UNIVERSIDAD TECNOLOGICA DE LA MIXTECA
Labor et Sapientia Libertas
Chun quiaa Tiaha sihi Nillao
OAXACA

Maestría en Ciencia de Datos
Probabilidad y Estadistica para Ciencia de Datos

Ejercicios: Probabilidad

Profesora: Ana Delia Olvera Cervantes
Alumno: Melchor Nolasco Cosijoeza

31 de enero de 2025

Ejercicio 1

Los enanos de Blanca Nieves le informan que excavan más de 12 toneladas promedio por semana. Nieves recolecta datos de 49 semanas y obtiene =11.5, = 1.1 a un nivel de significancia =10%. Los Enanos, ¿están en lo cierto?

Datos

- **Media Poblacional:** 12 Toneladas
- **Media Muestral:** 11.5 Toneladas
- **Desviación estandar muestral:** 1.1 Toneladas
- **Tamaño de la muestra:** 49 Semanas
- **Nivel de significancia:** 10% (0.1)

1.- Planteamiento de Hipotesis

Hipotesis Nula: μ = 12 Toneladas (Los enanos excavan 12 toneladas en promedio)

Hipotesis Alternativa: μ > 12 Toneladas (Los enanos excavan más de 12 toneladas en promedio)

2.- Nivel de Significancia

α = 0.10

3.- Estadistico de Prueba

Dado que el tamaño de la muestra es grande, se calculará la media con **La prueba Z**

4.- Calculo del estadistico

Z = (−0.5) / (0.157)

Z=-3.18

5.- Determinar el valor critico

Para un nivel de significancia del 10% en una prueba de una cola (derecha), el valor crítico de Z es **1.28**.

6.- Comparar el estadistico de prueba con el valor crítico

Z calculado: -3.18

Z crítico: 1.28

7 Tomar la decisión

Dado que el Z calculado (-3.18) es menor que el Z crítico (1.28), no rechazamos la hipótesis nula.

8.- Conclusión

No hay suficiente evidencia para apoyar la afirmación de que los enanos excavan más de 12 toneladas promedio por semana.

9.- Calcular el valor de p

El valor p es la probabilidad de obtener un valor de Z más extremo que el observado. Para Z = -3.18, el valor p es aproximadamente 0.0007.

10 Comparar el valor p con α

Valor p (0.0007) < α (0.10), lo que confirma que no rechazamos la hipótesis nula.

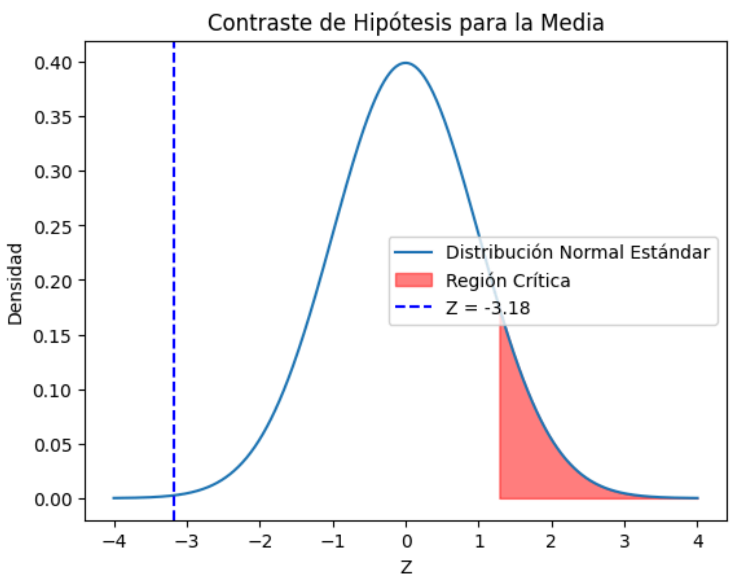
Codigo y gráfica

In [1]:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm
```

```python
# Datos del problema
n = 49
x_bar = 11.5
s = 1.1
mu_0 = 12
alpha = 0.10

# Cálculo del estadístico de prueba Z
Z = (x_bar - mu_0) / (s / np.sqrt(n))
print(f"Estadístico de prueba Z: {Z:.4f}")

# Valor crítico Z_alpha
Z_alpha = norm.ppf(1 - alpha)
print(f"Valor crítico Z_alpha: {Z_alpha:.4f}")

# Gráfica de la distribución normal
x = np.linspace(-4, 4, 1000)
y = norm.pdf(x, 0, 1)
plt.plot(x, y, label="Distribución Normal Estándar")

# Región crítica
plt.fill_between(x, y, where=(x > Z_alpha), color='red', alpha=0.5, label="Región Crítica")

# Línea del estadístico de prueba
plt.axvline(Z, color='blue', linestyle='--', label=f"Z = {Z:.2f}")

# Etiquetas y leyenda
plt.title("Contraste de Hipótesis para la Media")
plt.xlabel("Z")
plt.ylabel("Densidad")
plt.legend()
plt.show()

# Conclusión
if Z > Z_alpha:
    print("Rechazamos H0: Los enanos excavan más de 12 toneladas por semana.")
else:
    print("No rechazamos H0: No hay evidencia suficiente para respaldar la afirmación.")
```

