Teste de Hipóteses para Tomada de Decisão

Introdução

Um dos propósitos da Inferência Estatística é o de testar hipóteses estatísticas.

Uma hipótese estatística é uma afirmação feita sobre algum parâmetro de uma variável de interesse que, aqui, denotamos usualmente por X.

Insper

Exemplo da aula anterior

Comprar ou não comprar um lote de resistores de um determinado fabricante?

- A equipe recebeu uma amostra de resistores do lote de compra. Ou seja, com n=10, cada grupo gerou uma estimativa para a média \overline{x} .
- Deixando claro todos os resultados numéricos e possíveis suposições, descreveu a conclusão (tomada de decisão).

Introdução

Assim, formulada as hipóteses sobre a característica de interesse da população, desejamos saber se os resultados amostrais trazem evidências para refutar (ou não) tal afirmação.

É importante observar que não temos como evitar erros na tomada de decisão, uma vez que tal decisão é feita com base em informações provenientes de amostras. Todavia, a probabilidade de ocorrência do tal erro pode ser controlada ou mensurada. Assim, é possível associar uma medida de validade às conclusões obtidas.

Teste de Hipóteses

O procedimento básico de um teste de hipóteses consiste em supor verdadeira uma das hipóteses em questão e verificar se a amostra observada leva à rejeição ou não desta hipótese, ou seja, verificar se os dados coletados trazem evidências a favor ou não de uma hipótese formulada.

Iniciamos explicitando claramente qual a hipótese que estamos colocando à prova, e a chamamos de hipótese nula (H_n) . No nosso caso:

$$H_0: \mu = \mu_0$$

Em seguida, convém explicitar a hipótese alternativa $(H_A \quad ou \quad H_1)$. O uso de uma dessas hipóteses alternativas depende das informações que o problema traz. A hipótese alternativa será <u>uma</u> das seguintes:

H₁:
$$\mu > \mu_0$$
 H₁: $\mu < \mu_0$ H₁: $\mu = \mu_1$ Insper

Teste de Hipóteses

A teoria do teste de hipóteses fornece métodos para a tomada de decisão a respeito de hipóteses formuladas, informando também a probabilidade de erro que acompanha a decisão.

O erro de decisão não pode ser evitado, mas sua probabilidade de ocorrência pode ser controlada ou mensurada, obtendo-se uma medida de validade das conclusões obtidas.

Insper

Chamamos de α , nível de significância, probabilidade de cometer o erro complementar ao coeficiente de confiança γ , isto é,

 $\alpha = P(rejeitar H_0 | H_0 \text{ é verdadeira})$

Buscar uma regra de decisão que aponte quais resultados amostrais te levam a rejeição de H_0 , ou seja, te levam a concluir pelo descrito na hipótese alternativa H_1 .

A probabilidade α de cometer um erro de primeira espécie é um valor arbitrário e recebe o nome de nível de significância do teste.

O resultado da amostra é cada vez mais significante para rejeitar H_0 quanto menor for esse nível α . Usualmente, esses valores são fixados em 1%, 5% ou 10%.



Esta região é construída de modo que

$$P(\overline{X} \in RC | H_o \text{ é verdadeira})$$

seja igual a α, que é um número fixado.

RC recebe o nome de região crítica ou região de rejeição (convém observar que a construção da RC é SEMPRE feita sob a suposição de H₀ ser verdadeira).

Insper

Passos para Construção de um Teste de Hipóteses

Primeiro Passo: Fixe qual a hipótese nula, H_0 , a ser testada e qual a hipótese alternativa (H_A).

Segundo Passo: Use a teoria estatística e as informações disponíveis para decidir qual estatística será usada para julgar a hipótese H₀. Não se esqueça de levantar as propriedades dessa estatística.

Terceiro Passo: Fixe a probabilidade α de cometer um erro de primeira espécie, e use este valor para construir a região crítica RC. Lembre que esta região é construída para a estatística definida no segundo passo, usando os valores hipotetizados por H_0 .

Quarto Passo: Use as informações fornecidas pela amostra para encontrar o valor da estatística que definirá a decisão.

