

EPH505 TF Review

Assignment Project Exam Help

Session!

<https://powcoder.com>

Add WeChat Exam1 powcoder

Lecture 1: Getting Started in R

Masters will ...

- Conduct basic operations at the R command prompt
- Create an R-file for portable coding
- Interpret basic elements of the R interface

Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder



Version should be >=4.0.0

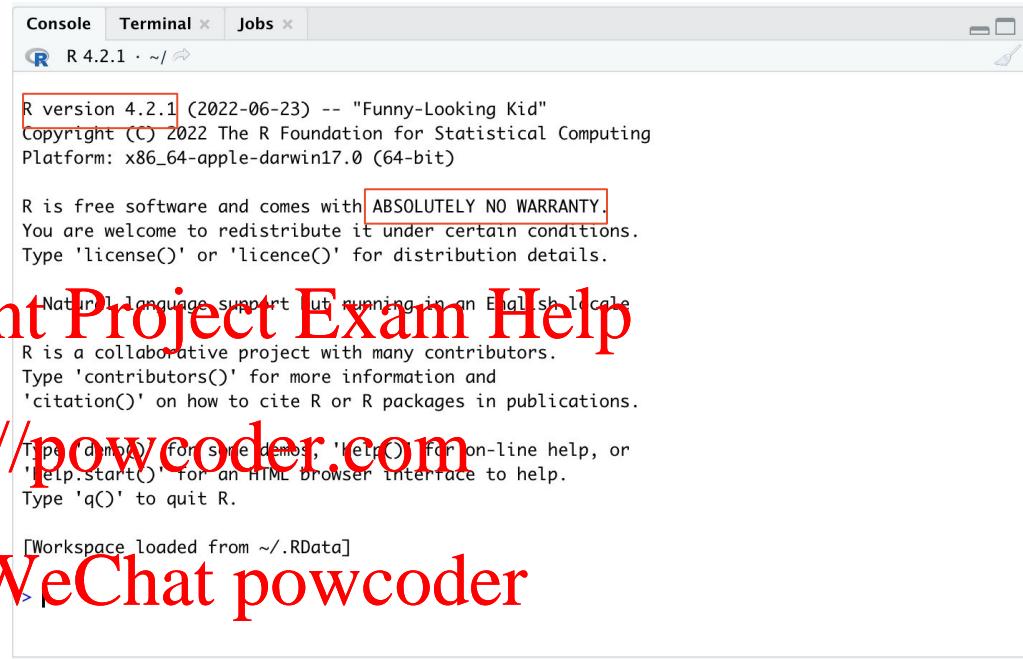
You can use R for anything,
and it's well-accepted ... but
it's your responsibility to
know how to apply
appropriate methodologies

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Command prompt!



R version 4.2.1 (2022-06-23) -- "Funny-Looking Kid"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support not running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

Command Prompt

```
>  
>  
> 09252022|
```

Assignment Project Exam Help

<https://powcoder.com>

Here, R is simply echoing what you inputted. A new command prompt is now ready.

Here, R has not yet registered the input.

Add WeChat powcoder

```
> 09252022  
[1] 9252022  
> |
```

Variable Assignment

```
> biostats=193  
> |
```

When you assign a variable in R, it will suppress the output and an empty command prompt is a good sign!

Assignment Project Exam Help

```
> biostats=193  
> biostats/5  
[1] 38.6  
>
```

<https://powcoder.com>

Now, we can perform basic arithmetic operations on this variable.

Add WeChat powcoder

```
> biostats  
[1] 193  
> |
```

Typing in the variable name and hit <ENTER> and you get the value

```
> ls()  
[1] "biostats"  
> |
```

ls() tells you the variables stored within R's memory

R-Files: why do we use them?

- They function like a text editor/word document where large amounts of code can be stored
- You can make changes to your code without receiving output each time you hit <ENTER>
 - You can execute your code from an R-file: just hit Edit > Execute
- Very useful for when you want to collaborate
 - If all collaborators have the dataset on their end (and make sure working directory is specified to your computer), then all you have to do is run the R-file code



```
source on  
1 biostats=193  
2 biostats  
3 |
```

Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder

Example code from
R-file → not live

R Miscellany: numbering

```
> claire=runif(17)
> claire
[1] 0.14002989 0.13528896 0.51604985 0.89091009 0.87817609 0.61731840 0.56144079 0.31745152
[9] 0.23953406 0.16145500 0.65324055 0.89381498 0.82376200 0.08894893 0.77072739 0.76358957
[17] 0.76435939
> claire[10]
[1] 0.161455
>
```

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

- R uses [square brackets] to index variables
- Reads the rows left to right, starting at 1
- Inputting claire[10] identified the 10th element

R Miscellany: over-writing & re-annotating

```
> chocolate=100  
> chocolate  
[1] 100  
> chocolate=200  
> chocolate  
[1] 200  
> |
```

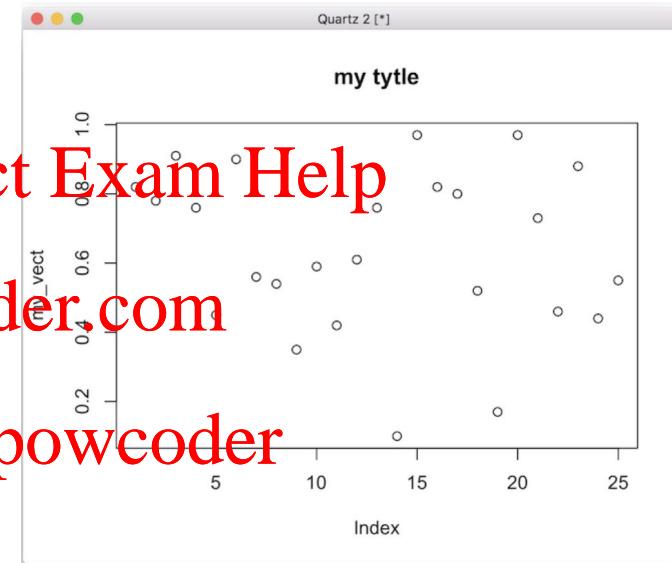
In R, you can overwrite a variable so that it contains new/different elements

But make sure it was intentional! You don't want to accidentally lose information

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



If you accidentally annotate your plot incorrectly, you can...

- Manually exit out of plot
- Type dev.off()
- Or issue new plot() command and overwrite your old one

R Miscellany: data frames & strings

Data frame: a special 2-dimensional data structure (specific to R) that allows us to mix data-types (e.g. strings & numbers) → different from a matrix because a matrix will only allow *same* data-types



```
R Console
>
> plot(my_vect,main="my tytle")
> plot(my_vect,main="my proper title")
>
>
> data=read.csv("Yale_Athletics.csv")
> head(data)
   Year Number      Name Position Class Height Weight
1 2008     11 Ryan Brenner        C Freshman    71   180
2 2008     34 Joe Castaldi       P Junior     74   190
3 2008      1 Gant Elmore       IF Sophomore   69   170
4 2008     40 Chris Finneran     P Junior     71   175
5 2008     29 Robert Gruber       P Junior     79   220
6 2008     17 Brandon Josselyn     P Senior     75   200
   TownHS Sport
1 Manhasset, N.Y. / Chaminade Baseball
2 Sunrise, Fla. / Pine Crest School Baseball
3 Bloomington, Ind. / Bloomington North Baseball
4 Suffern, N.Y. / Don Bosco Prep Baseball
5 Atlanta, Ga. / Chamblee Baseball
6 Hanson, Mass. / Whitman-Hanson Regional Baseball
>
> class(data)
[1] "data.frame"
```



What class-type would
class(data\$Class) yield?
What about
class(data\$Height)?

Does not support
arithmetic! Even if it is a
number, wrapped in ""

Strings: any information that is protected by ""

Can be characters or text, as long as it has quotes

<https://powcoder.com> Class-type is character



```
R Console
>
> my_varbl = 505
> my_varbl/2
[1] 252.5
>
>
> my_varbl = "505"
> my_varbl/2
Error in my_varbl/2 : non-numeric argument to binary
operator
>
```

R Miscellany: nesting

Nesting is embedding one functional expression as the argument of another functional expression

Assignment Project Exam Help

```
> vector=c(0,9,2,5,3,1) https://powcoder.com ← Nesting  
> length(which(vector>2))  
[1] 3  
>
```

Add WeChat powcoder

Some helpful basic R commands

- `ls()`
 - No arguments ; tells you variables stored in R's memory
- <ESCAPE>
 - Pressing this key will "interrupt" processes in R and return user to command prompt
- `dir()`
 - No arguments: tells you the contents of current working directory
 - With file path argument: tells you contents of the particular folder
- `setwd("file/folder path")` > setwd("/Users/claireszapary/Desktop/YaleMPH/EPH505/Lab")
- `getwd()`
 - No arguments ; tells you what your working directory currently is
- `read.csv("file path")` > yale=read.csv("YaleAthletics.csv")
 - Reads data in from csv file and will create a data-frame from it
- `head(name of dataframe, # of rows you want to see)`
 - Will show you the several rows of choice that you want to see from data-frame head(yale, 6)

