**A/B Testing Plan**

**Input Data:**

In our mobile application, users are offered a weekly subscription priced at **$4.99** that provides access to premium features immediately after onboarding.  
Currently, **17%** of users who receive this offer make a purchase.  
We have decided to test an alternative version of the subscription screen, which also offers a **$4.99** weekly subscription but informs the user that the price includes a **50% discount**.

*Approximately* ***2,000 users install*** *the app daily, and* ***34%*** *of them reach the subscription screen.*

**Testing Parameters:**

* **Significance level (α):** 0.05
* **Statistical power (1 – β):** 80% (probability of detecting a true effect if it exists)
* **Daily installations:** 2,000 → **680 users per day** reach the subscription screen (available for testing)
* **Baseline conversion rate:** 17%

**1. Hypothesis**

* **H₀ (Null Hypothesis):** The design change (displaying “50% off” while keeping the same $4.99 price) does not affect the subscription conversion rate.
* **H₁ (Alternative Hypothesis):** The new version increases the subscription conversion rate.

**2. Participant Population**

* **Included:** **All new users** who reach the subscription offer screen (approximately 34% of daily installs).
* **Excluded:** QA/dev teams, bot traffic, and internal corporate accounts.
* **Traffic allocation:** Random 50/50 split — Control (current screen) vs. Variant (“50% off” screen).

**3. Primary and Secondary Metrics**

**Primary metric:**

**Conversion Rate** on the subscription screen (share of users who purchased a subscription) — the key test metric.

**Secondary metric:**

* **ARPU (Average Revenue per Reached User)** — to assess the impact on revenue;
* Number of purchases within the first **7 and 30 days** (retention of paying users);
* **CTR** — share of users who clicked the discount button;
* **CTA** performance (main call-to-action on the screen);
* **Negative signals** — cancellations, complaints (safety check).

**Justification**: The primary metric (Conversion Rate) represents the direct business objective. Secondary metrics help prevent false-positive results that could negatively affect long-term value (LTV).

**4. Sample Size and Test Duration**

The **MDE (Minimum Detectable Effect)** should be chosen pragmatically. Using a normal approximation:

* **Baseline conversion (control):** p₀ = 0.17
* Parameters: α = 0.05, power = 0.8

| **MDE** | **Expected Variant CR** | **n** per group | **Total n** | **Duration (≈680 users/day)** |
| --- | --- | --- | --- | --- |
| +2 pp | 19% | ≈5,783 | ≈11,566 | ~ 17–18 days |
| +3 pp | 20% | ≈2,618 | ≈5,236 | ~ 8 days |
| +5 pp | 22% | ≈984 | ≈1,968 | ~ 3 days |

**Recommendation:** select **MDE = +3 pp**, which is both realistic and business-relevant.  
Thus, the **minimum sample size** is **≈2,618 users per group**, with an estimated test duration of **8 days under ideal conditions**.

**Practical adjustments:**

* Run the test for at least **14 days**, even if the calculated sample is reached earlier, to cover weekly behavioral cycles and seasonal variations.
* Ensure daily traffic stability; if there are fluctuations, extend the test until metrics stabilize.

**5. Success Criteria and Next Steps**

**Success criteria:**

* **Statistical significance:** p-value < 0.05 (two-tailed test) and observed difference ≥ MDE (e.g., +3 pp).
* **Business validation:** ARPU or ROMI must not decrease (total revenue should increase or remain stable).

**Actions:**

* If **successful** (p < 0.05 and business metrics stable or improved): deploy the new design for all users.
* If **statistically significant but revenue/LTV worsens:** do not roll out; perform root-cause analysis (e.g., short-term boosts followed by higher churn).
* If **not significant** (p ≥ 0.05): close the test as inconclusive and analyze by segments (platform, country, traffic source). Consider retesting with different creatives or a larger sample (lower MDE).

**6. Alternative Scenarios if the Test Fails**

If the “50% off” message does **not** increase conversions:

* Test a **free trial offer** (e.g., 7-day free trial followed by $4.99) to evaluate the impact on acquisition and retention.
* Test **different CTA and design variations:** change the button text, urgency messaging (“offer expires soon”), social proof (reviews), or screen layout (price order).
* Run **segmented experiments:** show the discount only to specific cohorts (e.g., new iOS users or certain countries) — a positive effect may exist within subsets.

**Pre-Launch Checklist**

1. Implement **50/50 randomization** and log group\_id.
2. Track key events: impression, view, click, purchase, revenue, user\_id, date, device, country.
3. Validate the **baseline metric stability** (daily reached users).
4. Run the test for **a minimum of 14 days** or until both groups reach the required sample size.
5. Monitor **stop criteria:** critical bugs, sudden churn increases, or technical issues.
6. Upon completion, perform a **two-proportion z-test**, analyze **secondary metrics and user segments**, and summarize actionable insights.