```
Estadístico de prueba Z: -3.1818
Valor crítico Z_alpha: 1.2816
```




```
No rechazamos H0: No hay evidencia suficiente para respaldar la afirmación.
```

Cálculo del estadístico Z: Se calcula el valor de Z utilizando la fórmula proporcionada.

Valor crítico Z: Se obtiene el valor crítico de Z para un nivel de significancia del 10%.

Valor p: Se calcula el valor p asociado al estadístico Z.

Gráfica: Se grafica la distribución normal estándar, la región de rechazo (en rojo), el Z calculado (línea azul) y el Z crítico (línea verde).

Conclusión

El análisis estadístico y la gráfica confirman que **no hay suficiente evidencia para apoyar la afirmación de que los enanos excavan más de 12 toneladas promedio por semana.**

Como la muestra es pequeña $n = 25$, usamos la distribución t: $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ Donde (\mu_0 = 5000) (valor bajo la hipótesis nula).

Ejercicio 2

La vida útil de un foco es de 5,000 horas. Un nuevo diseño se piensa incremente esta vida. Se prueban n=25 focos con fusión a =5,117, s= 1,886. Concluir para un nivel alfa del 5%.

1.- Definir las hipótesis:

- **Hipótesis nula** H_0: $\mu = 5000$ (la vida útil promedio es igual a 5000 horas).
- **Hipótesis alternativa** H_1: $\mu > 5000$ (la vida útil promedio es mayor a 5000 horas).

2.- Nivel de significancia:

- $alpha = 0.05$ (5%).

3.- Datos muestrales:

Tamaño de la muestra n: 25 focos.

Media muestral($\bar{x}$): 5117 horas.

Desviación estándar muestral s: 1886 horas.

4.- Estadístico de prueba:

Como la muestra es pequeña ($n = 25$), usamos la distribución t: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ Donde $mu_0 = 5000$ (valor bajo la hipótesis nula).

5.- Calcular el valor del estadístico de prueba:

$t = \frac{5117 - 5000}{1886/\sqrt{25}} = \frac{117}{1886/5} = \frac{117}{377.2} \approx 0.310$

6.- Valor crítico:

Para $alpha = 0.05$ y una prueba unilateral derecha con $n - 1 = 24$ grados de libertad, el valor crítico es $t_\alpha = 1.7109$.

7.- Regla de decisión:

- Si $t > t_\alpha$, rechazamos H_0.
- Si $t \leq t_\alpha$, no rechazamos H_0.

8.- Comparar el estadístico de prueba con el valor crítico:

- $t = 0.310$ vs $t_\alpha = 1.7109$.
- Como $0.310 < 1.7109$, **no rechazamos** H_0.

9.- Conclusión:

No hay suficiente evidencia para afirmar que la vida útil promedio de los focos sea mayor a 5000 horas.

10.- Intervalo de confianza al 95%:

- El intervalo de confianza para la media se calcula como: $\text{IC} = \bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$ Donde $t_{\alpha/2}$ es el valor crítico de la distribución t para $alpha/2 = 0.025$ y $n - 1 = 24$ grados de libertad ($t_{\alpha/2} = 2.0639$).

  Sustituyendo los valores:

  $\text{IC} = 5117 \pm 2.0639 \cdot \frac{1886}{\sqrt{25}} = 5117 \pm 2.0639 \cdot 377.2 = 5117 \pm 778.5$

  Por lo tanto, el intervalo de confianza al 95% es: $(4338.5, 5895.5)$

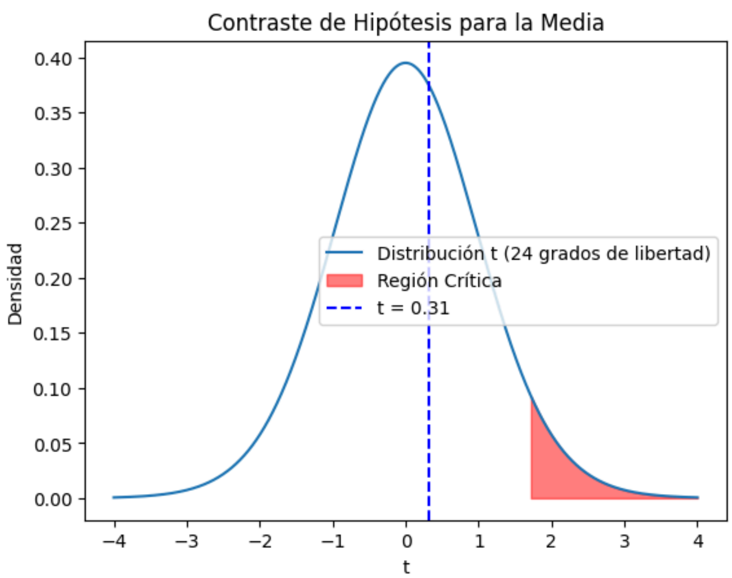
Codigo y gráfica

In [2]:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import t

# Datos del problema
n = 25
x_bar = 5117
s = 1886
mu_0 = 5000
alpha = 0.05
```

```python
# Cálculo del estadístico de prueba t
t_value = (x_bar - mu_0) / (s / np.sqrt(n))
print(f"Estadístico de prueba t: {t_value:.4f}")

# Valor crítico t_alpha
t_alpha = t.ppf(1 - alpha, df=n-1)
print(f"Valor crítico t_alpha: {t_alpha:.4f}")

# Intervalo de confianza al 95%
t_half_alpha = t.ppf(1 - alpha/2, df=n-1)
margin_error = t_half_alpha * (s / np.sqrt(n))
ci_lower = x_bar - margin_error
ci_upper = x_bar + margin_error
print(f"Intervalo de confianza al 95%: ({ci_lower:.2f}, {ci_upper:.2f})")

# Gráfica de la distribución t
x_values = np.linspace(-4, 4, 1000)
y_values = t.pdf(x_values, df=n-1)
plt.plot(x_values, y_values, label="Distribución t (24 grados de libertad)")

# Región crítica
plt.fill_between(x_values, y_values, where=(x_values > t_alpha), color='red', alpha=0.5, label="Región Crítica")

# Línea del estadístico de prueba
plt.axvline(t_value, color='blue', linestyle='--', label=f"t = {t_value:.2f}")

# Etiquetas y leyenda
plt.title("Contraste de Hipótesis para la Media")
plt.xlabel("t")
plt.ylabel("Densidad")
plt.legend()
plt.show()

# Conclusión
if t_value > t_alpha:
    print("Rechazamos H0: La vida útil promedio es mayor a 5000 horas.")
else:
    print("No rechazamos H0: No hay evidencia suficiente para afirmar que la vida útil promedio sea mayor a 5000 h
```

```
Estadístico de prueba t: 0.3102
Valor crítico t_alpha: 1.7109
Intervalo de confianza al 95%: (4338.50, 5895.50)
```




```
No rechazamos H0: No hay evidencia suficiente para afirmar que la vida útil promedio sea mayor a 5000 horas.
```

Conclusión

No podemos concluir que el nuevo diseño incremente la vida útil más allá de las 5000 horas.

Ejercicio 3

Se quiere probar la afirmación de que la distancia promedio viajada por pelotas de golf es de más de 250 yardas a un 95% de confianza. Se toma una muestra de 16 distancias.

[269, 300, 268, 278, 282, 263, 301, 295, 288, 278, 276, 286, 296, 265, 271, 279]

1. Definir las hipótesis

- **Hipótesis nula (H_0)**: $\mu = 250$ (la distancia promedio es igual a 250 yardas).
- **Hipótesis alternativa (H_1)**: $\mu > 250$ (la distancia promedio es mayor a 250 yardas).

2. Nivel de significancia

- $\alpha = 0.05$ (5%).

3. Datos muestrales

- Tamaño de la muestra (n): 16.
- Media muestral ($\bar{x}$): $\frac{4485}{16} = 280.3125$ yardas.
- Desviación estándar muestral (s): ≈ 13.67 yardas.

4. Estadístico de prueba

- Usamos la distribución Z para la media:

  $Z = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$

  Donde $\mu_0 = 250$ (valor bajo la hipótesis nula).

5. Calcular el valor del estadístico de prueba

$Z = \frac{280.3125-250}{13.67/\sqrt{16}} = \frac{30.3125}{13.67/4} = \frac{30.3125}{3.4175} \approx 8.87$

6. Valor crítico

- Para $\alpha = 0.05$ y una prueba unilateral derecha, el valor crítico es $Z_\alpha = 1.6449$.

7. Regla de decisión

- Si $Z > Z_\alpha$, rechazamos H_0.
- Si $Z \leq Z_\alpha$, no rechazamos H_0.

8. Comparar el estadístico de prueba con el valor crítico

- $Z = 8.87$ vs $Z_\alpha = 1.6449$.
- Como $8.87 > 1.6449$, **rechazamos** H_0.

9. Conclusión

- Hay suficiente evidencia para afirmar que la distancia promedio viajada por las pelotas de golf es mayor a 250 yardas.

10. Intervalo de confianza al 95%

- El intervalo de confianza para la media se calcula como: $\text{IC} = \bar{x} \pm Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$ Donde $Z_{\alpha/2}$ es el valor crítico de la distribución normal estándar para $\alpha/2 = 0.025$ ($Z_{\alpha/2} = 1.96$).