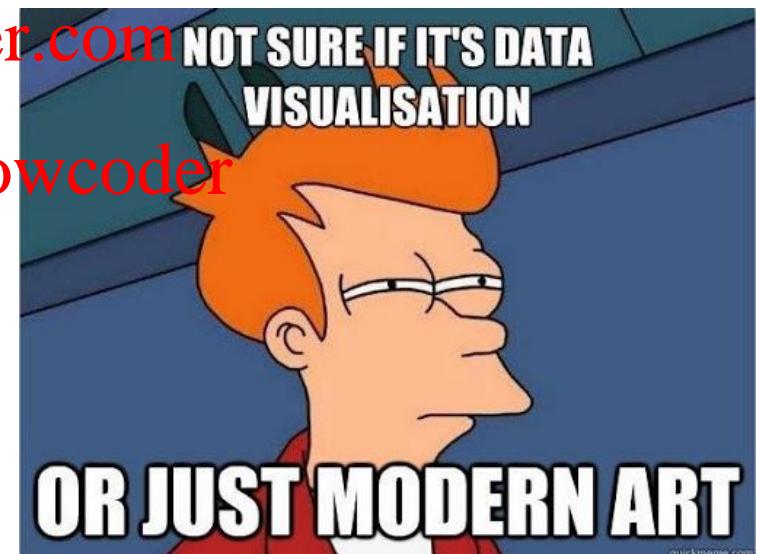
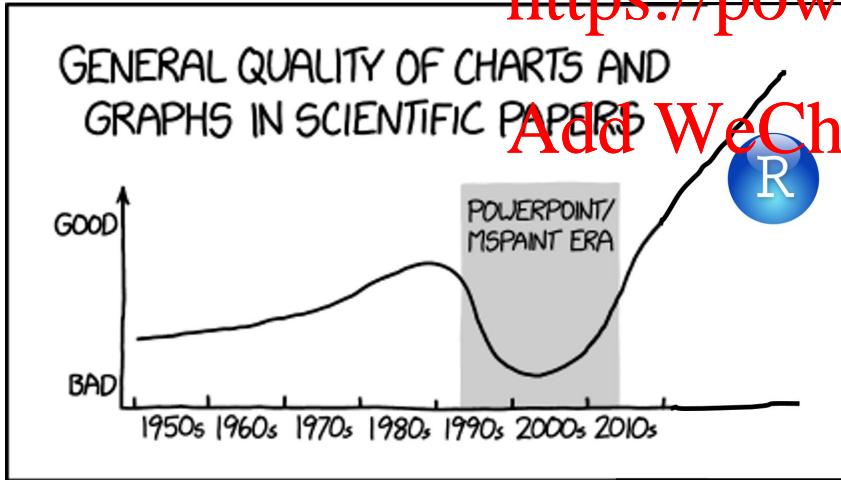
Lectures 2 Masters Will

- Differentiate data among the 5 main (statistical) data-types
- Predict the data-type in R (without even using R!)
- Illustrate data using the proper chart type

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Predict the Data Types!

```
> poke
```

	pokedex_number	name	height_m	weight_kg	generation	base_egg_steps	zipcode
1	1	Bulbasaur	1	6.9	1	5120	06515
2	2	Ivysaur	1	13.0	1	5120	06510
3	3	Venusaur	2	109.0	1	5120	06511
799	799	Guzzlord	6	888.0	7	30720	06503
800	800	Necrozma	2	230.0	7	30720	06505
801	801	Magearna	1	80.5	7	30720	06513

Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder



Predict the Data Types!

Nominal Nominal Continuous Continuous Ordinal Discrete Nominal

Assignment Project Exam Help

	pokedex_number	name	height_m	weight_kg	generation	base_egg_steps	zipcode
1	1	Bulbasaur	1	6.9	1	5120	06515
2	2	Ivysaur	1	13.0	1	5120	06510
3	3	Venusaur	2	100.0	1	5120	06511
799	799	Guzzlord	6	888.0	7	30720	06503
800	800	Necrozma	2	230.0	7	30720	06505
801	801	Maqearna	1	80.5	7	30720	06513

Add WeChat powcoder



Dataset from Kaggle: <https://www.kaggle.com/datasets/rounakbanik/pokemon>

Categorical Data Types

	Definition	Arithmetic supported?	Example
Nominal	Categories have no order Assignment Project Exam Help	No	Zipcodes, Names, IDs, Colors, Repe Pizza flavors, Sports teams, Gender
Ordinal	Categories have an order Add WeChat powcoder	No	Grade levels, T-shirt sizes, Difficulty level, Amazon ratings, Disease stages, Age Range,
Rank	Elements ordered in a <u>single</u> category, ties allowed	Somewhat*	Billboard hits, Top 10 leading causes of death,

<https://powcoder.com>

Add WeChat powcoder

Ordinal vs. Rank?

- **Ordinal:** assigns hierarchy to an overall category

Assignment Project Exam Help

- **Rank:** assigns hierarchy to each observation within a category
 - The rank that is assigned to every element of the dataset is sensitive to the composition of the dataset
 - Ranks are PRESERVED only if the ranks aren't changed

Example on next slide!

Rank the pokemon 1-6 with rank=1 being the lightest

- What pokemon is rank=1?

Assignment Project Exam Help

- What generation does the pokemon belong to?

<https://powcoder.com>

```
> poke
   pokedex_number      name height_cm weight_kg generation base_egg_steps zipcode
1           1 Bulbasaur     1       4.0        1             1      5120    06515
2           2 Ivysaur      1      13.0        1             1      5120    06510
3           3 Venusaur     2      100.0        1             1      5120    06511
799        799 Guzzlord     6     888.0        7            7     30720    06503
800        800 Necrozma     2     230.0        7            7     30720    06505
801        801 Magearna     1      80.5        7            7     30720    06513
```

Add WeChat powcoder

Rank the pokemon 1-6 with rank=1 being the lightest

- What pokemon is rank=1?

```
> poke$name  
[1] "Bulbasaur" "Ivysaur"  "Venusaur" "Guzzlord" "Necrozma"  "Magearna"  
> rank(poke$weight_kg)  
[1] 1 2 4 6 5 3  
> |
```

Assignment Project Exam Help

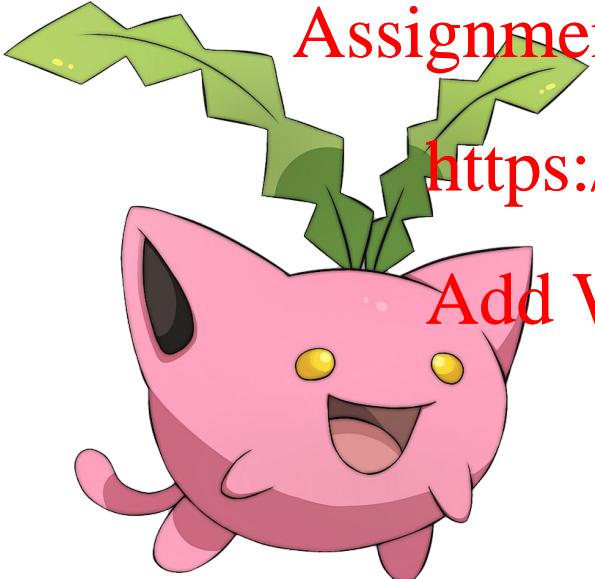
<https://powcoder.com>

Add WeChat powcoder

- What generation does the pokemon belong to?

```
++  
> poke$generation[which(poke$name=="Bulbasaur")]  
[1] 1  
|
```

NEW CATCH! HOPPIP



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

GRASS TYPE

GENERATION

2

WEIGHT

0.5 kg

HEIGHT

0.4 m

```
> poke2
```

	pokedex_number	name	height_m	weight_kg	generation	base_egg_steps	zipcode
1	1	Bulbasaur	1	6.9	1	5120	06515
2	2	Ivysaur	1	13.0	1	5120	06510
3	3	Venusaur	2	100.0	1	5120	06511
187	187	Hoppip	0	0.5	2	5120	06503
799	799	Guzzlord	6	888.0	7	30720	06505
800	800	Necrozma	2	230.0	7	30720	06513
801	801	Magearna	1	70.5	7	30720	06503

Assignment Project Exam Help

- Now, with Hoppip included in our dataset, what is Bulbasaur's rank by weight?

Add WeChat powcoder

```
> poke2$name
```

```
[1] "Bulbasaur" "Ivysaur" "Venusaur" "Hoppip" "Guzzlord" "Necrozma" "Magearna"
```

```
> rank(poke2$weight_kg)
```

```
[1] 2 3 5 1 7 6 4
```

Bulbasaur's rank changed
from #1 to #2,
His generation did NOT
change

Quantitative Data Types

	Definition	Arithmetic supported?	Example
Discrete	Counting process Typically integer* (no always!) “How many?”	Yes	Number of pokemon, gyms, coffee shops
Continuous	Measurement process Can be integer, fraction, decimal “How much”	Yes	Weight, Height, Temperature, Time*

<https://powcoder.com>

Add WeChat powcoder

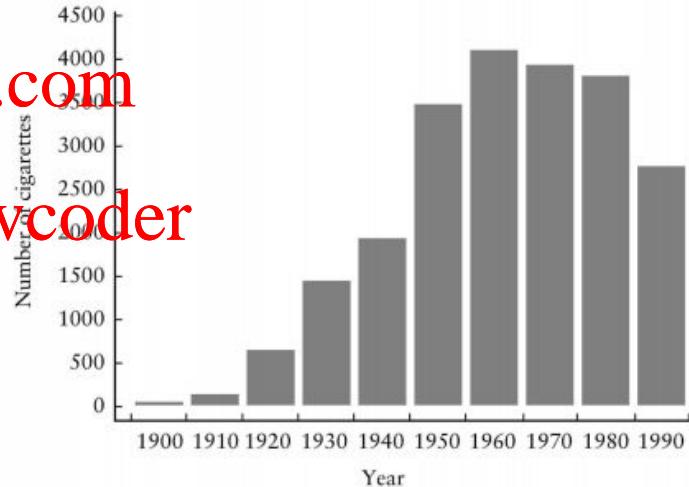
Warning! Data Presentation vs. Reality

- Data that is shown as a whole number can refer to categorical data
 - Zipcodes, IDs, Gender(1="Male", 2="Female")
- Data shown as whole numbers can also be continuous
 - Height, Weight often rounded to the nearest whole number but since they are measured they are continuous
- Data shown with decimals can be discrete
 - Seeing 5 pokemon per in 2 hours shown as a rate of 2.5
- Time is continuous but can be visualized as ordinal (see right figure)

