Analítica de datos

Intervalos de confianza y pruebas de hipótesis

Pontificia Universidad JAVERIANA Bogotá

Profesor: Nicolás Velásquez

Pruebas de hipótesis

¿Qué es una hipótesis?

- Es una afirmación sobre un(os) parámetro(s) poblacional(s) (aquí la media poblacional):

Ejemplo: La media del gasto mensual en servicios de telefonía celular en la ciudad es μ = $42mil al mes.

Hipótesis nula, H_0

- Especifica la afirmación que se quiere poner a prueba (queremos ver si hay suficiente evidencia para descartarla)

Ejemplo: La media del diámetro de los tornillos fabricados es 30mm ($H_0 : \mu = 30$).

- ¡Es sobre un parámetro poblacional, no un estadístico muestral!

$H_0: \mu = 30$

$\cancel{H_0: \bar{X} = 30}$

Hipótesis nula, H_0

- Comenzamos con el supuesto de que la hipótesis nula es cierta y queremos saber si hay suficiente evidencia para rechazarla (a favor de la alternativa).

- Siempre contiene signos “=“, o “≤”, o “≥”.

- Se rechaza o no se rechaza.

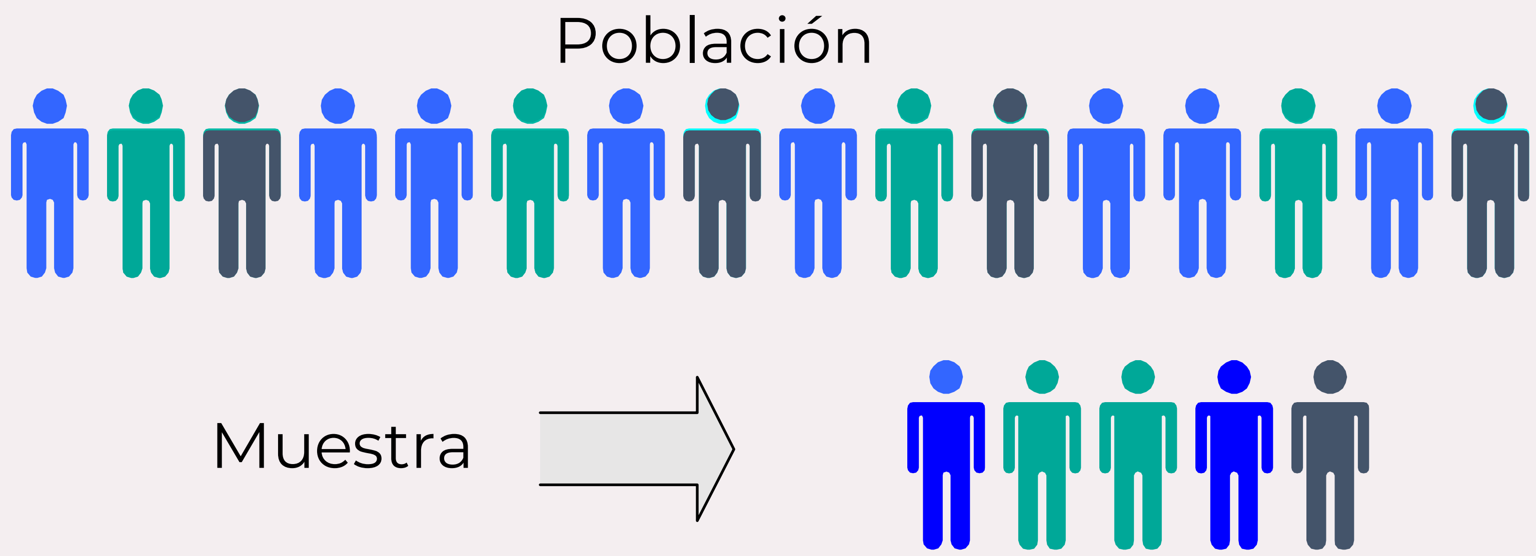
Hipótesis alternativa, H_1

- Lo contrario a la nula.
  - p.e., La media del diámetro de los tornillos fabricados no es 30mm(H_1: $\mu \neq 30$).

- En general, es la hipótesis que “queremos” probar cierta.

El proceso de prueba de hipótesis

- “La media (poblacional) de la edad es 50”.
  - H_0: $\mu = 50$, H_1: $\mu \neq 50$
- Tomamos una muestra y construimos la media muestral.



El proceso de prueba de hipótesis

- Supongamos que resulta que la media muestral $\overline{X} = 20$.
- Esto es mucho más bajo que los 50 de la hipótesis nula.
- Si la hipótesis nula fuera cierta, la probabilidad de observar un número tan alejado de 50 (osea 20) sería muy baja.
- Rechazamos la hipótesis nula, a favor de la alternativa.

