STAT 105 Exam I 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	$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Quantile Function $Q(p)$ For a dataset 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	$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(x_i-\bar{x})^2}}$
	$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2)}}$
coeffecient of determination	$R^2 = (r)^2$
	$\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2 - \sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$

Factorial Analysis (Two Factors)

Assuming

- Factor A with levels $1, 2, ..., I$,
- Factor B with levels $1, 2, ..., J$,
- n is the total number of observations,
- n_{ij} is the total number of observations with Factor A at level i and Factor B at level j,
- $n_{i.}$ is the total number of observations with Factor A at level i,
- $n_{.j}$ is the total number of observations with Factor B at level j.
- y_{ijk} is the kth observation where Factor A is at level i and Factor B is at level j.

$y_{..} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}$	$\bar{y}_{..} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}$
$\bar{y}_{i.} = \frac{1}{n_{i.}}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}$	$\bar{y}_{.j} = \frac{1}{n_{.j}}\sum_{i=1}^{I}\sum_{k=1}^{K} y_{ijk}$
Main effect of Factor A at level i	$a_i = \bar{y}_{i.} - \bar{y}_{..}$
Main effect of Factor B at level j	$b_j = \bar{y}_{.j} - \bar{y}_{..}$
Fitted Value	$\hat{y}_{ij} = a_i + b_j + \bar{y}_{..}$

Discrete Random Variables

Probability function	$P[X = x] = f_X(x)$
Cumulative probability function	$P[X \leq x] = F_X(x)$
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 Distributions and Related Distributions

Joint Probability Function	$P[X = x, Y = y] = f(x, y)$
Marginal Probability Function	$P[X = x] = f_X(x) = \sum_{\text{all } y} f(x, y)$ $P[Y = y] = f_Y(y) = \sum_{\text{all } x} f(x, y)$
Conditional Probability Function	$P[X = x \mid Y = y] = \frac{f(x,y)}{f_Y(y)}$ $P[Y = y \mid X = x] = \frac{f(x,y)}{f_X(x)}$

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	$\sigma^2 = Var(X) = np(1-p)$

Continuous Random Variables

Probability density function	$P[a \leq X \leq b] = \int_a^b f_X(x)dx$
Cumulative probability 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	$\sigma = \sqrt{Var(X)}$

Normal Random Variables

Let X be a normal random variable with mean μ and variance σ^2.

Probability density function	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
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.
If X is normal(μ,σ^2) then $P[a \leq X \leq b] = P\left[\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right]$

Probability density function	$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

Functions of 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 $W = a_0 + a_1 X_1 + \ldots + a_n X_n$:

- $E(W) = a_0 + a_1 E(X_1) + a_2 E(X_2) + \ldots + a_n E(X_n)$
- $Var(W) = a_1^2 Var(X_1) + a_2^2 Var(X_2) + \ldots + a_n^2 Var(X_n)$

Confidence Intervals and Hypothesis Tests

Confidence Intervals $n \geq 25$

$(1-\alpha) \cdot 100\%$ Confidence interval for population mean	$\bar{x} \pm z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}$
$(1-\alpha) \cdot 100\%$ Confidence lower bound	$\bar{x} - z_{1-\alpha}\sqrt{\frac{\sigma^2}{n}}$
$(1-\alpha) \cdot 100\%$ Confidence upper bound	$\bar{x} + z_{1-\alpha}\sqrt{\frac{\sigma^2}{n}}$

Confidence Intervals $n < 25$

$(1-\alpha) \cdot 100\%$ Confidence interval for population mean	$\bar{x} \pm t_{1-\alpha/2,n-1}\sqrt{\frac{\sigma^2}{n}}$
$(1-\alpha) \cdot 100\%$ Confidence lower bound	$\bar{x} - t_{1-\alpha,n-1}\sqrt{\frac{\sigma^2}{n}}$
$(1-\alpha) \cdot 100\%$ Confidence upper bound	$\bar{x} + t_{1-\alpha,n-1}\sqrt{\frac{\sigma^2}{n}}$

Test statistics in hypothesis tests for population mean

$n \geq 25$	$\frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}}$ $N(0,1)$
$n < 25$	$\frac{\bar{x}-\mu}{\sqrt{\sigma^2/n}}$ t with $\nu = n-1$ degrees of freedom