MATH-338 Midterm 1 Study Guide

THEORY

Day 2: Independent events happen same time, not affecting one another (P(A ∩ B) = P(A)P(B)). Disjoint is the opposite (P(A ∩ B) = 0). Probability Mass Function (PMF) is a dictionary mapping of events to positive probabilities. Over an infinite amount of iterations, RVs converge to a number.

Day 3: Law of Large Numbers: more times means more precise result.

Day 4: Parameter: any numerical quantity that characterizes a given population. Population proportion: a percentage value associated with a population. Sample proportion: the proportion of individuals in a sample sharing a certain trait ($\hat{p}$). Sample Mean($\bar{X}$). Sampling distribution: probability distribution of statistic obtained through a large number of samples drawn (sample **must** be know).

Day 5: We want low bias and high variability. Bias bad. Variability ↓ as the sample size ↑. Binomial Probability Distribution Conditions: Binary outcome (TF), Independent (previous outcomes do **not** affect next.), Number of outcomes, Success is equally likely. 'X' denotes the number of successes and 'n' is the number of elements in your sample. $\hat{P}$ does **NOT** have a binomial distribution.

Day 6: Interacting variables: one variable can affect the another variable (non-independent). Confounding variable: a factor that influences the results of an experiment. Block design: split sample initially based on traits (possibly confounding) then randomly assign in those groups. Matched Pairs Design: blocks sizes of two (only looking with two levels). Repeated Measures Design two similar subjects have the same tests and those results are compared. Hawthorne Effect: individuals know they are being experimented on.

Day 7: Sensitivity: proportion of actual positive. Specificity: proportion of actual negative. Positive Predictive Value: proportion of positive tests that were actually positive. Negative Predictive Value: same as above but for negative. Prevalence: base rate.

Day 8: Neyman-Pearson Testing: This test will allow us to make preemptive decisions based on conditions presented before the study is conducted. These are the theoretical outcomes WITHOUT taking any sample data. Null Hypothesis: nothing unexpected (original hypothesis, H_0). Alternate Hypothesis: "something is happening and we should change our minds" (H_a). Critical region: range of values that corresponds to the rejection of H_0 at some chosen probability level. Type I Error: occurs when a significance test results in the rejection of a true null hypothesis. Type II Error: the data do not provide strong evidence that the null hypothesis is false.

FORMULAS

- Mean of Probability Dist. : $\mu_x = \Sigma x \times p(x)$
- Variance : $\sigma^2{}_x = \Sigma[x^2 \times P(x)] - \mu^2{}_x$
- Standard Deviation : $\sigma_x = \sqrt{\sigma_x}$ and $\sigma_{x+y} = \sqrt{\sigma_x + \sigma_y}$
- Number successes : $X \sim B(n, p)$
- Mean of binomial RV: nP
- Variance of Bernoulli RV: $P(1-P)$
- Variance of binomial RV: $nP(1-P)$
- Standard deviation of binomial RV: $\sqrt{nP(1-P)}$
- Bayes' Rule: $\frac{P(B|A)P(A)}{P(B)}$
- P(B|A) = $\frac{number\,of\,outcomes\,in\,A \cap B}{number\,of\,outcomes\,in\,A} = \frac{P(A \cap B)}{P(A)} > 0$

- Population proportion: $\hat{P} = \frac{X}{n}$
- Variance($\hat{P}$) = $\frac{P(1-P)}{n}$
- Standard Deviation($\hat{P}$) = $\sqrt{\frac{P(1-P)}{n}}$
- Sensitivity: $\frac{TP}{TP+FN}$
- Specificity: $\frac{TN}{TN+FP}$
- PPV: $\frac{TP}{TP+FP}$
- NPV: $\frac{TN}{TN+FN}$
- Prevalence: $\frac{Actual\,Positive}{Actual\,Positive + Actual\,Negative}$
- $\alpha = P(1) -$ P(Concluded H_a | H_0 is true)
- Baseline $\alpha = 0.05$
- $\beta = P(2) -$ P(Concluded H_0 | H_a)
- Power: $1 - \beta$

ABBREVIATIONS AND MISC

- TP: True Positive
- TN: True Negative
- FP: False Positive
- FN: False Negative
- FTP: File Transfer Protocol

- Independent events: $P(A \cap B) = P(A) \times P(B)$
- Conditional probability: $P(A \cap B) = P(A) \times P(B|A)$ [Tree Mapping]