Assignment Project Exam Help

<https://powcoder.com>

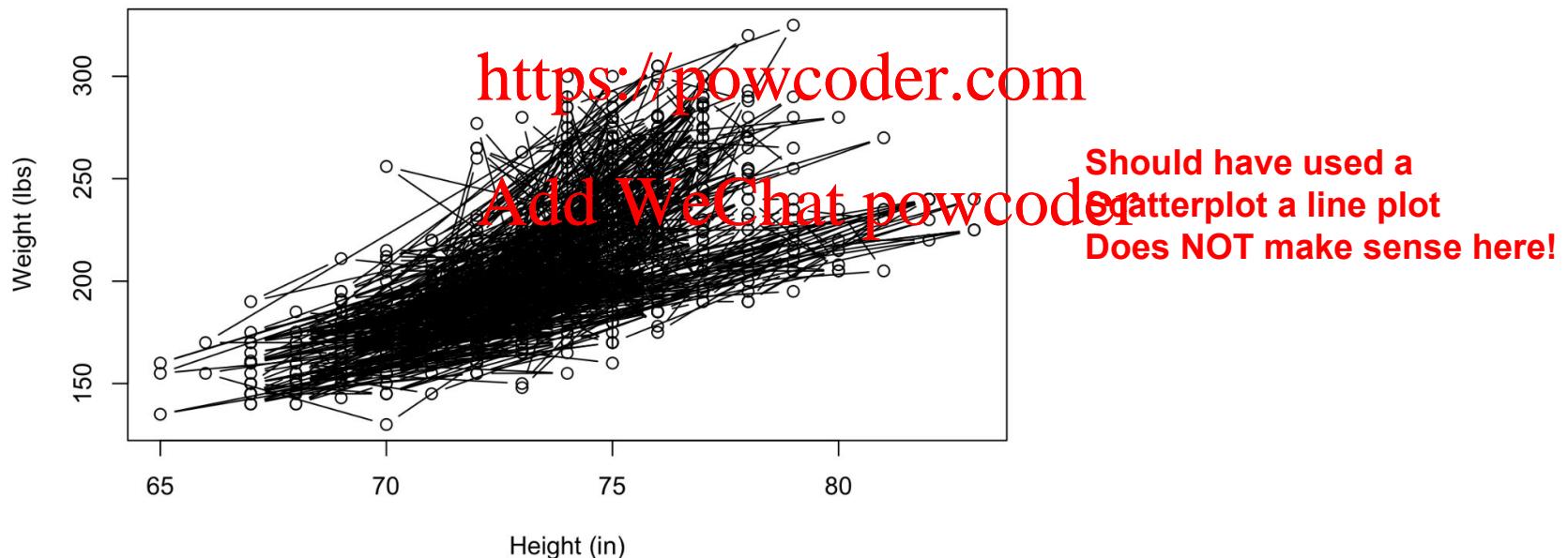
Add WeChat powcoder



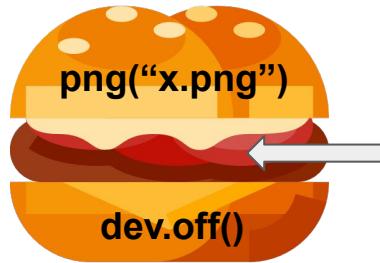
What's the big deal with data types?

- Data visualization and statistical analysis (more in Module 2)

Assignment Project Exam Help



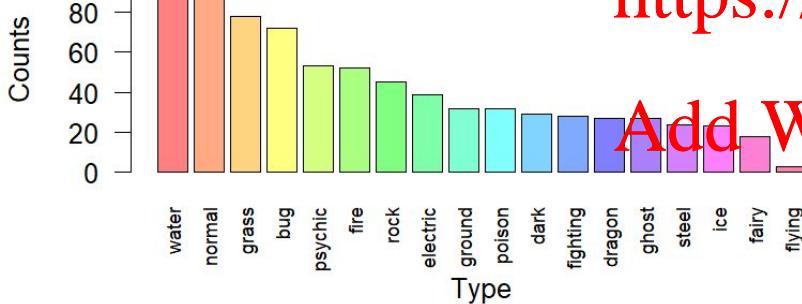
How to save a plot?



Analogy: Plot Sandwich

Plot code HERE

Pokemon Counts by Type



png("pokemon_counts.png")

```
barplot(sort(table(pokemon$type1), decreasing=TRUE),  
       las=2,  
       cex.names=0.7,  
       col=rainbow(18, s=0.5))  
title(main="Pokemon Counts by Type")  
title(ylab="Counts")  
title(xlab="Type")
```

<https://powcoder.com>

Add WeChat powcoder

dev.off()

Chart Types

Note: `plot(dfx,dfy)` is the same as `plot(df$y ~ df$y)`

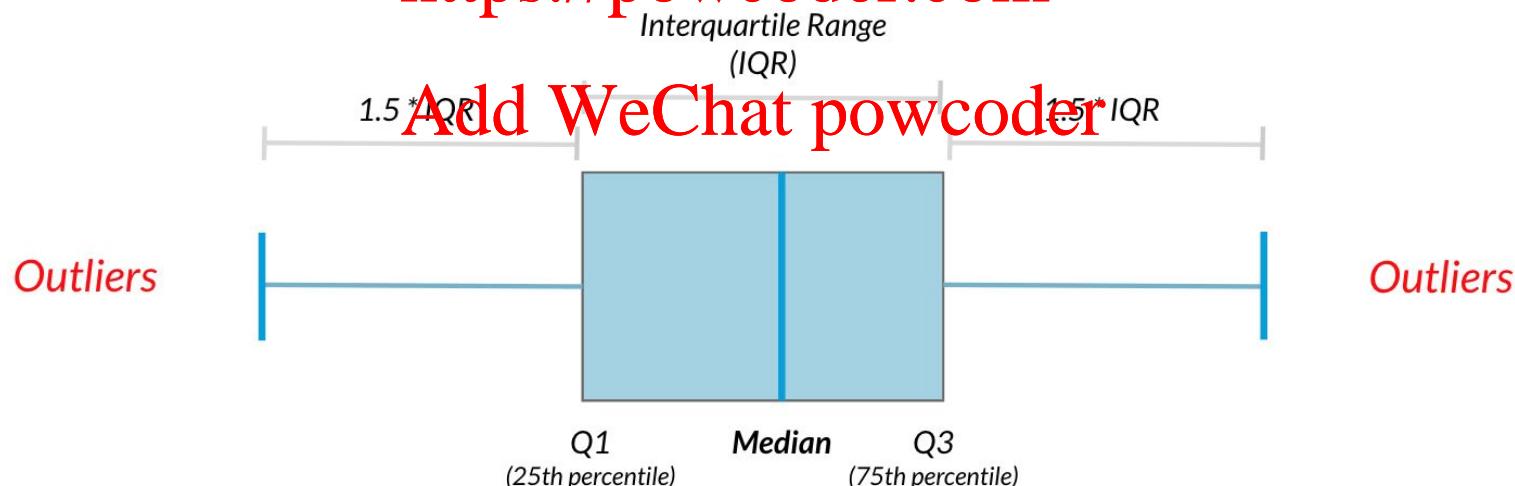
	Independent Variable (X)	Dependent variable (Y)	Code
Bar Chart	Nominal/Ordinal	Count	<code>barplot(table())</code>
Histogram Density	Discrete/Continuous	Count	<code>hist()</code> <code>hist(prob=TRUE)</code> <code>lines(density())</code>
Scatterplot	Continuous	Continuous	<code>plot()</code>
Lineplot	Ordinal	Continuous	<code>plot(line="l")</code>
Boxplot	Nominal	Continuous	<code>boxplot()</code>

Boxplot vocabulary (quick review from Primer)

- Whiskers = Central lines spanning 1.5 IQL from either quartile
- Outliers = Beyond whiskers
- Box = 25th and 75th percentile
- Horizontal line = Median

Assignment Project Exam Help

<https://powcoder.com>



Predict the class in R!

What did I forget to do?

```
pokemon = read.csv("../data/pokemon.csv")
my_cols = c("name", "height_m", "weight_kg", "generation", "base_egg_steps", "type1")
my_rows = c(1, 2, 3, 799, 800)
Assignment Project Exam Help
poke = subset(pokemon, subset=pokemon$pokedex_number %in% my_rows, select=my_cols)
poke$zipcode=c("06515", "06510", "06511", "06505", "06513")
poke
```

<https://powcoder.com>

> poke

		name	height_m	weight_kg	generation	base_egg_steps	type1	zipcode
1	Bulbasaur		0.7	6.9	1	5120	grass	06515
2	Ivysaur		1.0	13.0	1	5120	grass	06510
3	Venusaur		2.0	100.0	1	5120	grass	06511
799	Guzzlord		5.5	888.0	7	30720	dark	06503
800	Necrozma		2.4	230.0	7	30720	psychic	06505
801	Magearna		1.0	80.5	7	30720	steel	06513

Add WeChat powcoder

Double-check

```
> class(poke$name)
[1] "character"
> class(poke$height_m)
[1] "numeric"
> class(poke$weight_kg)
[1] "numeric"
> class(poke$generation)
[1] "integer"
> class(poke$base_egg_steps)
[1] "integer"
```

```
> table(poke$generation)
```

1	7
3	3

Assignment Project Exam Help

<https://powcoder.com>

```
> str(poke)
'data.frame':   6 obs. of  5 variables:
 $ name        : chr  "Bulbasaur" "Ivysaur" "Venusaur" "Guzzlord" ...
 $ height_m    : num  0.7 1 2 5.5 2.4 1
 $ weight_kg   : num  6.9 13 100 888 230 80.5
 $ generation  : int  1 1 1 7 7 7
 $ base_egg_steps: int  5120 5120 5120 30720 30720 30720
```

Add WeChat powcoder

	Default class in R:	Data Type
Integer	Vector contains all numbers, no decimals AND vector is read from file	Discrete
Numeric	Vector contains decimals or <u>vector created manually</u> and contains strictly numbers	Continuous R does NOT "Rank" class It's just a NUMERIC type
Character	Vector with string data https://powcoder.com Add WeChat powcoder	Need to convert it to factor or ordered factor to do analysis
Factor	Vector with string data that has levels with no particular order	Categorical
Ordered Factor	Vector with string data that has ordered levels	Ordinal

Questions

- I create a smaller dataframe with just the first generation Pokemon, which column(s) have changed their class?

