VE401 Recitation Class Note

Final Part1

Chen Siyi
siyi.chen_chicy@sjtu.edu.cn

1 Overview

1. The Fisher Test
2. Neyman-Pearson Decision Theory
3. Null Hypothesis Significance Testing
4. Single Sample Tests for the Mean and Variance
5. Non-Parametric Single Sample Tests for the Median
6. Inferences on Proportions
7. Comparison of Two Variances
8. Comparison of Two Means
9. Non-Parametric Comparisons; Paired Tests and Correlation
10. Categorical Data
11. Simple Linear Regression I: Basic Model and Inferences
12. Simple Linear Regression II: Predictions and Model Analysis
13. Multiple Linear Regression I: Basic Model
14. Multiple Linear Regression II: Inferences on the Model
15. Multiple Linear Regression III: Finding the Right Model
16. ANOVA I: Basic Model
17. ANOVA II: Further Model Inferences

2 Fisher's Null Hypothesis Testing

Goal:
Find statistical evidence that allows us to **reject** the null hypothesis H_0.

Null Hypothesis:

1. $\theta = \theta_0$
2. $\theta \leq \theta_0$
3. $\theta \geq \theta_0$

Steps:

1. Set H_0 as what you wish to reject
2. Gather data for a random sample.
3. Calculate P-value for the data.
4. If P-value is **small** enough, reject H_0 at the [P-value] level of significance.

2.1 P-value

Definition:

$$P[D|H_0] \leq P\text{-value}$$

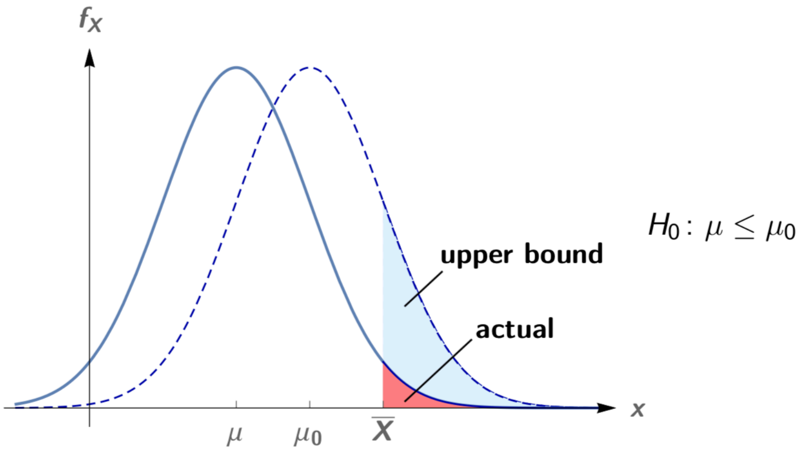
Interpretation:
D represents the statistical data.
But do notice here, take Z-test as example, D does not means obtaining this specific value of $\overline{x}$. D can be understood as the case "value of $\overline{X}$ being $\overline{x}$ or worse ones for supporting H_0 true". Simply means $\overline{X}$ being $\overline{x}$ or even further from what H_0 expects.

Example:
Z-test, $H_0 : \mu \leq \mu_0$, then

$$P\text{-value} = max\{P[\overline{X} \geq \overline{x}|\mu \leq \mu_0]\} = P[\overline{X} \geq \overline{x}|\mu = \mu_0]$$

$$P\text{-value} = P[\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}] = P[Z \geq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}]$$

Rejecting H_0:
Reject H_0 if P-value is small. Then we say we reject H_0 at the [P-value] level of significance.



Question: P-value

Suppose that a Fisher test of the null hypothesis

$$H_0\colon \mu \leq \mu_0$$

yields a very small P-value. Which of the following statements will be true?

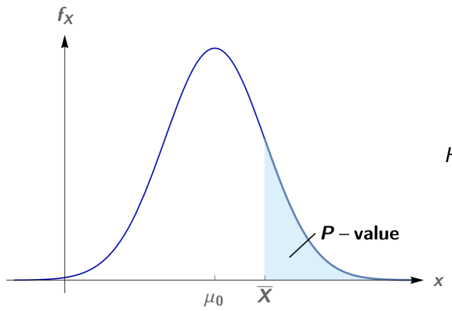
1. It is likely that the true value of μ is much larger than μ_0.
2. Data was obtained that was very unusual, if the assumption is made that H_0 is true.
3. It is unlikely that H_0 is true, given the data that was obtained.
4. The rejection of H_0 is unlikely to be a mistake.

