

2. Modelización de sistemas ambientales (2): Elementos de modelización y simulación.

Modelización de Sistemas Ambientales (07M1)
Dpto. de Ecología e Hidrología y Dpto. Geografía
Facultad de Biología
Universidad de Murcia

Índice

Curso 2006–07

1. Introducción	1
2. Estadística y azar (1:30 horas)	1
2.1. Tirando una moneda	1
2.2. Eligiendo un rumbo	2
3. Construyendo una tabla de la normal (2:00 horas)	3
3.1. Integración y cálculo numérico	3
3.2. Simulación de valores normales	4
4. Para entregar	5

1. Introducción

En esta segunda sesión trabajaremos aspectos relacionados con la estadística, la obtención de valores aleatorios, o pseudoaleatorios, y el cálculo numérico.

Una de las bases y el interés de la simulación estriban en la obtención de resultados de los modelos modificando las condiciones para la obtención: variación de los valores de los parámetros, de las condiciones iniciales, ...

Gracias a esta repetición puede describirse cuales son los resultados más esperados e incluso que frecuencia, disponiendo entonces de una estima de la probabilidad.

Durante esta sesión trabajaremos con la simulación de variables aleatorias discretas y continuas. También elaboraremos una tabla de la distribución normal por diversos métodos con el fin de contrastar los resultados.

2. Estadística y azar (1:30 horas)

2.1. Tirando una moneda

EJERCICIOS

Puede utilizarse la función `sample()` para elegir al azar los elementos de un conjunto dado: `sample(1:10)` lo que resulta equivalente a un sorteo. Esta función puede sustituir a la clásica urna de los problemas de estadística.

1. Comprobar que la expresión anterior produce resultados distintos cada vez que es evaluada por R. ¿Qué resultados proporciona `sample(1:10, 2)`?
2. De los resultados obtenidos anteriormente vemos que puede simularse el lanzamiento de una moneda mediante la expresión `sample(0:1, 1)`. Cada vez que se utilice la expresión devolverá 0 o 1 de forma aleatoria y equiprobable. ¿Cómo puede comprobarse esta afirmación?

Para modificar la probabilidad de aparición de cada elemento puede utilizarse, por ejemplo, un vector que indica el peso o probabilidad de cada elemento:

```
sample(0:1, 1, prob=c(3, 1)) o sample(0:1, 1, prob=c(0.75, 0.25)).
```

Repetir el proceso ideado anteriormente para comprobar que los resultados siguen la ley de frecuencias determinada.

3. Si se desea realizar un número repetido de lanzamientos es necesario considerar que la función considere este reemplazamiento: Para obtener diez elementos en el muestro: ¿Qué ocurre si utilizamos la expresión: `sample(0:1, 10)`? ¿Y si indicamos que deseamos el reemplazamiento: `sample(0:1, 10, replace=T)`?

4. La función `table()` proporciona una tabla de frecuencias a partir de un vector (opcionalmente más de uno). Por ejemplo: `table(rep(15:20, 3))` ¿Cómo se expresan los resultados?

Utilizando la función `table()` y la opción `replace=T` para `sample()` resulta muy fácil comprobar que R produce los resultados esperados como se había planteado anteriormente. Así considerando las siguientes expresiones:

- a) `table(sample(0:1, 1000, replace=T))`
- b) `table(sample(0:1, 1000, replace=T, prob=c(0.75, 0.25)))`

¿Cuáles son los valores teóricos en cada caso? ¿Qué diferencias hay entre los valores teórico y los obtenidos?

5. Recordando los test de hipótesis clásicos: ¿Hay alguna prueba estadística que permita comprobar si las diferencias son "importantes"?
6. Las variables discretas pueden "asimilarse" a un muestreo cualitativo. Sin embargo, la ley de aparición de los resultados es "fija": así tenemos variables de Poisson, binomiales, ... Para describir la proporción de hembras en una camada puede utilizarse una distribución de binomial positiva. En el caso del número de descendientes por camada puede ser una distribución de Poisson.
7. Para conocer la proporción camadas con 0, 1, 2, ..., 5 hembras, sabiendo que la probabilidad del nacimiento de una hembra es de 0.5 puede obtenerse mediante la función `dbinom()`, para el caso concreto de una binomial ($n=5, p=0.5$).

