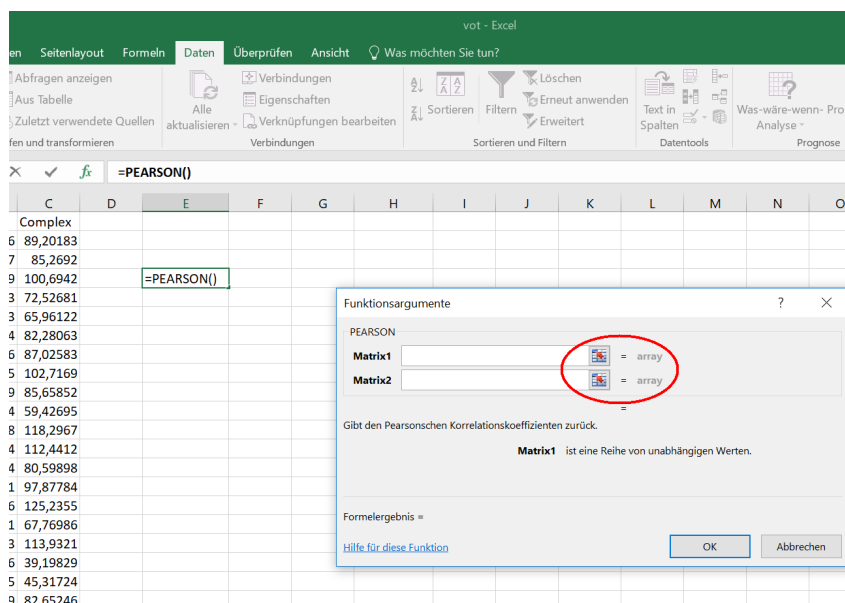


Figura 4: Calculando R de Pearson.

Pero también necesitaremos reportar el valor p en el informe. Para eso, tenemos que hacer algo un poco prehistórico (Excel no es lo mejor para análisis de estadísticas!): chequear el valor p en una tabla de valores críticos de R de Pearson (ver tabla adjunta).

Cómo leemos la tabla? Primero buscamos el número de pares que analizamos (en este caso, 15). Miramos en la fila correspondiente, en la columna .05. Si nuestro valor de R es mayor al que aparece ahí, entonces bingo! Nuestra correlación es significativa a un nivel de seguridad de 95 %.

Finalmente, el reporte se escribe así:

Se calculó un coeficiente R de Pearson para determinar si existe una correlación entre la velocidad de lectura y las horas frente a la TV. La prueba arrojó significados significativos: $r(13) = -0.56$ ($p < 0.05$). La cantidad de horas frente a la TV se encuentra inversamente relacionada con la velocidad de lectura.

IMPORTANTE:

- El reporte de un test estadístico DEBE TENER SIEMPRE:
 - El valor del test estadístico
 - Los grados de libertad (van entre paréntesis)
 - El valor p (se indica si es menor a 0.05, 0.01, o 0.001)
- Los grados de libertad para la prueba R de Pearson son simplemente el número de pares analizados -2.

2.2. Gráfico de la correlación

Para generar el gráfico de la correlación, vamos a necesitar un gráfico XY o scatterplot. Primero, seleccionamos todos los datos. Luego vamos a *Einfügen* > *Empholene Diagramme* > *Punkt (XY)* (Fig. 4). En otras versiones de Excel podemos ir a la pestaña *Einfügen* > *Punkt*.