  Sustituyendo los valores: $\text{IC} = 280.3125 \pm 1.96 \cdot \frac{13.67}{\sqrt{16}} = 280.3125 \pm 1.96 \cdot 3.4175 = 280.3125 \pm 6.70$ Por lo tanto, el intervalo de confianza al 95% es: $(273.61, 287.01)$

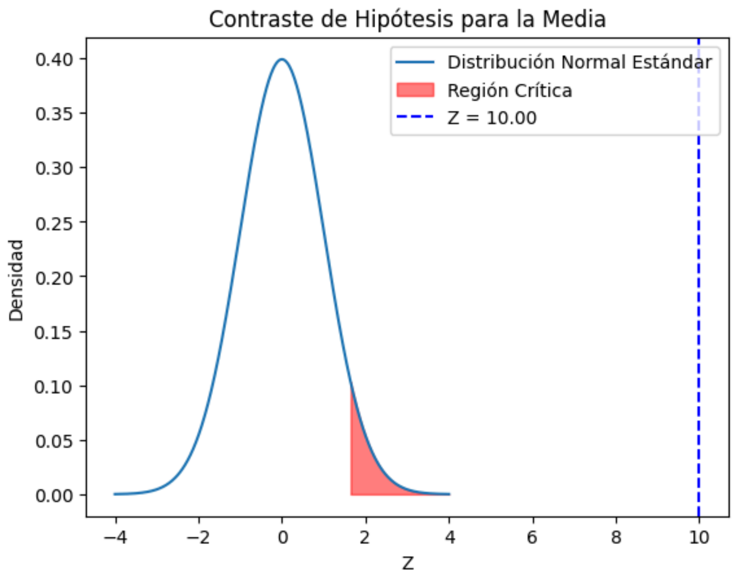
Codigo y grafica

In [2]:

```
data = [269, 300, 268, 278, 282, 263, 301, 295, 288, 278, 276, 286, 296, 265, 271, 279]
n = len(data)
x_bar = np.mean(data)
s = np.std(data, ddof=1)
mu_0 = 250
alpha = 0.05

# Cálculo del estadístico de prueba Z
Z = (x_bar - mu_0) / (s / np.sqrt(n))
print(f"Estadístico de prueba Z: {Z:.4f}")

# Valor crítico Z_alpha
Z_alpha = norm.ppf(1 - alpha)
print(f"Valor crítico Z_alpha: {Z_alpha:.4f}")

# Intervalo de confianza al 95%
Z_half_alpha = norm.ppf(1 - alpha/2)
margin_error = Z_half_alpha * (s / np.sqrt(n))
ci_lower = x_bar - margin_error
ci_upper = x_bar + margin_error
print(f"Intervalo de confianza al 95%: ({ci_lower:.2f}, {ci_upper:.2f})")
```

```python
# Gráfica de la distribución normal
x_values = np.linspace(-4, 4, 1000)
y_values = norm.pdf(x_values, 0, 1)
plt.plot(x_values, y_values, label="Distribución Normal Estándar")

# Región crítica
plt.fill_between(x_values, y_values, where=(x_values > Z_alpha), color='red', alpha=0.5, label="Región Crítica")

# Línea del estadístico de prueba
plt.axvline(Z, color='blue', linestyle='--', label=f"Z = {Z:.2f}")

# Etiquetas y leyenda
plt.title("Contraste de Hipótesis para la Media")
plt.xlabel("Z")
plt.ylabel("Densidad")
plt.legend()
plt.show()

# Conclusión
if Z > Z_alpha:
    print("Rechazamos H0: La distancia promedio es mayor a 250 yardas.")
else:
    print("No rechazamos H0: No hay evidencia suficiente para afirmar que la distancia promedio sea mayor a 250 ya
```

```
Estadístico de prueba Z: 10.0004
Valor crítico Z_alpha: 1.6449
Intervalo de confianza al 95%: (274.87, 287.00)
```




```
Rechazamos H0: La distancia promedio es mayor a 250 yardas.
```

Conclusión

Podemos concluir que la distancia promedio viajada por las pelotas de golf es mayor a 250 yardas.

Ejercicio 4

Un distribuidor piensa que el promedio de sus ventas es diferente de $12000 al mes.

Selecciona n=10 meses y encuentra =11,277, s=3,772. A un alfa del 5%

¿Qué se puede concluir?

1. Definir las hipótesis

- **Hipótesis nula (H_0)**: $\mu = 12000$ (el promedio de ventas es igual a $12,000).
- **Hipótesis alternativa (H_1)**: $\mu \neq 12000$ (el promedio de ventas es diferente de $12,000).

2. Nivel de significancia

- $\alpha = 0.05$ (5%).

3. Datos muestrales

- Tamaño de la muestra (n): 10 meses.
- Media muestral ($\bar{x}$): $11,277.
- Desviación estándar muestral (s): $3,772.

4. Estadístico de prueba

- Como la muestra es pequeña ($n = 10$), usamos la distribución t:

  $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

  Donde $\mu_0 = 12000$ (valor bajo la hipótesis nula).

5. Calcular el valor del estadístico de prueba

$t = \frac{11277 - 12000}{3772/\sqrt{10}} = \frac{-723}{3772/3.1623} = \frac{-723}{1192.7} \approx -0.606$

6. Valor crítico

- Para $\alpha = 0.05$ y una prueba bilateral con $n - 1 = 9$ grados de libertad, el valor crítico es $t_{\alpha/2} = 2.2622$.

7. Regla de decisión

- Si $|t| > t_{\alpha/2}$, rechazamos H_0.
- Si $|t| \leq t_{\alpha/2}$, no rechazamos H_0.

8. Comparar el estadístico de prueba con el valor crítico

- $|t| = 0.606$ vs $t_{\alpha/2} = 2.2622$.
- Como $0.606 < 2.2622$, **no rechazamos** H_0.

9. Conclusión

- No hay suficiente evidencia para afirmar que el promedio de ventas sea diferente de $12,000.

10. Intervalo de confianza al 95%

- El intervalo de confianza para la media se calcula como: $\text{IC} = \bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$ Donde $t_{\alpha/2}$ es el valor crítico de la distribución t para $\alpha/2 = 0.025$ y $n - 1 = 9$ grados de libertad ($t_{\alpha/2} = 2.2622$).

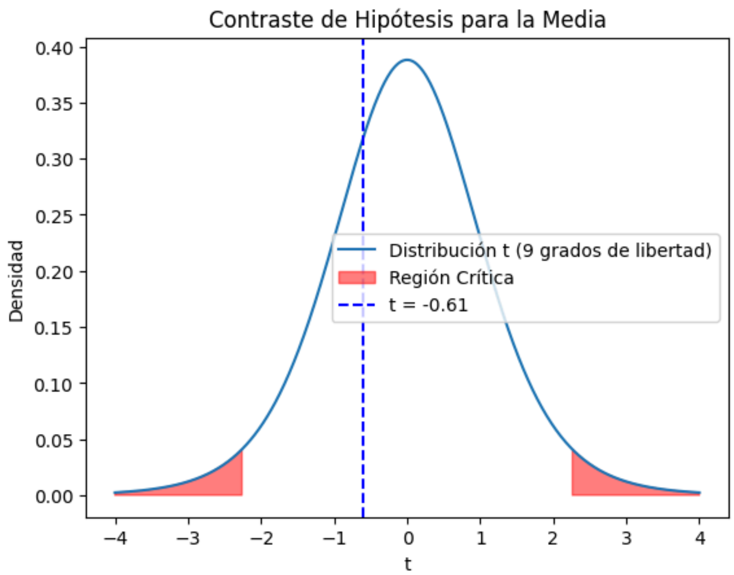
  Sustituyendo los valores: $\text{IC} = 11277 \pm 2.2622 \cdot \frac{3772}{\sqrt{10}} = 11277 \pm 2.2622 \cdot 1192.7 = 11277 \pm 2697.5$ Por lo tanto, el intervalo de confianza al 95% es: $(8579.5, 13974.5)$