Calcular las proporciones mediante `dbinom(0:5, 5, 0.5)` ¿Cuál es la probabilidad de obtener camadas sin hembras? ¿Y con sólo 2 hembras?

8. Para obtener un muestreo de una población binomial se utiliza la función `rbinom()`. Así, la expresión `rbinom(10000, 5, 0.5)` proporcionará 1000 valores procedentes de una distribución binomial con $p=0.5$ y $n=5$. ¿Coinciden los valores teóricos obtenidos en el ejercicio anterior con los muestrales? Para comparar los resultados divídase lo obtenido en la tabulación por 10000. La función `round()`, puede ayudar a expresar los resultados con más claridad.

2.2. Eligiendo un rumbo

A diferencia de variables cualitativas o discretas, como es el caso anterior, puede ser necesario recurrir a variables continuas, por ejemplo, elegir un valor entre 0 y 1. En este caso la función `runif()` puede proporcionar el resultado deseado.

¿Qué valores se obtienen con `runif(1000)`? ¿Qué ocurre al utilizar `hist(runif(1000))`? ¿Qué ocurre al elevar el número de valores solicitado: 10000, 100000 y 1000000? ¿Qué efectos tiene sobre el gráfico: hay alguna región entre 0 y 1 más probable que otra? ¿Qué efectos tienen sobre el tiempo de respuesta de R los distintos tamaños muestrales?

1. Para representar los valores en un gráfico puede recurrirse a la expresión: `plot(runif(100))` ¿Qué representa el gráfico obtenido? ¿Qué representamos en el eje de abscisas? ¿Y en el de ordenadas?

El resultado anterior "puede" mejorarse mediante `plot(runif(100), type="l")` ¿Qué diferencia hay con el gráfico precedente?

2. Dada una muestra de valores obtenida mediante la función `runif()`: ¿Qué media de los valores cabría esperar? ¿Cuanto valdrían sus cuartiles y mediana? ¿Por qué?
3. Comprobar con la expresión `summary(runif(1000))` que se cumple lo pronosticado en el ejercicio anterior. ¿Qué ocurre al considerar de un tamaño muestral mayor?

3. Construyendo una tabla de la normal (2:00 horas)

Como ya se ha dicho, una de las utilidades de la simulación es obtener una tabla de valores esperados y su frecuencia teórica. A continuación, utilizaremos una función bien conocida para explorar distintas técnicas de simulación y cálculo, para construir una tabla de probabilidad de los valores de una variable la normal, $N(0, 1)$, para valores iguales o menores que uno dado: z , que expresa las desviaciones típicas que nos alejamos de la media.

Tal como aparece en la bibliografía clásica debemos obtener los valores de probabilidad que corresponden a $z = 0, 0.01, 0.02, \dots, 3.9$.

Estos valores se encuentran tabulados tal como se adjunta en la tabla. Utilizando una tabla publicada en un manual de estadística ¿En qué falla la tabla adjunta? ¿qué ocurre si se cambia el número de decimales considerado?

3.1. Integración y cálculo numérico

EJERCICIOS

1. Sabiendo que la función de densidad de la normal es:

$$f(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{2\sigma^2}}$$

y que el área bajo esta curva permite calcular la probabilidad para un intervalo sin más que realizar la integral definida:

$$P(-\infty < z < a) = \int_0^a f(z)dz$$

De forma aproximada puede realizarse la integral utilizando la regla del trapecio: dividiendo el intervalo en varios subintervalos regulares de tamaño ϵ (siendo este pequeño), y aproximando el área a un trapecio definido por la base (ϵ), los dos lados ($f(z)$ y $f(x + \epsilon)$). Sumando el área aportada por todos los subintervalos obtendremos el valor de probabilidad deseado.

Esquematizar el procedimiento descrito.