12 Insber

Quinto Passo: Se o valor da estatística, observado na amostra, pertencer à região crítica, rejeite H_0 ; caso contrário, não rejeite H_0 .

É importante, sempre que fizermos testes de hipóteses, distinguir BEM estas cinco fases.

As latas de certa marca de refrigerante apresentam em seu rótulo o volume de 350 ml. Um órgão regulador admite um desvio padrão, máximo, igual a 10,5 ml em seu conteúdo.

Um órgão que atua na defesa do consumidor desconfia que tal marca está vendendo refrigerantes com conteúdo médio das latas abaixo de 350 ml contrariando o rótulo.

Exemplo 1 (cont.)

Para averiguar a desconfiança do órgão de defesa do consumidor, foi coletada uma amostra de 36 latas do refrigerante em pontos de comercialização e mediu-se o conteúdo destas latas.

O resultado obtido na amostra foi: $\bar{x} = 347 \text{ ml}$

Exemplo 1 (cont.)

```
data = [344.5, 352.9, 355.0, 348.8, 342.0, 344.7, 346.9, 340.9, 357.7, 341.2, 347.1, 365.3, 340.3, 343.2, 342.6, 333.7, 341.9, 351.0, 348.0, 354.3, 344.9, 348.0, 346.4, 358.0, 340.2, 354.4, 349.7, 339.4, 338.8, 344.9, 330.4, 341.7, 336.7, 363.2, 355.9, 357.4]
```

$$\bar{x} = 347 \text{ ml}$$

Escrevendo as hipóteses:

$$H_0: \mu = 350 \,\mathrm{ml}$$

$$H_A: \mu < 350 \,\mathrm{ml}$$

Note que as hipóteses são formuladas em termos do parâmetro da população. Isso ocorre porque há interesse em avaliar todo o processo de enchimento das latas, isto é, a população de todas as latas de refrigerante da marca avaliada.

Em palavras (no exemplo):

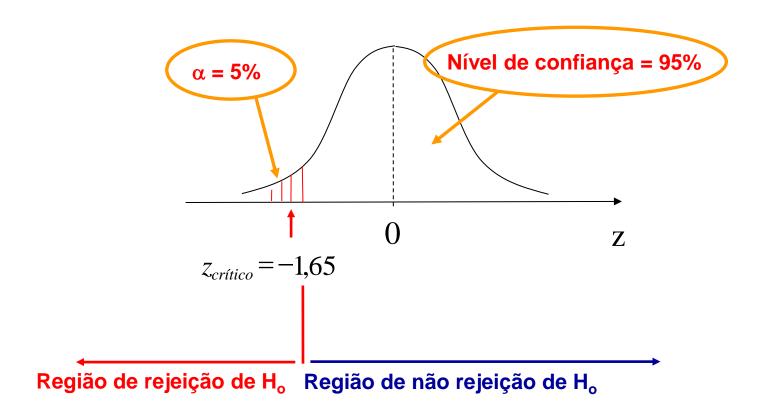
✓ Rejeitar a Hipótese Nula:

Os resultados estatísticos indicam que desconfiança do órgão de defesa do consumidor procede (ou seja, que as latas estão abaixo do conteúdo indicado no rótulo).

✓ Não Rejeitar a Hipótese Nula:

Não existe nenhuma evidência estatística que comprove que a desconfiança do órgão de defesa do consumidor procede (ou seja, que as latas estão abaixo do conteúdo indicado no rótulo).

Com base nas hipóteses, rejeita-se a hipótese nula para valores pequenos, ou seja, se temos:



Região crítica: se o valor da estatística z_{obs} for menor que -1,65, então rejeita-se a hipótese nula (ou seja, rejeita-se a hipótese de que o produto está de acordo com as especificações do fabricante).

Estatística do Teste (obtida da amostra)

Padronização dos dados amostrais sob a hipótese nula (H_0), ou seja, μ_0 = 350.