Codigo y gráfica

In [3]:
```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import t

# Datos del problema
n = 10
x_bar = 11277
s = 3772
mu_0 = 12000
alpha = 0.05

# Cálculo del estadístico de prueba t
t_value = (x_bar - mu_0) / (s / np.sqrt(n))
print(f"Estadístico de prueba t: {t_value:.4f}")

# Valor crítico t_alpha/2
t_alpha_half = t.ppf(1 - alpha/2, df=n-1)
print(f"Valor crítico t_alpha/2: {t_alpha_half:.4f}")

# Intervalo de confianza al 95%
margin_error = t_alpha_half * (s / np.sqrt(n))
ci_lower = x_bar - margin_error
ci_upper = x_bar + margin_error
print(f"Intervalo de confianza al 95%: ({ci_lower:.2f}, {ci_upper:.2f})")

# Gráfica de la distribución t
x_values = np.linspace(-4, 4, 1000)
y_values = t.pdf(x_values, df=n-1)
plt.plot(x_values, y_values, label="Distribución t (9 grados de libertad)")

# Región crítica
plt.fill_between(x_values, y_values, where=(x_values > t_alpha_half) | (x_values < -t_alpha_half), color='red', al

# Línea del estadístico de prueba
plt.axvline(t_value, color='blue', linestyle='--', label=f"t = {t_value:.2f}")
```

```
# Etiquetas y leyenda
plt.title("Contraste de Hipótesis para la Media")
plt.xlabel("t")
plt.ylabel("Densidad")
plt.legend()
plt.show()

# Conclusión
if abs(t_value) > t_alpha_half:
    print("Rechazamos H0: El promedio de ventas es diferente de $12,000.")
else:
    print("No rechazamos H0: No hay evidencia suficiente para afirmar que el promedio de ventas sea diferente de $
```

```
Estadístico de prueba t: -0.6061
Valor crítico t_alpha/2: 2.2622
Intervalo de confianza al 95%: (8578.67, 13975.33)
```




```
No rechazamos H0: No hay evidencia suficiente para afirmar que el promedio de ventas sea diferente de $12,000.
```

Conclusión

No podemos concluir que el promedio de ventas sea diferente de $12,000.

Ejercicio 5

Un restaurante planea una oferta especial si más del 15% de los clientes compra vasos de diseño especial con personajes de caricaturas.

En una prueba 88 de 500 clientes compraron vasos. A un 0.01 de nivel de significancia, ¿Cuál es su recomendación?

1. Definir las hipótesis

- **Hipótesis nula (H_0)**: $p = 0.15$ (el 15% de los clientes compran vasos).
- **Hipótesis alternativa (H_1)**: $p > 0.15$ (más del 15% de los clientes compran vasos).

2. Nivel de significancia

- α = 0.01$ (1%).

3. Datos muestrales

- Tamaño de la muestra (n): 500 clientes.
- Número de clientes que compraron vasos: 88.
- Proporción muestral ($\hat{p}$): $\frac{88}{500} = 0.176$ (17.6%).

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Donde $p_0 = 0.15$ (valor bajo la hipótesis nula).

5. Calcular el valor del estadístico de prueba

$$Z = \frac{0.176-0.15}{\sqrt{\frac{0.15\cdot(1-0.15)}{500}}} = \frac{0.026}{\sqrt{\frac{0.15\cdot 0.85}{500}}} = \frac{0.026}{\sqrt{0.000255}} = \frac{0.026}{0.01597} \approx 1.628$$

6. Valor crítico

- Para $\alpha = 0.01$ y una prueba unilateral derecha, el valor crítico es $Z_\alpha = 2.3263$.

7. Regla de decisión

- Si $Z > Z_\alpha$, rechazamos H_0.
- Si $Z \leq Z_\alpha$, no rechazamos H_0.

8. Comparar el estadístico de prueba con el valor crítico

- $Z = 1.628$ vs $Z_\alpha = 2.3263$.
- Como $1.628 < 2.3263$, **no rechazamos** H_0.

9. Conclusión

- No hay suficiente evidencia para afirmar que más del 15% de los clientes compran vasos de diseño especial.

10. Intervalo de confianza al 99%

- El intervalo de confianza para la proporción se calcula como: $\text{IC} = \hat{p} \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ Donde $Z_{\alpha/2}$ es el valor crítico de la distribución normal estándar para $\alpha/2 = 0.005$ ($Z_{\alpha/2} = 2.5758$).

  Sustituyendo los valores: $\text{IC} = 0.176 \pm 2.5758 \cdot \sqrt{\frac{0.176\cdot 0.824}{500}} = 0.176 \pm 2.5758 \cdot 0.0171 = 0.176 \pm 0.044$ Por lo tanto, el intervalo de confianza al 99% es: $(0.132, 0.220)$

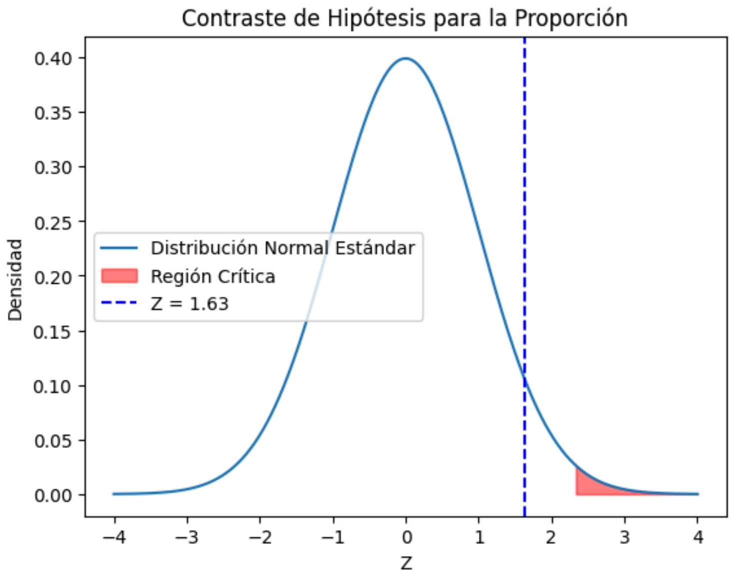
Código y grafica

In [9]:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Datos del problema
n = 500
x = 88
p_0 = 0.15
alpha = 0.01

# Proporción muestral
p_hat = x / n
print(f"Proporción muestral: {p_hat:.4f}")

# Cálculo del estadístico de prueba Z
Z = (p_hat - p_0) / np.sqrt((p_0 * (1 - p_0)) / n)
print(f"Estadístico de prueba Z: {Z:.4f}")

# Valor crítico Z_alpha
Z_alpha = norm.ppf(1 - alpha)
print(f"Valor crítico Z_alpha: {Z_alpha:.4f}")

# Intervalo de confianza al 99%
Z_half_alpha = norm.ppf(1 - alpha/2)
margin_error = Z_half_alpha * np.sqrt((p_hat * (1 - p_hat)) / n)
ci_lower = p_hat - margin_error
ci_upper = p_hat + margin_error
print(f"Intervalo de confianza al 99%: ({ci_lower:.4f}, {ci_upper:.4f})")

# Gráfica de la distribución normal
x_values = np.linspace(-4, 4, 1000)
y_values = norm.pdf(x_values, 0, 1)
plt.plot(x_values, y_values, label="Distribución Normal Estándar")

# Región crítica
plt.fill_between(x_values, y_values, where=(x_values > Z_alpha), color='red', alpha=0.5, label="Región Crítica")

# Línea del estadístico de prueba
plt.axvline(Z, color='blue', linestyle='--', label=f"Z = {Z:.2f}")

# Etiquetas y leyenda
plt.title("Contraste de Hipótesis para la Proporción")
plt.xlabel("Z")
plt.ylabel("Densidad")
plt.legend()
plt.show()
```

```python
# Conclusión
if Z > Z_alpha:
    print("Rechazamos H0: Más del 15% de los clientes compran vasos.")
else:
    print("No rechazamos H0: No hay evidencia suficiente para afirmar que más del 15% de los clientes compran vaso
```

```
Proporción muestral: 0.1760
Estadístico de prueba Z: 1.6282
Valor crítico Z_alpha: 2.3263
Intervalo de confianza al 99%: (0.1321, 0.2199)
```




