STAT 305 Exam II
Reference Sheet

Numeric Summaries

mean	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
population variance	$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
population standard deviation	$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
sample variance	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
sample standard deviation	$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Linear Relationships

Form	$y \approx \beta_0 + \beta_1 x$
Fitted linear relationship	$\hat{y} = b_0 + b_1 x$
Least squares estimates	$b_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$
	$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$
	$b_0 = \bar{y} - b_1\bar{x}$
Residuals	$e_i = y_i - \hat{y}_i$
sample correlation coeffecient	$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(x_i-\bar{x})^2}}$
	$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2)}}$
coeffecient of determination	$R^2 = (r)^2$
	$\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2 - \sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$

Multivariate Relationships

Form	$y \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$
Fitted relationship	$\hat{y} \approx b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$
Residuals	$e_i = y_i - \hat{y}_i$
Sums of Squares	$\text{SSTO} = \sum_{i=1}^{n}(y_i - \bar{y})^2$
	$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
	$\text{SSR} = SSTO - SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$
coeffecient of determination	$R^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2 - \sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$
	$R^2 = \frac{\text{SSTO}-\text{SSE}}{\text{SSTO}}$
	$R^2 = \frac{\text{SSR}}{\text{SSTO}}$
	$\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2 - \sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$

Functions

Quantile Function $Q(p)$ For a univariate sample consisting of n values that are ordered so that $x_1 \le x_2 \le \ldots \le x_n$ and value p where $0 \le p \le 1$, let $i = \lfloor n \cdot p + 0.5 \rfloor$. Then the quantile function at p is:

$$Q(p) = x_i + (n \cdot p + 0.5 - i)(x_{i+1} - x_i)$$

Basic Probability

Definitions

Random experiment	A series of actions that lead to an observable result. The result may change each time we perform the experiment.
Outcome	The result(s) of a random experiment.
Sample Space (S)	A set of all possible results of a random experiment.
Event (A)	Any subset of sample space.
Probability of an event ($P(A)$)	the likelihood that the observed outcome of a random experiment is one of the outcomes in the event.
A^C	The outcomes that are not in A.
$A \cap B$	The outcomes that are both in A and in B.
$A \cup B$	The outcomes that are either A or B.

General Rules

Probability A given B	$P(A\|B) = P(A \cap B)/P(B)$
Probability A and B	$P(A \cap B) = P(A\|B)P(B) = P(B\|A)P(A)$
Probability A or B	$P(A \text{ or } B) = P(A) + P(B) - P(A, B)$

Independence

Two events are called independent if $P(A, B) = P(A) \cdot P(B)$. Clever students will realize this also means that if A and B are independent then $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

Joint Probability

Joint Probability	The probability an outcome is in event A and in event $B = P(A, B)$.
Marginal Probability	If $A \subseteq B \cup C$ then $P(A) = P(A \cap B) + P(A \cap C)$.
Conditional Probability	For events A and B, if $P(B) \neq 0$ then $P(A\|B) = P(A \cap B)/P(B)$.

Discrete Random Variables

General Rules

Probability function	$f_X(x) = P(X = x)$
Cumulative probability function	$F_X(x) = P(X \leq x)$
Expected Value	$\mu = E(X) = \sum_x x f_X(x)$
Variance	$\sigma^2 = Var(X) = \sum_x (x - \mu)^2 f_X(x)$
Standard Deviation	$\sigma = \sqrt{Var(X)}$

Joint Probability Functions

Joint Probability Function	$f_{XY}(x, y) = P[X = x, Y = y]$
Marginal Probability Function	$f_X(x) = \sum_y f_{XY}(x, y)$ $f_Y(y) = \sum_x f_{XY}(x, y)$
Conditional Probability Function	$f_{X\|Y}(x\|y) = f_{XY}(x, y)/f_Y(y)$ $f_{Y\|X}(y\|x) = f_{XY}(x, y)/f_X(x)$

Geometric Random Variables

X is the trial count upon which the first successful outcome is observed performing independent trials with probability of success p.

