	SUPPLEMENT FOR STATISTICS 260	FORMULA REVIEW 1/5
	Flash -Card	**Formula Review**
Item	**Question side of Flash-Card**	**Answer Side of Flash-Card**
1	Histograms are used to check the data for	outliers, centrality and dispersion.
2	For an observed sample $x_1, x_2, \cdots, x_n$	
	sample mean $\overline{x} =$	$\frac{1}{n}\sum_{i=1}^{n} x_i$
	sample median $\widetilde{x} =$	middle ranked observation (n odd), or average of two middle ranked observations (n even)
	sample variance $s^2 =$	$\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$
	sample standard deviation $s =$	$\sqrt{s^2}$
3	$B_1, B_2, \cdots, B_n$ are mutually exclusive iff	$B_i \cap B_j = \varnothing$ for all $i \neq j$
4	$P(B) =$	the chance that event B will occur on any trial
5	$P(B') = P(\overline{B}) = P(B^c)$	$1 - P(B)$
6	$P(A \cup B) =$	$P(A) + P(B) - P(A \cap B)$
7	$P(A \cup B \cup C) =$	$P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C)$ $-P(B \cap C) + P(A \cap B \cap C)$
8	$P(A\|B) =$	$P(A \cap B)/P(B)$
9	$P(A \cap B) =$	$P(A)\, P(B\|A)$ and $P(B)\, P(A\|B)$
10	A & B are independent iff $P(A \cap B) =$	$P(A)P(B)$
11	The cdf of rv X is $F(x) =$	$P(X \leq x)$
12	The pmf of discrete rv X is $p(x) =$	$P(X = x)$
13	If rv X is discrete, $P(a \leq X \leq b) =$	$\sum_{a \leq x \leq b} P(X = x)$
14	If rv X is discrete, $E(X) = \mu_x =$	$\sum_{all\ x} x\ P(X = x)$
15	If rv X is discrete, $E[g(X)] =$	$\sum_{all\ x} g(x)\ P(X = x)$
16	$Var(X) = \sigma^2 = \sigma_x^2 =$	$E(X^2) - \mu_x^2 = E[(X - \mu)^2]$
17	$SD(X) = \sigma = \sigma_x =$	$\sqrt{Var(X)}$

		2/5
18	If X = total number of successes out of n independent trials where P(success)$= p$ on every trial, then the distribution of X is:	Binomial (n, p)
19	If $X \sim$ Binomial (n, p), then formulae for pmf, mean, value, and standard deviation are:	$\binom{n}{x} p^x(1-p)^{n-x}$, np, $\sqrt{np(1-p)}$
20	If events occur at random in time (or space) at the average rate of λ per unit time (or space), and X = total number of events that occur in a time (or space) window of size t, then the distribution of X is:	Poisson (λt)
21	If X ˜ Poisson(λ), then formulae for pmf, mean value, and standard deviation are:	$\frac{\lambda^x}{x!}e^{-\lambda}, \lambda, \sqrt{\lambda}$
22	If $X \sim$ Binomial (n, p), with n large, p small then the distribution of X is well approximated by:	Poisson($\lambda = np$)
23	If rv X is continuous with pdf f, then $P(a \leq X \leq b) = P(a < X < b) =$	$\int_a^b f\ (x)dx$
24	If rv X is continuous with pdf f, then $E(X) = \mu_x =$	$\int_{\infty}^{\infty} x\ f(x)dx$
25	If rv X is continuous with pdf f, then $E(g(X)) =$	$\int_{-\infty}^{\infty} g(x)f(x)dx$
26	To find η = the $(100p)$th percentile of a continuous distribution with cdf F, solve:	$F(\eta) = p$ for η
27	To find the median $\widetilde{\mu}$ of a continuous distribution with cdf F, solve:	$F(\tilde{\mu}) = 0.5$ for $\widetilde{\mu}$
28	If $X \sim Normal(\mu, \sigma)$, then the distribution $Z = \frac{X-\mu}{\sigma}$ is:	Standard Normal=$Normal(0, 1)$
29	The 100p'th percentile of $Normal(0,\ 1)$ is:	$\eta_z(p)$
30	The 100p'th percentile of $Normal(\mu,\ \sigma^2)$ is:	$\mu + \eta_z(p)\sigma$
31	If $X \sim$ Binomial(n, p) with $np \geq 5$, $n(1-p) \geq 5$ then the distribution of X is well approximated by:	$Normal(\mu = np, \sigma^2 = np(1-p))$ Can use cont. corr. here.
32	If events occur at random in time at the average rate of λ per unit time, and W = the waiting time till the next event occurs, then the distribution time of W is:	Exponential (λ)
33	If $X \sim$ Exponential (λ), then formulas for pdf, cdf, mean value, and standard deviation are:	$\lambda e^{-\lambda x}$ for $x > 0, 1 - e^{-\lambda x}$ for $x > 0$, $1/\lambda, 1/\lambda$