¿Cómo decidimos si rechazar o no la hipótesis nula?

- Vamos a utilizar un número, que se llama p-value, y nos dice el riesgo que corremos de rechazar, incorrectamente, la hipótesis nula.

  - Entre más bajo el p-value, menor el riesgo de cometer un error si rechazamos la hipótesis nula.

  - Los límites convencionales son: 0,01 (1%) 0,05 (5%) y 0,10 (10%).

  - Si el p-value es menor al límite de riesgo que nos impusimos, entonces rechazamos la hipótesis nula.

Posibles errores en prueba de hipótesis

Posibles resultados del test o prueba		
	Situación real	
Decisiones	H_0 Verdadera	H_0 Falsa
No rechazamos H_0	Decisión correcta	Error tipo II
Rechazamos H_0	Error tipo I	Decisión correcta

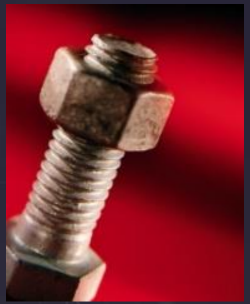
Ho: You are not pregnant.

Ejemplo

Queremos saber si hay evidencia para decir que el diámetro de los tornillos es distinto a 30mm.

1. Determinar hipótesis nula y alternativa:
   - H_0: $\mu = 30$ H_1: $\mu \neq 30$ (test de dos colas).

2. Especificar el nivel de significancia (o tu nivel de tolerancia al riesgo)
   - Por ejemplo 0.05.



Ejemplo

Obtenemos el p-value.

- Si p-value < 0,05, rechazamos la nula ☺
- Si p-value > 0,05 , no la rechazamos ☹

Pruebas de Hipótesis

La prueba de hipótesis se pude plantear de diversas formas:

- La media es distinta (2 colas):

$$\boxed{\begin{array}{l} H_0: \mu = 30 \\ H_1: \mu \neq 30 \end{array}}$$

- La media es mayor o menor que (1 cola):

$$\boxed{\begin{array}{l} H_0: \mu \geq 30 \\ H_1: \mu < 30 \end{array}} \qquad \boxed{\begin{array}{l} H_0: \mu \leq 30 \\ H_1: \mu > 30 \end{array}}$$

Ejemplo de prueba de Hipótesis

Hipótesis nula:

- **Gerente**: El cliente promedio realiza 30 transacciones por mes

Planteamiento de Hipótesis:

$H_0: \mu_T = 30$

$H_1: \mu_T \neq 30$

Tomo una muestra de 1000 clientes y obtengo los estadísticos:

$\bar{X} = 27.2$

$S = 5.5$

Defino el nivel de riesgo:

$\alpha = 5\%$

Calculamos el p-value :

- Si p-value < 0,05, rechazamos la nula
- Si p-value > 0,05 , no la rechazamos

Ejemplo de prueba de Hipótesis

$$H_0 : \mu = 30 \quad \text{vs} \quad H_1 : \mu \neq 30,$$

con $\bar{x} = 27.2$ y $s = 5.5$.

La estadística es

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{27.2 - 30}{5.5/\sqrt{n}},$$

Ejemplo de prueba de Hipótesis

y el **p-valor** (bilateral) es $p = 2\left[1 - F_{t_{n-1}}(|t|)\right].$

Dependiendo de la muestra n se define el p-value:
Asumiendo que n=30 p-value = 0.000926

$$t = \frac{27.2 - 30}{5.5/\sqrt{30}} = \frac{-2.8}{5.5/5.477} \approx \frac{-2.8}{1.004} \approx -2.79$$

- $n = 10$: $t = -1.61$, $p \approx 0.142$
- $n = 15$: $t = -1.97$, $p \approx 0.069$
- $n = 20$: $t = -2.28$, $p \approx 0.0346$
- $n = 30$: $t = -2.79$, $p \approx 0.00926$
- $n = 50$: $t = -3.60$, $p \approx 0.000741$
- $n = 100$: $t = -5.09$, $p \approx 0.00000170$

Ejemplo de prueba de Hipótesis

Con $gl = 29$, el p-valor bilateral es:

$$p \approx 0.0093$$

A cualquier nivel de significancia tradicional 10%, 5% y 1%, se rechaza la hipótesis nula en favor a la hipótesis alterna por lo que hay **evidencia estadistica para decir que el diámetro de los tornillos es distinto a 30mm.**

Ahora veremos prueba de hipótesis para dos grupos….

