STAT 235 FINAL EXAM FORMULA SHEET

Summaries:

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n}, \quad s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{1}{n-1}\left[\sum x_i^{\,2} - \frac{(\sum x_i)^2}{n}\right]$$

IQR = $Q_3 - Q_1$, outliers are observations 1.5*IQR below Q_1 or 1.5*IQR above Q_3.

Probability:

Conditional Probability: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$

Complement Law: $P(A') = 1 - P(A)$

Multiplication Law: $P(A \cap B) = P(A) \times P(B \mid A) = P(B) \times P(A \mid B)$

If A and B are independent, $P(A \cap B) = P(A) \times P(B)$ since $P(A \mid B) = P(A)$.

Addition Law: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

If A and B are mutually exclusive, $P(A \cup B) = P(A) + P(B)$

For any two events A and B: $P(A) = P(A \cap B) + P(A \cap B')$

Permutations: $P_k^n = n \times (n-1) \times (n-2) \times ... \times (n-k+1) = \dfrac{n!}{(n-k)!}$

Combinations: $C_k^n = \dbinom{n}{k} = \dfrac{n!}{k!(n-k)!}$, where $n! = n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$

Linear Combinations of Random Variables:

If $Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n + b$, $E(Y) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n) + b$

If $X_1, X_2, \ldots, X_n$ are independent, $Var(Y) = a_1^2 Var(X_1) + a_2^2 Var(X_2) + ... + a_n^2 Var(X_n)$

Discrete Variables:

$\mu = E(X) = \sum_{i=1} x_i p_i$, $F(x) = P(X \le x)$ = Sum of probabilities p_i for $x_i \le x$,

$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2 = \sum_{i=1} (x_i - \mu)^2 p_i = \sum_{i=1} x_i^2 p_i - \mu^2$

Distribution	Probability mass function	Mean	Variance
Binomial	$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$	$E(X) = np$	$Var(X) = np(1-p)$
Geometric	$f(x) = P(X = x) = (1-p)^{x-1} p$	$E(X) = \frac{1}{p}$	$Var(X) = \frac{(1-p)}{p^2}$
Negative Binomial	$f(x) = P(X = x) = \binom{x-1}{r-1} (1-p)^{x-r} p^r$	$E(X) = \frac{r}{p}$	$Var(X) = \frac{r(1-p)}{p^2}$
Poisson	$f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$	$E(X) = \lambda$	$Var(X) = \lambda$

Continuous Variables:

$\int_{-\infty}^{\infty} f(x)dx = 1$, $P(a \le X \le b) = \int_a^b f(x)dx$, $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$

$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$, $\sigma^2 = Var(X) = E(X^2) - [E(X)]^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$

Distribution	**Density function**	**Mean**	**Variance**	***F*(*x*)**
Uniform	$f(x) = \frac{1}{b-a}$	$E(X) = \frac{b+a}{2}$	$Var(X) = \frac{(b-a)^2}{12}$	$F(x) = \frac{x-a}{b-a}$
Exponential	$f(x) = \lambda e^{-\lambda x}$	$E(X) = \frac{1}{\lambda}$	$Var(X) = \frac{1}{\lambda^2}$	$F(x) = 1 - e^{-\lambda x}$

Normal: If X has a normal distribution with mean μ and variance σ^2, then $X \sim N(\mu, \sigma^2)$.

Standard normal: If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$ and $\Phi(z) = P(Z \le z)$.

Normal approximation to binomial: If X follows a binomial distribution with parameters n and p, then

$$P(a \le X \le b) \approx P(a - 0.5 \le Y \le b + 0.5),$$

where Y is normal with the mean np and the variance $np(1 - p)$, under the assumption that $np \ge 15$ and $n(1 - p) \ge 15$.

