etsinf

TABLE OF FORMULAS

Statistics – Bachelor in Computer Engineering - DEIOAC

Sample Parameters

Average:

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Median or C_2 :

If N is odd $\Rightarrow$ value at the position (N+1)/2

If N is even $\Rightarrow$ average of values at positions N/2 and (N/2+1)

Quartiles:

C_1 (*Lower quartile*) is the first quartile if:

No. data $\leq C_1$ is higher or equal to N/4

No. data $\geq C_1$ is higher or equal to 3N/4

C_3 (*Upper quartile*) is the third quartile if:

No. data $\leq C_3$ is higher or equal to 3N/4

No. data $\geq C_3$ is higher or equal to N/4

Variance: $S^2 = \sum_{i=1}^{N} \frac{(x_i - \bar{X})^2}{N-1}$

Standard deviation: $S = \sqrt{S^2}$

Range:

$R = X_{max} - X_{min}$

Interquartile range:

$RI = C_3 - C_1$

Coefficient of variation:

$CV = \frac{S}{\bar{X}}$

Skewness coefficient:

$$CA = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^3 / (N-1)}{S^3}$$

Standardized skewness coefficient (SSC):

If SSC < -2 $\Rightarrow$ negative skew

If SSC $\in$ [-2, 2] $\Rightarrow$ symmetric distribution

If SSC > 2 $\Rightarrow$ positive skew

Coefficient of kurtosis:

$$CK = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^4 \Big/ N-1}{S^4} - 3$$

Standardized kurtosis coefficient (SKC):

If SKC < -2 $\Rightarrow$ platykurtic data

If SKC $\in$ [-2, 2] $\Rightarrow$ mesokurtic data ("normal")

If SKC > 2 $\Rightarrow$ leptokurtic data

Covariance:

$$cov_{xy} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{N-1}$$

Correlation coefficient:

$$r_{xy} = \frac{cov_{xy}}{s_x \cdot s_y} \qquad r_{xy} \in [-1, +1]$$

Probability

Properties:

$0 \leq P(A) \leq 1$

If **A** and **B** are **exclusive** $\Rightarrow$ $P(A \cup B) = P(A) + P(B)$ $\Rightarrow$ $P(A \cap B) = 0$

Complementary event: $P(\bar{A}) = 1 - P(A)$

$P(\varnothing) = 0$ *being* ϕ *the empty set*

Rule of Laplace:

$$P(A) = \frac{\text{favorable cases}}{\text{possible cases}}$$

De Morgan's laws:

$\overline{A \cup B} = \bar{A} \cap \bar{B}$

$\overline{A \cap B} = \bar{A} \cup \bar{B}$

Union of events:

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

$P(A \cup B \cup C)=P(A)+P(B)+P(C)-P(A \cap B)-P(A \cap C)-P(B \cap C)+P(A \cap B \cap C)$

In general:

$P(A_1 \cup \ldots \cup A_n)=\sum(P(A_i))-\sum\left(P(A_i \cap A_j)\right)+\sum(P(A_i \cap A_j \cap A_k)) + \ldots + (-1)^{n+1}\left(\sum(P(A_1 \ldots A_n))\right)$

Conditional probability:	Intersection of events:	If A and B are independent ⇒
$P(A/B) = \dfrac{P(A \cap B)}{P(B)}$	$P(A \cap B) = P(A).P(B/A)$ $P(A \cap B) = P(B).P(A/B)$	$P(A \cap B) = P(A).P(B)$

Total probability theorem:	Bayes' Theorem
$P(B)=\sum_{j=1}^{n} P(A_j)P(B/A_j) = P(A_1)P(B/A_1)+\ldots+P(A_n)P(B/A_n)$	$P(A_i/B) = \dfrac{P(A_i \cap B)}{P(B)} = \dfrac{P(A_i)P(B/A_i)}{\sum_{j=1}^{n} P(A_j)P(B/A_j)}$

Probability distributions

Distribution function: $F(x)=P(X \le x)$	**Property:** $P(a < X \le b)=F(b)-F(a)$
Discrete random variables	**Continuous random variables**
Probability function: $P(X = x_i)$	**Density function:** $f(x)=\dfrac{dF(x)}{dx}$

Mathematical expectation and population parameters

Average: m=E(X) Discrete distributions: $E(X)=\sum_{i=1}^{n} x_i \cdot P(X=x_i)$ Continuous distr.: $E(X)=\int x \cdot f(x) \cdot dx$