Queremos saber si hay evidencia suficiente para decir que las medias (poblacionales) de los dos grupos son distintas.

Pruebas con dos muestras

Pruebas con dos muestras/grupos (AyB)

Muestras indep.

Muestras relac.

Grupo 1 vs. Grupo 2

Mismo grupo antes y después.

$$H_0: \mu_A - \mu_B = 0$$
$$H_1: \mu_A - \mu_B \neq 0$$

La hipótesis nula es que ambos grupos son iguales (en su media)

Muestras independientes y relacionadas

- Independientes: medición de la misma unidad antes y después de alguna intervención/suceso.

- Relacionadas: medición de unidades en distintos grupos.

¿Ventajas y desventajas de muestras relacionadas vs independientes?

Muestras relacionadas

- Medimos antes y después.
- Construimos una nueva variable, que es la diferencia entre ambas mediciones:

$$\boxed{D_i = X_{1i} - X_{2i}}$$

- Si ambos grupos son iguales entonces la diferencia será 0 en promedio. Si son distintos será distinta de 0. Es decir:

$$H_0: \mu_D = 0$$
$$H_1: \mu_D \neq 0$$

Ejemplo

Suponga que se quiere medir si los tiempos de entrega de una pizzería local son menores que los de una gran cadena. Los tiempos de entrega para 10 pedidos **en 10 horas diferentes** son los siguientes:

Hora	1	2	3	4	5	6	7	8	9	10
Local	16.8	11.7	15.6	16.7	17.5	18.1	14.1	21.8	13.9	20.8
Cadena	22	15.2	18.7	15.6	20.8	19.5	17	19.5	16.5	24

Utilizando un nivel de significancia del 5% ¿Existe evidencia de que la media del tiempo de entrega es menor para el restaurante local de pizzas que para la cadena nacional **en cada una de las horas**?

Ejemplo

Suponga que se quiere medir si los tiempos de entrega de una pizzería local son menores que los de una gran cadena. Los tiempos de entrega para 10 pedidos **en 10 horas diferentes** son los siguientes:

Hora	1	2	3	4	5	6	7	8	9	10
Local	16.8	11.7	15.6	16.7	17.5	18.1	14.1	21.8	13.9	20.8
Cadena	22	15.2	18.7	15.6	20.8	19.5	17	19.5	16.5	24

$$H_0 : \mu_{Local} - \mu_{Cadena} = 0 \quad \text{vs} \quad H_1 : \mu_{Local} - \mu_{Cadena} < 0$$

Ejemplo

```r
# Cargar la biblioteca necesaria
library(tibble) #Es de tidyverse

# Crear el dataframe para usar ahora
df <- tibble(
Local = c(16.8, 11.7, 15.6, 16.7, 17.5, 18.1, 14.1, 21.8, 13.9, 20.8),
Cadena = c(22, 15.2, 18.7, 15.6, 20.8, 19.5, 17, 19.5, 16.5, 24)
)

# Mostrar el marco de datos
print(df)
```

```r
# Realizar la prueba t para muestras relacionadas
resultado <- t.test(df$Local, df$Cadena, paired = TRUE, alternative = "less")

# Mostrar el resultado de la prueba
print(resultado)
```

TAREA

- Analizando datos históricos de retornos diarios de seis acciones entre 2020 y 2025, clasificadas en dos sectores: **Tecnología** (AAPL, NVDA, MSFT) y **Retail/Bebidas** (KO, PEP, WMT). Se tiene la sospecha de que el sector tecnologia tiene en promedio más retornos diarios que el sector Retail.

Preguntas

- Genere una tabla descriptiva con promedio de retorno diario,desviacion estandar diaria y promedio de retorno anual(ojo solo hay 252 dias habiles de la bolsa) ¿Cuál sector muestra un **promedio más alto** de retornos diarios y precios promedio?
- Realizando una prueba t, ¿hay evidencia estadística de que el promedio de retornos diarios es mayor para el sector de Tecnología que el de Retail?
- Realizando un analisis gráfico, ¿qué sector presenta una **mayor dispersión** en los retornos diarios? Sugerencia: Use un boxplot
- ¿Cuál sector presenta una **volatilidad diaria** más alta, y qué podría implicar esto para el riesgo de inversión?

Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment

Cohn, Fehr & Goette (MSci, 2015)

Introducción

- No es muy aventurado pensar que existen trabajadores que reciprocan un salario más alto con más esfuerzo.

- Sin embargo, no se conocen muy bien los determinantes (ni la prevalencia) del esfuerzo recíproco.

- Este proyecto investiga el papel de las percepciones de justicia y preferencias sociales como determinantes de la decisión de reciprocar salarios más altos.