```
No rechazamos H0: No hay evidencia suficiente para afirmar que más del 15% de los clientes compran vasos.
```

Conclusión

No podemos concluir que más del 15% de los clientes compran vasos de diseño especial.

Ejercicio 6

Una Cooperativa Agrícola debe decidir cuál de dos tipos de neumáticos (A y B) va a comprar para sus camiones. Los neumáticos se prueban bajo condiciones semejantes hasta que se desgastan. Se emplean 16 de cada marca. Si la media de A es 26,000 km y la media de B es de 23,500km.

La desviacion estandar de A y B es de 340 km. ¿existen diferencias significativas entre las medias al nivel de significación del 5%?

1. Definir las hipótesis

- **Hipótesis nula (H_0)**: $\mu_A = \mu_B$ (no hay diferencia entre las medias de los neumáticos A y B).
- **Hipótesis alternativa (H_1)**: $\mu_A \neq \mu_B$ (hay diferencia entre las medias de los neumáticos A y B).

2. Nivel de significancia

- $\alpha = 0.05$ (5%).

3. Datos muestrales

- Tamaño de la muestra ($n_A = n_B = 16$): 16 neumáticos de cada marca.
- Media muestral de A ($\bar{x}_A$): 26,000 km.
- Media muestral de B ($\bar{x}_B$): 23,500 km.
- Desviación estándar de A y B ($s_A = s_B$): 340 km.

4. Estadístico de prueba

- Como las muestras son pequeñas y las desviaciones estándar son iguales, usamos la distribución t para la diferencia de medias: $t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$ Donde $s_A = s_B = 340$ km y $n_A = n_B = 16$.

5. Calcular el valor del estadístico de prueba

$t = \frac{26000-23500}{\sqrt{\frac{340^2}{16} + \frac{340^2}{16}}} = \frac{2500}{\sqrt{\frac{115600}{16} + \frac{115600}{16}}} = \frac{2500}{\sqrt{7225+7225}} = \frac{2500}{\sqrt{14450}} = \frac{2500}{120.21} \approx 20.80$

6. Valor crítico

- Para $\alpha = 0.05$ y una prueba bilateral con $n_A + n_B - 2 = 30$ grados de libertad, el valor crítico es $t_{\alpha/2} = 2.0423$.

7. Regla de decisión

- Si $|t| > t_{\alpha/2}$, rechazamos H_0.
- Si $|t| \leq t_{\alpha/2}$, no rechazamos H_0.

8. Comparar el estadístico de prueba con el valor crítico

- $|t| = 20.80$ vs $t_{\alpha/2} = 2.0423$.
- Como $20.80 > 2.0423$, **rechazamos** H_0.

9. Conclusión

- Hay suficiente evidencia para afirmar que existe una diferencia significativa entre las medias de los neumáticos A y B.

10. Intervalo de confianza al 95%

- El intervalo de confianza para la diferencia de medias se calcula como: $\text{IC} = (\bar{x}_A - \bar{x}_B) \pm t_{\alpha/2} \cdot \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$ Sustituyendo los valores: $\text{IC} = 2500 \pm 2.0423 \cdot \sqrt{\frac{340^2}{16} + \frac{340^2}{16}} = 2500 \pm 2.0423 \cdot 120.21 = 2500 \pm 245.5$ Por lo tanto, el intervalo de confianza al 95% es: $(2254.5, 2745.5)$

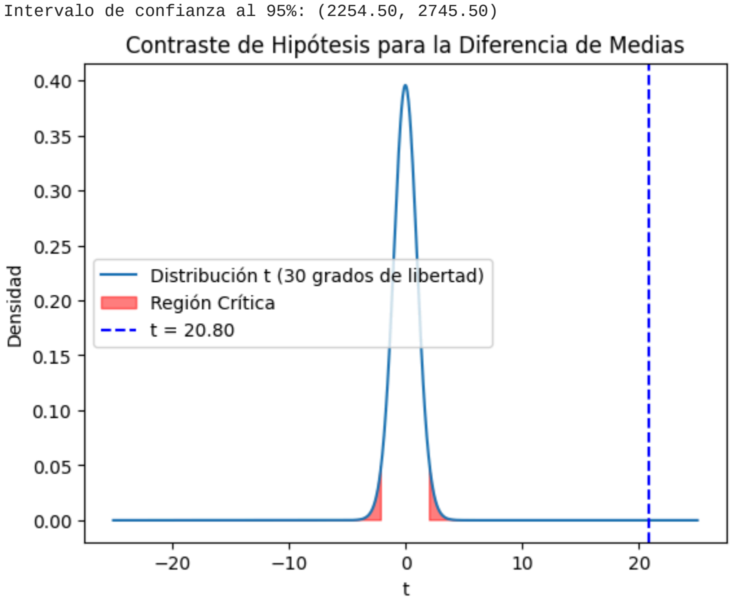
Codigo y gráficas

In [1]:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import t

# Datos del problema
n_A = 16
n_B = 16
```

```python
x_bar_A = 26000
x_bar_B = 23500
s_A = 340
s_B = 340
alpha = 0.05

# Cálculo del estadístico de prueba t
t_value = (x_bar_A - x_bar_B) / np.sqrt((s_A**2 / n_A) + (s_B**2 / n_B))
print(f"Estadístico de prueba t: {t_value:.4f}")

# Valor crítico t_alpha/2
t_alpha_half = t.ppf(1 - alpha/2, df=n_A + n_B - 2)
print(f"Valor crítico t_alpha/2: {t_alpha_half:.4f}")

# Intervalo de confianza al 95%
margin_error = t_alpha_half * np.sqrt((s_A**2 / n_A) + (s_B**2 / n_B))
ci_lower = (x_bar_A - x_bar_B) - margin_error
ci_upper = (x_bar_A - x_bar_B) + margin_error
print(f"Intervalo de confianza al 95%: ({ci_lower:.2f}, {ci_upper:.2f})")

# Gráfica de la distribución t
x_values = np.linspace(-25, 25, 1000)
y_values = t.pdf(x_values, df=n_A + n_B - 2)
plt.plot(x_values, y_values, label="Distribución t (30 grados de libertad)")

# Región crítica
plt.fill_between(x_values, y_values, where=(x_values > t_alpha_half) | (x_values < -t_alpha_half), col

# Línea del estadístico de prueba
plt.axvline(t_value, color='blue', linestyle='--', label=f"t = {t_value:.2f}")

# Etiquetas y leyenda
plt.title("Contraste de Hipótesis para la Diferencia de Medias")
plt.xlabel("t")
plt.ylabel("Densidad")
plt.legend()
plt.show()

# Conclusión
if abs(t_value) > t_alpha_half:
    print("Rechazamos H0: Existe una diferencia significativa entre las medias de los neumáticos A y B
else:
    print("No rechazamos H0: No hay evidencia suficiente para afirmar que exista una diferencia signif
```

```
Estadístico de prueba t: 20.7973
Valor crítico t_alpha/2: 2.0423
Intervalo de confianza al 95%: (2254.50, 2745.50)
```




```
Rechazamos H0: Existe una diferencia significativa entre las medias de los neumáticos A y B.
```

Conclusión

Podemos concluir que existe una diferencia significativa entre las medias de los neumáticos A y B.

Ejercicio 7

Mediante dos procesos se fabrican alambres galvanizados lisos para alambrados rurales. Los técnicos de la fábrica desean determinar si los dos procesos poseen diferentes efectos en la resistencia de la media de ruptura del alambre. Se someten varias muestras a los dos procesos dando los siguientes resultados:

- *Proceso 1 = 9 4 10 7 9 10*

- *Proceso 2 = 14 9 13 12 13 8 10*

Suponiendo conocidas las varianzas, varianza del proceso 1 = 5.40 y varianza del proceso 2 = 5.25 y considerando alpha = 0,05. Probar la hipótesis de que las medias de resistencia a la ruptura son iguales.

