MA 222

FINAL EXAM FORMULA SHEET

Probability & Counting	
Addition Rule ($E \cap F \neq \emptyset$): $P(E\ or\ F) = P(E) + P(F) - P(E\ and\ F)$	
Addition Rule for mutually exclusive events where ($E \cap F = \emptyset$): $P(E\ or\ F) = P(E) + P(F)$	
Multiplication Rule for dependent events: $P(E\ and\ F) = P(E) \cdot P(F\|E) = P(F) \cdot P(E\|F)$	
Multiplication Rule for independent events: $P(E\ and\ F) = P(E) \cdot P(F)$	Combination (order doesn't matter & repeats not allowed) ${}_nC_k = \frac{n!}{k!\,(n-k)!}$
Expected Value $E[x] = x_1p_1 + x_2p_2 + \cdots + x_np_n$	Permutation (order matters & repeats not allowed) ${}_nP_k = \frac{n!}{(n-k)!}$
Conditional Probability: $P(F\|E) = \frac{P(E\ and\ F)}{P(E)}$	Location: $L = n \cdot \frac{P}{100}$ Pth %ile: $P = \frac{L}{n} \cdot 100$

Sample Mean: $\bar{x} = \frac{\sum x}{n}$

Sample Standard Deviation: $s = \sqrt{\frac{\sum x^2 - \frac{1}{n}(\sum x)^2}{n-1}}$

Discrete Random Variables:

$\mu = E(X) = \sum_{i=1}^{n} x_i\, p(x_i)$ $\qquad$ $\sigma^2 = E(X^2) - \mu^2 = \sum_{i=1}^{n} x_i^2\, p(x_i) - \mu^2$

Binomial Random Variables:

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$\mu = E(X) = np$ $\qquad$ $\sigma_X = \sqrt{np(1-p)}$

Poisson Random Variables:

$$P(X = x; \mu) = \frac{e^{-\mu}\mu^x}{x!} \qquad E(X) = \mu \qquad \sigma_X = \sqrt{\mu}$$

Continuous Random Variables:

$$P(a \le X \le b) = \int_a^b f(x)dx$$

$$\mu = E(X) = \int_{-\infty}^{\infty} f(x)dx \qquad Var(X) = E(X^2) - \mu^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

Uniform Distribution: $U[a, b]$

$$\mu = E(X) = \frac{a+b}{2} \qquad Var(X) = \frac{(b-a)^2}{12}$$

Normal Distribution:

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

Jointly Distributed Random Variables: <u>**Discrete Case**</u>

1) Joint probability mass function (pmf)

$$p(x, y) = P(X = x \text{ and } Y = y)$$

$$p(x, y) \ge 0$$

$$\sum_x \sum_y p(x, y) = 1$$

$$P[(X, Y) \in A] = \sum_{(x,y)\in A} \sum p(x, y)$$

2) Marginal probability mass function of X

$$p_X(x) = \sum_{y:\, p(x,y) > 0} p(x, y)$$

- Expected value of X

$$\mu_X = E(X) = \sum_X [x \cdot p_X(x)]$$

- Variance of X

$$Var(X) = \sigma_X^2 = \sum_x [x^2 \cdot p_X(x)] - \mu_X^2$$

- Standard deviation of X

$$\sigma_X = \sqrt{Var(X)} = \sqrt{\sum_X [x^2 \cdot p_X(x)] - \mu_X^2}$$

3) Marginal probability mass function of Y

$$p_Y(y) = \sum_{x:\, p(x,y) > 0} p(x, y)$$

- Expected value of Y

$$\mu_Y = E(Y) = \sum_y [y \cdot p_Y(y)]$$

- Variance of Y

$$Var(Y) = \sigma_Y^2 = \sum_y [y^2 \cdot p_Y(y)] - \mu_Y^2$$

- Standard deviation of Y

$$\sigma_Y = \sqrt{Var(Y)} = \sqrt{\sum_y [y^2 \cdot p_Y(y)] - \mu_Y^2}$$

4) Independent Random Variables

    - X and Y are independent if

$$p(x, y) = p_X(x) \cdot p_Y(y) \qquad \text{for all values of x and y}$$

5) Expected value of a function $h(X, Y)$

$$E[h(X, Y)] = \sum_x \sum_y [h(x, y) \cdot p(x, y)]$$

Jointly Distributed Random Variables: **Continuous Case**

6) Joint probability density function (pdf)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$$

$$P[(X,Y) \in A] = \int_A\int f(x,y)dxdy$$

$$P(a \le X \le b,\ c \le Y \le d) = \int_a^b\int_c^d f(x,y)dydx$$

7) Marginal probability density function of X

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

- Expected value of X

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$$

- Variance of X

$$Var(X) = \sigma_X^2 = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)\, dx - \mu_X^2$$

- Standard deviation of X

$$\sigma_X = \sqrt{Var(X)} = \sqrt{\int_{-\infty}^{\infty} x^2 \cdot f_X(x)\, dx - \mu_X^2}$$

