Stat 305 Final Exam
Reference Sheet

Numeric Summaries

Basic 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 $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$

Quantiles

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

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

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 $r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$

$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right)}}$

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 $R^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2 - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$

$R^2 = \frac{\text{SSTO} - \text{SSE}}{\text{SSTO}}$

$R^2 = \frac{\text{SSR}}{\text{SSTO}}$

$\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2 - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$

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	$P[X = x] = f_X(x) = p(1-p)^{x-1}$
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 $N(\mu, \sigma^2)$ 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 $U = a_0 + a_1X_1 + \ldots + a_nX_n$:

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

$(1-\alpha)\cdot 100\%$ Confidence Intervals

Notation and Definitions

$z_{1-\#}$: the value such that for a standard normal $P(Z \le z_{1-\#}) = 1-\#$.

$t_{k,1-\#}$: the value such that for a t-distribution with k degrees of freedom $P(T \le t_{k,1-\#}) = 1-\#$.

Pooled Variance $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$

Two-Sided [Sample Estimate] $\pm$ [Distribution]$\sqrt{\frac{[\text{Variance}]}{[\text{Sample Size}]}}$

Upper Bound [Sample Estimate] + [Distribution]$\sqrt{\frac{[\text{Variance}]}{[\text{Sample Size}]}}$

Lower Bound Sample Estimate] − [Distribution]$\sqrt{\frac{[\text{Variance}]}{[\text{Sample Size}]}}$

Two-Sided Intervals for μ

Large sample size, σ known $\bar{x} \pm z_{1-\alpha/2}\sqrt{\sigma^2/n}$

Large sample size, σ unknown $\bar{x} \pm z_{1-\alpha/2}\sqrt{s^2/n}$

Small sample size $\bar{x} \pm t_{n-1,1-\alpha/2}\sqrt{s^2/n}$

Two-Sided Intervals for $\mu_1 - \mu_2$

Large sample size, σ_1 known, σ_2 known $\bar{x}_1 - \bar{x}_2 \pm z_{1-\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

Large sample size, σ_1, σ_2 unknown $\bar{x}_1 - \bar{x}_2 \pm z_{1-\alpha/2}\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

Small sample size, normal population, $\sigma_1 = \sigma_2$ $\bar{x}_1 - \bar{x}_2 \pm t_{n_1+n_2-2,1-\alpha/2}\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$

Hypothesis Tests

Test Statistics for μ

Large sample size, σ known $Z = \dfrac{\bar{x}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$

Large sample size, σ unknown $Z = \dfrac{\bar{x}-\mu}{\sqrt{s^2/n}} \sim N(0,1)$

Small sample size, $T = \dfrac{\bar{x}-\mu}{\sqrt{s^2/n}} \sim t_{n-1}$

P-value table

Situation	K	$\text{H}_a: \mu \ne \mu_0$	$\text{H}_a: \mu < \mu_0$	$\text{H}_a: \mu > \mu_0$
$n \ge 25, \sigma$ known	$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$P(\lvert Z\rvert > K)$	$P(Z < K)$	$P(Z > K)$
$n \ge 25, \sigma$ unknown	$\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	$P(\lvert Z\rvert > K)$	$P(Z < K)$	$P(Z > K)$
$n < 25, \sigma$ unknown	$\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	$P(\lvert T\rvert > K)$	$P(T < K)$	$P(T > K)$

Test Statistics for $\mu_1 - \mu_2$

Large sample size, σ_1 known, σ_2 known $Z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

Large sample size, σ_1 unknown, σ_2 unknown $Z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1)$

Small sample size, normal population, $\sigma_1 = \sigma_2$ $T = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \sim N(0,1)$

Common Values for z

z	$P(Z \le z)$	Two-Sided Confidence	One Sided Confidence
1.28	0.90	0.80	0.90
1.64	0.95	0.90	0.95
1.96	0.975	0.95	0.975
2.24	0.9875	0.975	0.9875
2.33	0.99	0.98	0.99
2.58	0.995	0.99	0.995

Quantiles of t-distributions

df	$Q(0.8)$	$Q(0.85)$	$Q(0.9)$	$Q(0.95)$	$Q(0.975)$	$Q(0.99)$	$Q(0.995)$
1	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.860	1.064	1.325	1.725	2.086	2.528	2.845