ESTADISTICA.AR

Comparar grupos

Estadística · Semana 11 · Unidad 6: Pruebas de hipótesis

1 / 23

NOTAS · 1

Pensada para 90 a 120 minutos. La lógica es la de la semana pasada; lo nuevo es la pregunta: ya no si un parámetro vale tanto, sino si dos grupos son distintos, o si dos variables están relacionadas.

PREGUNTA PARA LA CLASE

En la EPH de Gran Catamarca, quienes tienen secundario completo ganan en promedio $779,5 mil, y quienes tienen superior completo, $1.332,8 mil. ¿Alcanza con comparar las dos medias para decir que en la población ganan distinto?

2 / 23

NOTAS · 2

No: las medias de muestra varían de muestra en muestra, y una diferencia podría ser casualidad. Hace falta saber cuánta variabilidad hay dentro de cada grupo y cuántos casos se tienen: es lo que hace la prueba t.

Que arriesguen si esta diferencia parece casualidad. La respuesta en la parte 2: t = 5,10, una diferencia imposible de explicar por azar.



NOTAS · 3

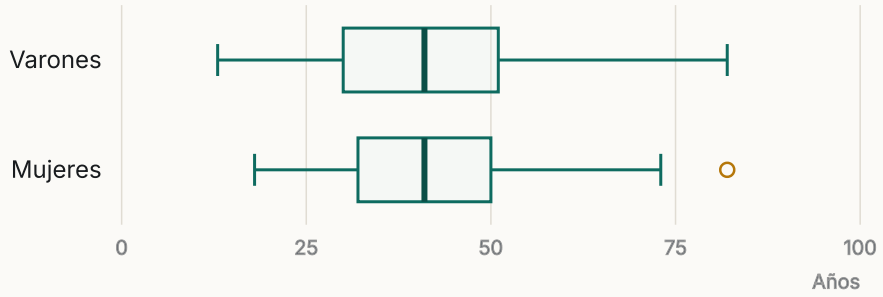
Las tres pruebas de hoy (t, Welch, chi-cuadrado) son las que más aparecen en los artículos de cualquier disciplina.



SEMANA 11 · DOS MEDIAS

La edad de las personas ocupadas

Personas ocupadas encuestadas en Gran Catamarca, por sexo.



Fuente: INDEC, Encuesta Permanente de Hogares, primer trimestre de 2026. Gran Catamarca.

5 / 23

NOTAS · 5

Antes de cualquier prueba, el gráfico: las dos cajas son casi iguales en posición y en ancho. Lo que se espera es que la prueba no encuentre diferencia.

SEMANA 11 · DOS MEDIAS

Los resúmenes de cada grupo

Grupo	n	Media (años)	Desvío (años)
Varones	347	40,97	12,87
Mujeres	284	41,36	12,03

H_0: $\mu_1 = \mu_2$ (la edad media de varones y mujeres ocupados es la misma); H_1: $\mu_1 \neq \mu_2$.

6 / 23

NOTAS · 6

Las hipótesis hablan de las medias de la población, no de estas dos: estas ya se sabe que son distintas. La pregunta es si la diferencia de 0,39 años se puede explicar por el azar del muestreo.

SEMANA 11 · DOS MEDIAS

La prueba t con varianza combinada

La varianza combinada

$s^2p = (346 \cdot 12{,}87^2 + 283 \cdot 12{,}03^2) / 629 = 156{,}23$

Un promedio de las dos varianzas, pesado por los grados de libertad.

El error estándar

$EE = sp \cdot \sqrt{(1/347 + 1/284)} = 1{,}000$

El estadístico

$t = (40{,}97 - 41{,}36) / 1{,}000 =$ **−0,39**

Con 629 grados de libertad: p = 0,70.

7 / 23

NOTAS · 7

$|t| = 0{,}39$ es mucho menor que el valor crítico 1,964: no se rechaza H_0. Los datos son compatibles con que varones y mujeres ocupados tengan la misma edad media.

El intervalo del 95 % para la diferencia va de −2,35 a 1,57 años: contiene al cero, y dice además que, si hay diferencia, es de menos de dos años y medio.

SEMANA 11 · DOS MEDIAS

El supuesto de varianzas iguales

Para qué hace falta

La varianza combinada supone que los dos grupos tienen la misma varianza en la población.

