MATH 111 FORMULA SHEETS

SAMPLES:

Sample mean: $\overline{x} = \frac{\sum x}{n}$

Sample variance:

$$s^2 = \frac{\sum (x-\overline{x})^2}{n-1} = \frac{\sum (x^2) - \frac{(\sum x)^2}{n}}{n-1}$$

Sample standard deviation:

$$s = \sqrt{s^2}$$

Chebyshev's Rule: For all samples

- At least $\frac{3}{4}$ of the data lie between $\overline{x} - 2s$ and $\overline{x} + 2s$;
- At least $\frac{8}{9}$ of the data lie between $\overline{x} - 3s$ and $\overline{x} + 3s$;
- In general, at least $1 - \frac{1}{n^2}$ of the data lie between $\overline{x} - ns$ and $\overline{x} + ns$.

Empirical Rule: For roughly normal data sets

- Approximately 68% of the data lie between $\overline{x} - s$ and $\overline{x} + s$;
- Approximately 95% of the data lie between $\overline{x} - 2s$ and $\overline{x} + 2s$;
- Approximately 99.7% of the data lie between $\overline{x} - 3s$ and $\overline{x} + 3s$.

Sample $z-$score: $z = \frac{x-\overline{x}}{s}$

Population $z-$score: $z = \frac{x-\mu}{\sigma}$

PROBABILITY:

For any event S, $0 \leq P(S) \leq 1$

Complement Rule: $P(A^C) = 1 - P(A)$

Addition Rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$; if A and B are disjoint, $P(A \cup B) = P(A) + P(B)$.

Multiplication Rule: $P(A \cap B) = P(A) \cdot P(B|A)$; if A and B are independent, $P(A \cap B) = P(A) \cdot P(B)$

Conditional Probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Independence: A and B are independent if $P(A) = P(A|B)$ or $P(B) = P(B|A)$

COUNTING:

Multiplicative Rule: If you draw one element from each of sets with size $n_1, n_2, \ldots n_k$, there are $n_1 n_2 n_3 \cdots n_k$ different results possible.

Combinations/Partitions: If you draw n elements from a set of N elements without regard for order, the number of different results is

$$\binom{N}{n} = {}_N C_n = \frac{N!}{n! \cdot (N-n)!}$$

More generally, if you partition a set of N elements into k groups, with n_1 elements in the first group, n_2 elements in the second group, etc., the number of different results is

$$\frac{N!}{n_1! \cdot n_2! \cdots n_k!}$$

Permutations: If you draw n elements from a set of N elements and arrange the elements into a distinct order, the number of different results is

$${}_N P_n = \frac{N!}{(N-n)!}$$

DISCRETE VARIABLES:

Mean (or expected value): $\mu = E(X) = \sum x \cdot p(x)$

Variance: $\sigma^2 = E(X-\mu)^2 = \sum (x-\mu)^2 \cdot p(x)$ or $\sigma^2 = E(X^2) - \mu^2 = (\sum x^2 \cdot p(x)) - \mu^2$

Standard deviation: $\sigma = \sqrt{\sigma^2}$

Binomials: If X is binomial with parameters n and p then $\mu = np$, $\sigma^2 = npq$, and $p(x) = \binom{n}{x} p^x q^{n-x}$

Poisson: If X is a Poisson variable with parameter λ then $\mu = \lambda$, $\sigma^2 = \lambda$, and $p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$

Hypergeometrics: If X is hypergeometric, where n elements are drawn from a population of size N that has r successes initially, then $\mu = \frac{nr}{N}$, $\sigma^2 = \frac{r(N-r)n(N-n)}{N^2(N-1)}$, and

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

CONTINUOUS VARIABLES:

Uniform: If X has a uniform distribution over the interval $[a, b]$, then $\mu = \frac{a+b}{2}$, $\sigma = \frac{b-a}{\sqrt{12}}$, and $f(x) = \frac{1}{b-a}$

Exponential: If X has an exponential distribution with parameter θ then $\mu = \theta$, $\sigma = \theta$, and $f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$ Also, $P(X \geq a) = e^{-\frac{x}{a}}$

Normal: If X is normal then $z = \frac{x-\mu}{\sigma}$

Centiles: For normal variable X, the centile is $x_\alpha = \mu + z_\alpha \sigma$

SAMPLING DISTRIBUTIONS:

Central Limit Theorem: For a random sample from a (large) population the sampling ditribution of $\overline{x}$ is approximately normal.

Sample mean: For a random sample of n elements from a population with mean μ and standard deviation σ, the sampling distribution has $\mu_{\overline{x}} = \mu$ and $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$

Normal approximation to binomial: If X is binomial with parameters n and p, and $np \pm 3\sqrt{npq}$ is in the interval $[0, n]$,(same as $9q \leq np$ and $9p \leq nq$), then $P(a \leq X \leq b) = P(a-1 < X < b+1)$ is approximately $P(a - \frac{1}{2} < Y < b + \frac{1}{2})$ where Y is normal with $\mu = np$ and $\sigma = \sqrt{npq}$

CONFIDENCE INTERVALS:

For μ with known σ or large random sample: $\overline{x} \pm z_{\frac{\alpha}{2}}(\frac{\sigma}{\sqrt{n}})$