2. Escribir la fórmula abreviada (usando $f(z)$ en la expresión) para calcular el área dado el intervalo y el límite inferior del intervalo.
3. La función `pnorm(z, mu, sigma)` proporciona la probabilidad de encontrar valores menores que z en una distribución normal de media μ y desviación típica σ . Por defecto se considera una distribución normal $\mu = 0$ y $\sigma = 1$.
Calcular los probabilidades asociadas a los valores $-1, 0, 0.5, 1.96, 2, 3, 3.8$ y compararlas con las de la tabla. ¿Hay diferencias?
4. La siguiente función permite calcular el área asociada a un trapecio cuyos vértices son $(z, 0)$, $(z + \epsilon, 0)$, $(z, f(z))$ y $(z + \epsilon, f(z + \epsilon))$. ¿Está ese trapecio bajo la curva correspondiente a la función de densidad de una normal siendo ϵ un valor muy pequeño?

```
areanormal<-function (x,i,mu=0,sigma=1){  
  (1/(sqrt(2*pi)*sigma)*exp(-(x - mu)^2/(2*sigma^2)) +  
  1/(sqrt(2*pi)*sigma)*exp(-((x+i) - mu)^2/(2*sigma^2)))/2*i}
```

Calcular `areanormal(0, 0.01)` y `pnorm(0.01) - 0.5` ¿Calculan, teóricamente, el mismo valor? ¿Proporcionan resultados similares? ¿De qué orden es el error?

5. Realizar el gráfico que se indica a continuación:

```
z<-c(0,0.01)  
plot(z,dnorm(z),type="l")  
points(z,dnorm(z),pch=15,cex=1.5)  
zp<-seq(0,0.01,0.001)  
points(zp,dnorm(zp))
```

¿Qué representan los puntos marcados con un cuadrado negro? ¿y la línea que los une? ¿Qué representan los puntos representados por círculos vacíos? ¿Qué supone el área encerrada entre la línea entre los cuadrados negros y la poligonal descrita por los círculos vacíos? ¿Hay algún método rápido para calcular este área?

6. La función `abline()` puede resultar útil para representar líneas o conjuntos de ellas. Podemos generar una maya de densidad conocida mediante:

```
abline(v=seq(0,0.01,0.001))
abline(h=seq(dnorm(0.01),dnorm(0),0.00001))
```

Tenemos, en vertical una maya de densidad 0.001 y en horizontal de 0.00001. ¿Cuál es la superficie de cada celdilla de la maya? ¿Cuántas celdillas, aproximadamente están en el área que no interesa? ¿Cuál es su superficie?

7. Un elemento importante en el cálculo es la amplitud del intervalo. ¿Qué ocurre si el intervalo considerado es más pequeño?

Al utilizar muchos intervalos ajustamos mejor el cálculo al comportamiento de la curva:

```
sum(areanormal(seq(0,0.01,.001),0.001))
```

¿Qué diferencia hay con respecto al valor que nos proporciona la tabla `pnorm(0.01)-0.5`? ¿De qué orden es el error?

La función `cumsum()` proporciona una suma acumulativa de los valores de un vector o serie (por ej: `cumsum(1:5)`). Si utilizamos:

```
cumsum(areanormal(seq(0,1,0.001),0.001))+0.5
```

¿A qué se corresponden los valores? ¿El valor décimo, vigésimo, ... son valores que aparecen en la tabla? Sugerencia: presentar sólo los valores de interés mediante:

```
cumsum(areanormal(seq(0,3.9,0.001),0.001))[seq(1,3900,10)]+0.5
```

¿Ayudaría en la comparación disminuir el número de cifras decimales mediante la función `round()`?

8. Construyamos ahora la tabla mediante, de los valores necesarios con la ayuda de la función `matrix()`

```
i<-0.0001 #intervalo
c(0,cumsum(areanormal(seq(0,4,i),i)))[seq(1,40000,100)]+0.5->N
N<-round(N,4) #redondeo de los valores del vector a 4 cifras decimales
N<-t(matrix(N,10)) #construcción de una tabla de 10 columnas
colnames(N)<-(0:9)/100
rownames(N)<-(0:39)/10
Nint<-N #Tabla resultante por integración
```

¿Es el resultado correcto considerando la tabla adjunta? ¿Cuáles son las principales diferencias? ¿A qué pueden atribuirse?