Cómo se revisa

Mirando las cajas, comparando los desvíos o con la prueba de Levene.

Una regla práctica: que el desvío mayor no duplique al menor.

Acá

12,87 / 12,03 = 1,07: se cumple con holgura.

8 / 23

NOTAS · 8

Los otros supuestos: muestras aleatorias e independientes entre sí, y datos normales en cada grupo o muestras grandes. Con cientos de casos, el teorema central del límite se ocupa de la normalidad.

2

La prueba de Welch

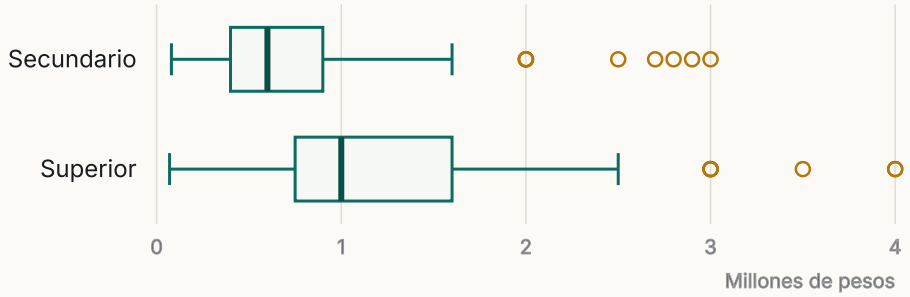
Sin suponer varianzas iguales

9 / 23

SEMANA 11 · LA PRUEBA DE WELCH

El ingreso según el nivel educativo

Ingreso de la ocupación principal de las personas ocupadas que lo declararon, con secundario completo y con superior completo.



Fuente: INDEC, Encuesta Permanente de Hogares, primer trimestre de 2026. Gran Catamarca. Otros 6 ingresos atípicos, de más de 4 millones, no entran en el gráfico.

10 / 23

NOTAS · 10

La caja del superior completo está más a la derecha: su mediana, 1.000 mil pesos, supera al tercer cuartil del secundario, 900 mil. Más de la mitad de quienes tienen superior completo gana más que tres de cada cuatro con secundario. Los dos grupos son asimétricos, con atípicos altos, como todo ingreso.

SEMANA 11 · LA PRUEBA DE WELCH

Los resúmenes, en miles de pesos

Grupo	n	Media	Mediana	Desvío
Secundario completo	217	779,5	600	1.089,7
Superior completo	163	1.332,8	1.000	1.012,2

Fuente: INDEC, Encuesta Permanente de Hogares, primer trimestre de 2026. Gran Catamarca.

11 / 23

NOTAS · 11

La media queda bastante por encima de la mediana en los dos grupos: la asimetría que se vio en la semana 3. Con estos tamaños de muestra la prueba t sigue siendo válida para comparar medias, por el teorema central del límite.

SEMANA 11 · LA PRUEBA DE WELCH

La prueba de Welch

Sin suponer varianzas iguales

EE = $\sqrt{(s_1^2/n_1 + s_2^2/n_2)}$ = 108,4 mil pesos

Cada grupo con su propia varianza.

El estadístico

t = (1.332,8 − 779,5) / 108,4 = **5,10**

Con unos 361 grados de libertad, de una fórmula aproximada.

La decisión

$p < 0{,}001$: se rechaza H_0.

Una diferencia así no se explica por azar.

12 / 23

NOTAS · 12

Los grados de libertad de Welch salen de una fórmula que no hace falta saber: la calcula el software, y por eso suelen tener decimales.

Muchos programas (R entre ellos) usan Welch por defecto: si las varianzas son iguales da casi lo mismo que la t combinada, y si no lo son, protege. Es una buena opción por defecto.

Acá los desvíos son parecidos (1.089,7 y 1.012,2), y la t combinada daría casi el mismo resultado. Welch importa cuando un grupo es mucho más disperso que el otro y los tamaños son distintos.

3

Dos variables cualitativas

La prueba chi-cuadrado de independencia

13 / 23

SEMANA 11 · DOS VARIABLES CUALITATIVAS

La tabla de la semana 4, otra vez

Personas de 25 a 64 años: observadas y, entre paréntesis, esperadas si el nivel educativo y la ocupación fueran independientes.

