– Excel —

spreadsheet program

using for statistics

First of all, for algebra equations:

↓ means multiply

=(2*3/4) this is actually $\frac{2 \times 3}{4}$

↑ divide

↑ use parenthesis to tell Excel the formula

↑ put equal sign at beginning

other uses

↓ square root

=sqrt(2*3/4)

or another

=((2*3/4)^2)

↑ squared

to name the sheet you are working on (right click) at bottom where it says Sheet #1 etc.

2 Kinds of Statistics

① Descriptive = to describe data

② Inferential = to infer, forecast, make predictions (based on data)

Now for statistics uses with "Data Analysis", which is called an "Add in"

add - also can use insert

↓

Function

↓

then

methodology

↙ we will use this later.

Add Ins

↓

Analysis Tool pak

↓

then for Histogram

Tools

↓

Data Analysis

↓

Histogram

↓

To Graph

Data for the two columns

① click & select "highlight" then w/ mouse highlight data with the mouse (include labels

input range

Bin range

② OK

☑ labels

output range (will put graph on the same sheet)

click OK

To get rid of spaces on the histogram ....

click on the bar then options

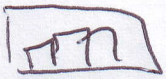
left click, then right click, format data series.

gap width 0

↑ for no spaces

Can also select cumulative frequency.

☑ cumulative (what % each Bin range is, the second data collumn)

Also pie charts  on tool bar

ex: line graphs to show the ~~trend~~ over time.

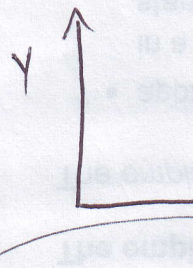
For # of decimals "Format cells"

decimal places

---

ex: Correlation

price	size
✓	✗
✓	✗
✓	✗
✓	✗
✓	✗
✓	✗

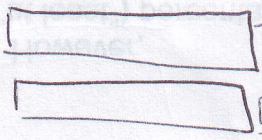


Format Axis ↙ to set range (click on y axis)

options → "show formula"

Tools → data Analysis → Regression

select areas



① click

② click and select

Xy scatter plot

☑ line fit plots

click on a dot then, right click, add trend line

Excel

Probability

multiply (and)
add + (or)

ex: dice roll

chance for two 4's

4 and 4

so

$4 \times 4 = 16$

so

$\frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$ probability

or

$0.0625 = 6.25\%$

or

How about -

the probability of rolling a 1 or 2 or 3

$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2} = 50\%$

Discrete probability = whole #'s ex: 1, 2, 3, 4

Continuous = decimals, fractions

0.02

$\frac{1}{2}$

$X^0 = 1$ anything to the power of zero is 1.

$X \wedge Y = X^Y$ ← exponent

$_nC_x = \frac{n!}{x!(n-x)!}$

↑ Combinations formula (How many different ways)

! means factorial (ex: $3! = 3 \times 2 \times 1$)

n = how many observations

Insert Function Statistical Binomdstr

Binomial, "probability of something"

$$P(X) = {}_nC_x \, \pi^x (1-\pi)^{n-x}$$

True if cumulative False if not cumulative.

n = trials (ex: # of questions) x = random variable π = prob. of success

Excel

"we sample because it is impossible to study every item in a population"

Confidence interval

ex: 90% of the time we are confident the population mean will be within . . .

$t <$ 30 samples

$z >$ 30 samples

For z

$$t = \frac{\bar{x} - M}{s/\sqrt{n}}$$

$\bar{x}$ = sample mean; M = "mu" = the population average or pop. mean; t = for test statistic; for tests ✱

for 95% Confidence $\bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$

99% Confidence $\bar{x} \pm 2.58 \frac{s}{\sqrt{n}}$

↑ from Appendix

"central limit theorem"
95% is 1.96 std. dev from $\bar{x}$ mean
99% is 2.58 std dev from $\bar{x}$.

For T

$$\bar{x} \pm t \frac{s}{\sqrt{n}}$$

↑ for t go to Appendix using degrees of freedom and % confidence.

---

for deciding on level of significance for tests. (α)

.10 is used in begining stage
.05 is used for more accuracy
.01 when ready to rollout product.

P value is "the measure of the strength of the test"
If close to α then not so sure.

Excel

* Alternate Hypothesis shows which way > or < the critical value is place (which side of the curve)

(no difference) → $=$ or $\geq$ one tailed (one side)

$\neq$ or $<$ two tailed (two sides)

(does make a difference) ↑

The equal sign is always in H_0. (null)

For ANOVA - "A filter" "How effective is the independent variable?"

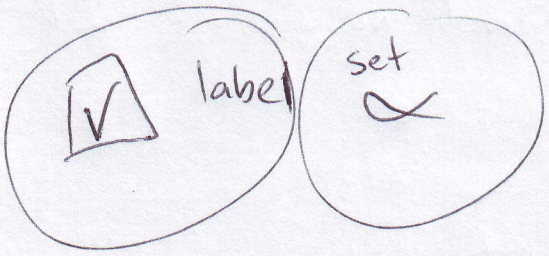
"Splits the source of variation into two parts"

tools

Data Analysis

ANOVA (when one independent variable)
single factor

input all the data



Week 4

MBA-510

ANOVA

chap 12

analysis of variance. (always two tailed)

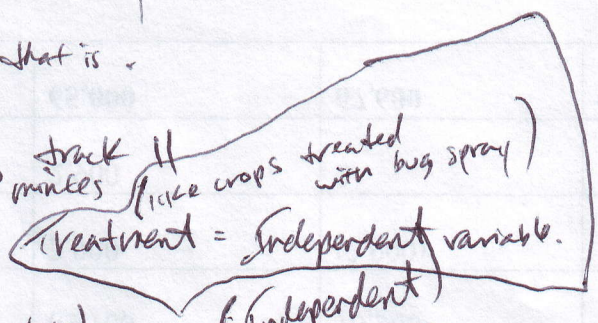
(something that is) Independent and	Dependent variable	only equal or not.
(control) weight of car education	gas mileage. (controlled by independent) wages	
↑ are these related. ↑		

H_0: = (no difference)

H_1: ≠ (makes a difference)

dep. variable that is.

eg:



pg 401 Chap 12 #8

Starting 13 track minutes

(Independent)

Treatment 1 (AKA 0 hrs/wk)	Treatment 2 (1 hr/wk)	Treatment 3 (4 hrs/wk)
9	13	10
7	20	9
11	14	15
9	13	14
12		15
10		

↙ dependent variable.