2.2 One-tailed and Two-tailed Test

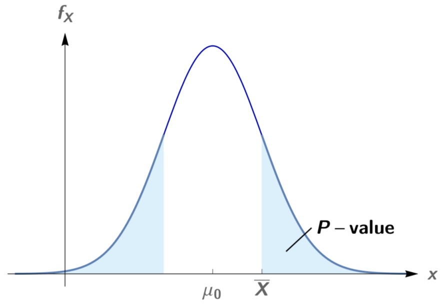
One-tailed:
$H_0 : \theta \leq \theta_0$ or $H_0 : \theta \geq \theta_0$

Two-tailed:
$H_0 : \theta = \theta_0$

Be careful with the different P-value for one-tailed and two-tailed test. For two-tailed test, there is a double.



$H_0 : \mu \leq \mu_0$



$H_0 : \mu = \mu_0$

2.3 Remarks

1. Need multiple, independent significant tests.
2. $P[H_0|D]$ is wanted instead of $P[D|H_0]$, but we obtain $P[D|H_0]$.
3. Two-tailed tests are pointless. H_0 can always be rejected if the sample size n is chosen large enough.

3 Neyman–Pearson Decision Theory

Goal:
Seek to reject H_0, in which case we accept H_1.

1. H_0: null hypothesis
2. H_1: research hypothesis, or alternative hypothesis.

Steps:

(i) Select appropriate hypotheses H_1 and H_0 and a test statistic;

(ii) Fix α and β for the test;

(iii) Use α and β to determine the appropriate the sample size;

(iv) Use α and the sample size to determine the critical region;

(v) Obtain the sample statistic; if the test statistic falls into the critical region, reject H0 at significance level α and accept H_1. Otherwise, accept H_0.

3.1 Type I, Type II Errors and Power

	Actual situation	
Decision	H_0 true	H_1 true
Reject H_0	Type I error (probability α)	Correct decision
Fail to reject H_0	Correct decision	Type II error (probability β)

1. $\alpha := P[\text{ Type I error }] = P[\text{ reject } H_0 \mid H_0 \text{ true }] = P[\text{ accept } H_1 \mid H_0 \text{ true }]$
2. $\beta := P[\text{ Type II error }] = P[\text{ fail to reject } H_0 \mid H_1 \text{ true }] = P[\text{ accept } H_0 \mid H_1 \text{ true }]$
3. Power $:= 1 - \beta = P[\text{ reject } H_0 \mid H_1 \text{ true }] = P[\text{ accept } H_1 \mid H_1 \text{ true }]$

3.2 α and the Critical Region

Definition:
If H_0 is true, then the probability of the test statistic's values falling into the critical region is $\leq \alpha$.

Rejecting H_0:
If the value of the test statistic falls **into** the critical region, then we **reject** H_0.

Example:
Z-test, $H_0 : \mu = \mu_0$, $\quad (H_1 : |\mu - \mu_0| \geq \delta)$, then

$$\alpha = P[\overline{X} \text{ in the critical region} \mid \mu = \mu_0]$$

Notice $P[\frac{|\overline{X}-\mu_0|}{\sigma/\sqrt{n}} > z_{\alpha/2}] = \alpha$
So the critical region is set as:

$$\bar{x} \neq \mu_0 \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

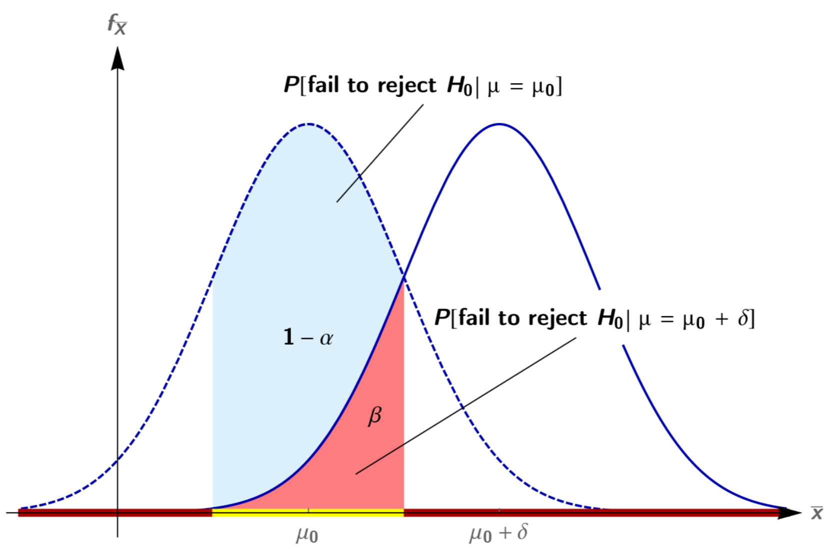
Comment:
The critical region depends on H_0, α, and always the sample size n.
But it has no relation to H_1.

Question: Clarify CR and CI

1. What's the difference between critical region and confidence interval?
2. Clarify CR and CI in Z-test.
3. Recommend: Review the homework problem 7.2 7.6.

3.3 β and the Sample Size

β is decided by H_1, α, and sample size n. (Of course also related to the distribution of the statistic you use.)



1. Intuitively, you can calculate β by integration.

Example:

Z-test, $H_0 : \mu = \mu_0, \quad H_1 : |\mu - \mu_0| \geq \delta$, then

$$\beta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-z_\beta} e^{-t^2/2} dt$$

$$-z_\beta \approx z_{\alpha/2} - \delta\sqrt{n}/\sigma$$

$$n \approx \frac{\left(z_{\alpha/2} + z_\beta\right)^2 \sigma^2}{\delta^2}$$

2. When using a statistic of a typical distribution, read n from OC curves is another efficient way.

3.4 OC Curves

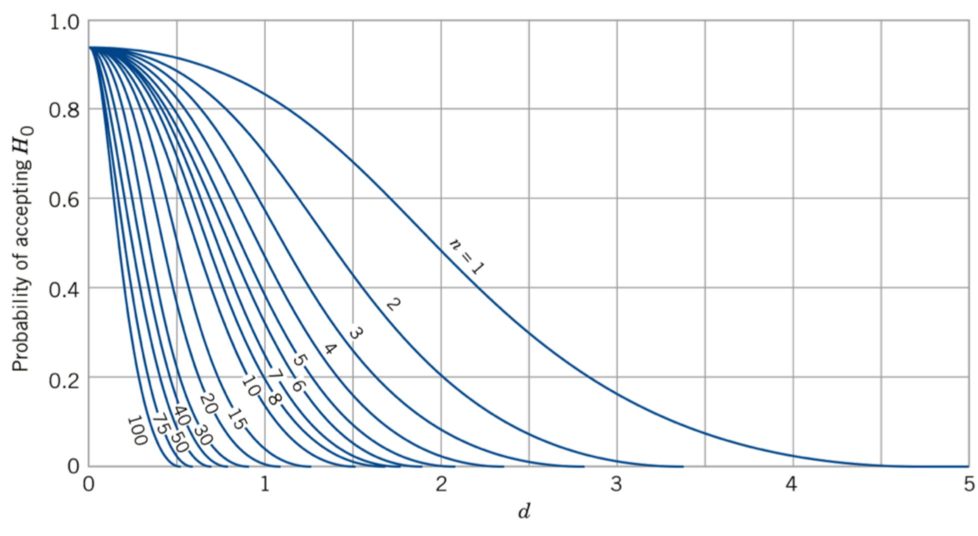
α, β, n, H_1, when any two of them are fixed, then the left two will have a relationship described by a "function", where you need a parameter representing the effect of H_1.

We prefer standardized curves.

Always, in one graph, we fix α, and choose several n. The horizontal ordinate uses a standardized parameter representing the effect of H_1, and the vertical coordinate is $1 - \beta$.

Example for two-tailed Z-test, where the abscissa is standardized as:

$$d = \frac{|\mu - \mu_0|}{\sigma}$$



Basic Reading:

1. Known H_0, H_1, α, β. Determine the needed n.
2. Known H_0, H_1, α, n. Obtain β.
3. Known H_0, α, β, n. Make inferences on H_1.