Contexto

- Proyecto en colaboración con una casa editorial durante el lanzamiento de un nuevo diario.

- Se promociona el diario repartiendo copias gratuitas a lo largo de 3 meses (Zurich)

- El papel de los trabajadores (temporales) es distibuir dichas copias.

Contexto

- Trabajadores tienen que indicar su disponibilidad en cuanto a días.
- Solamente de lunes a viernes, de 4p a 7p.
- El salario acordado es de CHF 22 por hora.
- Líderes de equipo:
  - Aseguran abasto de diarios
  - Aseguran que no se haga trampa
  - Registran cuantas copias distribuye cada vendedor en un turno.

Diseño experimental

- Control: trabajadores reciben salario regular por hora de CHF 22.
- Tratamiento: trabajadores reciben salario de CHF 27.

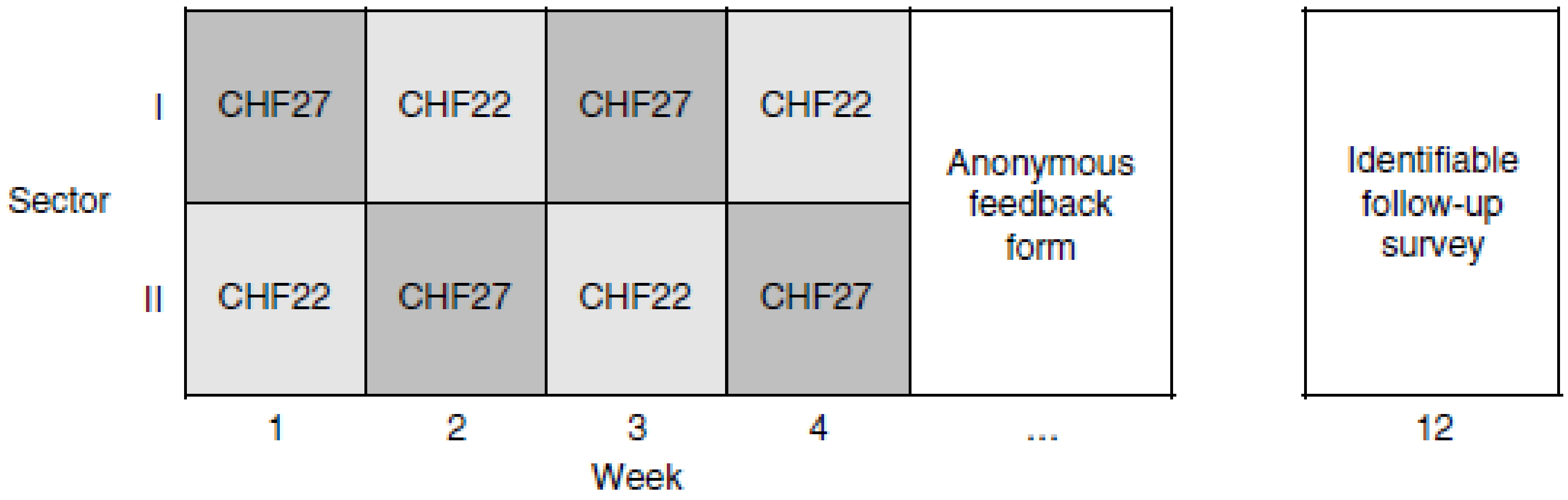
- Se les informa el salario a trabajadores poco antes de iniciar un turno mediante postal y mensaje de texto.

- Se analizan datos de 4 semanas.

- Se aleatoriza la aplicación del tratamiento a nivel semana-sector.

Diseño experimental

Figure 1 **Timing of Events**



¿Por qué?

Diseño experimental

- Tres meses después de haber terminando su participación , trabajadores reciben encuesta en la que se les pregunta <u>qué salario por hora consideran justo para cada uno del los trabajos realizados en los últimos 6 meses.</u>