$H_0: M_1 = M_2$ etc

$H_1: M_1 \neq M_2 \neq_3$

↑ what makes a difference

Tools

Add ins

Analysis tool pack

Data Analysis →

Anova: Single factor (one independent variable)

MBA-510 Week 4 7A

ANOVA
— is like a coffee filter

Input all data in table
(include label)

✓ label in first row.

α 0.05

New worksheet

"make sense of data"

OK

Std dev. = $\sqrt{\text{variance}}$.

click sheet if need to go back

At this point just descriptive.

levels are 3 how many different treatments

ANOVA

SS (sum of square) variation in dep variable. df MS (mean square) $\frac{SS}{df}$

Source of variation			
Between Groups	70.4	(3-1) 2	35.2
Within Groups (dep. variable)	82.5333	14-2 12 (all else)	6.87778

Continue.

Continue.

numerator vs denominator F stat = $\frac{MS_{Between}}{MS_{within}}$ ratio = $\frac{35.2}{6.87}$ ← if numerator

so large numerator, independent variable makes a difference.

> 1

< 1 larger denominator. dep. variable

ex: exercise

O weight loss

or O ← that exercise makes just a bit of difference

$H_0: M_1 = M_2 = M_3$

$H_1: M \neq M_2 \neq M_3$ ← makes a difference

Week 4

MBA-510

7B

Are we crossing the critical value — ?

pg 723

Std. dev. cannot be negative.

F table pg 723 the other page is .01

Appendix DF (numerator)

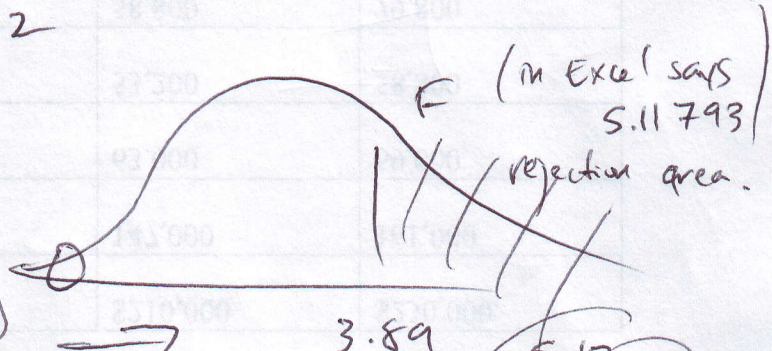
α .05

(Between groups)

Numerator df | 2

denom. df | 12

DF denom. | 3.89



Excel gives up the critical value

if larger than F stat reject H_0

F crit

Excel also calculates P value to compare with α

So reject the null H_0

whether Independent variable effects the dependant.

positive or negative correlation

larger

SS	MS
SST	MST
SSE	MSE

"error means noise"

something that you don't want. want error to be small compared to Independent

Regression

Excel

Does the data correlate?
Is it a positive or negative correlation?
How strong is the correlation?

"relationship" between independent and dependent variables.

$H_0 : M_1 = M_2 = M_3$ (no difference)

$H_1 : M_1 \neq M_2 \neq M_3$ (significant difference)

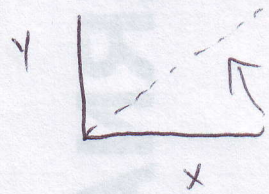
Correlation Coefficient -1 or $+1$ (strong correlation)

$\varnothing$ (insignificant)

$-.5$ $+.5$ (medium correlation)

if r is high ex: 0.95 (positive correlation)

Excel insert → chart → xy scatter → select data



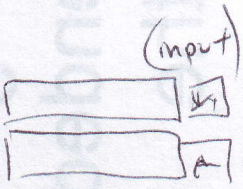
"line of best fit"

For Excel (has to be)

X (is on left) | Y (on right)

(for the program to recognize)

Insert → function → Statistical → Correl



☑ labels

☑ line fit plots

- New worksheet
  or
- output (click on square)

r^2 is coefficient of determination ex: 0.137 = 14 %

Regression

for trend line

① right click line ② left click ③ add trendline [options] display equation

Dep. variable

ex: $y = 0.0703x + 64.793$

change in y for 1 unit change in x — Slope — co-efficient

+ positive corr.

− neg. corr

y intercept — where line intercepts y at $x = \emptyset$

Can use the formula to find the result for a new variable input.

Can only use equation if the correlation co-efficient is high.

r = co-efficient of correlation

r^2 = co-efficient of determination

Excel Summary does not show + or − in co-efficient.

always a + number is shown.

when highlighting, include only the cells with #'s

For formulas ex: y' "estimate" plugging variable into equation

put formula w/ #'s

do one, then

① click on x and drag to bottom of column

② or click in cell and drag to bottom

for 1st on $y' = 0.0703x + 64.793$

plug in each variable by dragging

Standard error of estimate is how far observed values are around the line of regression.

$$S_{y \cdot x} = \sqrt{\frac{\sum (y - y')^2}{n - 2}}$$