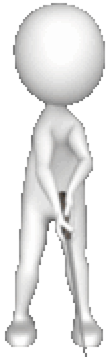
Hypothesis Testing

Learning objectvies

- Implement a z-test/t-test for a population mean.
- Use Python to perform a t-test.

What's the Process? Example

- Jack tells Jean that his average drive of a golf ball is 275 yards. Jean is skeptical and asks for substantiation. To that end, Jack hits 25 drives.
- The (sample) mean of Jack's 25 drives is only 264.4 yards. Jack still maintains that, on average, he drives a golf ball 275 yards and that this relatively poor performance can reasonably be attributed to chance.
- Do the data provide sufficient evidence to conclude that Jack's mean driving distance is less than 275 yards?



- What are the hypotheses?

  H_0: $\mu = 275$ yards (Jack's claim)

  H_a: $\mu < 275$ yards (Jean's suspicion).
- Is the observed value "too small" or "too unusual" to be due to chance variation?
- i.e. Is it possible that, even though we observed $\bar{x}$=264.4 , the true μ is in fact 275?
- If this sample is due to random chance, (just a bad sample) then we do not reject the null hypothesis.
- However, if the sample mean driving distance is "too much smaller" than 275 yards, we would reject the null hypothesis in favor of the alternative hypothesis.

What’s the Process? Significance Level

- We need to define two things:
  - **significance level alpha (α)**
  - **P-value**
- α: the risk you’re willing to take in rejecting the null hypothesis. (You might reject the null hypothesis but it might in fact be true. You don’t want the probability of this be more than α).

Rejecting the null hypothesis H_0 when it is true is defined as a **type I error.**

- Therefore, you could say α is the probability of making a Type I error.

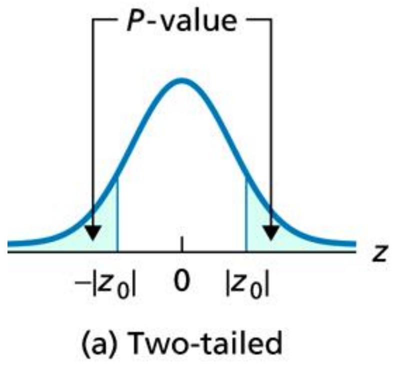
What's the Process?
P-value

_P_-Value

The ***P*-value** of a hypothesis test is the probability of getting sample data at least as inconsistent with the null hypothesis (and supportive of the alternative hypothesis) as the sample data actually obtained. We use the letter ***P*** to denote the *P*-value.

- In other words: the probability computed, assuming that H_0 is true, that the test statistic would take a value as or more extreme than observed value is called the P-value of the test.
- The **smaller** the P-value, the weaker or stronger the evidence against H_0 provided by the data in our sample?
- **Stronger!**

P-Value



P-value	Evidence against H_0
$P > 0.10$	Weak or none
$0.05 < P \leq 0.10$	Moderate
$0.01 < P \leq 0.05$	Strong
$P \leq 0.01$	Very strong

What's the Process?

P-VALUE APPROACH TO HYPOTHESIS TESTING

Step 1	State the null and alternative hypotheses.
Step 2	Decide on the significance level, α.
Step 3	Compute the value of the test statistic.
Step 4	Determine the P-value, P.
Step 5	If $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.
Step 6	Interpret the result of the hypothesis test.

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Recall that $\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$

Example-revisited

- Jack tells Jean that his average drive of a golf ball is 275 yards. Jean is skeptical and asks for substantiation. To that end, Jack hits 25 drives.
- The (sample) mean of Jack's 25 drives is only 264.4 yards. Jack still maintains that, on average, he drives a golf ball 275 yards and that this relatively poor performance can reasonably be attributed to chance.
- Do the data provide sufficient evidence to conclude that Jack's mean driving distance is less than 275 yards? (Assume σ=20)

- **Hypotheses:**
- **α = 0.05**
- **Test Statistic:**

H_0: $\mu = 275$ yards (Jack's claim)
H_a: $\mu < 275$ yards (Jean's suspicion).

$$\mu = 275 \text{ yards} \qquad \text{and} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{25}} = 4$$

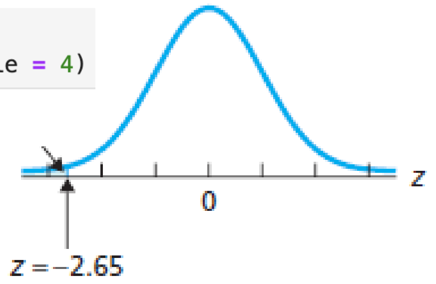
- $$z = \frac{\bar{x} - 275}{4} = \frac{264.4 - 275}{4} = -2.65.$$

```python
from scipy.stats import norm
norm.cdf(-2.65)
```

0.004024588542758306

```python
# loc is mu, scale is sigma
norm.cdf(x = 264.4,loc = 275, scale = 4)
```

0.004024588542758239



- **P < α** => We reject the null.
- **Conclusion**: At the 5% significance level, the data provide sufficient evidence to conclude that Jack's mean driving distance is less than his claimed 275 yards.

