STAT 105 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 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:

$$Q(p) = \begin{cases} x_i & \lfloor n \cdot p + 0.5 \rfloor = n \cdot p + 0.5 \\ x_i + (n \cdot p - i + 0.5)(x_{i+1} - x_i) & \lfloor n \cdot p + 0.5 \rfloor \neq n \cdot p + 0.5 \end{cases}$$

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	$\sigma = \sqrt{Var(X)}$

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	$\sigma = \sqrt{Var(X)}$