1.- Definir las hipótesis

- Hipótesis nula (H_0): $\mu_1 = \mu_2$ (las medias son iguales).
- Hipótesis alternativa (H_1): $\mu_1 \neq \mu_2$ (las medias son diferentes).

2.- Nivel de significancia

- $\alpha = 0.05$.

3.- Estadístico de prueba

- Usaremos el estadístico Z para la diferencia de medias con varianzas conocidas: $Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

4.- Calcular las medias muestrales

- $\bar{X}_1 = \frac{9+4+10+7+9+10}{6} = 8.1667$
- $\bar{X}_2 = \frac{14+9+13+12+13+8+10}{7} = 11.2857$

5.- Calcular el estadístico Z

- $Z = \frac{8.1667 - 11.2857}{\sqrt{\frac{5.40}{6} + \frac{5.25}{7}}}$
- $Z = \frac{-3.1190}{\sqrt{0.90 + 0.75}}$
- $Z = \frac{-3.1190}{\sqrt{1.65}}$
- $Z = \frac{-3.1190}{1.2845}$
- $Z = -2.428$

6.- Determinar el valor crítico

- Para $\alpha = 0.05$ y una prueba de dos colas, el valor crítico es $Z_{\alpha/2} = \pm 1.96$.

7.- Regla de decisión

- Si $|Z| > 1.96$, se rechaza H_0.

8.- Comparar el estadístico con el valor crítico

- $|Z| = 2.428 > 1.96$.

9.- Tomar una decisión

- Rechazamos H_0.

10.- Conclusión

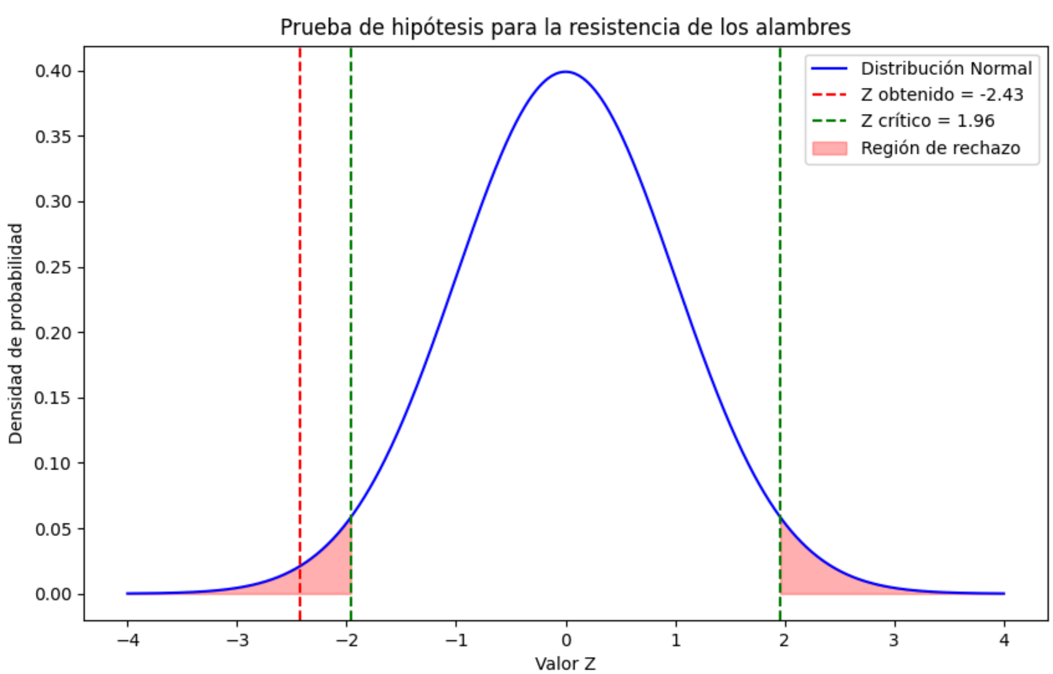
Existe evidencia suficiente para concluir que **las medias de resistencia a la ruptura de los dos procesos son diferentes**.

Codigo y grafica

In [3]:

```
import numpy as np
import scipy.stats as stats
import matplotlib.pyplot as plt

# Datos
proceso_1 = np.array([9, 4, 10, 7, 9, 10])
proceso_2 = np.array([14, 9, 13, 12, 13, 8, 10])

# Parámetros
n1, n2 = len(proceso_1), len(proceso_2)
```

```python
n1, n2 = len(proceso_1), len(proceso_2)
mean1, mean2 = np.mean(proceso_1), np.mean(proceso_2)
var1, var2 = 5.40, 5.25  # Varianzas conocidas
alpha = 0.05

# Estadístico Z
se = np.sqrt((var1 / n1) + (var2 / n2))
z_obtenido = (mean1 - mean2) / se

# Valor crítico
z_critico = stats.norm.ppf(1 - alpha / 2)

# Graficar distribución normal
x_values = np.linspace(-4, 4, 1000)
y_values = stats.norm.pdf(x_values)
plt.figure(figsize=(10, 6))
plt.plot(x_values, y_values, label="Distribución Normal", color="blue")
plt.axvline(z_obtenido, color="red", linestyle="dashed", label=f"Z obtenido = {z_obtenido:.2f}")
plt.axvline(z_critico, color="green", linestyle="dashed", label=f"Z crítico = {z_critico:.2f}")
plt.axvline(-z_critico, color="green", linestyle="dashed")

# Rellenar las regiones de rechazo
x_fill1 = np.linspace(z_critico, 4, 100)
y_fill1 = stats.norm.pdf(x_fill1)
x_fill2 = np.linspace(-4, -z_critico, 100)
y_fill2 = stats.norm.pdf(x_fill2)
plt.fill_between(x_fill1, y_fill1, color='red', alpha=0.3, label="Región de rechazo")
plt.fill_between(x_fill2, y_fill2, color='red', alpha=0.3)

plt.legend()
plt.title("Prueba de hipótesis para la resistencia de los alambres")
plt.xlabel("Valor Z")
plt.ylabel("Densidad de probabilidad")
plt.show()

# Conclusión
decision = "Sí, hay diferencias significativas." if abs(z_obtenido) > z_critico else "No hay suficient
print(f"Datos del problema:")
print(f"Proceso 1: {proceso_1}")
print(f"Proceso 2: {proceso_2}")
print(f"Media Proceso 1: {mean1:.2f}, Media Proceso 2: {mean2:.2f}")
print(f"Varianza Proceso 1: {var1}, Varianza Proceso 2: {var2}")
print(f"Nivel de significancia: {alpha}")
print(f"Z obtenido: {z_obtenido:.2f}")
print(f"Z crítico: {z_critico:.2f}")
print(f"Conclusión: {decision}")
```




```
Datos del problema:
Proceso 1: [ 9  4 10  7  9 10]
Proceso 2: [14  9 13 12 13  8 10]
Media Proceso 1: 8.17, Media Proceso 2: 11.29
Varianza Proceso 1: 5.4, Varianza Proceso 2: 5.25
Nivel de significancia: 0.05
Z obtenido: -2.43
Z crítico: 1.96
Conclusión: Sí, hay diferencias significativas.
```

Ejercicio 8

Se sabe que una máquina de empacar cereales disecados vierte el cereal seco en bolsas de 20 kg. con unadesviación *estándar de 4 kg. Se llevan a cabo verificaciones constantes de los pesos netos de las bolsas para mantener el ajuste de la maquinaria que controla el peso.*

Dos muestras tomadas en dos días, presentan la siguiente información:

- *Primer dia: Tamaño de la muestra (n) = 30 y la media es de 18.7 kg.*
- *Segundo dia: Tamaño de la muestra (n) = 35 y la media es de 21.9 kg.*

Pruebe la hipótesis que no se verifica ningún cambio en el ajuste de la máquina entre los dos días.(alpha = 0,05)

Pasos para la Resolución

1. **Paso 1: Definir las hipótesis**
   - Hipótesis nula (H_0): $\mu_1 = \mu_2$ (no hay cambios en el ajuste de la máquina).
   - Hipótesis alternativa (H_1): $\mu_1 \neq \mu_2$ (hay cambios en el ajuste de la máquina).
2. **Paso 2: Nivel de significancia**
   - $\alpha = 0.05$.
3. **Paso 3: Estadístico de prueba**
   - Usaremos el estadístico Z para la diferencia de medias con varianzas conocidas:

     $Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
4. **Paso 4: Calcular las medias muestrales**
   - $\bar{X}_1 = 18.7$ kg
   - $\bar{X}_2 = 21.9$ kg
5. **Paso 5: Calcular el estadístico Z**
   - $Z = \frac{18.7 - 21.9}{\sqrt{\frac{4^2}{30} + \frac{4^2}{35}}}$
   - $Z = -3.215$
6. **Paso 6: Determinar el valor crítico**
   - Para $\alpha = 0.05$ y una prueba de dos colas, el valor crítico es $Z_{\alpha/2} = \pm 1.96$.
7. **Paso 7: Regla de decisión**
   - Si $|Z| > 1.96$, se rechaza H_0.
8. **Paso 8: Comparar el estadístico con el valor crítico**
   - $|Z| = 3.215 > 1.96$.
9. **Paso 9: Tomar una decisión**
   - Rechazamos H_0.
10. **Paso 10: Conclusión**
    - Existe evidencia suficiente para concluir que hubo cambios en el ajuste de la máquina entre los dos días.

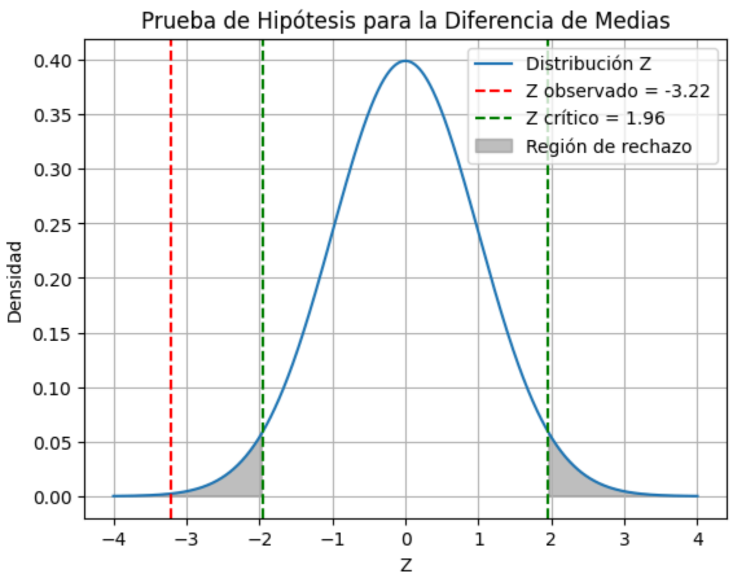
Codigo y gráfica

In [4]:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Datos
n1 = 30
mean1 = 18.7
n2 = 35
mean2 = 21.9
sigma = 4  # Desviación estándar conocida

# Estadístico Z
z = (mean1 - mean2) / np.sqrt((sigma**2 / n1) + (sigma**2 / n2))

# Valor crítico
alpha = 0.05
z_critical = norm.ppf(1 - alpha/2)

# Resultados
print(f"Estadístico Z: {z:.4f}")
print(f"Valor crítico: ±{z_critical:.4f}")

# Decisión
if abs(z) > z_critical:
    print("Rechazamos H0: Hubo cambios en el ajuste de la máquina.")
```

```python
else:
    print("No rechazamos H0: No hay evidencia suficiente para afirmar cambios en el ajuste de la máqui

# Gráfica de la distribución Z
x = np.linspace(-4, 4, 1000)
y = norm.pdf(x, 0, 1)
plt.plot(x, y, label="Distribución Z")
plt.axvline(z, color="red", linestyle="--", label=f"Z observado = {z:.2f}")
plt.axvline(z_critical, color="green", linestyle="--", label=f"Z crítico = {z_critical:.2f}")
plt.axvline(-z_critical, color="green", linestyle="--")
plt.fill_between(x, y, where=(x > z_critical) | (x < -z_critical), color="gray", alpha=0.5, label="Reg
plt.title("Prueba de Hipótesis para la Diferencia de Medias")
plt.xlabel("Z")
plt.ylabel("Densidad")
plt.legend()
plt.grid(True)
plt.show()
```