		3/5
34	If $X \sim$ Exponential (λ), then for all $x, y \geq 0$ we have $P(X > x + y \mid X > x) =$	$P(X > y)$ Memoryless property
35	The joint pmf of X, Y is $p(x, y) =$	$P(X = x, Y = y)$
36	Discrete rv's X, Y are independent iff:	$p(x, y) = P(X = x)P(Y = y)$ for all x, y
37	For discrete rv's X, Y $\quad E(h(X, Y))$	$\sum_{all\ x,y} h(x, y)p(x, y)$
38	$\text{Cov}(X, Y) = E((X - \mu_x)(Y - \mu_y)) =$	$E(XY) - E(X)E(Y)$
39	Correlation coefficient $\rho =$	$\frac{Cov(X,Y)}{\sigma_X \sigma_Y}$
40	If X, Y are independent rv's, then $E(XY) =$	$E(X)E(Y)$
41	The sequence of rv's $X_1, X_2, \cdots, X_n$ constitutes a random sample provided:	these rv's are independent and identically (iid) distributed
42	$E(aX + bY + c) =$	$aE(X) + bE(Y) + c$
43	$Var(aX + bY + c) = Var(aX + bY)$	$a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$
44	If $X_1, X_2 \cdots, X_n$ are independent rv's then $Var(a_1X_1 + a_2X_2 + \cdots + a_nX_n) =$	$a_1^2Var(X_1) + a_2^2Var(X_2) + \cdots + a_n^2Var(X_n)$
45	If $X_1, X_2 \cdots, X_n$ is a random sample from a distribution with mean μ and standard deviation σ, then the sample total T has a mean value, standard deviation: the sample mean $\overline{X}$ has mean value, standard deviation:	 $n\mu, \sqrt{n}\sigma$ $\mu, \frac{\sigma}{\sqrt{n}}$
46	the Central Limit Theorem states that if n is sufficiently large ($n \geq 30$ usually will suffice), the approximate distribution of $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{T - n\mu}{\sqrt{n}\sigma}$ is:	Standard Normal
47	linear combination of normally distributed rv's is:	normally distributed