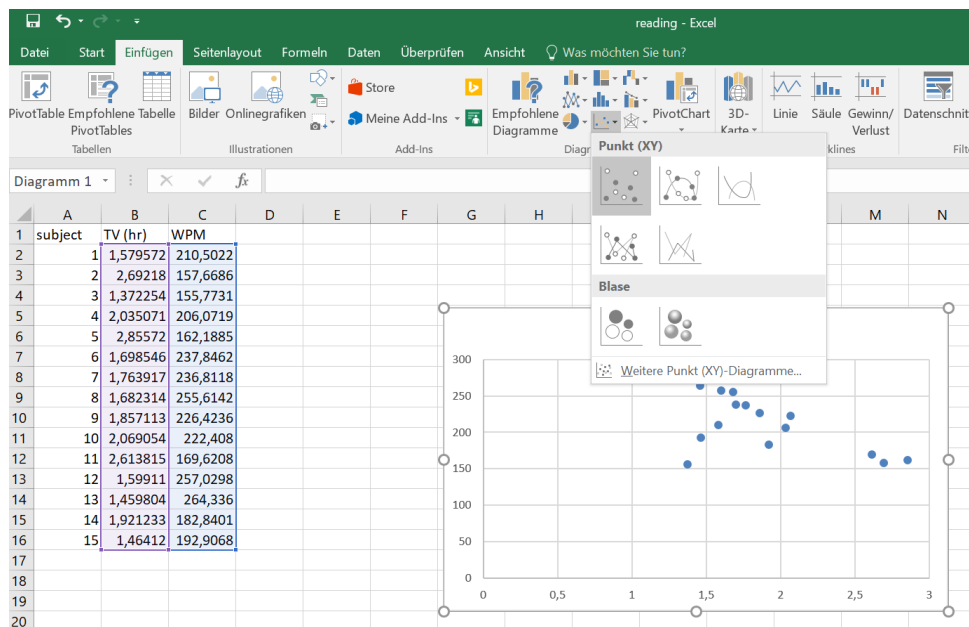


Figura 5: Creando un gráfico de tipo XY.

Pero esto nos dará solamente los puntos en el gráfico, y no la línea de correlación. Para eso vamos a *Entwurf* > *Diagrammelement hinzufügen* > *Trendlinie* > *Linear* (Fig. 6). En versiones anteriores de Excel, vamos a *Layout* > *Trendlinie* > *Lineare Trendlinie*. Luego podemos (¡y debemos!) personalizar nuestro gráfico haciendo clic en él y usando las opciones de la pestaña *Entwurf*.

3. Comparación entre dos grupos (variable numérica vs. variable categórica)

Ahora suponemos que queremos comparar la duración del VOT en ataques (onsets) silábicos simples (esto es, solamente una consonante al comienzo de la sílaba) vs. en

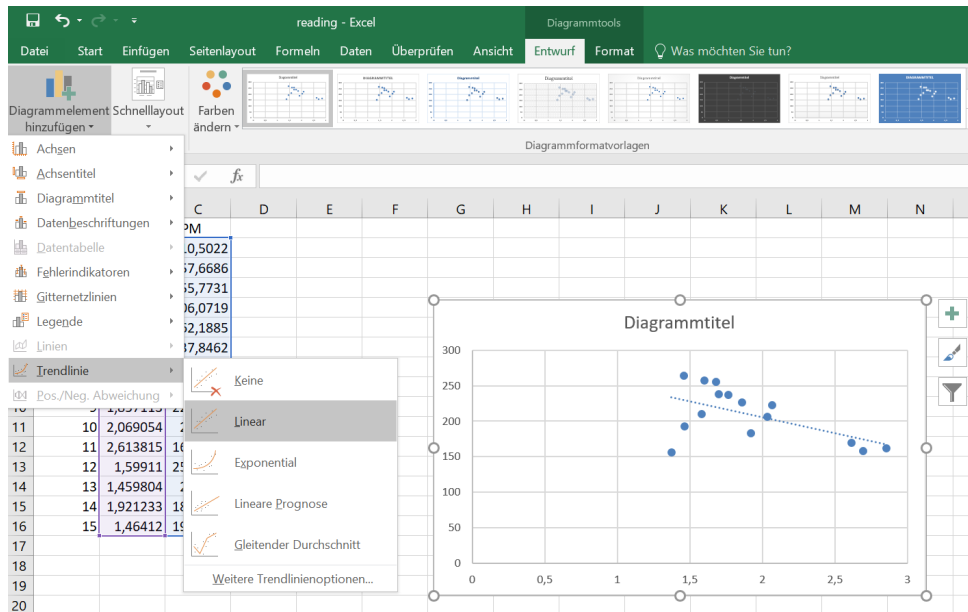


Figura 6: Línea de correlación.

ataques complejos (dos o más consonantes al comienzo de la sílaba) en un idioma X. El archivo `VOT.csv` tiene una pequeña muestra.

3.1. Cálculo de t-test

Entonces queremos comparar: ¿hay una diferencia significativa entre el VOT promedio de los ataques simples vs. el VOT de los ataques complejos? Para esto necesitamos un **t-test**, o T de Student. Este test compara los promedios y nos da un valor p.