Variance: $\sigma^2 = E(X-m)^2$ Discrete distr.: $\sigma^2 = \sum_{i=1}^{n}(x_i - m)^2 \cdot P(X=x_i)$ Standard deviation: $\sigma = \sqrt{\sigma^2}$

Properties of the average:

If $Y = a_0 \pm a_1 \cdot X_1 \pm a_2 \cdot X_2 \pm \ldots \pm a_n \cdot X_n \Rightarrow m_Y = a_0 \pm a_1 \cdot m_{X1} \pm a_2 \cdot m_{X2} \pm \ldots \pm a_n \cdot m_n$

Particular cases:

If $Y = a + b \cdot X \Rightarrow m_Y = a + b \cdot m_X$	If $Y = a - b \cdot X \Rightarrow m_Y = a - b \cdot m_X$
If $Y = X_1 + X_2 \Rightarrow m_Y = m_{X1} + m_{X2}$	If $Y = X_1 - X_2 \Rightarrow m_Y = m_{X1} - m_{X2}$

Properties of the variance: If $Y = a_0 \pm a_1 \cdot X_1 \pm a_2 \cdot X_2 \Rightarrow \sigma_Y^2 = a_1^2 \cdot \sigma_{X_1}^2 + a_2^2 \cdot \sigma_{X_2}^2 \pm 2 \cdot a_1 \cdot a_2 \cdot \operatorname{cov}_{X_1 X_2}$

Particular cases:

If $Y = a + b \cdot X \Rightarrow \sigma_Y^2 = b^2 . \sigma_X^2$	If $Y = a - b \cdot X \Rightarrow \sigma_Y^2 = b^2 . \sigma_X^2$
If X_1 y X_2 are **independent**: $Y = X_1 \pm X_2 \Rightarrow \sigma_Y^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2$	

Variation coefficient: $CV_x = \sigma_x / m_x$	**Interquartile range:** $C_3 - C_1$	**Range:** $X_{max} - X_{min}$

Covariance: $\sigma_{X_1 X_2}^2 = E\left((X_1 - m_{X_1})(X_2 - m_{X_2})\right)$	**Correlation coefficient:** $\rho_{x_1 x_2} = \dfrac{\operatorname{cov}_{x_1 x_2}}{\sigma_{x_1} \cdot \sigma_{x_2}}$

Most important distributions

Binomial: X~B(n, p) (X = 0, 1, ..., n)

Probability function:	Distribution function:	Average:	Variance:
$P(X=x)=\binom{n}{x}\cdot p^x\cdot(1-p)^{n-x}=\dfrac{n!\cdot p^x\cdot(1-p)^{n-x}}{x!\cdot(n-x)!}$	$P(X\le x)=\sum_{x_i=0}^{x}P(X=x_i)$	$m_X=E(X)=n.p$	$\sigma_X^2=n.p.(1-p)$

Properties:

$$\left.\begin{array}{l} X_1\approx B(n_1,p)\\ \ldots\ldots\ldots\ldots\\ X_N\approx B(n_N,p)\end{array}\right\}\Rightarrow Y=X_1+\ldots+X_N\approx B(n_1+\ldots+n_N,p)$$

Poisson: X ~ Ps (λ) (X = 0, 1, ..., ∞)

Probability function:		Average:	Variance:
$P(X=x)=e^{-\lambda}.\dfrac{\lambda^x}{x!}$	$P(X\le x)=\sum_{x_i=0}^{x}P(X=x_i)\Rightarrow$ Abacus of Poisson	$m_X=E(X)=\lambda$	$\sigma_X^2=\lambda$

Properties:

$$\left.\begin{array}{l} X_1\approx Ps(\lambda_1)\\ \ldots\ldots\ldots\ldots\\ X_N\approx Ps(\lambda_N)\end{array}\right\}\Rightarrow Y=X_1+\ldots+X_N\approx Ps(\lambda_1+\ldots+\lambda_N)$$

Uniform: X ~U (a,b) (a < X < b)

Density function:		Average:	Variance:
$f(x)=\dfrac{1}{b-a}\quad a\le x\le b$	$P(X\le x)=\dfrac{x-a}{b-a}\quad a<x\le b$	$m_X=E(X)=\dfrac{a+b}{2}$	$\sigma_X^2=\dfrac{(b-a)^2}{12}$