Possible Values	$x = 1, 2, 3, \ldots$
Probability function	$P[X = x] = f_X(x) = p^x(1 - p)^{x-1}$
Expected Value	$\mu = E(X) = \frac{1}{p}$
Variance	$\sigma^2 = Var(X) = \frac{1-p}{p^2}$

Binomial Random Variables

X is the number of successful outcomes observed in n independent trials with probability of success p.

Possible Values	$x = 0, 1, 2, \ldots, n$
Probability function	$P[X = x] = f_X(x) = \frac{n!}{(n-x)!x!} p^x(1 - p)^{n-x}$
Expected Value	$\mu = E(X) = np$
Variance	$\sigma^2 = Var(X) = np(1 - p)$

Poisson Random Variables

X is the number of times a rare event occurs over a predetermined interval (an area, an amount of time, etc.) where the number of events we expect is λ.

Possible Values	$x = 0, 1, 2, 3, \ldots$
Probability function	$P[X = x] = f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$
Expected Value	$E(X) = \lambda$
Variance	$Var(X) = \lambda$

Continuous Random Variables

General Rules

Probability density function	$P[a \leq X \leq b] = \int_a^b f_X(x)dx$
Cumulative density function	$P[X \leq x] = F_X(x) = \int_{-\infty}^x f_X(t)dt$
Expected Value	$\mu = E(X) = \int_{-\infty}^{\infty} x f_X(x)dx$
Variance	$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx$
Standard Deviation	$\sigma = \sqrt{Var(X)}$

Joint Probability Density Functions

Joint Probability Density Function	$f_{XY}(x,y)$ is the joint density of both X and Y.
	$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x,y)dydx$
Marginal Probability Density Function	$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy$
	$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx$
Conditional Probability Density Function	$f_{X\|Y}(x\|y) = f_{XY}(x,y)/f_Y(y)$
	$f_{Y\|X}(y\|x) = f_{XY}(x,y)/f_X(x)$

Uniform Random Variables

Used when we believe an outcome could be anywhere between two values a and b but have no other beliefs.

Probability density function	$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & o.w. \end{cases}$
Cumulative density function	$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{1}{b-a}x - \frac{a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$
Expected Value	$E(X) = \frac{1}{2}(b+a)$
Variance	$Var(X) = \frac{1}{12}(b-a)^2$

Exponential Random Variables

Used when we an outcome could be anything greater than 0 but the likelihood is concentrated on smaller values.

Probability density function	$f_X(x) = \begin{cases} \frac{1}{\alpha}\exp\left(-\frac{x}{\alpha}\right) & x \geq 0 \\ 0 & o.w. \end{cases}$
Cumulative density function	$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - \exp\left(-\frac{x}{\alpha}\right) & x \geq 0 \end{cases}$
Expected Value	$E(X) = \alpha$
Variance	$Var(X) = (\alpha)^2$

Normal Random Variables

Used when we believe an outcome could be above or below a certain value μ but we also believe it is more likely to be close to μ than it is to be far away.

Probability density function	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Cumulative density function	There is no general formula.
Expected Value	$E(X) = \mu$
Variance	$Var(X) = \sigma^2$

Standard Normal Random Variables (Z)

A normal random variable with mean 0 and variance σ^2.

Probability density function	$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$
Cumulative density function	There is no general formula.
Expected Value	$E(Z) = 0$
Variance	$Var(Z) = 1$
Relationship with $X \sim N(\mu, \sigma^2)$	If X is normal(μ,σ^2) then $P[a \leq X \leq b] = P\left[\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right]$

Functions of Random Variables

Linear Combinations of Independent Random Variables

For $X_1, X_2, \ldots, X_n$ independent random variables and $a_0, a_1, a_2, \ldots, a_n$ constants if $U = a_0 + a_1X_1 + \ldots + a_nX_n$:

- $E(U) = a_0 + a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n)$
- $Var(U) = a_1^2 Var(X_1) + a_2^2 Var(X_2) + \ldots + a_n^2 Var(X_n)$