Hay distintos tipos de t-test:

- **Pareado o no pareado** (Ein Stichprobe (= Gepaart) / Zwei Stichproben): Se hizo la medición en el mismo grupo de personas o en dos grupos distintos? Si medimos algo en el mismo grupo de personas (como en este caso), entonces hacemos un t-test pareado. Por ejemplo, si los datos en `VOT.csv` fueran de 8 sujetos, donde 4 corresponden a la condición Ataque Simple y 4 a la condición Ataque Complejo, entonces necesitamos un t-test no pareado.
- De **una o dos colas** (einseitiger / zweiseitiger): ¿Tenemos una hipótesis sobre la dirección en la cual se da la diferencia? (i.e. si la condición Ataque Simple tiene un promedio MENOR al de la condición Ataque Complejo)? Si tenemos una idea sólida, entonces podemos correr un t-test de una cola. Si no, la mejor opción es un test de dos colas (y esto es lo que siempre se sugiere).

Para calcular el t-test, tenemos dos alternativas:

- Podemos ir a *Daten* > *Datenanalyse* y luego seleccionar el t.test (ver ítem 2) correspondiente. Seleccionamos los datos y tenemos la tabla con todos los valores necesarios en un solo paso. Para reportar necesitamos el valor de la prueba misma (Fig. 7, celda amarilla), los grados de libertad (celda naranja) y el valor p (celda verde).

- Hacemos clic en una celda vacía (por ejemplo, E5), vamos a ir al botón de funciones (Fig. 3a), y buscamos T.TEST. Aquí tenemos que llenar los campos como lo hicimos con la R de Pearson, pero además tenemos que establecer si queremos un test de 1 o 2 colas (einseitiger / zweiseitiger), y luego si es pareado (gepaart), no pareado homocedástico (zwei Stichproben, gleiche Varianz - homoskedastisch), o no pareado heterocedástico (zwei Stichproben, ungleiche Varianz - heteroskedastisch) (Fig. 7). En este caso, haremos un t-test con dos colas (Escribimos 2 en el campo *Seiten*) y pareado (Escribimos 1 en el campo *Typ*). En la celda aparecerá solamente el valor p: $4.88E-10$. ($4.88 \cdot 10^{-10} = 0.000000004.88 < 0.05$, o sea, tenemos un resultado significativo, ¡yay!).

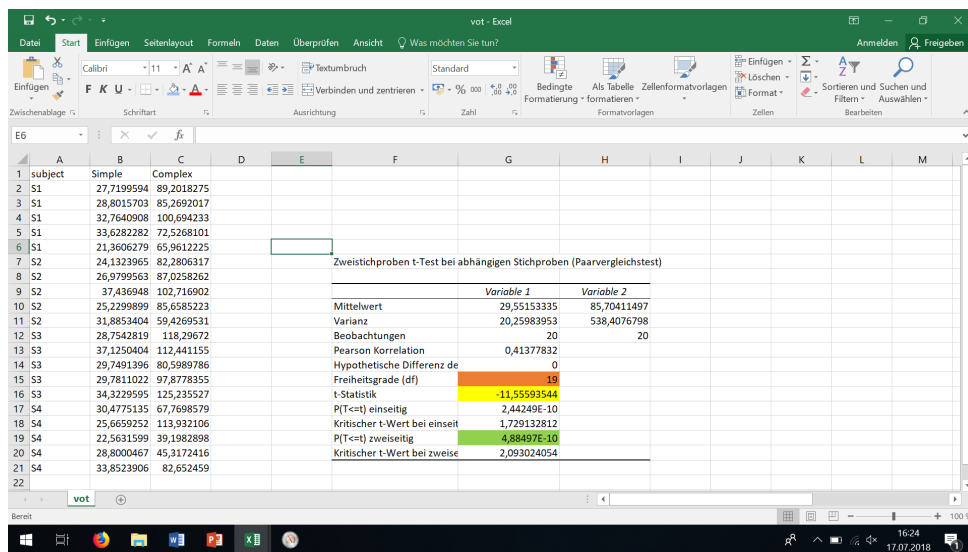


Figura 7: Calculando un t-test.

Ahora, necesitamos el valor de t propiamente tal. Hacemos clic en una celda vacía y escribimos lo siguiente:

$$=T.INV.2T(E5, 19)$$

... Que da un valor de 11.55. En esta función usamos el valor p (que estaba en la celda E5) y los grados de libertad, que en este caso son 19 (número de observaciones - 1).