Exponential: X ~ Exp (α) $(0\le X\le\infty)$

Density function:		Average:	Variance:
$f(x)=\alpha e^{-\alpha x}\quad x\ge 0$	$P(X\le x)=1-e^{-\alpha x}\quad x\ge 0\Rightarrow$	$m_X=E(X)=\dfrac{1}{\alpha}$	$\sigma_X^2=\dfrac{1}{\alpha^2}$

Normal: X ~ N(m, σ) $(-\infty<X<\infty)$

Density function:		Average:	Variance:
$f(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$	Z ~ N(0;1) standard Normal $P(Z\ge z)\Rightarrow$ Table	$m_X=E(X)=m$	$\sigma_X^2=\sigma^2$

Standard Normal distribution: $Z\sim N(m_Z=0,\sigma_Z=1)$ $\quad$ $If\ \ X\approx N(m,\sigma)\Rightarrow\dfrac{X-m}{\sigma}\approx N(0;\ 1)$

$$P[N(m,\sigma)>x_i]=P\left[N(0;\ 1)>\frac{x_i-m}{\sigma}\right]=P[Z>z]\Rightarrow Table$$

Properties:

$$\left.\begin{array}{l} X_1\approx N(m_{x_1},\sigma^2_{x_1})\\ \ldots\ldots\ldots\ldots\\ X_N\approx N(m_{x_N},\sigma^2_{x_N})\end{array}\right\}\Rightarrow Y=X_1+\ldots+X_N\approx N(m_Y=m_{x_1}+\ldots+m_{x_N};\sigma_Y^2=\sigma^2_{X_1}+\ldots+\sigma^2_{X_N})$$

Standard deviation of Y: $\sigma_Y=\sqrt{\sigma^2_{X_1}+\cdots+\sigma^2_{X_N}}$

Particular cases	
If $Y = a + b \cdot X \Rightarrow Y \sim N(m_Y = a + b.m_X;\ \sigma_Y^2 = b^2.\sigma_X^2)$	If **X** $\sim N(\mathbf{m_x}, \boldsymbol{\sigma_x})$ and **Y** $\sim N(\mathbf{m_y}, \boldsymbol{\sigma_y})$ **independent** $Z = X \mp Y \sim N(m_z = m_x \mp m_y,\ \sigma_z^2 = \sigma_x^2 + \sigma_y^2)$
If **X** $\sim N(\mathbf{m_x}, \boldsymbol{\sigma_x}) \Rightarrow \begin{cases} 68{,}26\% \text{ of X values} \in [m-\sigma,\ m+\sigma] \\ 95{,}44\% \text{ of X values} \in [m-2\sigma,\ m+2\sigma] \\ 99{,}73\% \text{ of X values} \in [m-3\sigma,\ m+3\sigma] \end{cases}$	
Normal approximations	
Central Limit Theorem: $\left.\begin{array}{l} X_1 \approx g_1(m_{X_1}, \sigma_{X_1}^2) \\ \dots\dots\dots \\ X_N \approx g_1(m_{X_N}, \sigma_{X_N}^2) \end{array}\right. \Rightarrow Y = X_1 + \dots + X_N \approx N(m_Y = m_{X_1} + \dots + m_{X_N};\ \sigma_Y^2 = \sigma_{X_1}^2 + \dots + \sigma_{X_N}^2)$	Being: **g** $\rightarrow$ any distribution (Binomial, Poisson, etc.) **N** $\rightarrow \infty$ (N very high)
X~B(n, p) If $\sigma^2 \geq 9 \Rightarrow X \sim N(m = np, \sigma^2 = np(1-p))$	**X ~ Ps (λ)** If $\sigma^2 \geq 9 \Rightarrow X \sim N(m = \lambda, \sigma^2 = \lambda)$

Previous teaching material from R. Alcover (DEIOAC - UPV)