Question: Neyman–Pearson Decision Theory

Recommend: Review the homework problem 7.3.
Hint:
1. $n_1 \neq n_2$, but in the comparison test for variance, reading from OC curve requires $n_1 = n_2$. Some estimations:... But do remember to indicate your estimation.
2. Calculate by the definition of power. Not required, but can try and ask if you are interested.

4 Null Hypothesis Significance Testing

Steps:

1. Two hypotheses, H_0 and H_1 are set up, but H_1 is always the logical negation of H_0
2. Then either a "hypothesis test" is performed, whereby a critical region for given α is defined, the test statistic is evaluated and H_0 is either rejected or accepted.
3. Alternatively (and more commonly), the test statistic is evaluated immediately, a P-value is found, and H_0 is either rejected or accepted based on that value.
4. In either case, there is no meaningful discussion of β, since H_1 is exactly the negation of H_0.

5 Parametric Test for Mean

Distribution of X_i	Sample size n	Variance σ^2	Statistic
$X_i \sim \mathcal{N}(\mu, \sigma)$	any	known	$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$
$X_i \sim$ any distribution	large	known	$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$
$X_i \sim$ any distribution	large	unknown	$\frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} \sim \mathcal{N}(0,1)$
$X_i \sim \mathcal{N}(\mu, \sigma)$	small	unknown	$\frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$
$X_i \sim$ any distribution	small	known or unknown	Go home!

5.1 Z-test

Test Statistic:

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

Reject at significance level α:

(i) $H_0 : \mu = \mu_0$ if $|Z| > z_{\alpha/2}$

(ii) $H_0 : \mu \leq \mu_0$ if $Z > z_\alpha$

(iii) $H_0 : \mu \geq \mu_0$ if $Z < z_{-\alpha}$

Abscissa of OC Curves:

$$d = \frac{|\mu - \mu_0|}{\sigma}$$

5.2 T-test

Test Statistic:

$$T_{n-1} = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

Reject at significance level α:

(i) $H_0 : \mu = \mu_0$ if $|T_{n-1}| > t_{\alpha/2}$

(ii) $H_0 : \mu \leq \mu_0$ if $T_{n-1} > t_\alpha$

(iii) $H_0 : \mu \geq \mu_0$ if $T_{n-1} < t_{-\alpha}$

Abscissa of OC Curves:

$$d = \frac{|\mu - \mu_0|}{\sigma}$$

Estimate σ:

(i) Use prior experiments to insert a rough estimate for σ

(ii) Express the difference $\delta = |\mu - \mu_0|$ relative to σ

(iii) Substitute the sample standard deviation s for σ.

6 Parametric Test for Variance

Distribution of X_i	Sample size n	Mean μ	Statistic
$X_i \sim \mathcal{N}(\mu, \sigma)$	any	known or unknown	$\frac{s^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}$

6.1 Chi-squared Test

Test Statistic:

$$\chi^2_{n-1} = \frac{(n-1)S^2}{\sigma_0^2}$$

Reject at significance level α:

(i) $H_0 : \sigma = \sigma_0$ if $\chi^2_{n-1} > \chi^2_{\alpha/2,n-1}$ or $\chi^2_{n-1} < \chi^2_{1-\alpha/2,n-1}$

(ii) $H_0 : \sigma \leq \sigma_0$ if $\chi^2_{n-1} > \chi^2_{\alpha,n-1}$

(iii) $H_0 : \sigma \geq \sigma_0$ if $\chi^2_{n-1} < \chi^2_{1-\alpha,n-1}$

Abscissa of OC Curves:

$$d = \frac{\sigma}{\sigma_0}$$

7 Non-parametric Test for Median

1. Non-parametric statistics: do not assume the dependence on any parameter.

2. Distribution-free statistics: do not assume that X follows any particular distribution (such as the nor- mal distribution).

7.1 Sign Test

Number of Signs:

Let $X_1, \ldots, X_n$ be a random sample of size n from an arbitrary continuous distribution and let

$$Q_+ = \#\{X_k : X_k - M_0 > 0\}, \quad Q_- = \#\{X_k : X_k - M_0 < 0\}$$

Q_+ is the number of "positive signs" and Q the number of "negative signs".
If $X_i - M_0 = 0$, usual practice is to exclude X_i from the analysis.