Assignment Project Exam Help

```
> poke_small = subset(poke, subset=poke$generation==1)
> poke_small
```

<https://powcoder.com>

		name	height_m	weight_kg	generation	base_egg_steps	type1	zipcode
1	Bulbasaur		0.7	6.9	1	5120	grass	06515
2	Ivysaur		1.0	13.0	1	5120	grass	06510
3	Venusaur		2.0	100.0	1	5120	grass	06511
.								

Add WeChat powcoder

No subsetting a dataframe does not change its class.

Questions

- How can I change “type1” from character to factor?

```
poke$type1 = as.factor(poke$type1)
pokemon = read.csv("../data/pokemon.csv", stringsAsFactors = TRUE)
```

```
> class(poke$type1)
[1] "factor"
```

<https://powcoder.com>

- If I created a column called pokemon strength with 3 levels which one would R see as the first level?

```
strength = c("weak", "neutral", "strong")
strength = ordered(as.factor(strength))
```

Need to include a levels argument for it to order correctly
Rather than alphanumerical order

```
> strength
[1] weak    neutral strong
Levels: neutral < strong < weak
```

Questions

- I decided to round my height_m to whole number, has it changed its class?

```
name height_m weight_kg generation base_egg_steps type1 zipcode
1 Bulbasaur      1       6.9             1           5120 grass    06515
2 Ivysaur        1      13.0             1           5120 grass    06510
3 Venusaur       2     100.0             1           5120 grass    06511
> |
```

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

No! It's still numeric because it
Was read in from file AND contained a
decimal

```
> class(poke_small$height_m)
[1] "numeric"
```

Warning! Note these distinctions

- String vs. character?
 - Character is a class, strings are CONTAINED in the character class
- Character vs. Factor?
 - Characters are just simple strings
 - Factors are seen as categories in R that have different levels
- Rank vs. Order() vs. Sort()
 - **Rank()** returns the RANK, either a decimal if there are ties or a whole number
 - **Order()** returns INDICES that would yield a sorted set, has to be a whole number
 - **Sort()** returns sorted VALUES, has to be a whole number

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

```
> v<-c(2,1,3,4,5,1)
> rank(v)
[1] 3.0 1.5 4.0 5.0 6.0 1.5
> order(v)
[1] 2 6 1 3 4 5
> v[order(v)]
[1] 1 1 2 3 4 5
> sort(v)
[1] 1 1 2 3 4 5
> |
```

Lecture 3 Masters Will

- Apply measures of centrality in their appropriate contexts
- Calculate group-wise summaries from a raw dataset
- Explain importance of accounting for survey weights and groupings

Assignment Project Exam Help



<https://powcoder.com>

Add WeChat powcoder



Measures of Central Tendency

	Conceptual Definition	Function in R	Use Case
Mean	VALUE that lies in the middle of the datasets' VALUES	mean() sum() / length()	Symmetrical and Single-peaked
Median	DATAPOINT that lies in the middle of the dataset's SORTED DATAPOINTS (if odd),	median() quantile(x, 0.5)	When NOT symmetrical or NOT singly-peaked
Mode	Most populated element	sort(table()) No designated function in R	Categorical data

What kind of distribution do you expect income will have?

More on this in Lecture 6

Report mean +/- standard deviation

Unimodal, symmetric

1 measure
ordinal
as the
smallest
other
observations
at mea-

Assignment Project Exam Help

<https://powcoder.com>

Report median and IQR

Unimodal, right-skewed

Sometimes reported as
Median (25th-75th)

is the
r equal

Report median and IQR

Unimodal, left-skewed

Add WeChat powcoder

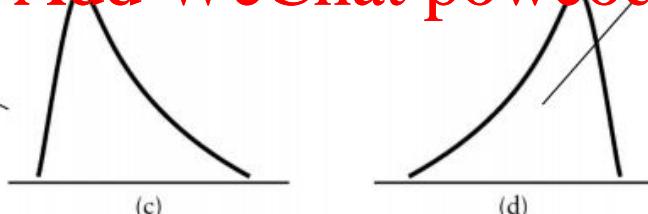


FIGURE 3.1

Possible distributions of data values

Mean vs. (Median and Mode)

- Mean is an abstraction of data element
- Median (if its odd) and Mode are actual elements in the dataset

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Measures of Variation

	Definition	Function in R	Use Case
Variance	Average of the squared deviations of the mean	var()	For statisticians-units-squared Assignment Project Exam Help https://powcoder.com
Standard Deviation	Square root of the variance	sd()	Same units as the mean Add WeChat powcoder
Range	Maximum value minus the minimum value	diff(range(x))	Table 1

Measures of Variation

	Definition	Function in R	Use Case
Interquartile Range	Difference between the upper and lower quartiles	IRQ() quantile(x, 0.75) quantile(x, 0.25)	Report with the median
Coefficient of Variation	Mean divided by the standard deviation, unitless	mean()/sd() No designated function in R	Compare variation between two things that have different units

<https://powcoder.com>

Add WeChat [powcoder](#)

Sigma Notation - Mean

SIGMA NOTATION
(represents the sum of a sequence)

$$\sum_{k=1}^n x_k$$

Upper Limit (where to stop) → n
Greek letter "Sigma" (summation symbol) → \sum
Lower Limit (where to start) → $k=1$

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

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

$$\begin{aligned}\bar{x} &= \frac{1}{13} \sum_{i=1}^{13} x_i \\ &= \left(\frac{1}{13} \right) (2.30 + 2.15 + 3.50 + 2.60 + 2.75 + 2.82 + 4.05 \\ &\quad + 2.25 + 2.68 + 3.00 + 4.02 + 2.85 + 3.38) \\ &= \frac{38.35}{13} \\ &= 2.95 \text{ liters.}\end{aligned}$$

Sum of the values of X from X_1 through X_n

$$\sum_{i=1}^n a_i = a_1 + a_2 + \dots + a_n$$

Sigma Notation - Variance

- Average of the squared deviations of the mean

Assignment Project Exam Help

**“Mean” term

Deviation term

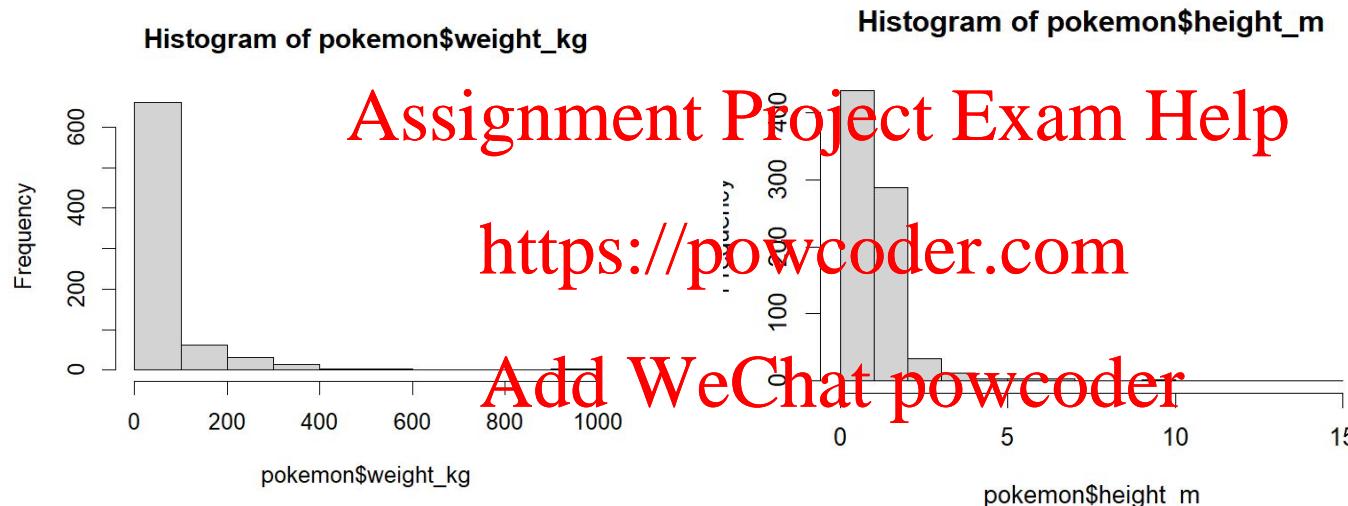
$$s^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2$$

https://powcoder.com

Add WeChat powcoder

Why do we square the deviations?

Does height or weight have more variation?



```
> sd(pokemon$weight_kg, na.rm=TRUE)
[1] 109.3548
> sd(pokemon$height_m, na.rm=TRUE)
[1] 1.080326
>
```

Does height or weight have more variation?

Use the Coefficient of Variation to compare two measures with different units. Even though we achieved the same answer, this won't always be the case, see Problem Set #3

$$CV = \frac{s}{\bar{x}} \times 100\%$$

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder
cv_wt = sd(pokemon\$weight_kg, na.rm=TRUE)/mean(pokemon\$weight_kg, na.rm=TRUE)
cv_ht = sd(pokemon\$height_m, na.rm=TRUE)/mean(pokemon\$height_m, na.rm=TRUE)

```
-- 
> cv_ht
[1] 0.9282011
> cv_wt
[1] 1.781658
```

Weighted Mean/Global Mean vs. Mean

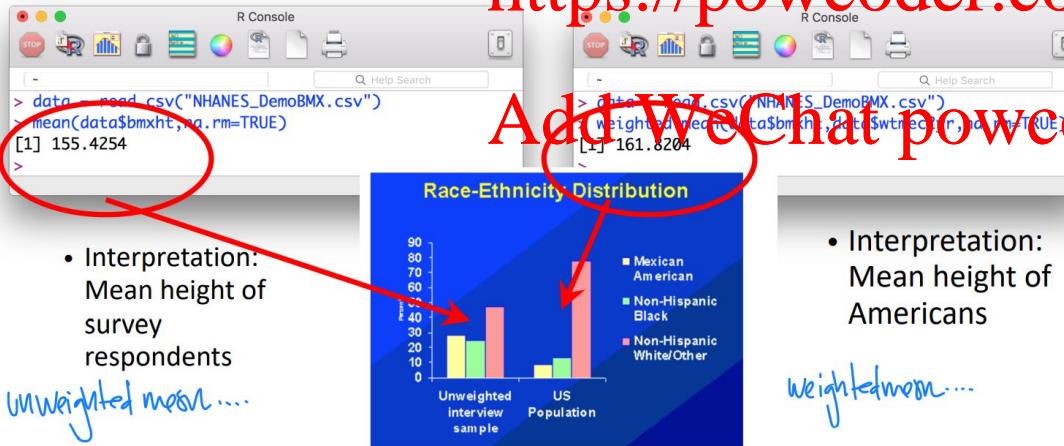
- Weighted Mean/Group Mean/Global average = each person in the group represents a proportion of the population
- Mean = each person represents themselves

Assignment Project Exam Help

weighted.mean () has 3 arguments
1. The values you want to take the mean of
Weights
3. na.rm=TRUE

Mean vs Group Mean in NHANES

<https://powcoder.com>



Add Wechat powcoder

Why is weighted mean important?