Distributions in the sampling of normal populations	
$X \sim N(m,\sigma)$ and $\bar{x}$ is the average of a sample with size N	$\dfrac{\bar{x}-m}{\sigma/\sqrt{N}} \sim N(0,1)$
$X \sim N(m,\sigma)$ and s^2 is the variance of a sample with size N	$(N-1)\dfrac{s^2}{\sigma^2} \sim \chi^2_{N-1}$
$X \sim N(m,\sigma)$ and $\bar{x}$ and s^2 are the mean and variance of a sample with size N	$t = \dfrac{\bar{x}-m}{s/\sqrt{N}} \sim t_{N-1}$
$X_1 \sim N(m_1,\sigma_1)$, $X_2 \sim N(m_2,\sigma_2)$ independent. $S^2{}_1$ and $S^2{}_2$ are the sample variances of X_1 and X_2 (sizes N_1 and N_2)	$\dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{N_1-1,N_2-1}$

Inference in normal populations

α: significance level (type-I risk) ; **1-α**: confidence level ; β: type-II risk ; **1-β**: power

Hypothesis test for the mean (t - test)

$\left.\begin{array}{l} H_0: m = m_0 \\ H_1: m \neq m_0 \end{array}\right\}$	If $\left\lvert \dfrac{\bar{x}-m_0}{s/\sqrt{N}} \right\rvert \leq t_{N-1}^{\alpha/2} \Rightarrow \text{Accept } H_0$	If ***p-value*** $> \alpha \Rightarrow$ Accept H_0
	If $\left\lvert \dfrac{\bar{x}-m_0}{s/\sqrt{N}} \right\rvert > t_{N-1}^{\alpha/2} \Rightarrow \text{Reject } H_0$	If ***p-value*** $< \alpha \Rightarrow$ Reject H_0
	$t_{N-1}^{\alpha/2} \Rightarrow$ Value in table	

Confidence interval for the mean (CI_m)

$$CI_m \Rightarrow \left[\bar{x} - t_{N-1}^{\alpha/2}\frac{s}{\sqrt{N}},\ \bar{x} + t_{N-1}^{\alpha/2}\frac{s}{\sqrt{N}}\right]$$

$t_{N-1}^{\alpha/2} \Rightarrow$ Value in t table

Hypothesis test for the mean using the confidence interval

$$\left.\begin{array}{l} H_0: m = m_0 \\ H_1: m \neq m_0 \end{array}\right\} \qquad \begin{array}{l} \textit{If } m_0 \in CI_m \Rightarrow \textit{Accept } H_0 \\ \textit{If } m_0 \notin CI_m \Rightarrow \textit{Reject } H_0 \end{array}$$

Confidence interval for the variance (σ^2)

$$IC_{\sigma^2} \Rightarrow \left[\frac{(N-1)S^2}{g_2}, \frac{(N-1)S^2}{g_1} \right]$$

g_1 and $g_2 \Rightarrow$ values in Chi^2 table so that: $P(\chi^2_{N-1} > g_1) = 1 - \frac{\alpha}{2}$ and $P(\chi^2_{N-1} > g_2) = \frac{\alpha}{2}$

Hypothesis test for the variance (σ^2) using the confidence interval

$$\left.\begin{array}{l} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{array}\right\}$$

If $\sigma_0^2 \in CI_{\sigma^2} \Rightarrow$ Accept H_0

If $\sigma_0^2 \notin CI_{\sigma^2} \Rightarrow$ Re ject H_0

Hypothesis test for the comparison of means

$$\left.\begin{array}{l} H_0 : m_1 = m_2 \\ H_1 : m_1 \neq m_2 \end{array}\right\}$$

If $\left| \frac{\bar{x}_1 - \bar{x}_2}{s_{(\bar{x}_1 - \bar{x}_2)}} \right| \leq t^{\alpha/2}_{N_1+N_2-2} \Rightarrow$ Accept H_0

If $\left| \frac{\bar{x}_1 - \bar{x}_2}{s_{(\bar{x}_1 - \bar{x}_2)}} \right| > t^{\alpha/2}_{N_1+N_2-2} \Rightarrow$ Re ject H_0

If ***p-value*** $> \alpha \Rightarrow$ Accept H_0

If ***p-value*** $< \alpha \Rightarrow$ Reject H_0

$t^{\alpha/2}_{N_1+N_2-1} \Rightarrow$ Value in table

α = Significance level

$$S_{(\bar{x}_1 - \bar{x}_2)} = S \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

$$S = \sqrt{\frac{(N_1 - 1)\ S_1^2 + (N_2 - 1)\ S_2^2}{N_1 + N_2 - 2}}$$

Confidence interval for the difference of means ($m_1 - m_2$)

$$IC_{m_1 - m_2} \Rightarrow (\bar{x}_1 - \bar{x}_2) \pm t^{\alpha/2}_{N_1+N_2-2}\ S_{(\bar{x}_1 - \bar{x}_2)}$$