3.2. Simulación de valores normales

Puede utilizarse la generación de números aleatorios de procedentes de una distribución normal asociada a la función `rnorm()`. También es posible utilizar una función uniforme, `runif()`, para generar 12 valores ($x_1 \dots x$) y a partir de ellos obtener un valor aleatorio normal de una distribución $N(0,1)$:

$$z = \sum_{i=1}^{12} x_i - 6$$

EJERCICIOS

1. Representar el histograma para los valores aleatorios generados y la función de densidad de la normal mediante:

```
i<-seq(-5,5,0.5)
hist(rnorm(10000),i,prob=T)
lines(i,dnorm(i),col=2)
```

¿Se puede mejorar la figura para que la curva sea “más continua”? Sugerencia el intervalo más pequeño.
¿Qué se puede mejorar ahora?

2. Puesto que la distribución normal es simétrica cabe considerar todos los valores simulados como positivos y duplicar el tamaño de la muestra en la región deseada. ¿Qué significa lo siguiente?

```
i<-seq(-5,5,0.1)
hist(abs(rnorm(500000)),i,prob=T)
lines(i,dnorm(i)*2,col=2)
```

¿Qué efectos tiene sobre los resultados el uso de la función `abs()`?

3. Construir la tabla asociada a los valores obtenidos por un muestro de número aleatorios procedentes de una distribución normal mediante la función `rnorm()`

```
table(cut(abs(rnorm(500000)),seq(0,5,0.01)))->N
c(0.5,cumsum(N)/500000/2+0.5)[1:400]->N
N<-t(matrix(round(N,4),10))
colnames(N)<-(0:9)/100
rownames(N)<-(0:39)/10
Nrn<-N
```

¿Qué diferencias hay entre esta tabla y la anterior? ¿Cuál es la mayor de las diferencias? ¿y en terminos relativos?

4. Definiremos una función que automatice la tarea de generación de valores normales a partir de valores procedentes de una distribución uniforme mediante:

```
normale<-function() {(normale<-sum(runif(12))-6)}
```

Como en el caso anterior podemos representar el histograma mediante:

```
i<-seq(-5,5,0.5)
hist(replicate(500000,abs(normale())),i,prob=T)
lines(i,dnorm(i)*2,col=2)
```

La tabulación se consigue mediante:

```
table(cut(replicate(500000,abs(normale())),seq(0,5,0.01)))->N
c(0.5,cumsum(N)/500000/2+0.5)[1:400]->N
N<-t(matrix(round(N,4),10))
colnames(N)<-(0:9)/100
rownames(N)<-(0:39)/10
Nru12<-N
```

5. ¿Cuales han sido las principales diferencia en los resultados de esta tabla con las anteriores? ¿Pueden considerarse igualmente costosas en cálculo? ¿Aporta alguna idea el uso del último procedimiento?

4. Para entregar

Los resultados de la práctica se resumiran en un breve texto (max una cara de folio). En ellos se discutirá las principales diferencias que se producen para obtener una tabla correspondiente a las areas bajo la curva normal $(N(0,1))$ de 0 a 3.99 con los distintos métodos empleados.

Para la elaboración del texto se utilizará el esquema del procedimiento habitual¹. Se asume que los obejivos, planteados en el párrafo anterior, y los métodos emplados son conocidos por el lector y bastará con hablar de los resultados y establecer algunas conclusiones.

Este texto se grabará en formato plano (editable por nano o block de notas) en el la cuenta de deimos, en el directorio del usuario y con el nombre: `s02.txt`.

Sugerencias: Pueden cosiderarse la similitud/diferencias entre los métodos y adecuación por distintas características: tiempos de cálculo, tamaños muestrales, error, ...

Además, se grabará desde R la tabla mencionada anteriormente, y correspondiente a la simulación de valores normales mediante la función '`rnorm()`', utilizando:

¹Introducción, material y métodos, resultados, discusión y conclusiones.

```
write.table(Nrn, "s1.dat")
```

Nota: En caso de realizar la tarea en una máquina personal se utilizará el servicio ftp para dejar los fichero en el servidor.