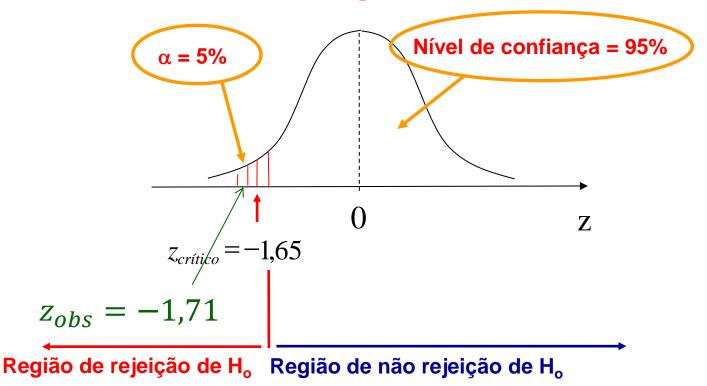
$$z_{obs} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

$$\bar{x} = 347 \, ml$$

$$\sigma = 10,5 \, ml$$

$$z_{obs} = \frac{347 - 350}{10,5 / \sqrt{36}} = -1,71$$

Insper



Conclusão: Rejeitamos a hipótese nula, isto é, existem evidências estatísticas de que desconfiança do órgão de defesa do consumidor procede (conteúdo das latas está abaixo das especificações do fabricante, ao nível de significância de 5% (ou com 95% de confiança).

Teste de hipóteses para uma média populacional

Caso 1

(variância populacional conhecida)

Vamos considerar a seguinte hipótese nula:

$$H_0: \mu = \mu_0$$

Vimos que um estimador com boas propriedades para o parâmetro μ é \overline{X} .

Caso a variável em estudo na população de interesse tenha média μ e variância σ^2 ,

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

se a variável em estudo tiver distribuição normal ou se o tamanho da amostra for suficientemente grande para usarmos o TLC.

Ainda, nesse caso,

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

Porém, a quantidade anterior não pode ser usada como estatística de teste pois μ é um parâmetro desconhecido. Mas, sob H_0

$$z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0, 1)$$

pode ser considerada uma estatística de teste.

Insper

Assim, sob a hipótese nula, podemos facilmente calcular o valor da estatística z. Para concluir o teste, precisamos construir a RC e verificar se o valor de z pertence ou não à RC, lembrando que esta região é construída de modo que

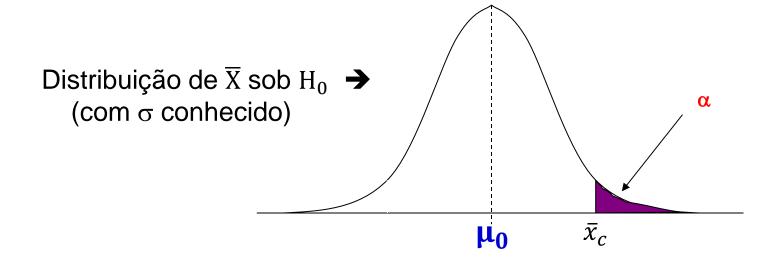
$$P(z \in RC \mid H_0 \text{ \'e verdadeira}) = \alpha.$$

A localização da RC depende das hipóteses formuladas (olhar para o sentido de H_△):

Unilateral ou unicaudal à direita

$$H_0: \mu = \mu_0 \text{ (ou } \mu \leq \mu_0)$$

$$H_A: \mu > \mu_0$$

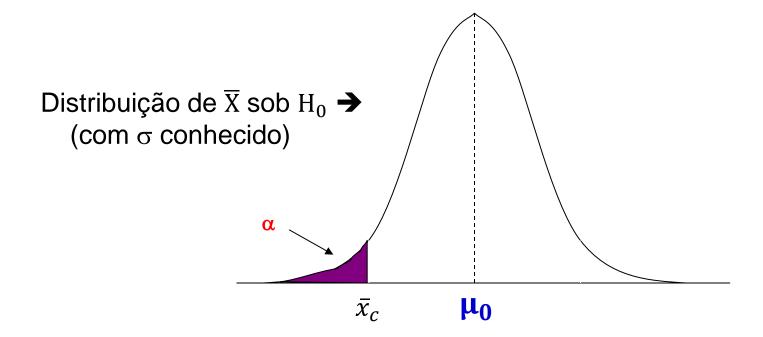


Região crítica: Rejeita-se H₀ para qualquer média amostral $\bar{x}_{obs} > \bar{x}_c$

