Statistics: Significance Testing

- Recall that in binary hypothesis testing, we had two hypotheses H_0 and H_1 and an observation Y generated by either $f_{Y|H_0}(y)$ or $f_{Y|H_1}(y)$ (in the continuous case). We used Y to decide whether H_0 or H_1 occurred.
  - → ML Rule: Decide H_1 if $f_{Y|H_1}(y) \geq f_{Y|H_0}(y)$ and H_0 otherwise.
- In significance testing, we only have an explicit model for the null hypothesis H_0. Based on our observation Y, we will either reject the null hypothesis or fail to reject the null hypothesis.
  - → Ex: Deploy an "A" and a "B" version of a website to different users. Null Hypothesis: mean click-through rate the same.
  - → Ex: Administer a new drug to a group of patients and a placebo to a control group. Null Hypothesis: mean cholesterol the same.
  - → Ex: Change reactor temperature, measure yield before and after. Null Hypothesis: mean yield is the same.

- The significance level α is used to determine when to reject the null hypothesis.
- Given a statistic calculated from a dataset, the p-value is the probability of observing a value at least this extreme under the null hypothesis.
  - → If p-value $< \alpha$, reject the null hypothesis.
  - → If p-value $\geq \alpha$, fail to reject the null hypothesis.
- We now introduce four significance tests for the mean.
  - → We assume that, under the null hypothesis, our data is i.i.d. Gaussian (or that this is a good approximation).
  - → In a one-sample test, we compare the mean of one dataset to a known baseline mean μ.
  - → In a two-sample test, we compare the means of two datasets to each other.

- **One-Sample Z-Test:**

→ Dataset: $X_1, \ldots, X_n$

→ Null Hypothesis: Data is i.i.d. Gaussian with known mean μ and known variance σ^2.

→ Informally, does the mean of the data differ significantly from the baseline μ?

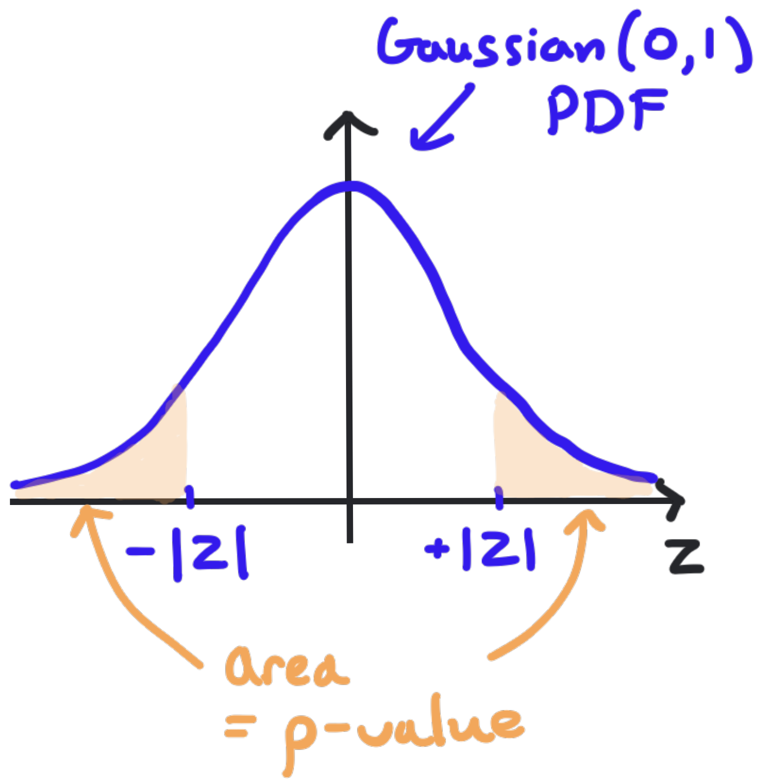
① Calculate the sample mean $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

② Calculate the Z-statistic $Z = \frac{\sqrt{n}(M_n - \mu)}{\sigma}$.

③ Calculate the p-value $= 2\Phi(-|Z|)$

④ If p-value $< \alpha$, reject the null.

If p-value $\geq \alpha$, fail to reject the null.



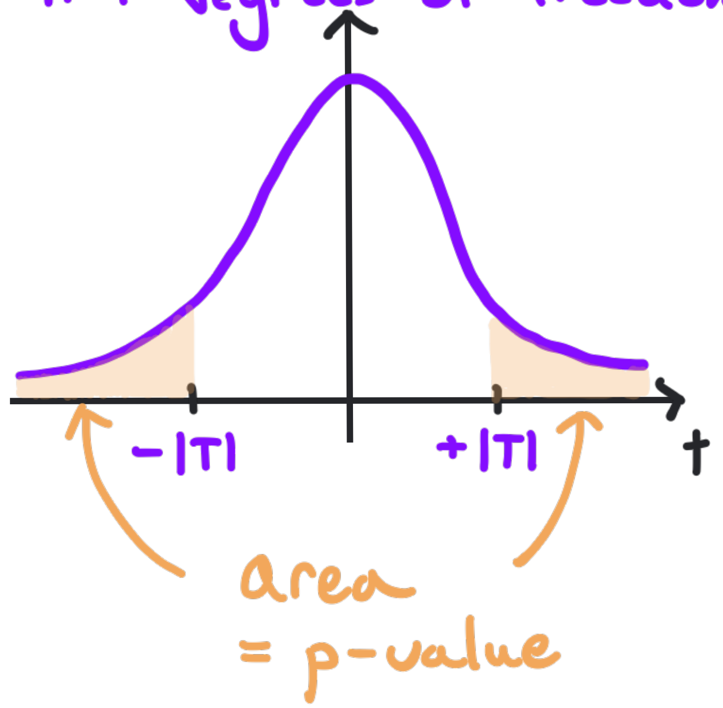
→ In practice, reasonable to use this test when $n > 30$, even if variance estimated from data by sample variance V_n. In this regime, Central Limit Theorem offers a good approximation.