Tabla

Áreas bajo la curva normal para valores menores o iguales que z :

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Los valores de la tabla pueden obtenerse en R mediante:

```
N<-pnorm(seq(0, 3.99, 0.01)) #Cálculo de la probabilidad para z de 0 a 3.99
N<-t(matrix(round(N, 4), 10)) #Tabulación de los valores, redondeando a diezmilésimas
rownames(N) <- (0:39)/10      #Etiquetado de las filas de la tabla
colnames(N) <- (0:9)/100     #Etiquetado de las columnas de la tabla
```

TABLE P. Areas of the normal curve

Standard deviation units	0	1	2	3	4	5	6	7	8	9	Standard deviation units
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359	0.0
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753	0.1
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141	0.2
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517	0.3
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879	0.4
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224	0.5
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549	0.6
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852	0.7
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133	0.8
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389	0.9
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621	1.0
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830	1.1
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015	1.2
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177	1.3
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319	1.4
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441	1.5
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545	1.6
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633	1.7
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706	1.8
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767	1.9
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817	2.0
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857	2.1
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890	2.2
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916	2.3
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936	2.4
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952	2.5
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964	2.6
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974	2.7
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981	2.8
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986	2.9
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990	3.0
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993	3.1
3.2	.4993	.4994	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995	3.2
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997	3.3
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998	3.4
3.5	.499767										
3.6	.499841										
3.7	.499892										
3.8	.499928										
3.9	.499952										
4.0	.499968										
4.1	.499979										
4.2	.499987										
4.3	.499991										
4.4	.499995										
4.5	.499997										
4.6	.499998										
4.7	.499999										
4.8	.499999										
4.9	.500000										

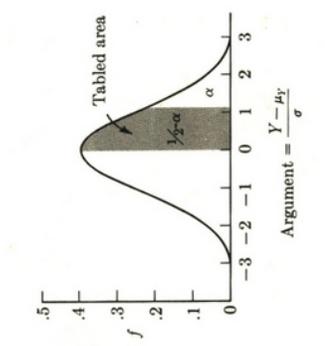


TABLE P. Areas of the normal curve.

Arguments are furnished for the right half of the normal curve from the mean up to 3.5 standard deviations in increments of 0.01 standard deviation units, and from there to 4.9 standard deviations from the mean in increments of 0.1 standard deviation units. The quantity given is the area under the standard normal density function between the mean and the critical point. The area is generally labeled $\frac{1}{2} - \alpha$ (as shown in the figure). The proportion of the area between the mean and the critical value is given to four significant decimal places up to 3.49 standard deviations, and from there to 4.9 standard deviations to six significant decimal places.

The area between the mean and 2.21 standard deviations, for example, is found from the table to comprise 0.4864 or 48.64% of the total area of the curve. From this figure one can easily compute two other quantities. The area beyond 2.21 standard deviations is the 0.5-complement of the function just looked up. Thus, it will be $0.5000 - 0.4864 = 0.0136$, which is α , the proportion of the area of the curve to the right of 2.21 standard deviations. To find $1 - \alpha$, the total area of the curve to the left of 2.21, add 0.5 (the entire area of the left half of the curve) to the area given as a function in the table. Thus, the area to the left of 2.21 standard deviations is 0.9864.

By inverse table look-up and interpolation one can find the number of standard deviations corresponding to a given area. Thus, an area of 0.3264 represents the area between the mean and 0.94 standard deviations.

This table is of wide application in statistics. It is employed whenever making probability statements about normally distributed variables (Section 7.9), when normalizing a frequency distribution (Section 6.5), and in a variety of other applications.

The table was generated by using the polynomial expression given in C. Hastings, *Approximations for Digital Computers* (Princeton University Press, 1955) and shown below, which approximates the integral of the normal probability distribution with a maximum error of 7.5×10^{-8} :

$$P(X) - \frac{1}{2} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} t(b_1 + t(b_2 + t(b_3 + t(b_4 + tb_5))))$$

where $t = 1/(1 + pX)$, $p = 0.2316419$, $b_1 = 0.319381530$, $b_2 = -0.356563782$, $b_3 = 1.781477937$, $b_4 = -1.821255978$, and $b_5 = 1.330274429$.