It saves time, \$\$\$, resources (we don't have to go out and find 300 million people)

Weighted Mean/Global Mean vs. Mean

Number (Hrs of Sleep)	Weighting factor (Number of weeks)	Number x Weighting factor
7	x 9	= 63
5	x 3	= 15
8	x 2	= 16
4	x 1	= 4
		15 98

Weighted Average =
$$\frac{\text{Sum of } (\text{Number} \times \text{Weighting Factor})}{\text{Sum of all the Weights}}$$

$$= \frac{98}{15} = 6.53 \text{ (Hrs)}$$

Assignment Project Exam Help

Add WeChat powcoder

```
> sleep_hours = c(7,5,8,4)
> wt_weeks = c(9,3,2,1)
> df = data.frame(sleep_hours, wt_weeks)
> mean(df$sleep_hours)
[1] 6
> weighted.mean(df$sleep_hours, df$wt_weeks)
[1] 6.533333
[((7*9)+(5*3)+(8*2)+(4*1))/sum(df$wt_weeks))
[1] 6.533333
```

The doBy Package

Functional notation “as a function of” dataframe

Function(s) you want to apply.
If multiple functions, use a vector to list them

Assignment Project Exam Help

Keep.names=FALSE
as a default

```
> summaryBy(Weight~Sport, data=y1, FUN=c(mean, sd, length))
```

	Sport	Weight.mean	Weight.sd	Weight.length
1	Baseball	191.9439	19.43049	107
2	Basketball	205.6520	21.16036	95
3	Football	220.9655	35.14439	580
4	Hockey	186.8974	15.58425	117
5	Lacrosse	185.3486	14.59565	175
6	Soccer	NA	NA	154

Add WeChat powcoder

What happens if you add keep.names =TRUE?

Long format

- **Long-format:**
 - SINGLE observations as rows, 1 column per variable
- Assignment Project Exam Help

Long-format

<https://powcoder.com>

name	date	weight_kg
Bulbasaur	9/23/2022	6.9
Bulbasaur	9/24/2022	8
Venusaur	9/23/2022	13
Venusaur	9/24/2022	11
Charmeleon	9/23/2022	8.5
Charmeleon	9/24/2022	10

Add WeChat powcoder

NOT Long-format

name	9/23/2022	9/24/2022
Bulbasaur	6.9	8
Venusaur	13	11
Charmeleon	8.5	10

Lecture 4: Probability

Masters Will...

Assignment Project Exam Help

- Detect probability structures in word problems
<https://powcoder.com>
- Calculate raw and conditional probabilities
Add WeChat powcoder
- Recognize sampling scenarios that lead to erroneous interpretations



Frequentist vs. Bayesian Statistics

Frequentist

- **Definition:** If an experiment is repeated n times under essentially identical conditions, and if the **event A occurs m times**, then as n grows large, the ratio m/n approaches a **fixed limit** that is the **probability of A** (Pagano & Gauvreau, pg 127)
 - $P(A) = m/n$
- Can also be described as the **proportion of times an event occurs**
- It is the **bedrock** of how **nearly all biostatistics** are based on this notion of frequentism

Assignment Project Exam Help

<https://powcoder.com>
Add WeChat powcoder

Bayes Theorem

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\text{not } A) \cdot P(\text{not } A)}$$

Where

- $P(B|A)$ = probability of B occurring, given A has occurred
- $P(A)$ = probability of A occurring
- $P(B|\text{not } A)$ = probability of B occurring, despite A not occurring
- $P(\text{not } A)$ = probability of A not occurring
- $P(A|B)$ = probability of A occurring, given B

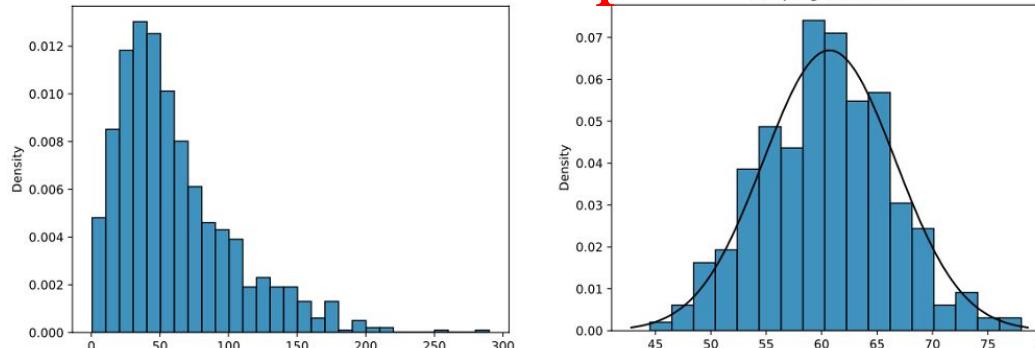
Frequentist Statistics

- If we conduct more and more experiments, the collection of results will converge to the true population result.
- The notion of the sampling distribution is grounded in this concept!
- 100 people go out → each find 10 people → the averages collected by each of the 100 people, should be closer to the true population mean than each individual average. If plotted, the averages should be approximately normally distributed around the true population mean.

Assignment Project Exam Help

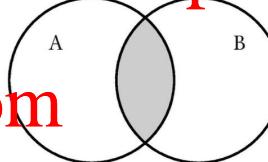
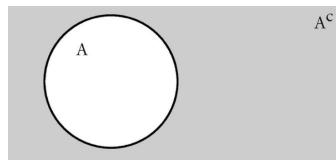
<https://powcoder.com>

Add WeChat powcoder



Set Theory Vocabulary

- ★ *Probability = “proportion”*
- ★ *Event
 - Usually indicated by upper-case letters
 - Either occurs, or does not occur*

Notations	Definition	Image
$A \cap B$	<u>“A intersect B”</u> Both events occurred https://powcoder.com	
$A \cup B$	<u>“A union B”</u> Either A occurred, or B occurred, or A & B both occurred	
A^c	<u>“A complement”</u> Event A not occurring	

Set Operations in R



Create a new vector called a

Create a new vector called b

Finds all of the values where they are the same in both vectors... in other words “intersect”
Note: you can also use intersect(b,a) to yield the same output

Finds values where either a,b, or both a and b are included
Note: you can also use union(b,a) to yield the same output

Finds the set differences in b relative to a (output is unique values in “a” vector)

Finds the set differences in a relative to b (output is unique values in “b” vector)

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

```
> a=c(8,6,0,1,2,3,4)
> print(a)
[1] 8 6 0 1 2 3 4
> b=c(2,0,3,5,6,7,8)
> print(b)
[1] 2 0 3 5 6 7 8
>
> intersect(a,b)
[1] 8 6 0 2 3
> union(a,b)
[1] 8 6 0 1 2 3 4 5 7
> setdiff(a,b)
[1] 1 4
> setdiff(b,a)
[1] 5 7
>
```

Raw vs. Conditional Probability

Raw

- **Definition:** $P(X)$
- What are the chances of X?
- $P(X) = m/n$
 - Dividing a portion (m) from the total (n)
- **Example:**
 - What are the chances that an athlete is taller than 72 inches?
 - $P(h>72)$

Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder

Conditional

- **Definition:** $P(X|Y)$. What are the chances of X, given Y. Alternatively, probability of Y given X
- $P(X|Y) = P(X \cap Y) \div P(Y)$
- The chances that an event occurred, given that another event already occurred
- Since every probability has some context, every probability is conditional
- **Example:**
 - What is the probability of being taller than 72 inches, given that you are a football player?
 - $P(h>72 | \text{football})$

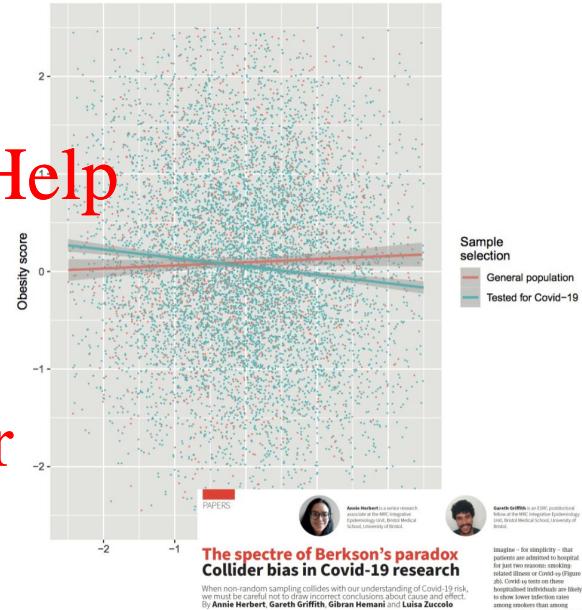
Berkson's Paradox / Collider Bias

- Inappropriate sampling causes events to appear correlated when they are actually not

- Example: Are smokers less likely to be hospitalized?