$t^{\alpha/2}_{N_1+N_2-1} \Rightarrow$ Value in t table

Hypothesis test for the comparison of means using the confidence interval

$$\left.\begin{array}{l} H_0 : m_1 = m_2 \\ H_1 : m_1 \neq m_2 \end{array}\right\}$$

If $0 \in CI_{m_1 - m_2} \Rightarrow$ Accept H_0

If $0 \notin CI_{m_1 - m_2} \Rightarrow$ Re ject H_0

Confidence interval for the ratio of variances (σ^2_1 / σ^2_2)

$$IC_{\sigma_1^2/\sigma_2^2} \Rightarrow \left(\frac{S_1^2}{S_2^2\ f_2}, \frac{S_1^2}{S_2^2\ f_1} \right)$$

f_1, $f_2 \Rightarrow$ Values in F table so that: $P(F_{(N_1-1),(N_2-1)} > f_1) = 1 - \frac{\alpha}{2}$ and $P(F_{(N_1-1),(N_2-1)} > f_2) = \frac{\alpha}{2}$

Hypothesis test for the comparison of variances using the confidence interval

$$\left.\begin{array}{l} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{array}\right\}$$

If $1 \in CI_{\sigma_1^2/\sigma_2^2}$ $\left[or\ if\ s_1^2 / s_2^2 < F^{\alpha}_{n_1-1, n_2-1} \right] \Rightarrow$ *Accept* H_0

If $1 \notin CI_{\sigma_1^2/\sigma_2^2}$ $\left[or\ if\ s_1^2 / s_2^2 > F^{\alpha}_{n_1-1, n_2-1} \right] \Rightarrow$ *Re ject* H_0

being $s_1^2 > s_2^2$

Analysis of Variance (ANOVA)

Nomenclature

N = total number of observations	$\mathbf{F_i}$ = factor i $\mathbf{F_j}$ = factor j $\mathbf{F_i}$ x $\mathbf{F_j}$ = interaction of $\mathbf{F_j}$ with $\mathbf{F_i}$	**I** = number of levels in factor $\mathbf{F_i}$ **J** = number of levels in factor $\mathbf{F_j}$	**SS** = Sum of Squares **df** = degrees of freedom

$\mathbf{SS_{TOTAL}}$ = total sum of squares $\mathbf{SS_{Fi}}$ = Sum of squares of factor i $\mathbf{SS_{Fj}}$ = Sum of squares of factor j $\mathbf{SS_{FixFj}}$ = SS of the interaction $\mathbf{F_i}$ x $\mathbf{F_j}$ $\mathbf{SS_{res}}$ = residual sum of squares	$\mathbf{df_{Tot}}$ = total degrees of freedom $\mathbf{df_{Fi}}$ = degrees of freedom associated to the SS of factor $\mathbf{F_i}$ $\mathbf{df_{Fj}}$ = deg. freeedom associated to the SS of factor $\mathbf{F_j}$ $\mathbf{df_{Fi}}$ = deg. fr. associated to the SS of the interaction $\mathbf{F_i x F_j}$ $\mathbf{df_{res}}$ = residual degrees of freedoms

df_{Tot} → (N-1)	$\mathbf{df_{Fi}}$ → (I-1)	$\mathbf{df_{Fj}}$ → (J-1)	$\mathbf{df_{FjxFj}}$ → (I-1)x(J-1)	$df_{res} = df_{total} - \left(\sum_{\forall Fact} df_{factors} + \sum_{\forall Fact} df_{int\,eractions}\right)$

Fundamental equation of ANOVA

$$SS_{total} = \sum_{\forall Fact} SS_{factors} + \sum_{\forall int} SS_{int\,eractions} + \sum SS_{resid}$$

Hypothesis test for ANOVA (F test)

$\left.\begin{array}{l} H_0 : m_1 = m_2 = \cdots m_k \\ H_1 : \exists i, j \quad i \neq j \;/\; m_i \neq m_j \end{array}\right\}$	$F_{ratio} = \dfrac{MS_{factor}}{MS_{resid}} \approx F_{df_F, df_{resid}}$	$\mathbf{MS_{Factor}}$ = Mean square associated to the effect of one factor or interaction $\mathbf{MS_{resid}}$ = mean square of residuals