8) Marginal probability density function of Y

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

- Expected value of Y

$$\mu_Y = E(Y) = \int_{-\infty}^{\infty} \mathrm{y} \cdot \mathrm{f_Y(y)}\, \mathrm{dy}$$

- Variance of Y

$$Var(Y) = \sigma_Y^2 = \int_{-\infty}^{\infty} y^2 \cdot f_Y(y)\, dy - \mu_Y^2$$

- Standard deviation of Y

$$\sigma_Y = \sqrt{Var(Y)} = \sqrt{\int_{-\infty}^{\infty} y^2 \cdot f_Y(y)\, dy - \mu_Y^2}$$

9) Independent Random Variables
   - X and Y are independent if

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

10) Expected value of a function $h(X, Y)$

$$E[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y) dx dy$$

Sampling Distribution of Sample Means: **Central Limit Theorem**

For any given population with mean μ and standard deviation σ, a sampling distribution of sample means will have the following three characteristics if <u>either</u> the sample size, *n*, is at least 30 <u>or</u> the population is normally distributed.

1. The mean of a sampling distribution of sample means equals

$$\mu_{\bar{X}} = \mu$$

2. The standard deviation of a sampling distribution of sample means equals

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

3. Formula for z-score:

$$z = \frac{\bar{x} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Sampling Distribution of Sample Proportions: **Central Limit Theorem**

1) For large samples, the sample proportion is approximately normally distributed, with the mean and standard deviation given by the following formula:

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

where p is the population proportion and n is the sample size.

- A sample is large if the interval $[p - 3\sigma_{\hat{p}}, p + 3\sigma_{\hat{p}}]$ lies wholly within the interval [0, 1].

2) Formula for z-score

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Types of Errors in Hypothesis Testing		**Reality**	
		H_0 is true	H_0 is false
Decision	Reject H_0	Type 1 Error	Correct Decision
	Fail to reject H_0	Correct Decision	Type 2 Error

c	$\alpha = 1 - c$	z_α	$\pm z_{\alpha/2}$
0.90	0.10	1.28	±1.645
0.95	0.05	1.645	±1.96
0.98	0.02	2.05	±2.33
0.99	0.01	2.33	±2.575

Inference about a Population Mean				
Inference Condition		Confidence Interval	Test Statistic	*df*
$n \geq 30$	σ known	$\bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$	$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$	N/A
	σ unknown	$\bar{x} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$	$z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$	N/A
$n < 30$ and normal population	σ known	$\bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$	$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$	N/A
	σ unknown	$\bar{x} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$	$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$	$n - 1$
Inference about a Population Proportion				
$\hat{p} \pm 3\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \in [0,1]$		$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$	N/A

Calculate p-values

- For a left-tailed test, *p*-value = $P(Z \leq z)$.
- For a right-tailed test, *p*-value = $P(Z \geq z)$.
- For a two-tailed test, *p*-value = $P(|Z| \geq |z|)$.

Conclusions Using *p*-Values

- If *p*-value $\leq \alpha$, then *reject* the null hypothesis.
- If *p*-value $> \alpha$, then *fail to reject* the null hypothesis.

<table>
<tr><th colspan="5">Inference about Two Population Means</th></tr>
<tr><td colspan="2">Inference Condition</td><td>Confidence Interval</td><td>Test Statistic</td><td>df</td></tr>
<tr><td rowspan="3">Independent samples</td><td>$n_1 \geq 30$
$n_2 \geq 30$</td><td>$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_1^2}{n_2}}$</td><td>$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$</td><td>N/A</td></tr>
<tr><td rowspan="2">$n < 30$
σ_1 and σ_2 unknown but assumed equal, normal populations</td><td>$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$</td><td>$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$</td><td rowspan="2">$n_1 + n_2 - 2$</td></tr>
<tr><td colspan="2">$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$</td></tr>
<tr><td colspan="2">Paired samples, normal population of differences</td><td>$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$</td><td>$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$</td><td>$n - 1$</td></tr>
</table>

Rejection Regions (in the case of *z*-test statistic)	• For left-tailed test: $z \leq -z_\alpha$ or $(-\infty, -z_\alpha)$ • For right-tailed test: $z \geq z_\alpha$ or (z_α, ∞) • For two-tailed test: $\lvert z \rvert \geq z_{\alpha/2}$ or $\left(-\infty, -z_{\alpha/2}\right) \cup \left(z_{\alpha/2}, \infty\right)$
Rejection Regions (in the case of *t*-test statistic)	• For left-tailed test: $t \leq -t_\alpha$ or $(-\infty, -t_\alpha)$ • For right-tailed test: $t \geq t_\alpha$ or (t_α, ∞) • For two-tailed test: $\lvert t \rvert \geq t_{\alpha/2}$ or $\left(-\infty, -t_{\alpha/2}\right) \cup \left(t_{\alpha/2}, \infty\right)$