- One-Sample T-Test:

→ Dataset: $X_1, \ldots, X_n$

→ Null Hypothesis: Data is i.i.d. Gaussian with known mean μ and unknown variance.

→ Informally, does the mean of the data differ significantly from the baseline μ?



① Calculate the sample mean $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and the sample variance $V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_n)^2$.

② Calculate the T-statistic $T = \frac{\sqrt{n}(M_n - \mu)}{\sqrt{V_n}}$

③ Calculate the p-value $= 2F_{T_{n-1}}(-|T|)$.

④ If p-value $< \alpha$, reject the null.

If p-value $\geq \alpha$, fail to reject the null.

→ In practice, reasonable to use this test when $n \leq 30$ and data is well-approximated by a Gaussian distribution.

- **Two-Sample Z-Test:**

→ Datasets: $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ (possibly with $n_1 \neq n_2$)

→ Null Hypothesis: $X_1, \ldots, X_{n_1}$ is i.i.d. Gaussian(μ, σ_1^2) and $Y_1, \ldots, Y_{n_2}$ is i.i.d. Gaussian(μ, σ_2^2) with known variances σ_1^2, σ_2^2. Mean μ is unknown.

→ Informally, do the datasets have the same mean?

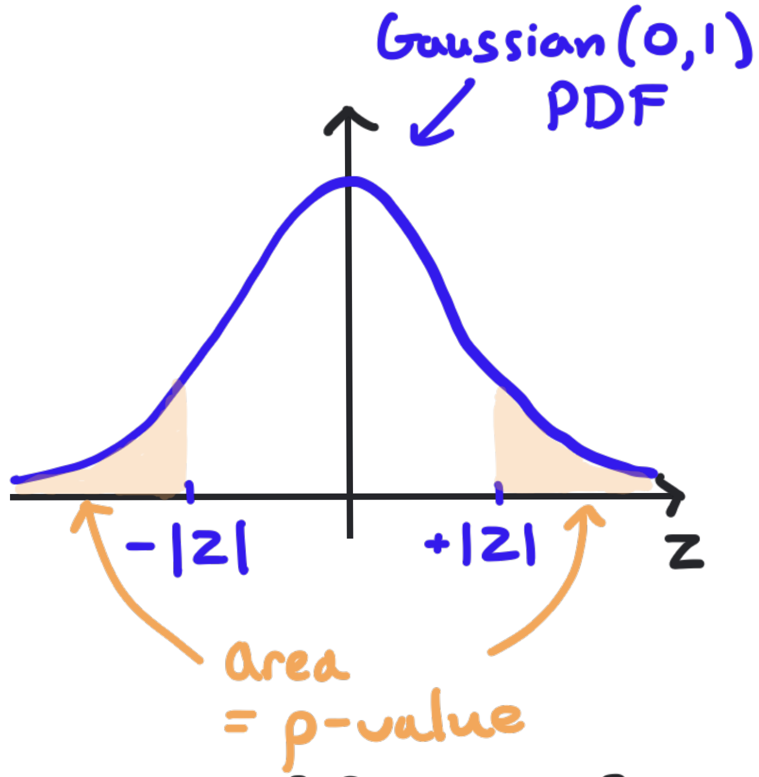
① Calculate the sample means $M_{n_1}^{(1)} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i \quad M_{n_2}^{(2)} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$.

② Calculate the Z-statistic $Z = \dfrac{M_{n_1}^{(1)} - M_{n_2}^{(2)}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$.

③ Calculate the p-value $= 2\Phi(-|Z|)$.

④ If p-value $< \alpha$, reject the null.

If p-value $\geq \alpha$, fail to reject the null.



→ In practice, reasonable to use this test for $n_1 > 30$, $n_2 > 30$, even if variances estimated from data.

- **Two-Sample T-Test:**

→ Datasets: $X_1, \dots, X_{n_1}$ and $Y_1, \dots, Y_{n_2}$ (possibly with $n_1 \neq n_2$)

→ Null Hypothesis: $X_1, \dots, X_{n_1}$ i.i.d. Gaussian(μ, σ^2) and $Y_1, \dots, Y_{n_2}$ i.i.d. Gaussian(μ, σ^2) with unknown, equal variance σ^2. Mean μ is unknown.

→ Informally, do the datasets have the same mean?

① Calculate the sample means $M^{(1)}_{n_1} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$ $M^{(2)}_{n_2} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$, sample variances $V^{(1)}_{n_1} = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - M^{(1)}_{n_1})^2$ $V^{(2)}_{n_2} = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - M^{(2)}_{n_2})^2$, and the pooled sample variance $\hat{\sigma}^2 = \frac{(n_1 - 1) V^{(1)}_{n_1} + (n_2 - 1) V^{(2)}_{n_2}}{n_1 + n_2 - 2}$.

② Calculate the T-statistic $T = \frac{M^{(1)}_{n_1} - M^{(2)}_{n_2}}{\hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$.

③ Calculate the p-value $= 2 F_{T_{n_1+n_2-2}}(-|T|)$.

④ If p-value $< \alpha$, reject the null.

If p-value $\geq \alpha$, fail to reject the null.