Sampling Distributions

Sample mean	Sample proportion
$E(\bar{X}) = \mu$	$E(\hat{p}) = p$
$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$	$Z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{\hat{p} \pm \frac{0.5}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$
Normal if $n \ge 30$ OR population is normal	Normal if $np \ge 15$ and $n(1 - p) \ge 15$

Inferences about μ (σ known)

$z_0 = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$, $\bar{x} \pm z_{\alpha/2} \times \left(\frac{\sigma}{\sqrt{n}}\right)$, $\mu \le \bar{x} + z_\alpha \times \frac{\sigma}{\sqrt{n}}$, $\bar{x} - z_\alpha \times \frac{\sigma}{\sqrt{n}} \le \mu$

Inferences about μ (σ unknown)

$t_0 = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$, $\bar{x} \pm t_{\alpha/2,\, n-1} \times \left(\frac{s}{\sqrt{n}}\right)$, $\mu \le \bar{x} + t_{\alpha, n-1} \times \frac{s}{\sqrt{n}}$, $\bar{x} - t_{\alpha, n-1} \times \frac{s}{\sqrt{n}} \le \mu$

Inferences about μ_d for paired data

$t_0 = \frac{\bar{d} - \delta_0}{s_d / \sqrt{n}}$, $\bar{d} \pm t_{\alpha/2,\, n-1} \times \left(\frac{s_d}{\sqrt{n}}\right)$

Inferences about $\mu_1 - \mu_2$ (independent samples, variances known)

$$z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \; (\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Inferences about $\mu_1 - \mu_2$ (independent samples, variances unknown, equal variances)

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}, \; s.e.(\bar{x}_1 - \bar{x}_2) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$t_0 = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{s.e.(\bar{x}_1 - \bar{x}_2)}$, $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,\,v} \times s.e.(\bar{x}_1 - \bar{x}_2)$, t-distribution with $v = df = n_1 + n_2 - 2$

Inferences about $\mu_1 - \mu_2$ (independent samples, variances unknown, variances not assumed equal)

$t_0 = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$, $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,\,v} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, t-distribution with $v = df = \min\{n_1, n_2\} - 1$

Simple Linear Regression (SLR)

Simple linear regression model: $Y = \beta_0 + \beta_1 x + ERROR$, $ERROR \sim N(0, \sigma^2)$.
Estimated Regression Line: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}, \; S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}), \; \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}, \; \hat{\sigma}^2 = \frac{SSE}{n-2},$$

$$SSE = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2$$

$$SST = SSR + SSE, \; F_0 = \frac{SSR/1}{SSE/(n-2)} = \frac{MSR}{MSE}, \; R^2 = \frac{SSR}{SST}$$

Inferences in Regression

$t_0 = \dfrac{\hat{\beta}_1 - b_1}{s.e.(\hat{\beta}_1)}$, $\hat{\beta}_1 \pm t_{\alpha/2,\,v} \times s.e.(\hat{\beta}_1)$, $s.e.(\hat{\beta}_1) = \sqrt{\dfrac{\hat{\sigma}^2}{S_{xx}}}$, t distribution with $v = n - 2$

$$\left(\hat{\beta}_0 + \hat{\beta}_1 x^*\right) \pm t_{\alpha/2,\,v} \times \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}\;;\left(\hat{\beta}_0 + \hat{\beta}_1 x^*\right) \pm t_{\alpha/2,\,v} \times \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

Inferences about *p*

$z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$, under the assumption that $np_0 \geq 15$ and $n(1 - p_0) \geq 15$

$\hat{p} \pm z_{\alpha/2} \times \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$, $p \geq \hat{p} - z_{\alpha}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$, $p \leq \hat{p} + z_{\alpha}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$,

under the assumption that $n\hat{p} \geq 15$ *and* $n(1-\hat{p}) \geq 15$.

Inferences about $p_1 - p_2$

$\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$, $z_0 = \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$, $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Assumptions: $n_1\hat{p}_1$, $n_1(1-\hat{p}_1)$, $n_2\hat{p}_2$, $n_2(1-\hat{p}_2)$ are all larger than 15.

One-Way ANOVA

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$, $SST = SSTr + SSE$

$$F_0 = \frac{SSTr/(k-1)}{SSE/(N-k)} = \frac{MSTr}{MSE} \sim F_{k-1,N-k}$$