**Chapter 8 Statistics – (Study) Formula Sheet**

**8.1 – Collecting Data**

**Chart, scatter chart, box and whisker chart

Description automatically generated**Graphical user interface

Description automatically generated

Sampling techniques

**8.2 – Displaying Data**

Frequency Tables

* Summarize datasets by counting the number of observations for each category, distinct value or interval.

Table

Description automatically generated

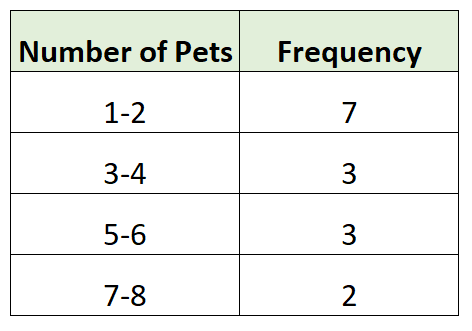
Examples

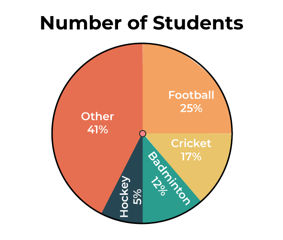
a) What percent of observations have between 1 and 4 pets inclusive?

b) What percent of students prefer Football and Hockey?

c) Bar graph – Which season has the highest frequency?

d) Histogram – How many items cost between $11 and $40 inclusive?





Graphical Displays of Data

* Pie charts (categorical data)
  + Compare parts to a whole

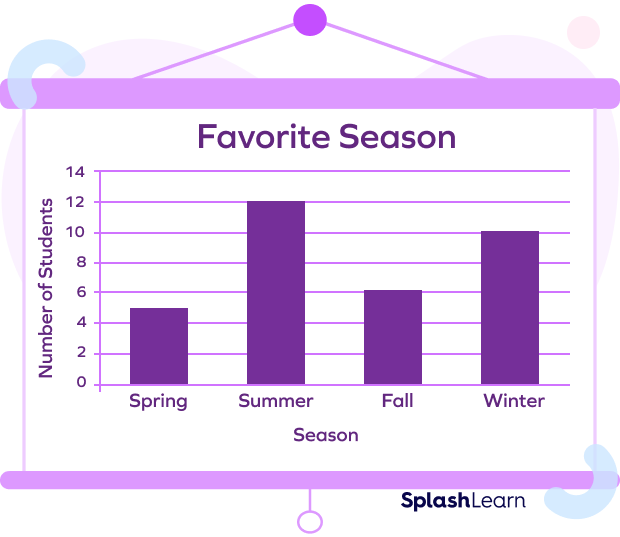
(slices are proportion of a category).

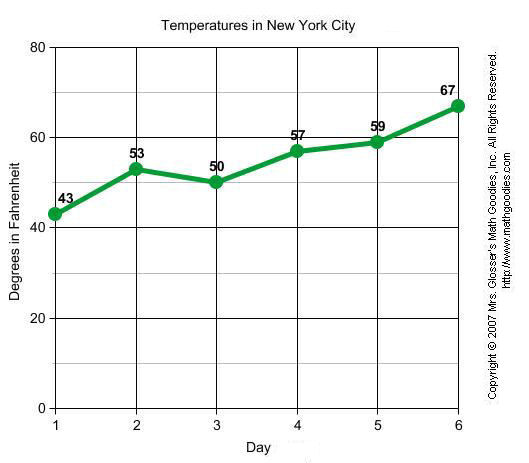
* Bar graphs (categorical data) and Histograms (numeric data)
  + Height of bar represents amount of data

in each category (counts or relative frequencies).

Chart, histogram

Description automatically generated



* Line graph

e) How many days was the temperature between 55 and 60 °F?

* + Shows changes in a numerical

variable over time.

**8.3 – Describing and Analyzing Data**

Example

Dataset: 1, 2, 7, 3, 6, 9, 1, 0, 4, 7

a) Find the mean. **Calc: 1-Var Stats**

(Data in L­1)

b) Find the median.

c) Find the mode.

d) Find the range.

e) Find the sample standard deviation.

Measures of Center

* **Mean** (average) =
  + NOT resistant 🡪 Affected by outliers
* **Median** (middle)
  + The middle value in an ordered list.
  + Resistant 🡪 NOT affected by outliers.
* **Mode** (most common)
  + The most frequently occurring value(s).
  + Resistant 🡪 NOT affected by outliers.
  + Only measure of center that can be used with categorical data.

Measures of Spread

* **Range** = Max – Min
* **Standard deviation**
  + Measures average distance from the mean.
  + (Don’t calculate by hand).

Chart, histogram

Description automatically generatedEmpirical Rule (68 – 95 – 99.7 Rule)

68% of the data lies within 1 st dev of the mean.

95% of the data lies within 2 st devs of the mean.

99.7% of the data lies within 3 st devs of the mean.

* Finding probabilities using the Empirical Rule.

Example

Oak tree heights are normally distributed with

mean 60 m and st dev 5 m.

a) Find the percent of trees between 50 m and 70 m tall.

b) Find the percent of trees greater than 65 m

* + Step 1 🡪 **Draw** and **label** curve.
  + Step 2 🡪 **Shade** curve.
  + Step 3 🡪 **Use empirical rule**.

A picture containing text, lamp

Description automatically generated

A picture containing text, lamp

Description automatically generated

**8.4 – The Normal Distribution**

Diagram

Description automatically generatedFinding probabilities based on the normal distribution

* Step 1 🡪 **Standardize** using the **z-score**.
  + Formula:
  + Ex) *X* has a normal distribution with mean 10 and st dev 2.

Find the z-score for .

Diagram

Description automatically generated with medium confidence

* Step 2 🡪 **Draw**, **label** and **shade** curve.
  + This is how you show your work!!!
* Step 3 🡪 Use ‘**Standard Normal Distribution’** table to find the probability for Z.
  + Table ALWAYS gives probability LESS THAN Z: P(Z < z).

A picture containing text, lamp

Description automatically generated

* + Examples
  + Left probability = TABLE (Directly)

A picture containing text, lamp

Description automatically generated

* + Right probability = 1 - LEFT

* + A picture containing text, lamp

    Description automatically generatedBetween probability = LEFT Z2 – LEFT Z1

Calendar

Description automatically generated**8.5 – Linear Regression**

Scatterplots:

* **Form**: Linear, lurved, or random scatter
* **Direction**: Positive, negative or no association
* Timeline

  Description automatically generated**Strength**: Weak, moderate or strong

Correlation (*r*):

* Interpreting correlation (LINEAR)
  + Sign = Direction
  + Absolute value |*r*| = Strength

Example

Dataset:

a) Calculate the correlation *r*.

b) Determine if *r* is significant

for α = 0.01.

c) Suppose we have different regression equation where

.

Predict *Y* for *X = 3*:

* Calculate using calculator
  + **LinReg(ax+b) or 2-Var Stats**
  + *L1 = X, L2 = Y*

Regression:

Table

Description automatically generated with medium confidence

* Step 1 🡪 Determine if there is a **significant correlation**

(**linear relationship**).

* + Compare |*r*| and Critical Value (CV)

for *n* (sample size) and significance level α.

* + If |*r*| > CV 🡪 statistically significant.
* Step 2 🡪 Once we have a significant correlation,

we can find the **regression line**.

* + (get results from correlation calculation)

*= slope x + intercept*

Chart, scatter chart

Description automatically generated

* Step 3 🡪 Make **predictions** using the regression line.
  + Just plug in the new *X* value to our equation

and this will give us the predicted *Y.*