		4/5
48	If $X \sim Binomial(n,p)$, then the sample proportion $\hat{p} = X/n$ has mean value, standard deviation:	p, $\sqrt{\frac{p(1-p)}{n}}$
49	the Central Limit Theorem states that if $np \geq 5$ and $n(1-p) \geq 5$, then the approximate distribution of $\frac{\widehat{p}-p}{\sqrt{p(1-p)/n}}$ is:	Standard Normal
50	Given two independent random samples from normally distributed populations having common variance, we estimate that variance as:	$s_p^2 = \frac{(m-1)s_1^2+(n-1)s_2^2}{m+n-2}$
51	The estimated standard error (ese) of $\overline{x}$ for estimating μ is:	$\frac{s}{\sqrt{n}}$
52	The estimated standard error (ese) of $\widehat{p}$ for estimating p is:	$\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$
53	the estimated standard error (ese) of $\widehat{p}_1 - \widehat{p}_2$ for estimating $p_1 - p_2$ is:	$\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{m} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n}}$
54	The estimated standard error (ese) of $\overline{x} - \overline{y}$ for estimating $\mu_1 - \mu_2$ is:	$\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$ or $\sqrt{s_p^2\left(\frac{1}{m} + \frac{1}{n}\right)}$
55	In item 54, the only case where the second formula for ese is used is when:	$m < 30$ and/or $n < 30$, both distributions are near normal and $\frac{\text{larger of } s_1,s_2}{\text{smaller of } s_1,s_2} \leq 1.4$, suggesting $\sigma_1 = \sigma_2$
56	The estimated standard error (ese) of $\overline{d}$ for estimating μ_D is:	$\frac{s_D}{\sqrt{n}}$
57	Critical value (cv) $z_{\alpha/2}$ satisfies $P(Z > z_{\alpha/2}) =$	$\alpha/2$
58	Critical value (cv) $t_{\alpha/2,\upsilon}$ satisfies $P(T_{(\nu)} > t_{\alpha/2,\nu}) =$	$\alpha/2$
59	100(1-α)% confidence intervals for $\mu, p, \mu_1 - \mu_2$, $p_1 - p_2, \mu_D$ have the form:	estimate $\pm(cv_{\alpha/2})(ese)$
60	For testing hypotheses about $\mu, \mu_1 - \mu_2, \mu_D$, and $p_1 - p_2$, the test statistics have the form:	$\frac{\text{estimate - parameter value under H}_o}{ese}$
61	For testing $H_0 : p = p_0$, the statistic has the form:	$\frac{\widehat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$

		5/5
62	To compute cv or p-value	
	(i) with large sample(s) use:	Standard Normal distribution
	(ii) with $n < 30$ and a near-normal distribution use:	t distribution with df=$n-1$
	(iii) with $m < 30$ and/or $n < 30$, near-normal distributions, $\frac{\text{larger of } s_1, s_2}{\text{smaller of } s_1 s_2} \leq 1.4$, use:	t distribution with df=$m+n-2$
	(iv) with $m < 30$ and/or $n < 30$, near-normal distributions, and $\frac{\text{larger of } s_1, s_2}{\text{smaller of } s_1, s_2} > 1.4$, use:	t distribution with df=ν on Formula List
63	p-value = the probability, under H_0, that a rerun of the experiment would yield:	a test statistic at least as extreme, in the direction of H_1, as observed
64	If p-value ≤ 0.01, there is:	very strong evidence against H_0 in favour of H_1
65	If $0.01 <$ p-value ≤ 0.05,there is:	strong evidence against H_0 in favour of H_1
66	If $0.05 <$ p-value ≤ 0.10,there is:	moderate evidence against H_0 in favour of H_1,
67	If $0.10 <$ p-value, there is:	little or no evidence against H_0 in favour of H_1
68	We should reject H_0 at significance level α if and only if the p-value is:	less than or equal to α
69	For paired data, analyze:	the single sample of differences
70	Assumption underlying large-sample Z procedures for analyzing a single sample or paired differences:	the data is an observed random sample
71	Assumption underlying large-sample Z procedures for analyzing two samples:	the data are two independent observed random samples
72	Assumptions underlying small-sample T procedures for analyzing a single sample or paired differences:	(i) the data is an observed random sample, and (ii) the distribution is near normal
73	Assumptions underlying small-sample; pooled variance estimate, T procedures for analyzing two samples:	(i) the data are two independent observed random samples, (ii) the distributions are near normal and (iii) the population variances are equal
74	Assumptions underlying small-sample unpooled variance estimate, T procedures for analyzing two samples:	(i) the data are two independent observed random samples, (ii) the distributions are near normal and (iii) the population variances are not equal

**Note that this is not a comprehensive list of all concepts and formulae. It is simply a study tool.