Unilateral ou unicaudal à esquerda

$$H_0: \mu = \mu_0 \text{ (ou } \mu \geq \mu_0)$$

$$H_A: \mu < \mu_0$$



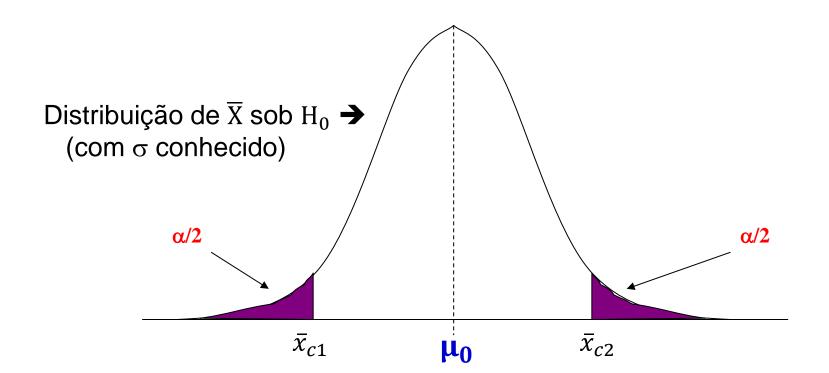
Região crítica: Rejeita-se H_0 para qualquer média amostral $\bar{x}_{obs} < \bar{x}_c$

Teste bilateral ou bicaudal

$$H_0: \mu = \mu_0$$

 $H_A: \mu \neq \mu_0$

$$H_A: \mu \neq \mu_0$$



Região crítica: Rejeita-se H_0 para qualquer média amostral $\bar{x}_{obs} < \bar{x}_{c1}$ ou $\bar{x}_{obs} > \bar{x}_{c2}$ Insper

Insper

Tipos de Erros

Erro Tipo I: rejeitar H0 quando H0 é verdadeira

Erro Tipo II: não rejeitar H0 quando H0 é falsa

Exemplo 1 (erros tipos I e II)

Qual a probabilidade de concluir que desconfiança do órgão de defesa do consumidor não procede se verdadeiro conteúdo médio for igual a 345 ml?

Resolva considerando nível de significância de 5%.



Erros do tipo I e II

Qualquer que seja a decisão tomada, estamos sujeitos a cometer erros.

Para facilitar a linguagem, necessitamos das seguintes definições:

Erro do tipo I – rejeitar H₀ quando esta é verdadeira.

Erro do tipo II – não rejeitar H₀ quando esta é falsa.

Voltando ao Exemplo 1

ERRO TIPO I: erro cometido ao concluir que desconfiança procede, quando na verdade desconfiança não procede.

ERRO TIPO II: erro cometido ao concluir que desconfiança não procede, quando na verdade procede.



Chamamos de α a probabilidade de cometermos o erro do tipo I, isto é,

```
\alpha = P(erro do tipo I) =
= P(rejeitar H<sub>0</sub> | H<sub>0</sub> é verdadeira) =
= P(\overline{X} \in RC \mid H_0 é verdadeira)
```

Chamamos de β a probabilidade de cometermos o erro do tipo II, isto é,

```
β = P(erro do tipo II) =
= P(não rejeitar H<sub>0</sub> | H<sub>0</sub> é falsa) =
= P(\overline{X} ∉ RC | H<sub>0</sub> é falsa)
```

Insper

Teste de Hipóteses para Tomada de Decisão

VALOR-P

Introdução

Nos slides anteriores, vimos como testar uma hipótese a respeito de uma média quando a variância populacional é conhecida.

Para a condução do teste de hipótese seguindo os passos discutidos, necessitamos fixar o nível de significância α para assim construirmos a região crítica.

Uma das críticas a esse procedimento é que o nível de significância é arbitrário e valores diferentes de a podem levar a conclusões diferentes.

Aqui, veremos uma forma de construção de um teste de hipóteses que não exige a fixação a priori do nível de significância.