Entonces reportamos de la siguiente manera:

Se realizó una prueba t de Student pareada de dos colas para determinar si la diferencia entre los valores de VOT en Ataque Simple es diferente a la de la condición Ataque Complejo. Los resultados son significativos: $t(19) = 11.55$ ($p > .05$). Esto sugiere que en el idioma X, los valores de VOT en ataques simples son significativamente más largos que los de los ataques complejos.

3.2. Gráfico de variables numérica vs. variable categórica

Aquí tenemos dos posibilidades. Podemos hacer un gráfico de cajas (*Kastengrafikdiagramm* o *Boxplot*) o un gráfico de barras (*Säulendiagramm*). Lamentablemente, sólo Excel 2016 tiene la posibilidad de hacer un gráfico de cajas; la otra posibilidad es hacer un gráfico de barras y disfrazarlo de gráfico de cajas (¡muy complicado!). Por razones de tiempo,

vamos simplemente a aprender a hacer un gráfico de barras (pero en Excel 2016 basta con seleccionar los datos, y luego ir a *Einfügen > Empholene Diagramme > Kastengrafik* (Fig. 8) para tener un diagrama de cajas).

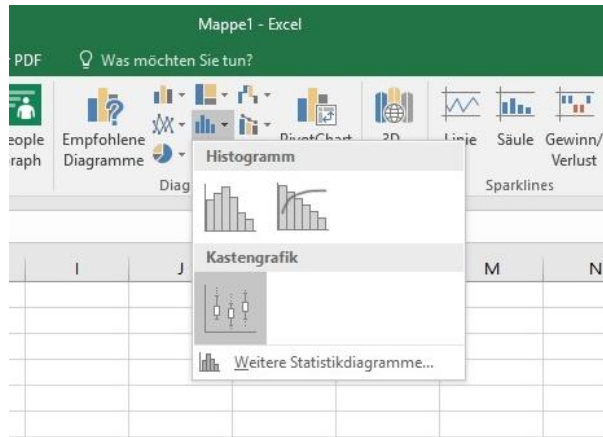
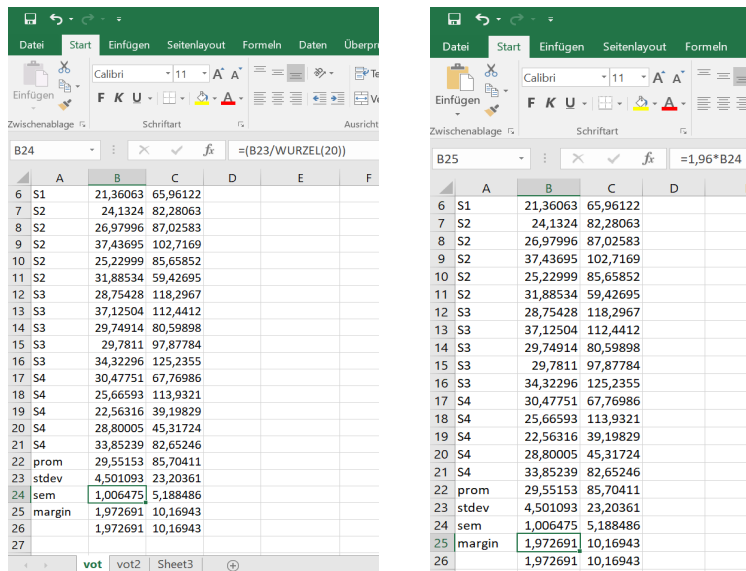


Figura 8: Diagrama de cajas (sólo en Excel 2016)

Para hacer un gráfico de barras, seleccionamos los datos de *vot.csv*, y luego vamos a *Einfügen > Empholene Diagramme > Säulendiagramm*. El diagrama aparecerá automáticamente. Luego tenemos que poner las barras de error: estas dan una idea de qué tan representativa *de la población entera* es el promedio de la muestra. Las barras de error son MUY importantes en los gráficos científicos. Para esto necesitamos calcular el margen de error al 5% (nuevamente, convención). Esto se hace de la siguiente manera:



(a) Cálculo del error estándar

(b) Margen de error

Figura 9: Calculando error estándar y margen de error. La fórmula aparece en la barra de funciones.

1. Calculamos el error estándar (Fig. 9a): Supongamos que tenemos la desviación estándar de la variable Simple en la celda E20. En una celda vacía (supongamos, E21) escribimos:

$$=E20/WURZEL(20)$$

Es decir, la desviación estándar dividida por la raíz cuadrada del número de observaciones. Repetimos con los valores de la variable Complejo.