$If \;\; F_{ratio} < F^{\alpha}_{df_F, df_{res}} \Rightarrow Accept \;\; H_0$ $If \;\; F_{ratio} > F^{\alpha}_{df_F, df_{res}} \Rightarrow Re\,ject \;\; H_0$	If ***p-value*** $> \alpha \Rightarrow$ Accept H_0 If ***p-value*** $< \alpha \Rightarrow$ Reject H_0	$\alpha = P\left(F_{df_F, df_{res}} > F^{\alpha}_{df_F, df_{res}}\right)$ $p - value = P\left(F_{df_F, df_{res}} > F_{ratio}\right)$

Introduction to Multiple Linear Regression models

Nomenclature

Model $E(Y/X_1{=}x_{1t},\ldots,X_I{=}x_{It}) = \beta_0 + \beta_1 x_{1t} + \ldots + \beta_I x_{It}$	**$E(Y/X_1{=}x_{1t},\ldots,X_I{=}x_{It})$** is the average of the conditional distribution of Y when $X_1{=}x_{1t} \ldots X_I{=}x_{It}$

N = Total number of observations	$\boldsymbol{\beta_i}$ = model parameters $\mathbf{b_i}$ = estimated model parameters **I** = number of explicative variables	$\mathbf{SS_{total} = SS_{expl} + SS_{resid}}$
$\mathbf{SS_{total}}$ = total Sum of Squares $\mathbf{SS_{expl}}$ = explained SS $\mathbf{SS_{resid}}$ = residual SS	$\mathbf{df_{tot}}$ = total degrees of freedom → (N-1) $\mathbf{df_{exp}}$ = d.f. associated to SS_{expl} → I $\mathbf{df_{res}}$ = residual degrees of fr. → (N-1) - I	$MS_{expl} = SS_{expl} / df_{exp}$ $MS_{resid} = SS_{resid} / df_{res}$

Global significance test of the model (ANOVA)

$H_0: \beta_1 = \beta_2 = \ldots = \beta_I = 0$ H_1: at least one $\beta_i \neq 0$ $if\ \dfrac{MS_{expl}}{MS_{resid}} > F^{\alpha}_{I,N-1-I} \Rightarrow reject\ H_0\ (p-value < \alpha)$	**Coefficient of determination:** $R^2 = \dfrac{SS_{expl}}{SS_{total}} \cdot 100$	**Residual variance:** $\sigma^2_{resid} = MS_{resid}$

Significance test for the effect of one X_i variable (*t* test)

$\left.\begin{array}{l} H_0: \beta_i = 0 \\ H_1: \beta_i \neq 0 \end{array}\right\}$	$t_{statistic} = \dfrac{b_i}{s_{b_i}} \approx t_{N-1-I}$	$if\ \lvert b_i/s_{b_i}\rvert < t^{\alpha/2}_{N-1-I} \Rightarrow accept\ H_0$ $if\ \lvert b_i/s_{b_i}\rvert > t^{\alpha/2}_{N-1-I} \Rightarrow reject\ H_0$	If ***p-value*** $> \alpha \Rightarrow$ Accept H_0 If ***p-value*** $< \alpha \Rightarrow$ Reject H_0

Predictions

$$(Y/X_1{=}x_{1t},\ldots,X_I{=}x_{It}) \sim \text{Normal}\,[\, m = E(Y/X_1{=}x_{1t},\ldots,X_I{=}x_{It})\ ;\ \sigma = \sigma_{res}\,]$$

Simple Linear Regression models

Model $E(Y/X{=}x_t) = \beta_0 + \beta_1 \cdot x_t$	**$E(Y/X{=}x_t)$** is the average of the conditional distribution of Y when $X{=}x_t$

Estimated regression line:	*Slope*	*Intercept*
$Y = a + b \cdot X$	$b = r_{xy} \cdot \dfrac{s_y}{s_x} = \dfrac{cov}{s_x^2}$	$a = \overline{Y} - b \cdot \overline{X}$

Coefficient of determination: $R^2 = (r_{xy})^2 \cdot 100 = \left(\dfrac{cov_{xy}}{s_x \cdot s_y}\right)^2 \cdot 100$	**Residual variance:** $S^2_{residual} = S^2_y \cdot (1 - r^2_{xy})$	**Residual:** $e_i = y_i - (a + b \cdot x_i)$