Supondo que a hipótese nula seja verdadeira, qual deve ser a probabilidade de observar um valor tão ou mais desfavorável à hipótese nula quanto a particular média amostral estimada?

Essa probabilidade pode ser útil na sua tomada de decisão? Em caso afirmativo, descreva como isso é possível.



Valor-p é o <u>menor nível de</u> <u>significância</u> que leva <u>à rejeição de</u> <u>H₀</u> com base na amostra.

Caso Unilateral a Esquerda:

Vamos considerar, por exemplo, o problema de testar H_0 : $\mu = \mu_0$

 H_1 : $\mu < \mu_0$

sendo μ_0 um valor conhecido e μ um valor desconhecido associado à média da variável de interesse, ou seja, E(X).

É razoável rejeitar H_0 se a média amostral observada (\bar{x}_{obs}) for muito menor que μ_0 .

Insper

Caso Unilateral a Esquerda:

Definição

Supor que H_0 é verdadeira (isto é, $\mu = \mu_0$). Qual é a probabilidade de se ter uma evidência tão (ou mais) desfavorável à hipótese nula do que a encontrada na amostra (valor-p)?

Em outras palavras, quanto vale a seguinte probabilidade:

$$P(\bar{X} \le \bar{x}_{obs} | \mu = \mu_0) = ?$$

Essa probabilidade é denominada valor-p (valor de probabilidade) ou *p-value* (probability of value).

Se o resultado do valor-p for "alto", significa que não é difícil ocorrer um valor tão desfavorável quanto a média amostral observada quando $\mu = \mu_0$. Logo, é razoável neste caso não rejeitar H_0 .

Por outro lado, se valor-p for um valor "pequeno", significa que é pouco provável ocorrer um valor tão desfavorável quanto a média amostral observada quando $\mu = \mu_0$. Logo, é razoável neste caso rejeitar H_0 .

Resultado Geral: Se for estabelecido como sendo α a probabilidade máxima de erro ao rejeitar H_0 quando H_0 é verdadeira, então, com base no valor-p, deve-se rejeitar H_0 se valor-p < α .

Caso Unilateral a Direita:

$$H_0: \mu = \mu_0$$

$$H_1: \mu > \mu_0$$

É razoável rejeitar H_0 se a média amostral observada (\bar{x}_{obs}) for muito maior que μ_0 .

Define-se o valor-p em um teste unilateral a direita como sendo

valor-p=
$$P(\bar{X} \ge \bar{x}_{obs} | \mu = \mu_0)$$

Caso Bilateral:

$$H_0$$
: $\mu = \mu_0$

$$H_1$$
: $\mu \neq \mu_0$

É razoável rejeitar H_0 se a média amostral observada (\bar{x}_{obs}) for muito maior ou muito menor que μ_0 .

Como calcular o valor-p neste caso?

Caso Bilateral: se $\overline{x}_{obs} > \mu_0$

Verifique se a média amostral observada é superior ou inferior a μ_0 .

Vamos supor que a média amostral observada (\bar{x}_{obs}) tenha sido superior a μ_0 , então, calculamos

$$p' = P(\bar{X} \ge \bar{x}_{obs} | \mu = \mu_0)$$

Define-se o valor-p em um teste bilateral como sendo

valor-p=
$$2p'$$

Caso Bilateral: se $\overline{x}_{obs} < \mu_0$

Verifique se a média amostral observada é superior ou inferior a μ_0 .

Vamos supor que a média amostral observada (\bar{x}_{obs}) tenha sido inferior a μ_0 , então, calculamos

$$p' = P(\bar{X} \le \bar{x}_{obs} | \mu = \mu_0)$$

Define-se o valor-p em um teste bilateral como sendo

valor-p=
$$2p'$$

Uma empresa de serviços na área de soluções de pagamentos eletrônicos fornece máquinas a estabelecimentos comerciais para processarem o pagamento por cartão de crédito efetuado por seus clientes. Entretanto, os donos desses estabelecimentos comerciais reclamaram que o tempo de processamento das transações de cartão de crédito nessas máquinas estava, em média, muito alto.