2. Calculamos el margen de error (Fig. 9b): En una celda vacía (por ejemplo, E22) escribimos

$$=E21*1,96$$

Repetimos con la otra variable.

Ahora volvemos al diagrama. Hacemos clic en el y vamos a *Diagrammelement hinzufügen > Fehlenindikatoren > Weitere Fehlerindikatorenoptionen...* Seleccionamos *Beide, Mit Abschluss*, y luego *Benutzerdefiniert*. Hacemos clic y en *Positiver Fehlerwert* hacemos clic al botón que lleva a la hoja de cálculo y seleccionamos donde tenemos los márgenes de error. Luego, en *Negativer Fehlerwert* hacemos exactamente lo mismo (mismas celdas) (Fig. 10). Pulsamos OK y el diagrama tendrá barras de error.

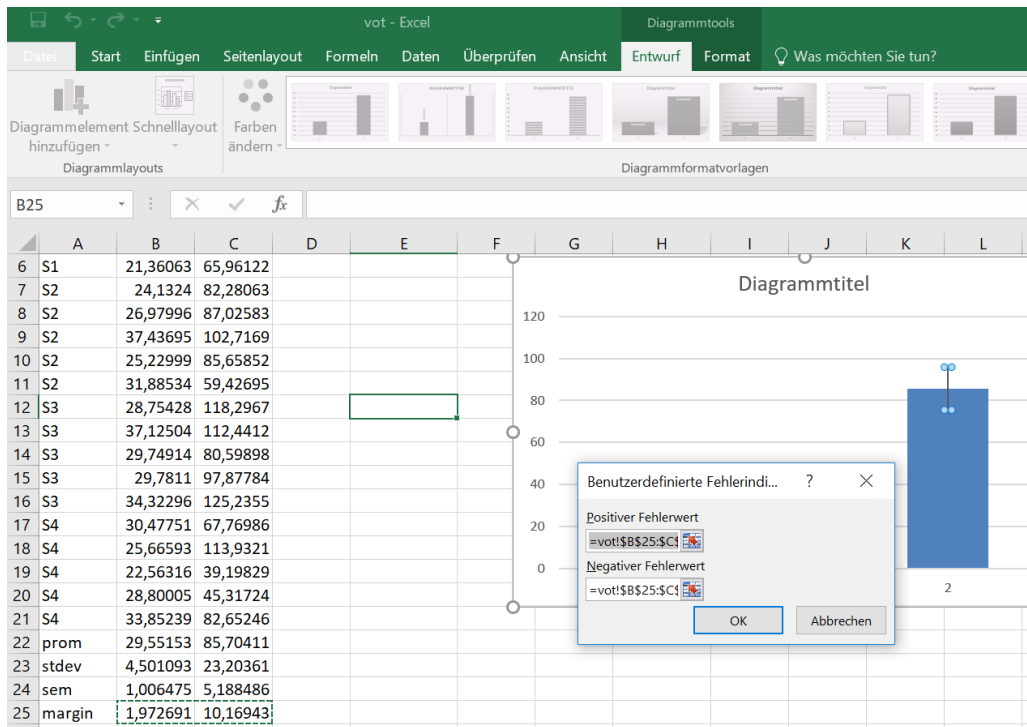


Figura 10: Llenando el campo de valores para las barras de error.

4. Comparación entre 3+ grupos (variable numérica vs. variable categórica)

Ahora vamos a suponer que tenemos los datos no solamente de la duración del VOT en ataques simples y complejos, sino también en coda (final de sílaba). En el archivo *vot2.csv* hay una pequeña muestra. Qué tan diferente entonces son los VOT en estos tres grupos?

4.1. Cálculo de ANOVA

Para medir esto necesitamos hacer un análisis de varianza (ANOVA). Si tenemos Excel 2016 o el add-on Analysis Toolpak, entonces podemos ir a *Data > Datenanalyse* y buscar la función ANOVA (*Varianzanalyse*). Pero hay varias posibilidades:

- **Unifactorial (einfaktorielle) o bifactorial (zweifaktorielle):** Queremos medir el efecto de UNA SOLA VARIABLE o DOS VARIABLES sobre 3+ grupos? Si es una sola variable, entonces usamos ANOVA unifactorial.
- **Con muestras repetidas (mit Messwiederholung):** ¿Queremos medir efectos sobre UN MISMO GRUPO que ha sido medido 3+ veces (como el t-test pareado con una muestra, pero con 3+ mediciones sobre los mismos sujetos)? Si es así, entonces necesitamos una prueba con muestras repetidas.