- Smokers are more likely to be present at hospitals (perhaps due respiratory illness)
- However, COVID leads to more non-smokers to be present at hospital
- Now, if you took a sample from the hospital during COVID, you are **more likely** to find **nonsmokers** in your sample → and **erroneously conclude that smokers are less likely to be hospitalized than non-smokers.**

Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder



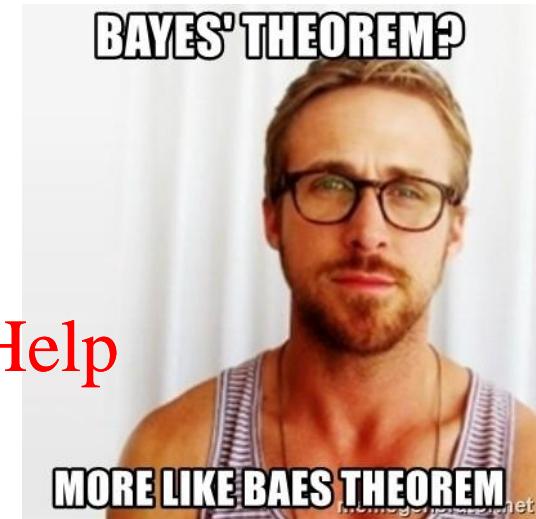
BAYES' THEOREM?

Lecture 5: Probabilistics

Masters Will...

Assignment Project Exam Help

- Identify conditional probabilities from word problems
<https://powcoder.com>
- Recognize True/False Positive/Negative from word problems
[Add WeChat powcoder](#)
- Calculate odds ratio and relative risk



Odds & Risk

- **Risk (Probability) = m/n**
 - The probability of an event (ie. an outcome is frequency of occurrence (m) over a large number of trials (n))
 - Always between 0 and 1

Assignment Project Exam Help

<https://powcoder.com>

- **Odds = $P / (1-P)$** Add WeChat powcoder
 - The probability that an event will occur, divided by the probability that the event will not occur
 - Not quite a probability, but derived from probability

Diagnostic Tests

		Disease	
		Disease (D)	No Disease - Healthy (H)
Test Result	Positive (+)	TP - Sensitivity $P(+ D)$	FP $P(+ H)$
	Negative (-)	FN $P(- D)$	TN - Specificity $P(- H)$

- **True-Negative:** Test **correctly** identifies healthy person (Specificity)
- **True-Positive:** Test **correctly** identifies sick person (Sensitivity)
- **False-Negative:** Test mistakes sick person as healthy
- **False-Positive:** Test mistakes healthy person as sick

Another Helpful Illustration...

Status of person
according to “gold standard”

Has the condition Does not have the condition

Result from screening test

<https://powcoder.com>

Add WeChat powcoder

		Positive	Negative
Positive	Has the condition	True positive	False positive
	Does not have the condition	False negative	True negative

↑ ↑

Column entries for determining sensitivity Column entries for determining specificity

Row entries for determining **positive predictive value**

Row entries for determining **negative predictive value**

Back to Bayes Theorem!

of existing cases/total population

$$P(D|+) = \frac{P(+|D) \cdot P(D)}{P(+|D) \cdot P(D) + P(+|H) \cdot P(H)}$$

Annotations:

- Sensitivity (red) points to $P(+|D)$ and $P(D)$.
- Prevalence (purple) points to $P(D)$ and $P(H)$.
- 1-Specificity (yellow) points to $P(+|H)$ and $P(H)$.
- 1-Prevalence (green) points to $P(H)$.
- A red arrow points from "Assignment Project Exam Help" to the term $P(+|D) \cdot P(D)$.
- A blue arrow points from "# of existing cases/total population" to the term $P(H)$.

Note: 1-Specificity=False Positive or $P(+|H)$

Beauty of Bayes

Allows us to revise existing predictions or theories (update probabilities)

given new or additional evidence

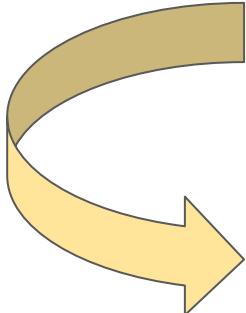
Assignment Project Exam Help

Posterior

Prevalence "Prior"

$$P(D|+) = \frac{P(+|D) \cdot P(D)}{P(+|D) \cdot P(D) + P(+|H) \cdot P(H)}$$

Add WeChat powcoder



Prevalence "Prior"

$$P(D|+) = \frac{P(+|D) \cdot Post_{1st\ test}}{P(+|D) \cdot Post_{1st\ test} + P(+|H) \cdot P(H)}$$

Relative Risk

- Definition: Probability that a member of a group receiving some exposure will develop the disease, relative to the probability that a member of a group that is unexposed will develop the disease

$$RR = \frac{P(\text{Disease} | \text{Exposure})}{P(\text{Disease} | \text{No exposure})}$$

- RR > 1: increased risk among those w/ exposure (compared to those unexposed)
- RR = 1: probabilities of disease risk in groups are identical
- RR < 1: decreased risk among those w/ exposure (compared to those unexposed)

Odds Ratio

- **Definition:** Odds of disease among members of the exposed group, relative to the odds of disease among members of the unexposed group

Assignment Project Exam Help

$$OR = \frac{P(\text{disease} | \text{exposed}) / [1 - P(\text{disease} | \text{exposed})]}{P(\text{disease} | \text{unexposed}) / [1 + P(\text{disease} | \text{unexposed})]}$$

<https://powcoder.com>
Add WeChat powcoder

How to Calculate Odds Ratio from a 2x2 Table

THREE WAYS

Disease of bones and organs of movement	
Yes	No
5	15
18	219
23	234
Totals	

Assignment Project Exam Help

<https://powcoder.com>

Cross-product

Relative odds (cross-products): 4.06

Vertically

Disease of bones and organs of movement	
Yes	No
5	15
18	219
23	234
Totals	

Relative odds (cross-products): 4.06

Horizontally

Disease of bones and organs of movement	
Yes	No
5	15
18	219
23	234
Totals	

5	15	20
18	219	237
23	234	257

Relative odds (cross-products): 4.06

Simpson's Paradox

A trend or result that is present when data is put into groups, but reverses or disappears when the data is combined. Count/proportion data can lead to different conclusions depending on whether a co-factor is introduced.

	Study 1		Study 2	
			https://powcoder.com	
Option A	12 / 48	0.250	183 / 582	0.314
Option B	104 / 411	0.253	45 / 140	0.321

- #1: $(0.250 \times 48 + 0.314 \times 582) \div (48+582) = 195 \div 630 = 0.310$
- #2: $(0.253 \times 411 + 0.321 \times 140) \div (411+140) = 149 \div 551 = 0.270$

At first glance, Option B seems superior (shows efficacy across two studies)

Add WeChat powcoder



However, if we account for sample size, Option A is actually superior overall!

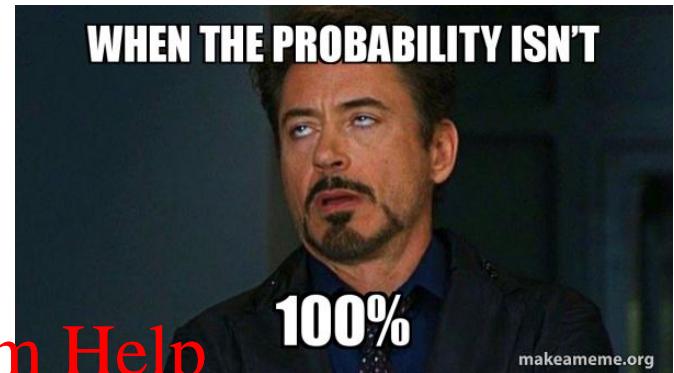


	Study 1		Study 2		Weighted Average	
Option A	12 / 48	0.250	183 / 582	0.314	See #1	0.310
Option B	104 / 411	0.253	45 / 140	0.321	See #2	0.270

Lecture 6 - Probability (canonical) Distributions

Masters Will...

Assignment Project Exam Help



- Infer distribution shape or type from knowledge of underlying process
- Convert raw data to standardized and normalized form
- Translate raw data into percentiles, and vice-versa

Add WeChat [powcoder](https://powcoder.com)

What is a probability (canonical) distribution

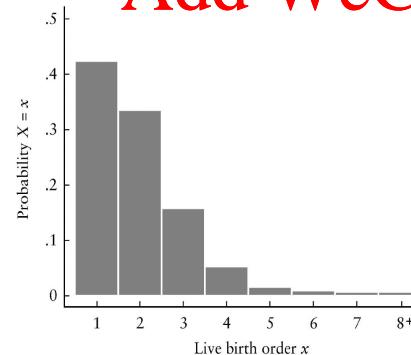


Frequency Distribution

Listing of all the frequencies of outcomes in an experiment that you actually observed during the experiment

Assignment Project Exam Help
<https://powcoder.com>

Ask 1000 children in New Haven: "What is your birth order, if you have any siblings?" and compute your results in the form of a histogram



Probability distribution

Listing of all the probabilities of possible outcomes that could occur if the experiment was done



Add WeChat powcoder

In our scenario, a probability distribution lists each possible outcome and its corresponding probability. The probabilities would represent the relative frequency of each occurrence, if the sample size was infinite. All possibilities are taken into account, so the sum of all their probabilities would be 1

Binomial Distribution

1. Dichotomous = two outcomes (*smoker vs. non-smoker*)
2. Each trial's outcome is independent
3. The probability of an outcome is constant in each trial
4. Fixed number of trials

- How do we compute these distributions? R does most of the work!

Assignment Project Exam Help
`dbinom(x, size, prob)` - returns the probability of getting a certain number of successes (x) in a certain number of trials (size) where the probability of success on each trial is fixed

<https://powcoder.com>

- Serge flips a fair coin 20 times. What is the probability that the coin lands on heads exactly 12 times? `dbinom(12, 20, 0.5)`

Add WeChat powcoder

- What does `dbinom(17,35,0.7)` translate to in plain English?