- Para evitar a perda de clientes, a empresa estabeleceu como meta reduzir o tempo médio de processamento das transações de cartão de crédito para menos do que 4 segundos até o fim de março de 2014 (assuma σ conhecido e igual a 1 segundo).
- Após a realização de diversas modificações, a empresa gostaria de checar se a meta foi atingida, ao nível de significância de 5%.
- Para isso, ela mediu o tempo de processamento de uma amostra aleatória de 49 transações, cuja média amostral foi igual a 3,6 s_{Insber}

Um estudo foi desenvolvido para avaliar o salário de empregadas domésticas na cidade de São Paulo. Foram sorteadas e entrevistadas 121 trabalhadoras. Admita que o desvio padrão dessa variável na cidade é de 0,88 salários mínimos.

- a) Você conhece a distribuição do estimador \overline{X} ? Se não, é possível fazer alguma suposição?
- b) Desconfia-se que a média salarial seja superior a 2 salários mínimos. Baseado nessa informação, formule as hipóteses e interprete os erros do tipo I e do tipo II.
- c) Para um nível de significância de 2%, construa a RC.
- d) Se a amostra forneceu média de 2,38 salários mínimos, qual a conclusão deste teste?
- e) Que suposições você fez para resolver os itens anteriores?

Insper

A duração de pilhas falsificadas segue uma distribuição normal com média de 15 ut e variância 16 ut². Pilhas autênticas têm uma duração média maior e mesmo desvio-padrão. Um lote de pilhas apreendido será leiloado e para definir seu preço é preciso decidir se são produtos falsificados ou não. O leiloeiro adotou como regra de decisão que se a duração média de uma amostra aleatória de 16 pilhas for maior que 18 ut, o lote será considerado autêntico. Critique essa regra.

Um estudo foi desenvolvido para avaliar a renda de pedreiros autônomos na cidade de São Paulo, o qual desconfia que a renda média seja superior a 3 salários mínimos. Admita que o desvio padrão dessa variável na cidade é de 0,7 salários mínimos.

Foram sorteados e entrevistados 100 trabalhadores, cuja amostra forneceu média de 3,2 salários mínimos.

- a) Baseado nessa informação, formule as hipóteses H_0 e H_1 e interprete o erro do tipo I e erro do tipo II, sendo (erro I = rejeitar H_0 | H_0 verdade) e (erro II = não rejeitar H_0 | H_0 falsa).
- b) Conclua com base na RC com 1% de significância. $RC = \{\bar{x}_{obs} > 3,1631\}$
- c) Supondo que a hipótese nula seja verdadeira, qual deve ser a probabilidade de observar um valor tão ou mais desfavorável à hipótese nula quanto a particular média amostral estimada. Qual a conclusão deste teste com 1% de significância? Valor-p=1-0,9979=0,0021=0,21%
- d) Compare as conclusões obtidas no item b (com base na RC) e no item c (no valor-p). Há diferenças? Por quê?

Insper

Exemplo 5

Em 2012, antes dos supermercados deixarem de distribuir sacolas plásticas gratuitamente para os consumidores, o consumo mensal de sacolas plásticas na cidade de São Paulo por adulto era, em média, de 25 sacolas plásticas, com desvio padrão de 6 unidades (valores populacionais).

No mesmo ano, os estabelecimentos passaram a cobrar pelo uso das mesmas e depois voltaram atrás. Alguns anos se passaram e uma ONG da área de sustentabilidade deseja avaliar se essas idas e vindas deixaram o consumidor um pouco mais consciente com o meio ambiente.

Para avaliar se consumo está mais consciente, verifique se o consumo de sacolas plásticas reduziu, em média, em mais do que 15% após todo esse período, ao nível de significância de 1%?

A ONG monitorou, por um mês, 100 adultos moradores da cidade e observou que o consumo médio amostral de sacolas plásticas entre eles foi de 20,01 unidades.

Segunda decimal de z 3 O 2 4 5 6 8 7 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 0.9993 0.9993 3.2 0.9993 0.9993 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 0.9998 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 3.7 0.9999 0.9999 0.9999 0.9999 0.9999 0.9999 0.9999 0.9999 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