Nivel educativo	Ocupada	No ocupada	Total
Hasta primario completo	58 (68,6)	33 (22,4)	91
Secundario, incompleto o completo	259 (266,2)	94 (86,8)	353
Superior, incompleto o completo	244 (226,2)	56 (73,8)	300
Total	561	183	744

Esperada = total de la fila · total de la columna / total. Por ejemplo, 91 · 561 / 744 = 68,6.

14 / 23

NOTAS · 14

En la semana 4 se vio que el porcentaje de ocupados crece con el nivel educativo. La pregunta de hoy: ¿esa relación existe en la población, o puede ser casualidad de la muestra?

Las esperadas reparten a cada nivel la misma proporción de ocupados que el total: si las variables fueran independientes, el nivel no cambiaría esa proporción.

SEMANA 11 · DOS VARIABLES CUALITATIVAS

La prueba chi-cuadrado de independencia

Hipótesis

H_0: las dos variables son independientes. H_1: están relacionadas.

Estadístico

$\chi^2 = \Sigma$ (observada − esperada)^2 / esperada

Suma la distancia entre la tabla real y la de independencia.

Grados de libertad

(filas − 1) · (columnas − 1) = 2 · 1 = 2

Condición

Todas las esperadas de 5 o más.

Acá la menor es 22,4.

15 / 23

NOTAS · 15

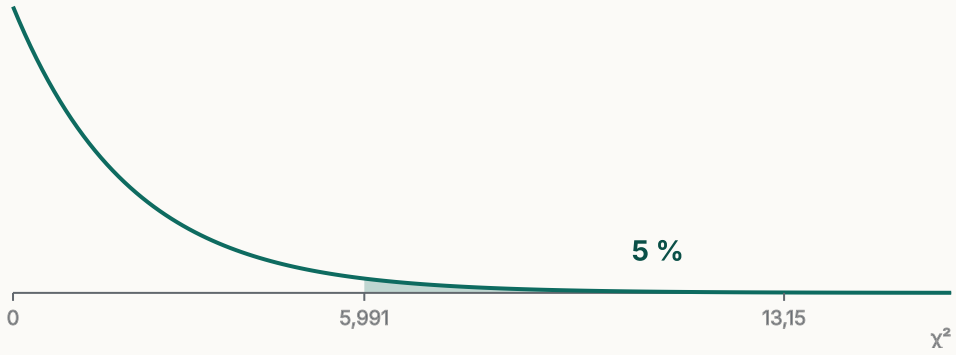
Un χ^2 grande quiere decir que la tabla observada está lejos de la de independencia. Cero sería una tabla idéntica a la esperada.

Con esperadas chicas, la aproximación chi-cuadrado falla: se juntan categorías o se usa la prueba exacta de Fisher (la del té, generalizada).

SEMANA 11 · DOS VARIABLES CUALITATIVAS

La distribución chi-cuadrado con 2 grados de libertad

Con $\alpha = 0{,}05$, el valor crítico es 5,991. El χ^2 de la tabla es 13,15.



$p = P(\chi^2 \geq 13{,}15) = 0{,}0014$: se rechaza la independencia.

16 / 23

NOTAS · 16

A diferencia de la normal y la t, la chi-cuadrado es asimétrica y solo toma valores positivos. La prueba es siempre de cola derecha: cualquier alejamiento de la independencia agranda χ^2.

Con 2 grados de libertad el p-valor tiene una fórmula exacta: e^(−χ^2/2) = e^(−13,15/2) = 0,0014. En general sale de la tabla o de la calculadora.

4

Significativo e importante

El tamaño del efecto

17 / 23

SEMANA 11 · SIGNIFICATIVO E IMPORTANTE

La asociación es real, pero débil

$\chi^2 = 13{,}15$ con $p = 0{,}0014$: muy significativo. ¿Qué tan fuerte es la relación?

$V = 0{,}133$

la V de Cramér: $\sqrt{(\chi^2 / (n \cdot (k - 1)))} = \sqrt{(13{,}15 / 744)}$

0 a 1

va de independencia total a asociación perfecta; menos de 0,2 es débil

18 / 23

NOTAS · 18

El p-valor depende del tamaño de muestra: con 744 personas, una asociación débil ya es significativa. La V de Cramér no depende de n: mide la fuerza.