Calculating P-value:

$$P[Q_- \leq k \mid M = M_0] = \sum_{x=0}^{k} \binom{n}{x} \frac{1}{2^n}$$

Rejecting H_0:

We reject at significance level α:

(i) $H_0 : M \leq M_0$ if $P[Q_- < k \mid M = M_0] < \alpha$

(ii) $H_0 : M \geq M_0$ if $P[Q_+ < k \mid M = M_0] < \alpha$

(iii) $H_0 : M = M_0$ if $P[\min(Q_-, Q_+) < k \mid M = M_0] < \alpha/2$

Question: Sign Test

Ex7.1: The diameter of a ball bearing was measured by an inspector using a new type of caliper. The results were as follows (in mm):
0.265, 0.263, 0.266, 0.267, 0.267, 0.265, 0.267, 0.267, 0.265, 0.268, 0.268, 0.263.
1. Use the sign test to evaluate the claim that the median ball diameter is equal to 0.265 mm.

7.2 Wilcoxon Signed Rank Test

Sum of Ranks:

Let $X_1, \ldots, X_n$ be a random sample of size n from a symmetric distribution.

Order the n absolute differences $|X_i - M_0|$ according to magnitude, so that $X_{R_i} - M_0$ is the $R_i{}^t h$ smallest difference by modulus.

If ties in the rank occur, the mean of the ranks is assigned to all equal values.

Let

$$W_+ = \sum_{R_i > 0} R_i, \quad |W_-| = \sum_{R_i < 0} |R_i|$$

W_+ is the sum of "positive ranks" and W the modulus of the sum of "negative signs".

If $X_i - M_0 = 0$, usual practice is to exclude X_i from the analysis.

The Critical Value:

For small n, for example ≤ 30, critical values are given in tables.

Rejecting H_0:

We reject at significance level α:

(i) $H_0 : M \leq M_0$ if W_- is smaller than the critical value for α

(ii) $H_0 : M \geq M_0$ if W_+ is smaller than the critical value for α

(iii) $H_0 : M = M_0$ if $W = min(W_+, |W_-|)$ is smaller than the critical value for $\alpha/2$

Calculating P-Value:

For large n, we use the normal distribution to approximately calculate P-value, with parameters

$$\mu = \frac{n(n+1)}{4}, \quad \sigma^2 = \frac{n(n+1)(2n+1)}{24}$$

Rejecting H_0:

We reject at the [P-value] significance level if [P-value] is small.

Question: Wilcoxon Signed Rank Test

Ex7.1:

0.265, 0.263, 0.266, 0.267, 0.267, 0.265, 0.267, 0.267, 0.265, 0.268, 0.268, 0.263.

2. Use the Wilcoxon signed-rank test to evaluate the claim that the median ball diameter is equal to 0.265 mm.

3. Comment on and interpret the results of your tests.

8 Inferences on Proportion

8.1 Estimator and Statistic

Define the random variable:

$$X = \begin{cases} 1 & \text{has trait,} \\ 0 & \text{does not have trait.} \end{cases}$$

If we take a random sample X_1,...,X_n of X, the sample mean is an (unbiased) estimator for p:

$$\widehat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

By the central limit theorem, when n is large enough, $\widehat{p}$ is approximately normally distributed with mean p and variance p(1 − p)/n. Hence,

$$Z = \frac{\widehat{p} - p}{\sqrt{p(1-p)/n}}$$

8.2 Interval Estimation and Sample Size

It follows immediately that the following is a 100(1 − α)% confidence interval for p:

$$\widehat{p} \pm z_{\alpha/2}\sqrt{p(1-p)/n}$$

But the interval depends on the unknown parameter p, which we are actually trying to estimate! One solution to the problem is to replace p by $\widehat{p}$, i.e., to write

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\widehat{p}(1-\widehat{p})/n}$$

We may want to be able to claim that "with 100(1 − α)% probability, $\widehat{p}$ differs from p by at most d."