Example

- The Census Bureau reports that households spend an average of 31% of their total spending on housing.
- A homebuilders association in Cleveland wonders if the national finding applies in their area.
- They interview a sample of 40 households in the Cleveland metropolitan to learn what percent of their spending goes toward housing. The sample mean turns out to be 28.6% among all Cleveland households.
- Take μ to be the mean percent of spending devoted to housing among all Cleveland households. Assume that σ = 9.6%.

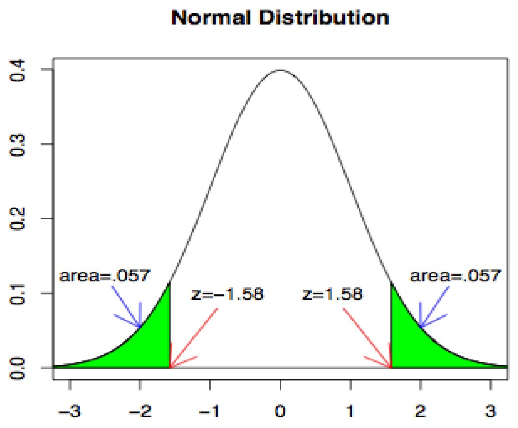
$$H_0 : \mu = 31\%$$
$$H_a : \mu \neq 31\%$$

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{28.6 - 31}{9.6 / \sqrt{40}} = -1.58$$

- Because the alternative is two-sided, values of z far from 0 in either direction count as evidence against H_0 and in favor of H_a.
- So the P-value is:
- $P(z < -1.58) + P(z > 1.58)$
- Alternatively:

```
from scipy.stats import norm
2*norm.cdf(-1.581139)
```

0.11384625916429064



- P > 0.05 => We don't reject the null.
- At the 5% significance level, the data does not provide sufficient evidence to conclude that the mean percent of spending devoted to housing is different than 31%.

Example

- A manufacturer of cereal wants to test the performance of one of its filling machines
- The machine is designed to discharge a mean amount of $\mu = 12$ ounces per box
- The manufacturer wants to detect any departure from this setting
- Suppose the sample yields the following results
  - $n = 100$ observations (boxes)
  - $\bar{x} = 11.85$ ounces
  - $\sigma = 0.5$ ounces
- Perform a test of significance and summarize your results at alpha = 0.01.

Solution

$$H_0 : \mu = 12 \text{ vs. } H_a : \mu \neq 12 \qquad \alpha = 0.01$$

$$\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{11.85 - 12}{0.5 / \sqrt{100}} = -3$$

$$P - value = 2 * P(z \geq |-3.0|) \cong 0.0027$$

```
2*stats.norm.cdf(11.85, loc = 12, scale = 0.5/10)
```

```
0.0026997960632601232
```

- p-value=0.0027 < 0.01 →Reject H_0.
- Conclusion: At this significance level, the sample provides enough evidence to believe that mean amount of cereal the machines discharge is different from 12 ounces per box.

Hypothesis Test with Unknown Population Sigma

- We discussed how to perform a hypothesis test for one population mean when the population standard deviation, σ, is known. However, the population standard deviation is usually not known.
- In such cases, instead of a *z*-test we perform t-test where the test statistic is calculated from $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ where s is the sample standard deviation.
- You can use stats.ttest_1samp from scipy import stats in Python

t-test - Example

- Acid rain from the burning of fossil fuels has caused many of the lakes around the world to become acidic.
- A lake is classified as nonacidic if it has a pH greater than 6.
- This table shows the pH levels obtained by the researchers for 15 lakes.
- At the 5% significance level, do the data provide sufficient evidence to conclude that, on average, high mountain lakes in the Southern Alps are nonacidic?

7.2	7.3	6.1	6.9	6.6
7.3	6.3	5.5	6.3	6.5
5.7	6.9	6.7	7.9	5.8

Solution

H_0: $\mu = 6$ (on average, the lakes are acidic)

H_a: $\mu > 6$ (on average, the lakes are nonacidic)

```
from scipy import stats
import numpy as np
x = np.array([7.2, 7.3, 6.1, 6.9, 6.6,7.3, 6.3, 5.5, 6.3, 6.5, 5.7, 6.9, 6.7, 7.9, 5.8])
stats.ttest_1samp(x, 6, alternative = 'greater')
```

```
Ttest_1sampResult(statistic=3.4586160984497423, pvalue=0.0019191606907199745)
```

Interpretation: At the 5% significance level, the data provide sufficient evidence to conclude that, on average, high mountain lakes in the Southern Alps are nonacidic.