Débil no quiere decir irrelevante: 18 puntos de diferencia en la tasa de ocupación entre el nivel más bajo y el más alto importan para una política de empleo. La V resume toda la tabla en un número, y por eso conviene mirar también los porcentajes.

SEMANA 11 · SIGNIFICATIVO E IMPORTANTE

El tamaño de la diferencia de ingresos

340 a 767

miles de pesos más por mes con superior completo: el intervalo del 95 % para la diferencia de medias

d = 0,52

la d de Cohen: la diferencia medida en desvíos estándar; alrededor de 0,5 es un efecto mediano

19 / 23

NOTAS · 19

El p-valor dice que hay diferencia; el intervalo dice cuánta, y eso es lo que importa para decidir. Informar solo «p < 0,001» esconde lo más útil.

Es una diferencia observacional: no dice cuánto ganaría una persona si terminara la universidad. Los que estudian más pueden diferir en muchas otras cosas, como se vio en la semana 4.

SEMANA 11 · SIGNIFICATIVO E IMPORTANTE

Cuidado con las comparaciones múltiples

Una prueba

Con $\alpha = 0{,}05$, si no hay ningún efecto, la probabilidad de un falso positivo es 5 %.

Veinte pruebas

La probabilidad de al menos un falso positivo es $1 - 0{,}95^{20} =$ **64 %**.

El remedio

Plantear antes qué se va a comparar e informar todas las pruebas hechas.

O ajustar α, como hace Bonferroni: α / cantidad de pruebas.

20 / 23

NOTAS · 20

Comparar por sexo, por edad, por nivel, por barrio... hasta que algo da significativo, y publicar solo eso, se llama p-hacking. Un resultado así es casi seguro un falso positivo.

El simulador de p-hacking de estadistica.ar muestra cómo, sin ningún efecto real, siempre aparece algo «significativo» si se buscan suficientes comparaciones.

5

Qué prueba usar

Según la pregunta y las variables

21 / 23

SEMANA 11 · QUÉ PRUEBA USAR

Qué prueba usar

Pregunta	Variables	Prueba
¿Una media vale tanto?	Una cuantitativa	t para una media
¿Una proporción vale tanto?	Una cualitativa	z para una proporción
¿Dos grupos tienen la misma media?	Cuantitativa y cualitativa de 2 grupos	t para muestras independientes (Welch)
¿Cambió de antes a después, en las mismas personas?	Dos mediciones apareadas	t para datos apareados
¿Dos variables cualitativas están relacionadas?	Dos cualitativas	Chi-cuadrado de independencia
¿Tres o más grupos tienen la misma media?	Cuantitativa y cualitativa de 3 o más grupos	ANOVA (semana 12)
¿Dos cuantitativas están relacionadas?	Dos cuantitativas	Correlación y regresión (semana 12)

Con muestras chicas y datos muy asimétricos, hay versiones no paramétricas: Mann-Whitney en lugar de la t, por ejemplo.

22 / 23

NOTAS · 22

La clave es la misma de la semana 4: qué tipo de variables hay y qué pregunta se hace. La herramienta «Qué prueba usar» de estadistica.ar guía la elección con preguntas.

Los datos apareados son un caso que se confunde seguido: si las dos mediciones son de las mismas personas, no son muestras independientes, y la prueba correcta trabaja con las diferencias.

SEMANA 11 · COMPARAR GRUPOS

Para llevarse

- Para comparar dos medias, la t para muestras independientes; la de Welch no supone varianzas iguales.
- Para dos variables cualitativas, la chi-cuadrado compara la tabla observada con la de independencia.
- El p-valor dice si hay diferencia; el intervalo y el tamaño del efecto, cuánta.
- Muchas comparaciones sin plan producen falsos positivos.
- La prueba se elige por la pregunta y por el tipo de variables.

Guía de trabajo práctico 11
estadistica.ar/docentes/trabajos-practicos/semana-11

Qué prueba usar
estadistica.ar/herramientas/selector-de-prueba

Calculadora de chi-cuadrado
estadistica.ar/herramientas/calc-chi-cuadrado

Calculadora de prueba t
estadistica.ar/herramientas/calc-prueba-t

23 / 23

NOTAS · 23

Para el práctico: las tablas de la t y de chi-cuadrado, y calculadora. El ejercicio con software usa la EPH.