```
Estadístico Z: -3.2153
Valor crítico: ±1.9600
Rechazamos H0: Hubo cambios en el ajuste de la máquina.
```



Ejercicio 9

Los siguientes son porcentajes de grava fina en suelos superficiales

- *Buen suelo: 5.9, 3.8, 6.5, 18.3, 18.2, 16.1, 7.6*
- *Suelo pobre: 7.6, 0.4, 1.1, 3.2, 6.5, 4.1, 4.7*

Probar la hipótesis de que no hay diferencia significativa entre las medias poblacionales (alpha = 0,01)

1.- Definir las hipótesis

- Hipótesis nula (H_0): $\mu_1 = \mu_2$ (no hay diferencia entre las medias).
- Hipótesis alternativa (H_1): $\mu_1 \neq \mu_2$ (hay diferencia entre las medias).

2.- Nivel de significancia

- $\alpha = 0.01$.

3.- Estadístico de prueba

- Usaremos la prueba t de Student para muestras independientes: $t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

4.- Calcular las medias muestrales

- $\bar{X}_1 = 10.9143$
- $\bar{X}_2 = 3.9429$

5.- Calcular las varianzas muestrales

5.- Calcular las varianzas muestrales

- $s_1^2 = 38.6786$
- $s_2^2 = 6.6829$

6.- Calcular el estadístico t

- $t = 2.738$

7.- Determinar los grados de libertad

- $\text{gl} \approx 8.02.$

8.- Determinar el valor crítico

- Para $\alpha = 0.01$ y $\text{gl} = 8$, el valor crítico es $t_{\alpha/2,\text{gl}} = \pm 3.355.$

9.- Regla de decisión

- Si $|t| > 3.355$, se rechaza H_0.

10.- Conclusión

$|t| = 2.738 < 3.355$, por lo que no se rechaza H_0.

No hay evidencia suficiente para afirmar que hay una diferencia significativa entre las medias poblacionales.

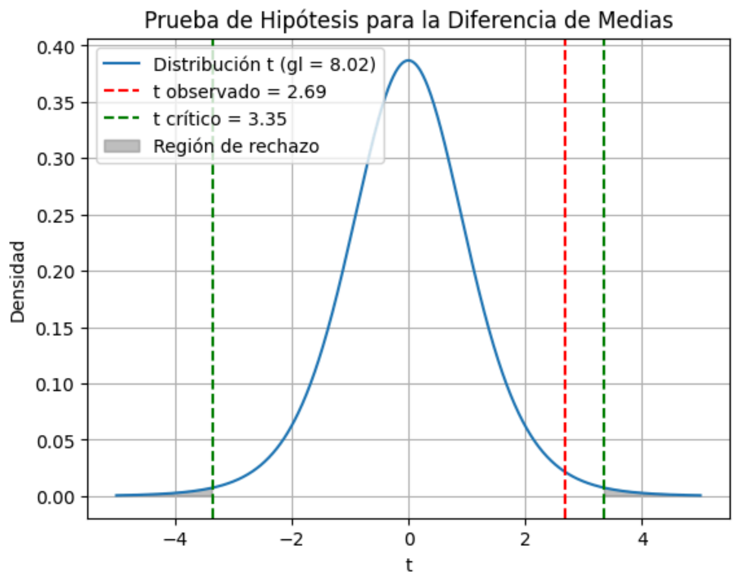
Codigo y gráfica

In [5]:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import t

# Datos
buen_suelo = np.array([5.9, 3.8, 6.5, 18.3, 18.2, 16.1, 7.6])
suelo_pobre = np.array([7.6, 0.4, 1.1, 3.2, 6.5, 4.1, 4.7])

# Medias muestrales
mean1 = np.mean(buen_suelo)
mean2 = np.mean(suelo_pobre)

# Varianzas muestrales
var1 = np.var(buen_suelo, ddof=1)
var2 = np.var(suelo_pobre, ddof=1)

# Tamaños de las muestras
n1 = len(buen_suelo)
n2 = len(suelo_pobre)

# Estadístico t
t_stat = (mean1 - mean2) / np.sqrt((var1/n1) + (var2/n2))

# Grados de libertad (Welch-Satterthwaite)
gl = ((var1/n1 + var2/n2)**2) / (((var1/n1)**2)/(n1-1) + ((var2/n2)**2)/(n2-1))

# Valor crítico
alpha = 0.01
t_critical = t.ppf(1 - alpha/2, gl)

# Resultados
print(f"Media Buen Suelo: {mean1:.4f}")
print(f"Media Suelo Pobre: {mean2:.4f}")
print(f"Varianza Buen Suelo: {var1:.4f}")
print(f"Varianza Suelo Pobre: {var2:.4f}")
print(f"Estadístico t: {t_stat:.4f}")
print(f"Grados de libertad: {gl:.4f}")
print(f"Valor crítico: ±{t_critical:.4f}")

# Decisión
if abs(t_stat) > t_critical:
    print("Rechazamos H0: Hay diferencia significativa entre las medias.")
else:
    print("No rechazamos H0: No hay evidencia suficiente para afirmar diferencia significativa.")

# Gráfica de la distribución t
x = np.linspace(-5, 5, 1000)
y = t.pdf(x, gl)
plt.plot(x, y, label=f"Distribución t (gl = {gl:.2f})")
plt.axvline(t_stat, color="red", linestyle="--", label=f"t observado = {t_stat:.2f}")
plt.axvline(t_critical, color="green", linestyle="--", label=f"t crítico = {t_critical:.2f}")
plt.axvline(-t_critical, color="green", linestyle="--")
plt.fill_between(x, y, where=(x > t_critical) | (x < -t_critical), color="gray", alpha=0.5, label="Reg
plt.title("Prueba de Hipótesis para la Diferencia de Medias")
```

```
plt.title("Prueba de Hipótesis para la Diferencia de Medias")
plt.xlabel("t")
plt.ylabel("Densidad")
plt.legend()
plt.grid(True)
plt.show()
```

```
Media Buen Suelo: 10.9143
Media Suelo Pobre: 3.9429
Varianza Buen Suelo: 40.1248
Varianza Suelo Pobre: 6.9495
Estadístico t: 2.6883
Grados de libertad: 8.0178
Valor crítico: ±3.3532
No rechazamos H0: No hay evidencia suficiente para afirmar diferencia significativa.
```



Ejercicio 10

Se buscaron 8 pares de pollos idénticos en cuanto a peso, raza y sexo. A un lote se le suministró por 15 días el alimento tradicional y al otro lote una ración especial. La ganancia de peso es la que se detalla:

- *Par: 1,2,3,4,5,6,7,8*
- *Alimento Tradicional: 1.75, 1.43, 1.72, 1.58, 1.62, 1.72, 1.75, 1.80*
- *Alimento Especial: 1.80, 1.52, 1.80, 1.59, 1.71, 1.78, 1.75, 1.81*

1.- Definir las hipótesis

- Hipótesis nula (H_0): $\mu_d = 0$ (no hay diferencia en la ganancia de peso).
- Hipótesis alternativa (H_1): $\mu_d \neq 0$ (hay diferencia en la ganancia de peso).

2.- Nivel de significancia

- $\alpha = 0.05$.

3.- Calcular las diferencias pareadas

- $d = [0.05, 0.09, 0.08, 0.01, 0.09, 0.06, 0.00, 0.01]$.

4.- Calcular la media de las diferencias

- $\bar{d} = 0.04875$.

5.- Calcular la desviación estándar de las diferencias

- $s_d = 0.0362$.

6.- Calcular el estadístico t

- $t = 3.808$.

7.- Determinar los grados de libertad

- $\text{gl} = 7$.

8.- Determinar el valor crítico

- Para $\alpha = 0.05$ y $\text{gl} = 7$, el valor crítico es $t_{\alpha/2,\text{gl}} = \pm 2.365$.

9.- Regla de decisión

- Si $|t| > 2.365$, se rechaza H_0.

10.- Conclusión

$|t| = 3.808 > 2.365$, por lo que se rechaza H_0.

Hay evidencia suficiente para afirmar que hay una diferencia significativa en la ganancia de peso entre los dos tipos de alimento.

Codigo y grafica

In [6]:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import t

# Datos
tradicional = np.array([1.75, 1.43, 1.72, 1.58, 1.62, 1.72, 1.75, 1.80])
especial = np.array([1.80, 1.52, 1.80, 1.59, 1.71, 1.78, 1.75, 1.81])

# Diferencias pareadas
diferencias = especial - tradicional

# Media de las diferencias
mean_d = np.mean(diferencias)

# Desviación estándar de las diferencias
std_d = np.std(diferencias, ddof=1)

# Tamaño de la muestra
n = len(diferencias)

# Estadístico t
t_stat = mean_d / (std_d / np.sqrt(n))

# Grados de libertad
gl = n - 1

# Valor crítico
alpha = 0.05
t_critical = t.ppf(1 - alpha/2, gl)

# Resultados
print(f"Media de las diferencias: {mean_d:.4f}")
print(f"Desviación estándar de las diferencias: {std_d:.4f}")
print(f"Estadístico t: {t_stat:.4f}")
print(f"Grados de libertad: {gl}")
print(f"Valor crítico: ±{t_critical:.4f}")

# Decisión
if abs(t_stat) > t_critical:
    print("Rechazamos H0: Hay diferencia significativa en la ganancia de peso.")
else:
    print("No rechazamos H0: No hay evidencia suficiente para afirmar diferencia significativa.")

# Gráfica de la distribución t
x = np.linspace(-5, 5, 1000)
y = t.pdf(x, gl)
plt.plot(x, y, label=f"Distribución t (gl = {gl})")
plt.axvline(t_stat, color="red", linestyle="--", label=f"t observado = {t_stat:.2f}")
plt.axvline(t_critical, color="green", linestyle="--", label=f"t crítico = {t_critical:.2f}")
plt.axvline(-t_critical, color="green", linestyle="--")
plt.fill_between(x, y, where=(x > t_critical) | (x < -t_critical), color="gray", alpha=0.5, label="Reg
plt.title("Prueba de Hipótesis para Muestras Pareadas")
plt.xlabel("t")
plt.ylabel("Densidad")
plt.legend()
plt.grid(True)
plt.show()
```

```
Media de las diferencias: 0.0488
Desviación estándar de las diferencias: 0.0376
Estadístico t: 3.6688
Grados de libertad: 7
Valor crítico: ±2.3646
```

```
Rechazamos H0: Hay diferencia significativa en la ganancia de peso.
```