Università degli Studi di Verona
A.A. 2017-2018

APPUNTI DI PROBABILITÀ E STATISTICA

Creato da: Davide Zampieri

PROCEDIMENTI PER RISOLVERE GLI ESERCIZI (TEORIA E LABORATORIO)

ESERCIZIO 1 (medie, mediana, moda, varianza, simmetria):

$M(x) = \frac{\sum x \cdot f}{\sum f}$	$M_2(x) = \sqrt{\frac{\sum x^2 \cdot f}{\sum f}}$	$M_a(x) = \frac{\sum f}{\sum \frac{f}{x}}$	$\ln M_g(x) = \frac{1}{\sum f} \cdot \sum \ln x \cdot f$	
$me = x_{50\%}$ posiz. centrale	$moda = x_{f_{MAX}}$ freq. maggiore	$M_a(x) = IMP$ se ci sono $x = 0$	$M_g(x) = e^{\ln M_g(x)}$	$M_g(x) = 0$ se ci sono $x = 0$
$V(x) = M(x^2) - M(x)^2 = \frac{\sum x^2 \cdot f}{\sum f} - \left(\frac{\sum x \cdot f}{\sum f}\right)^2$ media dei quadrati – quadrato della media			$Sk = \frac{M(x) - moda}{\sigma}$ $\sigma = \sqrt{V(x)}$	Sk < 0: asimm. a sx Sk = 0: simmetria Sk > 0: asimm. a dx
summary(nome_base_dati)	length(nome_base_dati)		boxplot(nome_base_dati)	

ESERCIZIO 2 (tabelle a doppia entrata, regressione lineare, calcolo combinatorio):

Calcolare i totali: f_R: tot. di riga f_C: tot. di colonna f_{TOT}: tot. tabella	Frequenze teoriche: $f^* = \frac{f_R \cdot f_C}{f_{TOT}}$	Chi-quadrato calcolato: $X_C^2 = \sum \frac{(f - f^*)^2}{f^*}$	$X_T^2 = \text{vedi tavola}$ $\alpha = 1\% \text{ o } 5\%$ $gdl = (R-1)(C-1)$	$X_C^2 < X_T^2$: indip. $X_C^2 > X_T^2$: conn.
$b = \frac{cov(x;y)}{V(x)}$ $a = M(y) - b \cdot M(x)$	$cov(x;y) = M(xy) - M(x) \cdot M(y)$ media dei prodotti – prodotto medie		$r = \frac{cov(x;y)}{\sigma_x \cdot \sigma_y}$ r = -1: rel. inversa	r = 0: indipend. r^2 = 0: pessimo r = 1: rel. diretta r^2 = 1: perfetto
$D_{n,s} = \frac{n!}{(n-s)!}$ $D_{n,s}^* = n^s$	$P_n = n!$ $P_n^{*(n_1,\dots,n_k)} = \frac{n!}{n_1! \cdot n_k!}$	$C_{n,s} = \binom{n}{s} = \frac{n!}{s! \cdot (n-s)!}$	$C_{n,s}^* = \binom{n+s-1}{s}$	$P(E) = \frac{\text{casi favorevoli}}{\text{casi possibili}}$
conta l'ordine e elementi diversi	conta solo l'ordine	contano solo gli elementi diversi	n: tot. elementi s: el. per gruppo n_k: n. ripetizioni	0 <= P(E) <= 1 P(E) = 0: imp. P(E) = 1: certo

ESERCIZIO 3 (VC discrete, VC continue, verifica di ipotesi):

VC binomiale (reinserimento): $P(x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$ $M(x) = n \cdot p\,;\ V(x) = n \cdot p \cdot q$	0 <= x <= n q = 1 – p Controllo: $\sum P(x) = 1$	k=c(0:n) dbinom(k, n, p) barplot(dbinom, names.arg=k, xlab="X", ylab="P(X)")
VC di Poisson (eventi rari): $P(x) = \frac{e^{-m} \cdot m^x}{x!}$ $M(x) = m\,;\ V(x) = m$	0 <= x <= ∞ P(x<=3) = P(0)+P(1)+P(2)+P(3) P(x>=4) = 1 – P(x<=3)	k=c(0:valore_richiesto) dpois(k, m)

Standardizzazione: $u = \frac{x - \mu}{\sigma}$ nuova VC normale con $\mu = 0$ e $\sigma^2 = 1$	P(u<0)=P(u>0)=0,5 / P(u>3.29)=0 P(-a<u<0)=P(0<u<a)=tavola P(u<-a)=P(u>a)=0,5–P(0<u<a) P(u<a)=0,5+P(0<u<a) P(a<u<b)=P(0<u<b)–P(0<u<a)	x=seq(0, dato, by = 0.01) normale=dnorm(x, μ, σ) plot(x, normale, type = "l", xlab="x", ylab = "densità …") pnorm(a, μ, σ, lower.tail=FALSE)
i-esimo percentile (x): i<50: cerco u → 50-i e lo nego i=50: mediana → μ dato i>50: cerco u → i-50 sostituisco nella formula della standardizzazione	Verifica di ipotesi sulla media: $\mu_C = \frac{m - \mu_0}{\sigma/\sqrt{n}}$ $\mu_T = \pm$ tavola μ → 50 – $^{\alpha}/_{2}$ μ_C esterno $\pm \mu_T$: rifiuto $H_0: \mu = \mu_0$	dati=c(…, …, …) *# > media:* t.test(dati, mu=media, conf.level=liv_conf, alternative="greater") *# = media:* alt…="two.sided"