The probability of obtaining exactly 17 events in 35 trials, where the probability of the event occurring is 0.7 (or 70%)

YEAH

BUT SHOW ME NUMBERS
menegenator.net

What does the binomial distribution look like

Assignment Project Exam Help



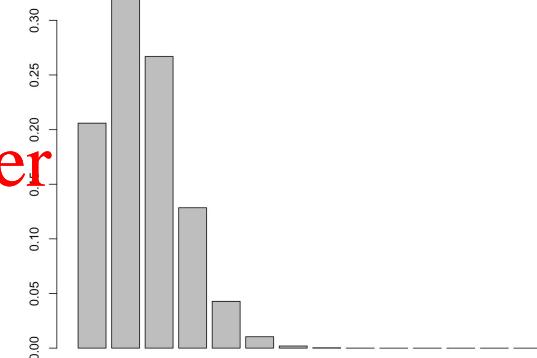
```
barplot(dbinom(0:20, 20, 0.5))
```

SYMMETRIC



```
barplot(dbinom(0:15, 15, 0.8))
```

LEFT SKEW



```
barplot(dbinom(0:15, 15, 0.1))
```

RIGHT SKEW

Poisson Distribution

Rare events

1. The average number of successes (μ) that occurs in a specified region is known
2. The probability that a success will occur is proportional to the size of the region
3. The probability that a success will occur in an extremely small region is virtually zero

Assignment Project Exam Help

- The `dpois(x,lambda)` function finds the probability that a certain number of successes (x) occur based on an average rate of success (lambda)
<https://powcoder.com>
- Add WeChat powcoder
- It is known that 20 cars drive in front of my house per hour. In a given hour, what is the probability that exactly 16 cars drive by? `dpois(16,20)`
- *Important: Slide 45/65 Lecture 6*

Normal Distribution

1. Symmetric data is not necessarily normal!
2. The normal distribution fits specific criteria
3. AUC is always 1, shape of curve changes depending on the characteristics (mean, sd)

- What if we have a continuous variable (e.g., height, temperature, serum cholesterol level)?

Assignment Project Exam Help

dnorm(x, mean, sd) - returns a value from the probability density function given a chosen random variable x , a population mean μ and population standard deviation σ

<https://powcoder.com>

- What is the probability of an individual having a systolic blood pressure of exactly 130, given that the average person has a SBP of 120 and the standard deviation in the population is 12.5? **dnorm(130,120,12.5)**

Standard Normal Distribution

- What if I want to compare apples to apples?

You calculate the Z score!

Assignment Project Exam Help

<https://powcoder.com>

- Z score is obtained by subtracting the mean from every value, and dividing the result by the standard deviation →

$$Z = \frac{X - \mu}{\sigma}$$

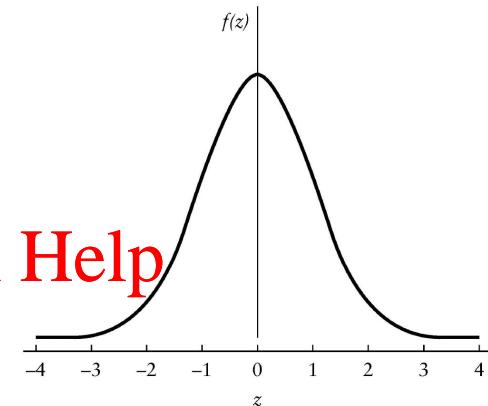


FIGURE 7.8

The standard normal curve for which $\mu = 0$ and $\sigma = 1$

Add WeChat powcoder

- We can also use the scale() function on R

Discrete Process		Continuous Process
Binomial	Poisson	Normal
Dichotomous Variables (Married/Not Married, Heads/Tails, Alive/Dead)	In frequent Events	Equal chance around Central Tendency
Fixed number of trials Outcomes of the trials are independent Probability of outcome is constant	Rate information (per hr, per day, per yr) Independent events λ - expectation of how many events occur $\lambda = np$	mean (μ) is location parameter sd (σ) is width parameter
<pre>> # probability of 1 person not having health insurance > dbinom(1, 10, 0.085) [1] 0.3821279</pre>	<pre>> # probability of 1 case of tetanus reported > lambda = 4.5 > poiss_vals = dpois(1,lambda) > poiss_vals [1] 0.04999048 > # yields 0.05</pre>	Binomial distribution with a large number of trials (or a large λ in Poisson) becomes normal

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

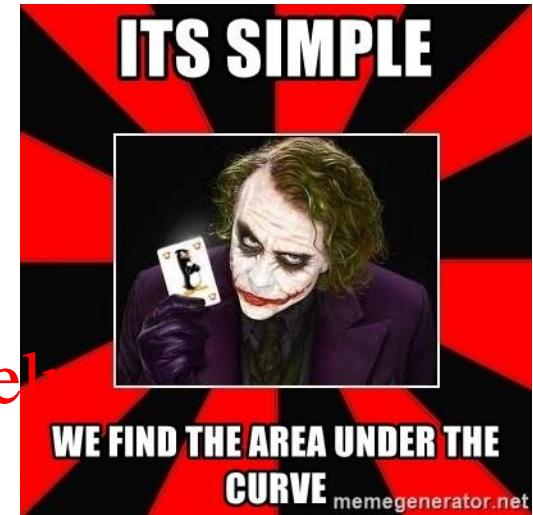
Lecture 7: Area Under the Curve

Assignment Project Exam Help

Masters will...

<https://powcoder.com>

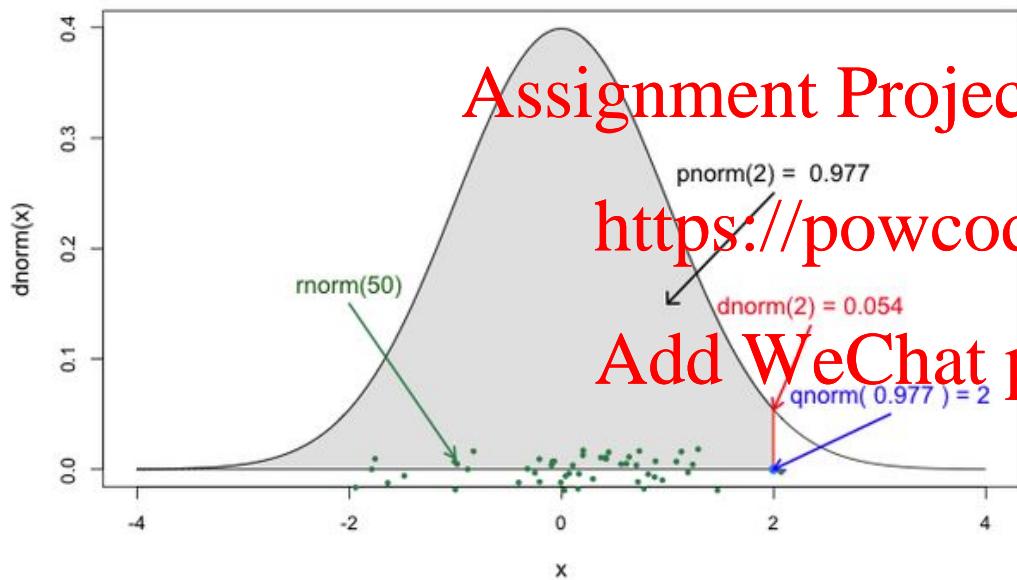
- Translate conventional probability curve tables into R code
- Calculate the area under the curve for a canonical distribution
- Obtain an index value for a canonical distribution, given the area



Two things to keep strait

- **Canonical probability distributions**
 - Normal: continuous variables with no behavioral or environmental constraints
 - E.g. Height, weight
 - Binomial: variables with dichotomous outcomes
 - E.g. results of coin flip, height if classified as tall/short
 - Poisson: variable measured with a rate
 - E.g. # of people who buy coffee during a certain time period
- **P/Q/D/R family functions**
 - Add WeChat powcoder
 - p [...] takes a value and tells you the cumulative probability
 - q [...] takes a quantile (aka cumulative probability) and tells you the corresponding value
 - d [...] takes a value and tells you the instantaneous probability
 - r [...] generates random numbers according to the given distribution

Visual Aid



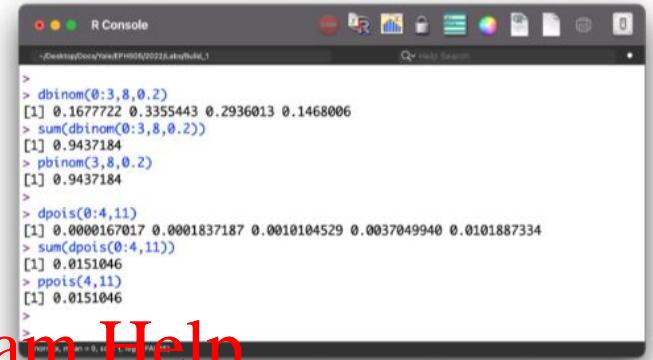
- $\text{pnorm}(2, \text{mean}=0, \text{sd}=1)$
 - Probability of getting 2 or less (entire area under curve to the left of 2)
- $\text{dnorm}(2, \text{mean}=0, \text{sd}=1)$
 - Probability of getting exactly 2
- $\text{qnorm}(0.977, \text{mean}=0, \text{sd}=1)$
 - What value is at the 97.7th percentile?

Borrowed this [image](#) from Prof. Jack Weiss @ UNC

How are functions related?

- `d[...]` gives you the *instantaneous* probability
- `p[...]` gives you the *cumulative* probability
 - Summing instantaneous probabilities gives you the cumulative probability
 - `sum(dbinom(0:2,2,.4))` <https://powcoder.com>
 - BEWARE: this not the case for continuous variables, so your best bet is to always use `p[...]` when looking for the cumulative probability
- `q[...]` takes a cumulative probability and gives you the value that corresponds with it
 - `p[...]` takes a value and gives you the cumulative probability that corresponds with it

Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder



```
> dbinom(0:3,8,0.2)
[1] 0.1677722 0.3355443 0.2936013 0.1468006
> sum(dbinom(0:3,8,0.2))
[1] 0.9437184
> pbinom(3,8,0.2)
[1] 0.9437184
>
> dpois(0:4,11)
[1] 0.0000167017 0.0001837187 0.0010104529 0.0037049940 0.0101887334
> sum(dpois(0:4,11))
[1] 0.0151046
> ppois(4,11)
[1] 0.0151046
>
>
```



```
>
>
> pnorm(0.4)
[1] 0.6554217
> qnorm(0.6554217)
[1] 0.3999999
>
> ppois(2,5)
[1] 0.124652
> qpois(0.124652,5)
[1] 2
>
> pbinom(3,10,0.1)
[1] 0.9872048
> qbinom(0.9872048,10,0.1)
[1] 3
>
```

SPOILER ALERT!