Una vez que ya sabemos qué test realizar (en el caso de *vot2.csv* sería unifactorial) Entonces simplemente buscamos la función y luego ingresamos los datos, pulsando el botón que lleva a la hoja de cálculo y luego seleccionando todos los datos. Esta función no está disponible en el botón de funciones; solo es parte del botón *Datenanalyse*.

Aquí necesitamos reportar: el valor de F (Fig. 11, celda naranja), los grados de libertad dentro y entre grupos (celda azul, y celda blanca sobre la azul, respectivamente), y el valor p (celda verde). Como ya sabemos, si el valor de p es menor a 0.05, entonces tenemos una diferencia estadísticamente significativa.

Gruppen	Anzahl	Summe	Mittelwert	Varianz
Simple	20	591,030667	29,55153335	20,2598395
Complex	20	1714,0823	85,70411497	538,40768
Coda	20	1021,75273	51,08763634	90,4757502

ANOVA						
Streuungsursache	dratsummen	ihheitsgrade (re	Quadratsumme	Prüfgröße (F)	P-Wert	itischer F-Wert
Unterschiede zw:	32101,445	2	16050,72249	74,1780278	1,3676E-16	3,15884272
Innerhalb der Gr:	12333,7221	57	216,3810899			
Gesamt	44435,1671	59				

	Dif. Prom	n1	n2	SE	q
simple-Complex	56,1525816	20	20	3,28923312	17,0716333
simple-coda	21,536103	20	20	3,28923312	6,54745414
complex-coda	34,6164786	20	20	3,28923312	10,5241792

Figura 11: Resultados de la prueba ANOVA.

4.2. Prueba post-hoc: Tukey-Kramer

¿Qué hacemos si la prueba ANOVA es estadísticamente significativa? ¡Recordemos que la prueba ANOVA sólo nos dice que hay una diferencia entre los tres grupos, pero no nos dice entre cuáles! Entonces ahora tenemos que encontrar qué pares son diferentes. Para eso usamos un prueba Tukey-Kramer de diferencias significativas (Ver tabla inferior, Fig. 11). Vamos a dividir el cálculo en dos pasos:

- Calculamos el error estándar de un par, usando esta fórmula. Se puede ver escrita en la barra de fórmulas en la Fig. 11 :

$$SE = \sqrt{\frac{1}{2}MS_w\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

Donde:

- MS_w es el cuadrado promedio (*Mittlere Quadratsumme*) entre grupos.
 - n_i es el número de observaciones (N) de uno de los grupos.
 - n_j es el número de observaciones (N) del otro grupo.
- Calculamos el *valor q*, con esta fórmula:

$$q = |M_1 - M_2|/SE$$

...y luego comparamos el valor en una tabla de valores críticos de q (Ver adjunto), buscando en la primera columna los grados de libertad (Fig. celda azul) y en la primera fila el número de grupos. Si nuestro valor q es mayor al valor crítico, entonces el par que analizamos es diferente. Procedemos a hacer lo mismo con todos los pares y así sabremos cuál es el correcto. En este caso, todos los pares son estadísticamente significativos.

Finalmente, reportamos. En el caso de ANOVA, debemos dar los grados de libertad entre grupos (Unterschiede zwischen den Gruppen) y dentro del grupo (Innerhalb der Gruppen).

Se realizó una prueba ANOVA unifactorial para verificar si existen diferencias entre el VOT de ataques simples, complejos, y en coda. Los resultados muestran que la diferencia es estadísticamente significativa ($F= 74.17 (2, 57), p<0.05$). Se realizó un test post-hoc Tukey-Kramer, el cual demostró que las tres condiciones presentan valores de VOT diferentes entre sí.

4.3. Graficando resultados

Este caso es exactamente igual que el anterior. Podemos usar el gráfico de cajas, o bien un gráfico de barras. Nuevamente, es importante tener las barras de error. NOTA: también se puede usar el error estándar o la desviación estándar para los gráficos de datos analizados con t-test o ANOVA (no necesariamente el margen de error del 5%). Eso sí, es importante mencionar qué presentan las barras de error.