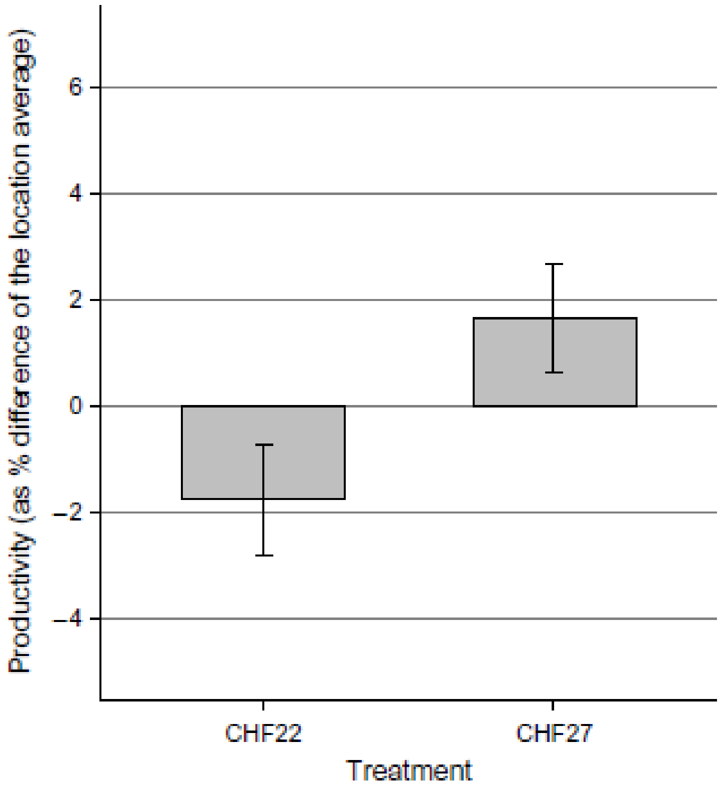
- Trabajadores y líderes de grupo no saben que están participando en un experimento.

Balance

	Treatment				
	CHF22		CHF27		
Variable	Mean	SD	Mean	SD	p-value
Perceived underpayment (in CHF)	1.088	2.099	1.081	2.142	0.69
Age (in years)	23.370	5.257	23.344	5.397	0.77
Male	0.281	0.450	0.267	0.443	0.68
Foreigner	0.161	0.368	0.172	0.378	0.70
Number of siblings	1.376	0.849	1.367	0.854	0.91
Secondary school	0.648	0.478	0.633	0.483	0.69
Apprenticeship/vocational school	0.331	0.471	0.308	0.462	0.52
Additional, further education	0.248	0.432	0.242	0.429	0.85
Baccalaureate	0.618	0.487	0.658	0.475	0.27
Technical school	0.251	0.434	0.211	0.409	0.22
University	0.245	0.431	0.211	0.409	0.29
Points returned if 1st mover proposed (18, 6)	−0.651	1.017	−0.663	1.004	0.98
Points returned if 1st mover proposed (12, 12)	0.251	0.726	0.248	0.684	0.88
Points returned if 1st mover proposed (6, 18)	0.811	0.902	0.857	0.891	0.46

Resultados

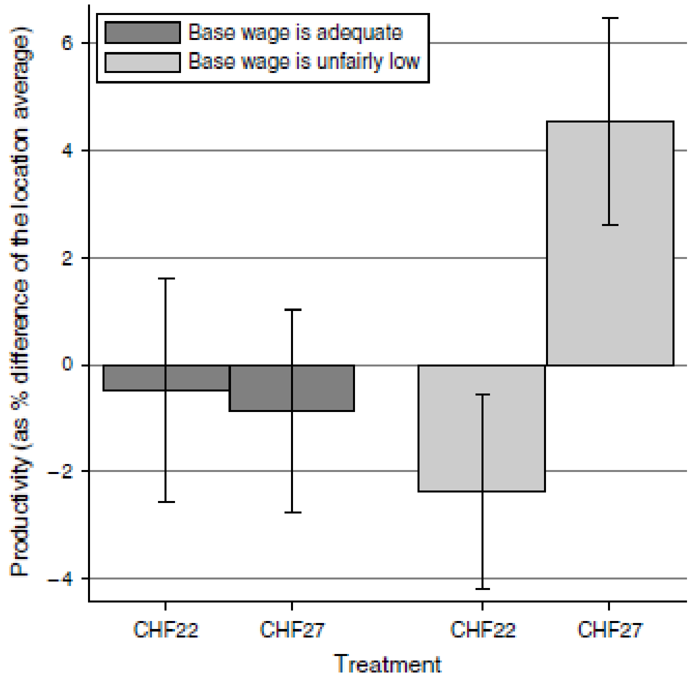
Figure 4 **Average Treatment Effect of the Wage Increase on Workers' Performance**



Notes. This figure shows workers' average performance and standard error of the mean in each treatment. Performance is measured as the hourly copies distributed normalized by the average hourly copies distributed at the location.

Resultados

Figure 5 **Heterogeneous Treatment Effect of the Wage Increase on Workers' Performance**



Notes. This figure displays workers' average performance and standard error of the mean in each treatment, shown separately for the workers who felt adequately paid at the base wage and those who felt underpaid. Workers' performance is the number of hourly copies distributed normalized by the average productivity at the location.