Masters will...

- Calculate the area under the curve for a **Cartonial** distribution
 - **p[...]** family function
 - `pnorm(value, mean=, sd=)`
 - `pbinom(# of successes, # of trials, probability of success)`
 - `ppois(observed value, lambda aka expected value)`
- Obtain an index value for a **Cartonial** distribution, given the area
 - **q[...]** family function
 - `qnorm(quantile, mean=, sd=)`
 - `qbinom(quantile, # of trials, probability of success)`
 - `qpois(quantile, lambda=)`

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat **powcoder**

Translating tables into R (Standard Normal)

TABLE A.3

Areas in the upper tail of the standard normal distribution

<i>z</i>	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
0.0	0.500	0.496	0.492	0.488	0.484	0.480	0.476	0.472
0.1	0.460	0.456	0.452	0.448	0.444	0.440	0.436	0.433
0.2	0.421	0.417	0.413	0.409	0.405	0.401	0.397	0.394
0.3	0.382	0.378	0.374	0.371	0.367	0.363	0.359	0.356
0.4	0.345	0.341	0.337	0.334	0.330	0.326	0.323	0.319
0.5	0.309	0.305	0.302	0.298	0.295	0.291	0.288	0.284
0.6	0.274	0.271	0.268	0.264	0.261	0.258	0.255	0.251
0.7	0.242	0.239	0.236	0.233	0.230	0.227	0.224	0.221
0.8	0.212	0.209	0.206	0.203	0.200	0.198	0.195	0.192
0.9	0.184	0.181	0.179	0.176	0.174	0.171	0.168	0.166
1.0	0.159	0.156	0.154	0.152	0.149	0.147	0.145	0.142
1.1	0.136	0.133	0.131	0.129	0.127	0.125	0.123	0.121
1.2	0.115	0.113	0.111	0.109	0.107	0.106	0.104	0.102
1.3	0.097	0.095	0.093	0.092	0.090	0.089	0.087	0.085
1.4	0.081	0.079	0.078	0.076	0.075	0.074	0.072	0.071
1.5	0.067	0.066	0.064	0.063	0.062	0.061	0.059	0.058
1.6	0.055	0.054	0.053	0.052	0.051	0.049	0.048	0.047
1.7	0.045	0.044	0.043	0.042	0.041	0.040	0.039	0.038
1.8	0.036	0.035	0.034	0.034	0.033	0.032	0.031	0.031
1.9	0.029	0.028	0.027	0.027	0.026	0.026	0.025	0.024

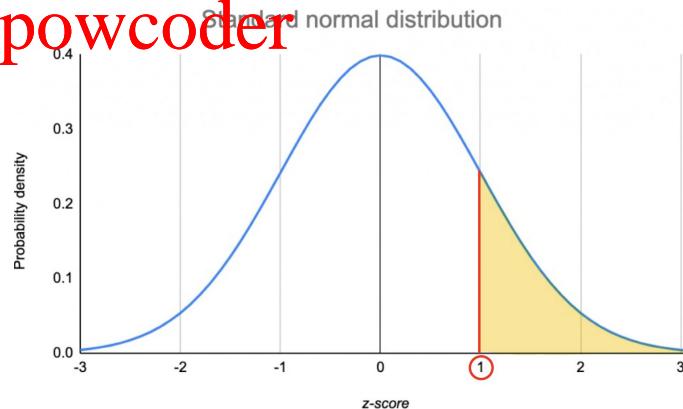
- Standard normal

- mean=0
- sd=1

Reporting the upper tail's cumulative probability

> pnorm(1, lower.tail=FALSE)
[1] 0.1586553

Assignment Project Exam Help
<https://powcoder.com>
Add WeChat powcoder



Translating tables into R (Poisson)

TABLE A.2
Poisson probabilities

k	0.5	1.0	1.5	2.0	2.5	3.0	3.5
0	0.6065	0.3679	0.2231	0.1355	0.0821	0.0498	0.0302
1	0.3033	0.3679	0.3347	0.2707	0.2052	0.1494	0.1057
2	0.0758	0.1839	0.2510	0.2707	0.2565	0.2240	0.1850
3	0.0126	0.0613	0.1255	0.1804	0.2138	0.2240	0.2158
4	0.0016	0.0153	0.0471	0.0902	0.1336	0.1680	0.1888
5	0.0002	0.0031	0.0141	0.0361	0.0668	0.1008	0.1322
6	0.0000	0.0005	0.0035	0.0120	0.0278	0.0504	0.0771
7	0.0000	0.0001	0.0008	0.0034	0.0099	0.0216	0.0385
8	0.0000	0.0000	0.0001	0.0009	0.0031	0.0081	0.0169
9	0.0000	0.0000	0.0000	0.0002	0.0009	0.0027	0.0066
10	0.0000	0.0000	0.0000	0.0000	0.0002	0.0008	0.0023

- $k = \# \text{ observed}$
- $\mu = \# \text{ expected}$
- Reporting the instantaneous probability

<https://powcoder.com>

Add WeChat powcoder

dpois(7,3)
[1] 0.02160403

Translating tables into R (Binomial)

TABLE A.1
(continued)

<i>n</i>	<i>k</i>	.05	.10	.20	.50
2	.0515	.1488	.2376	.2936	.3115
3	.0054	.0331	.0839	.1468	.2076
4	.0004	.0046	.0185	.0565	.0865
5	.0000	.0004	.0026	.0092	.0231
6	.0000	.0000	.0002	.0011	.0038
7	.0000	.0000	.0000	.0001	.0004
8	.0000	.0000	.0000	.0000	.0000
9	0	.6302	.3874	.2316	.1342
	1	.2985	.3874	.3679	.3020
	2	.0629	.1722	.2597	.3020
	3	.0077	.0446	.1069	.1762
	4	.0006	.0074	.0283	.0661

Assignment Project Exam Help

<https://powcoder.com>
Add WeChat powcoder

- n = # of trials
- k = # of successes
- Numbers across top = probability of success
- Reporting the instantaneous probability

```
> dbinom(3,9,.2)
[1] 0.1761608
>
```

Let's practice!

I know that the average height for a 10 year old boy is in the US is 54" and the standard deviation is 2.6". I have a patient who is very short for his age, but I only need to be concerned if he is at or below the 10th percentile.

What's the maximum height I would be concerned about?

Assignment Project Exam Help

Add WeChat powcoder

<https://powcoder.com>

Let's practice!

I know that the average height for a 10 year old boy is in the US is 54" and the standard deviation is 2.6". I have a patient who is very short for his age, but I only need to be concerned if he is at or below the 10th percentile.

What's the maximum height I would be concerned about?

```
> qnorm(0.1, mean=54, sd=2.6)
[1] 50.66797
```

Add WeChat powcoder

Let's practice!

I have another 10 year old male patient who is 53" tall. What are the chances of a 10 year old boy being as short, or shorter, than my patient?

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Let's practice!

I have another 10 year old male patient who is 53" tall. What are the chances of a 10 year old boy being as short, or shorter, than my patient?

Assignment Project Exam Help

```
> pnorm(53, mean=54, sd=2.6)
[1] 0.3502612
```

Add WeChat powcoder

Let's practice!

We had a picnic last week. Unfortunately, we found out afterwards that a batch of cookies were contaminated with something that causes illness in 75% of people who consume it. What are the chances that 4 of the 8 attendees who ate cookies will get sick?

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Let's practice!

We had a picnic last week. Unfortunately, we found out afterwards that a batch of cookies were contaminated with something that causes illness in 75% of people who consume it. What are the chances that 4 of the 8 attendees who ate cookies will get sick?

Assignment Project Exam Help

<https://powcoder.com>

> `dbinom(4, size=8, prob=0.75)`
[1] 0.08651733

Add WeChat powcoder

Let's practice!

The ICU at YNHH typically has 72 people admitted per day. Today, only 50 people were admitted. What are the chances of seeing 50 or fewer people be admitted?

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Let's practice!

The ICU at YNHH typically has 72 people admitted per day. Today, only 50 people were admitted. What are the chances of seeing 50 or fewer people be admitted?

Assignment Project Exam Help

<https://powcoder.com>

```
> ppois(50, lambda=72)  
[1] 0.00393046
```

Add WeChat powcoder

Let's practice!

What about the probability of *exactly* 50 people being admitted?

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Let's practice!

What about the probability of *exactly* 50 people being admitted?

Assignment Project Exam Help

<https://powcoder.com>

```
> dpois(50, lambda=72)  
[1] 0.00130121
```

Add WeChat powcoder

r[...] and set.seed

```
> set.seed(11590)
> rnorm(10)
[1]  1.0703727  1.6045851  0.2293789  0.1194001  1.8804929  1.4644473  1.3002455  -0.7975833 -1.0425974  0.9148733
>
> rnorm(10)
[1]  1.1937004  0.9176466 -1.5553064 -1.3183972 -0.2377616  1.2919886 -0.0295214  1.0410222 -0.4632907  1.1141515
>
> set.seed(11590)
> rnorm(10)
[1]  1.0703727  1.6045851  0.2293785  0.1669406  1.8804929  1.4644473  1.3002455 -0.7975833 -1.0425974  0.9148733
>
>
> set.seed(06511)
> runif(10)
[1] 0.7250691 0.8705999 0.6388562 0.5820337 0.9797588 0.8695976 0.6831646 0.7267082 0.4364206 0.5922392
>
> runif(10)
[1] 0.27186783 0.02425375 0.25697566 0.48761221 0.77959453 0.08130511 0.38432013 0.21832387 0.16653665 0.71729859
>
> set.seed(06511)
> runif(10)
[1] 0.7250691 0.8705999 0.6388562 0.5820337 0.9797588 0.8695976 0.6831646 0.7267082 0.4364206 0.5922392
>
```

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help
Good Luck!
<https://powcoder.com>

We're all rooting for you :)