5. Variables categóricas

Hasta ahora, hemos asumido que la variable dependiente (i.e. la que cambia de acuerdo a una condición) siempre es numérica y continua. Pero qué pasa cuando tenemos simplemente cuentas? Por ejemplo, Si le hemos pedido a un grupo de personas que escuchen dos sonidos y digan si esos sonidos son iguales o diferentes, entonces solo tendremos un

número de respuestas “igual” y otro número de respuestas “diferente”. En `discr.csv` tenemos las respuestas igual y diferente de un grupo de hablantes de español que escucharon las vocales [y] y [ø] separadas por un intervalo de 1200 ms (largo), y luego una vez más pero separadas por un intervalo de 300 ms (corto). Queremos saber si este intervalo entre los estímulos (*Interstimulus Interval* o ISI) juega un papel en la discriminación de estos sonidos en hablantes no nativos.

5.1. Cálculo de Chi cuadrado

Para este tipo de datos necesitamos calcular el **Chi cuadrado**. Hacemos con estas respuestas una **tabla de contingencia** como la siguiente (Tabla 1).

Tabla 1: Tabla de contingencia para calcular Chi cuadrado.

	Largo	Corto	Total
Igual	35	15	50
Diferente	5	25	30
Total	40	40	80

Para hacer nuestra tabla, Usaremos la función ZÄHLENWENN. Hacemos clic en una celda vacía (p.ej. en H17) y escribimos en ella lo siguiente:

```
=ZÄHLENWENN(C2:C41, ‘‘igual’’)
```

Y apretamos Enter, lo cual nos dará el número de respuestas “igual” dadas por los sujetos cuando el ISI es largo (Fig. 6). Ahora, en la celda inmediatamente debajo de donde hicimos esto (H18), escribimos:

```
=ZÄHLENWENN(C2:C41, ‘‘diferente’’)
```

Y apretamos Enter. Con esto tenemos el número de respuestas “diferente” cuando el ISI es largo. Luego, hacemos clic en la celda inmediatamente debajo (H19) y luego hacemos clic en el botón de sumatoria (Fig. 7). Eso sumará automáticamente las celdas superiores. Con esto tenemos la primera columna de la tabla.

Para generar la columna siguiente, solamente necesitamos repetir el proceso, pero considerando el rango de celdas C42:C81, que es donde están las respuestas con ISI corto.

Ahora, para calcular el Chi cuadrado necesitamos crear una segunda tabla con los **valores esperados** para cada respuesta. Esto podemos hacerlo de la siguiente manera: Seleccionamos cuatro celdas vacías (p.ej. M17, N17, M18 y N18) y vamos a la barra de fórmulas, donde escribimos:

```
=MMULT(J17:J18, H19:I19)/J19
```

Y luego apretamos **Ctrl + Shift + Enter**. Esto nos dará una tabla con los valores:

15	15
25	25

Lo que hicimos es un cálculo de probabilidad si asumimos H_0 , y como podemos ver, este cálculo asume que da igual si el ISI es corto o largo. Podemos también hacer que la

	Largo	Corto	Total
Igual	15	15	30
Diferente	25	25	50
Total	40	40	80

tabla se vea de la siguiente manera (y así sabemos qué números son éstos cuando volvemos a trabajar después de un recreo):

Ahora, por fin, calculamos el Chi cuadrado. Vamos a una celda vacía (p.ej. J23), luego al botón de funciones, y buscamos CHIQU.TEST. En el campo de valores obtenidos (*Beob. Messwerte*) seleccionamos las celdas correspondientes, y luego hacemos lo mismo en el campo de valores esperados (*Erwart. Werte*). En la celda aparecerá nuestro valor p, que debería ser: 3.85962E-06 ($=3.85963^{-6} = 0.00000385963$). Claramente, el valor es menor a 0.05, que es el valor p que asumimos como significativo. Por lo tanto, podemos rechazar H_0 .

Para reportar esto en el informe, necesitaremos también calcular el valor de Chi cuadrado en sí mismo. Lo hacemos haciendo clic en una celda vacía y escribimos lo siguiente:

=CHIQU.INV.RE(J23,1)

Lo cual nos dará un resultado de 21.333. Pero qué fue lo que acabamos de hacer? Simple: J23 es la celda donde tenemos el valor p. Luego le hemos dado a Excel los grados de libertad, que en este caso es 1 (Cuando tenemos una tabla de contingencia, se calcula: número de columnas -1 * Número de filas -1).

Entonces, el reporte de la prueba se hace así:

Se calculó una prueba de Chi cuadrado comparando las respuestas a estímulos de ISI largo e ISI corto. La prueba arrojó resultados significativos ($\chi^2(1) = 21.33$, $p < 0.05$): los pares de estímulos con ISI de 300 ms tienen mayor probabilidad de ser percibidos como diferentes que los pares con ISI de 1200 ms.

