MATH1005 Cheat Sheet

Sampling without Replacement

We have a population of size N = $N_1 + N_2 + \ldots + N_k$. The probability that a sample of size *n* of the composition n = $n_1 + n_2 + \ldots + n_k$, probability = $\dfrac{\binom{N_1}{n_1}\binom{N_2}{n_2}\cdots\binom{N_k}{n_k}}{\binom{N}{n}}$.

Independence

A and B are independent if and only if $P(A \cap B) = P(A)P(B)$.

Binomial Random Variables

If X ~ B(n, p), the mean of X is np and the variance of X is np(1 – p).\

Distribution Function

$F(x) = P(X \le x)$

The mean and variance of Y = aY + b are E(Y) = aE(Y) + b, and var(Y) = a^2var(X).

Central Limit Theorem

For large n, $\overline{X}$ is approximately $N\left(\mu, \frac{\sigma^2}{n}\right)$.

Hypothesis Testing

1. Null hypothesis
2. Alternative hypothesis
3. Test statistic
4. Sampling distribution of test statistic
5. What observed values of test statistic argue against H_0?
6. Find the observed value of the test statistic
7. Find the P-value
8. Findings

Test for p from B(n, p)

- Test statistic = X
- Under H_0, X ~ B(n, p_0)

Sign test

- Test statistic = X = number of + signs

Two-sample binomial tests

- Test statistic = $\tau = \frac{X_1}{n_1} - \frac{X_2}{n_2}$
- Under H_0, test statistic is approximately $N\left(0, p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$.

Tests for μ when *σ is known*

- Test statistic = $\overline{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$ (normally distributed)

Tests for μ when *σ is unknown*

- Test statistic = $\tau = \frac{\overline{X} - \mu_0}{S / \sqrt{n}} \sim t_{n-1}$

Tests of mean difference for *paired data*

- Test statistic = $\tau = \frac{\overline{X} - 0}{S / \sqrt{n}} \sim t_{n-1}$
- Show the mean difference is zero

Two-sample normal tests

- Test statistic = $\tau = \frac{\overline{X} - \overline{Y}}{\sigma\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \sim N(0,1)$
- Shows whether two means are equal or not

Confidence intervals

- When standard deviation is known: Confidence interval for μ is $\overline{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
- Meaning: the random interval calculated contains the real mean with a probability 1 − α
- When standard deviation is not known: Confidence interval for μ is $\overline{x} \pm t' \frac{s}{\sqrt{n}}$
- Probability p from B(n, p)
  - Approximate confidence interval: $p \pm z\sqrt{\frac{p(1-p)}{n}}$
  - Conservative confidence interval: $p \pm z\sqrt{\frac{1}{4n}}$

Chi-squared tests

- $\tau_{obs} = \sum_i \frac{O_i^2}{E_i} - \sum_i E_i$