Sample size (margin of error b) for μ :

$$n \geq \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{b}\right)^2$$

For μ with unknown σ and small sample from a normal population: $\overline{x} \pm t_{\frac{\alpha}{2}}(\frac{s}{\sqrt{n}})$ for t with $n-1$ degrees of freedom

For proportion p : If $\hat{p} \pm 3\sqrt{\frac{\hat{p}\hat{q}}{n}}$ is within the interval (0,1),(same as $9\hat{q} < n\hat{p}$ and $9\hat{p} < n\hat{q}$),then $\hat{p} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$

Sample size (margin of error b) for p :

$$n \geq \left(\frac{z_{\frac{\alpha}{2}}}{b}\right)^2 pq$$

ONE-POPULATION TESTS:

One-sample $z-$test for μ with known σ or large random sample: The test statistic is $z = \frac{\overline{x}-\mu_0}{(\frac{\sigma}{\sqrt{n}})}$

H_1	Reject H_0 if	$p-$value
$\mu \neq \mu_0$	$\lvert z\rvert > z_{\frac{\alpha}{2}}$	$2P(Z > \lvert z\rvert)$
$\mu > \mu_0$	$z > z_\alpha$	$P(Z > z)$
$\mu < \mu_0$	$z < -z_\alpha$	$P(Z < z)$

One-sample $t-$test for μ with unknown σ and small random sample from a normal population: The test statistic is $t = \frac{\overline{x}-\mu_0}{(\frac{s}{\sqrt{n}})}$ and we use t with $n-1$ degrees of freedom

H_1	Reject H_0 if	$p-$value
$\mu \neq \mu_0$	$\lvert t\rvert > t_{\frac{\alpha}{2}}$	$2P(T > \lvert t\rvert)$
$\mu > \mu_0$	$t > t_\alpha$	$P(T > t)$
$\mu < \mu_0$	$t < -t_\alpha$	$P(T < t)$

One-sample test for p : If $\hat{p} \pm 3\sigma_{\hat{p}}$ is within the interval (0,1) then the test statistic is $z = \frac{\hat{p}-p_0}{\sigma_{\hat{p}}}$ where $\sigma_{\hat{p}} = \sqrt{\frac{p_0 q_0}{n}}$

H_1	Reject H_0 if	$p-$value
$p \neq p_0$	$\lvert z\rvert > z_{\frac{\alpha}{2}}$	$2P(Z > \lvert z\rvert)$
$p > p_0$	$z > z_\alpha$	$P(Z > z)$
$p < p_0$	$z < -z_\alpha$	$P(Z < z)$

TWO-POPULATION TESTS

Two-sample $z-$test for means with known $\sigma's$ or large random samples: The test statistic is $z = \frac{(\overline{x_1}-\overline{x_2})-D_0}{\sigma}$ where $\sigma = \sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}$

H_1	Reject H_0 if	$p-$value
$\mu_1 - \mu_2 \neq D_0$	$\lvert z\rvert > z_{\frac{\alpha}{2}}$	$2P(Z > \lvert z\rvert)$
$\mu_1 - \mu_2 > D_0$	$z > z_\alpha$	$P(Z > z)$
$\mu_1 - \mu_2 < D_0$	$z < -z_\alpha$	$P(Z < z)$

Two-sample $t-$test for means with unknown $\sigma's$ and small random samples, where variances are equal: The test statistic is $t = \frac{(\overline{x_1}-\overline{x_2})-D_0}{s}$ for $s = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ and $s_p = \sqrt{\frac{(n_1-1)(s_1)^2+(n_2-1)(s_2)^2}{n_1+n_2-2}}$ and we use t with $n_1 + n_2 - 2$ degrees of freedom

H_1	Reject H_0 if	$p-$value
$\mu_1 - \mu_2 \neq D_0$	$\lvert t\rvert > t_{\frac{\alpha}{2}}$	$2P(T > \lvert t\rvert)$
$\mu_1 - \mu_2 > D_0$	$t > t_\alpha$	$P(T > t)$
$\mu_1 - \mu_2 < D_0$	$t < -t_\alpha$	$P(T < t)$

Paired sample $t-$test: The test statistic is $t = \frac{\overline{x_D}-D_0}{\sigma}$ where $\sigma = \frac{\sigma_D}{\sqrt{n}}$ and use t with $n-1$ degrees of freedom

H_1	Reject H_0 if	$p-$value
$\mu_D \neq D_0$	$\lvert t\rvert > t_{\frac{\alpha}{2}}$	$2P(T > \lvert t\rvert)$
$\mu_D > D_0$	$t > t_\alpha$	$P(T > t)$
$\mu_D < D_0$	$t < -t_\alpha$	$P(T < t)$

Two-sample $z-$test for proportions: The test statistic is

$$z = \frac{\hat{p_1} - \hat{p_2}}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}}$$

where $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$

H_1	Reject H_0 if	$p-$value
$p_1 - p_2 \neq 0$	$\lvert z\rvert > z_{\frac{\alpha}{2}}$	$2P(Z > \lvert z\rvert)$
$p_1 - p_2 > 0$	$z > z_\alpha$	$P(Z > z)$
$p_1 - p_2 < 0$	$z < -z_\alpha$	$P(Z < z)$