ALTRI ARGOMENTI DI LABORATORIO

Simmetria, curtosi:

```
# INDICE DI SIMMETRIA γ (gamma) DI FISHER:
gamma = function(x) {
  m3 = mean((x-mean(x))^3)
  skew = m3/(sd(x)^3)
  skew
  }
gamma(dati)
```

γ < 0: asimm. negativa
γ = 0: simmetria
γ > 0: asimm. positiva

```
# INDICE DI CURTOSI β (beta) DI PEARSON:
beta = function(x) {
  m4 = mean((x-mean(x))^4)
  curt = m4/(sd(x)^4)
  curt
 }
beta(dati)
```

β < 3: platicurtica
β = 3: mesocurtica
β > 3: leptocurtica

```
# INDICE DI CURTOSI γ2 (gamma2) DI FISHER:
gamma2 = function(x) {
  m4 = mean((x-mean(x))^4)
  curt = m4/(sd(x)^4)
  curt - 3
 }
gamma2(dati)
```

$\gamma 2$ < 0: platicurtica
$\gamma 2$ = 0: mesocurtica
$\gamma 2$ > 0: leptocurtica

Chi-quadrato, V di Cramer:

```
matrice=matrix(c(dati_per_riga), nrow=numero_righe, byrow=TRUE)
nomi_righe=c("…","…",…)
nomi_colonne=c("…","…",…)
dimnames(matrice)=list(nomi_righe, nomi_colonne)
matrice
```

```
# CALCOLO DEL CHI-QUADRATO:
testchiq=chisq.test(matrice)
testchiq
chiquadrato=testchiq$statistic
chiquadrato
# CALCOLO DEL "V" DI CRAMER:
N=sum(matrice)
V=sqrt( chiquadrato / ( N*(min(R,C)-1) ) )
```

Regressione lineare:

```
x=c(dati_x)
y=c(dati_y)
# GRAFICO:
plot(x,y)
# RETTA DI REGRESSIONE (ricordarsi che va prima y e dopo x):
retta=lm(y ~ x)
abline(retta, col="blue")
segments(x, fitted(retta), x, y, lty=2)
title(main="Regressione lineare fra y e x")
# PARAMETRI a E b:
summary(retta)
# ANALISI DEI RESIDUI:
plot(fitted(retta), residuals(retta))
abline(0,0)
# COEFFICIENTE DI CORRELAZIONE LINEARE:
R=cor(x,y)
# COEFFICIENTE DI DETERMINAZIONE:
R2=R^2
```

Calcolo delle probabilità su VC discrete:

```
# CALCOLO LA PROBABILITÀ DI n=1:
dbinom(1, n, p)
# CALCOLO LA PROBABILITÀ DI n COMPRESO FRA 0 E 10:
pbinom(10, n, p)
# CALCOLO LA PROBABILITÀ DI n MAGGIORE DI 10 (DA 11 A n):
1-pbinom(10, n, p)
pbinom(10, n, p, lower.tail=FALSE)
# CALCOLO LA PROBABILITÀ DI n COMPRESO FRA 7 E 12:
pbinom(12, n, p) - pbinom(6, n, p)
# CALCOLO IL VALORE DI k CORRISPONDENTE AD UNA CERTA PROBABILITA' (ES. 0,5):
qbinom(0.5, n, p)
```

```
# CALCOLO I VALORI DELLA VARIABILE DI POISSON:
k=c(0:10)
dpois(k,2)
barplot(dpois(k,2), names.arg = k)
# CALCOLO LA PROBABILITÀ DI n COMPRESO FRA 0 E 3:
ppois(3,2)
# CALCOLO LA PROBABILITÀ DI n COMPRESO FRA 4 E 10:
1 - ppois(3,2)
ppois(3,2,lower.tail = FALSE)
# CALCOLO IL VALORE DI k CORRISPONDENTE AD UNA CERTA PROBABILITA' (ES. 0,75):
qpois(0.75,2)
```

Calcolo dei percentili su VC continue:

```
# VALORE CHE INCLUDE IL n%:
qnorm(%_desiderata, media, sd)
# LA MEDIANA E':
qnorm(0.5, media, sd)
# IL PRIMO QUARTILE CORRISPONDE AL 25% DELLA DISTRIBUZIONE:
qnorm(0.25, media, sd)
# IL TERZO QUARTILE CORRISPONDEAL 75% DELLA DISTRIBUZIONE:
qnorm(0.75, media, sd)
```

Verifica di ipotesi:

```
# EFFETTUO IL TEST UNILATERALE PER VERIFICARE L'IPOTESI H0: mu=media, H1: mu<media
t.test(dati, mu=media, conf.level=liv_conf, alternative="less")
# CONFRONTO FRA DUE MEDIE CON VARIANZE UGUALI H0: M(x) = M(y), H1: M(x) != M(y)
t.test(x, y, var.equal=TRUE, conf.level=liv_conf)
# CONFRONTO FRA DUE MEDIE CON VARIANZE DIVERSE H0: M(x) = M(y), H1: M(x) != M(y)
t.test(x, y, var.equal=FALSE, conf.level=liv_conf)
# VERIFICA DI IPOTESI PER DATI APPAIATI H0: diff(prima-dopo)<=0, H1: diff(prima-dopo)>0
t.test(prima, dopo, alternative = "greater", paired = TRUE, conf.level=liv_conf)
```