Given a 100(1 − α)% confidence interval p = $\widehat{p}$ ± z know with 100(1 − α)% confidence that

$$d = z_{\alpha/2}\sqrt{\widehat{p}(1-\widehat{p})/n}$$

Given d, this means that we should choose

$$n = \frac{z_{\alpha/2}^2\widehat{p}(1-\widehat{p})}{d^2}$$

which requires us to have an estimate $\widehat{p}$ of p beforehand.

8.3 Test for Proportion

Let X_1, ... , X_n be a random sample of (large) size n from a Bernoulli distribution with parameter p and let $\widehat{p} = \text{X}$ denote the sample mean. Then any test based on the statistic

$$Z = \frac{\widehat{p} - p_0}{\sqrt{p_0\left(1 - p_0\right)/n}}$$

is called a large-sample test for proportion.
We reject at significance level α:

(i) $H_0 : p = p_0$ if $|Z| > z_{\alpha/2}$

(ii) $H_0 : p \leq p_0$ if $Z > z_\alpha$

(iii) $H_0 : p \geq p_0$ if $Z < z_\alpha$

Question: Neyman–Pearson Test for "Proportion"

Your factory is ordering a large number of widgets from a supplier. Each widget can be either functional or defective. The supplier guarantees that at most 3% of the widgets are defective. Since the widgets are cheap, you are actually willing to accept a rate of defectives as high as 8% before there is cause for concern.

The widgets are shipped in batches of $N = 10,000$ items. A sample of size $n = 100$ be taken from each batch to ensure that not too many widgets are defective. You and the supplier will agree on a *defective number* d so that

- If there are at least d defectives in the sample, the supplier agrees that the defective rate is greater than 3% and the batch can be rejected;
- If there are fewer than d defectives, you (the buyer) accepts the batch.

In the following, state all assumptions and/or approximations that you are making.

1. Set up a Neyman-Pearson test to decide between accepting and rejecting a batch.
2. How large does d need to be so that any shipment that is returned has a 99% chance of containing more than 3% of defectives?
3. Given this value of d, what is the probability that you end up accepting a batch with more than 8% of defectives?

9 Comparison of Two Proportions

For large sample size:

$$\overline{X}^{(1)} \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right), \quad \overline{X}^{(2)} \sim N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right)$$

So for large sample size:

$$\widehat{p_1 - p_2} = \widehat{p}_1 - \widehat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

Similarly we deduce the following $100(1-\alpha)\%$ confidence interval for $p_1 - p_2$:

$$\widehat{p}_1 - \widehat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}$$

9.1 Large-sample Test for Differences in Proportions

Suppose two random samples of (large) sizes n_1 and n_2 from two Bernoulli distributions with parameters p_1 and p_2 are given. Denote by $\widehat{p}_1$ and $\widehat{p}_2$ the means of the two samples.

Let $(\widehat{p}_1 - \widehat{p}_2)_0$ be a null value for the difference $p_1 - p_2$. Then the test based on the statistic

$$Z = \frac{\widehat{p}_1 - \widehat{p}_2 - (p_1 - p_2)_0}{\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}}$$

is called a large-sample test for differences in proportions.

We reject at significance level α:

(i) $H_0 : p_1 - p_2 = (p_1 - p_2)_0$ if $|Z| > z_{\alpha/2}$

(ii) $H_0 : p_1 - p_2 \leq (p_1 - p_2)_0$ if $Z > z_\alpha$

(iii) $H_0 : p_1 - p_2 \geq (p_1 - p_2)_0$ if $Z < -z_\alpha$

9.2 Pooled Test for Equality of Proportions

Suppose two random samples of (large) sizes n_1 and n_2 from two Bernoulli distributions with parameters p_1 and p_2 are given. Denote by $\widehat{p}_1$ and $\widehat{p}_2$ the means of the two samples.

Let $\widehat{p}$ be the pooled estimator for the proportion, which is defined as

$$\widehat{p} := \frac{n_1\widehat{p}_1 + n_2\widehat{p}_2}{n_1 + n_2}$$

Then the test based on the statistic

$$Z = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

is called a pooled large-sample test for equality of proportions.
We reject at significance level α:

(i) $H_0 : p_1 = p_2$ if $|Z| > z_{\alpha/2}$

(ii) $H_0 : p_1 \leq p_2$ if $Z > z_\alpha$

(iii) $H_0 : p_1 \geq p_2$ if $Z < -z_\alpha$