MATH-338 Final Cheat Sheet [Exam 1]

TERMS

Probability Model: a mathematical representation of a random event. **Sample Space:** a range of values of a random variable. **Probability:** a quantifiable amount for the likelihood of an event occurring. **Independent:** one event does not influence another. In conditional probability, this means $P(A \cap B) = P(A) \cdot P(B)$. **Disjoint:** the probability of two specific events occurring is 0. **Neither (example):** a person does not eat red meat & is vegetarian. **Both (example):** because you passed your driver's test, you have a driver's license. **PMF:** $P(X = x) \rightarrow$ mapped value. **Case:** elements in the collection and has one or more attributes. **Collection (Population):** $\forall \in S$ (S denotes set). **Collection (Sample):** $\subset \in S$. **Parameter:** any numerical quantity that characterizes a given population or some aspect of it (quantitative). **Statistic:** a characteristic of a sample (qualitative). **Distribution (Variable):** a description of the relative number of times each possible outcome will occur in a number of trails. **Distribution (Sampling):** a description of the given sample observed. **Bias:** the process to over-or-under estimate the value of a population. **Biased Estimator:** difference between expected value and the true value of the parameter being estimated. (residual?) **BINS: B**inary outcome, **I**ndependent trials, **N**umber of trials is fixed, **S**ame value of p for all trials. **Explanatory variable:** a type of independent variable (X value) **Response variable:** dependent variable (y-value) **Observational study:** not interacting with test subjects **Experimental study:** manipulate environmental variables that affect test subjects. **Levels of a Factor:** the number of variation of a given factor that were used in an experiment. **Interacting variables:** two variables that interact with each other to produce an interaction effect (internal agent). **Confounding variables:** a variable that influences both the dependent variable and independent variable, causing a spurious action (external agent). **Placebo Effect:** a beneficial effect produced by a placebo treatment, which cannot be attributed to the properties of the placebo itself. This is generally used in the control group. **Control:** a group that is placed in conditions that are seen as inert, not affecting the outcome of the study. **Randomization:** does not lead to bias **Replication:** the ability to do the study with other experimenters and get similar results **Repetition:** do something over and over by the same experimenter getting similar results. **Random Design:** anyone in the sample is likely to get a particular treatment and there are no predetermined groups. **Block Design:** the experimenter divides the groups into based on a criteria then those groups are given specific treatments. **Matched Pairs:** two subjects of similar characteristics are given two different treatments and the differences are compared. **Single-Blind:** information is concealed from the patients in the study **Double-Blind:** information is concealed from both experimenters and the patients in the study. **Conditional Probability:** an events outcome probability is dictated by previous events ¡++¿: ¡++¿

FORMULAS

- $\mu_x = \Sigma x \cdot p(x)$ [expected value]
- $\sigma_x^2 = \Sigma[x^2 \cdot P(x)] - \mu_x^2$ [variance]
- $\sigma_x = \sqrt{\Sigma[x^2 \cdot P(x)] - \mu_x^2}$ [standard devitation]
- $E(X \pm Y) = E(X) \pm E(Y)$ [expected value (transformation)]
- $Var(X \pm Y) = Var(X) \pm Var(Y)$ [variance (transformation)]
- $E(\hat{p}) = p$ [expected value (sample proportion)]
- $\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$ [variance (sample proportion)]
- $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ [standard deviation (sample proportion)]
- $E(X) = \mu_x = nP$ [expected value (BRV)]
- $Var(X) = n \cdot P \cdot (1 - P)$ [variance (BRV)]

TWO WAY TABLE

Validity	Do not reject H_o	Reject H_o
H_o is true	Correct decision	Type I Error α
H_o is false	Type II Error β	Correct decision

- Type I Error: seeing a wolf when there is not a wolf
- Type II Error: not seeing a wolf when there is a wolf

DIAGNOSTIC TEST TABLE

	(+)	(-)	
+	TP	FN	Sensitivity
-	FP	TN	Specificity
	PPV	NPV	

- $PPV = \frac{TP}{TP+FP}$ $\quad NPV = \frac{TN}{TN+FN}$
- $SN = \frac{TP}{TP+FN}$ $\quad SP = \frac{TN}{TN+FP}$

FRAMEWORKS

Neyman-Pearson Hypothesis Testing

- H_o: $\mu = 0$
- H_a: $\mu = n$
- Requires a rejection region, a small area where the null hypothesis should be rejected
- If the observed value falls in the region, H_a is true, reject H_o, vice versa.

Fisher's Significance Testing

- Select an appropriate test
- Set up H_o
- Calculate the theoretical proabability of the results under H_o (p)
- If $p = \alpha \therefore$ statistically significant
- If $p > \alpha \therefore$ statistically insignificant

Null Hypothesis Significance Testing

- H_o: $\theta = \theta_1$ (if candy causes cancer, assume candy does not caus cancer and find counter arguments)
- H_a: $\theta[<, >, \neq]\theta_1$
- Find its distribution under H_o
- Define a critical region such that if in critical region, reject H_o.
- Else fail to reject H_o