5.2. Graficando variables categóricas

La manera más sencilla de graficar este tipo de información es con un gráfico de columnas apiladas *Gestapelte Säulen*. Simplemente seleccionamos nuestra tabla de contingencia con los valores observados, y luego vamos a *Einfügen > Empholene Diagramme > Gestapelte Säule (100%)*. Esto nos dará un gráfico prácticamente listo con los valores en forma de porcentajes. Ahora solo tenemos que poner los nombres a los ejes y un título.

6. Algunos datos útiles

1. Las funciones en Excel pueden ejecutarse a través de cuadros de diálogo (como lo hicimos con la R de Pearson) o escribiendo directamente la función en la celda (como lo hicimos con Chi cuadrado).
2. La versión 2016 de Office tiene todas las funciones estadísticas en el botón *Datennalyse*. Esto incluye t-test, Anova, R de Pearson, etc. y además nos da el valor de la prueba estadística, los valores p, los grados de libertad, etc. Es muy útil, pero lamentablemente las versiones anteriores de Excel no lo tienen.

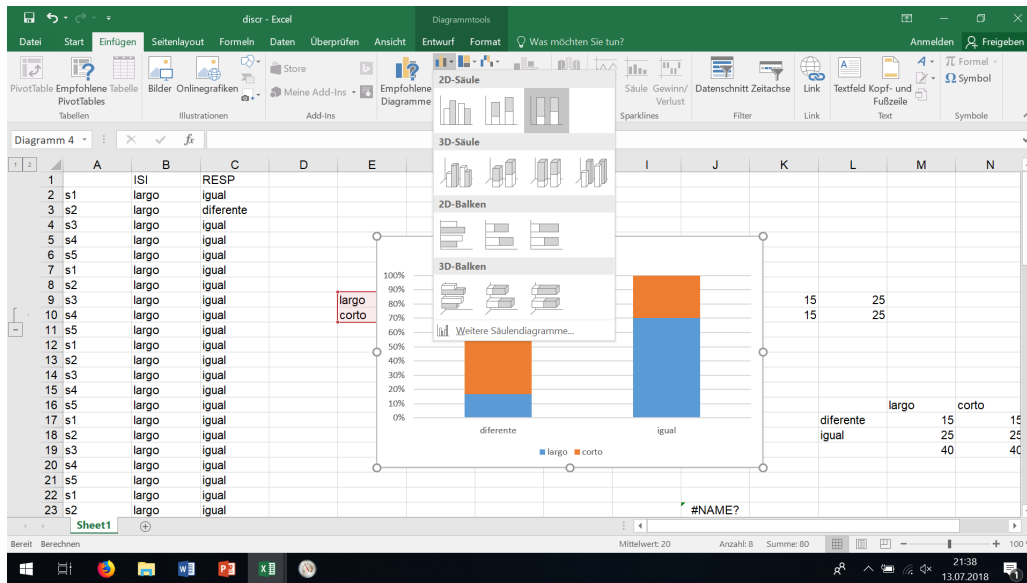


Figura 12: Gráfico de barras apilado en porcentajes.

3. Si usas Excel en Windows, se puede descargar el add-on Analysis Toolpak (*Analyse-Funktionen*). Lamentablemente, esto no está disponible para Excel Mac 2011.
4. Las pruebas estadísticas t-test y Anova asumen un par de supuestos que no hemos analizado aquí:
 - ¿Tienen los datos una distribución normal?
 - ¿Tienen los datos igual varianza?

...pero esto lo veremos caso por caso.
5. En estos tiempos donde todo está en Internet, podemos googlear algo como *wie kann ich ein Boxplotdiagramm in Excel erstellen* y encontraremos cientos de páginas, tutoriales en Youtube, etc.
6. **R** es un programa de libre acceso (open source) y aunque parece complicado, hace las cosas muy simples una vez que entendemos cómo funciona. Si tienes tiempo e interés en el análisis de datos, esta alternativa es siempre la mejor. Otras alternativas (pagada\$\$\$) son SPSS y SAS.
7. El gráfico de cajas era lo ideal para los datos en vot.csv. Si de verdad deseas hacerlo en una versión de Excel anterior a 2016, aquí hay un buen tutorial:

<https://www.youtube.com/watch?v=jaXwtPKUGL8>