Statistical testing II

Terms so far…

- Mean
- **Sum-of-Squares**
- **Variance**
- Standard Deviation
- Standard Error of the Mean
- 95% confidence interva

Tests

- one-way t-test
- between group t-test
- within group t-test

- histograms
- normal distribution
- empirical rule
- sampling distribution
- z-score
- t-score
- degrees of freedom (df)
- one-tailed & two-tailed
- Type 1 & Type 2 errors
- alpha (α) and power ($1-\beta$)

Overview

1. One-way ANOVA
   - determines differences between means of two or more levels of an independent variable
   - effect size (η^2, f)
   - post-hoc comparisons
   - contrast analysis (alternative?)

2. Two-way ANOVA
   - determines differences between means for the independent effects and interactions of two variables

Testing Knowledge (ANOVA):

- True/False: Factor is another term for independent or treatment variable
- You conduct an experiment where the independent variable has five levels. A one-way ANOVA reveals a significant F-ratio. Can you tell which level of IV is different from which other level?
- True/False: A post hoc test can be undertaken when an ANOVA's F ratio is not significant

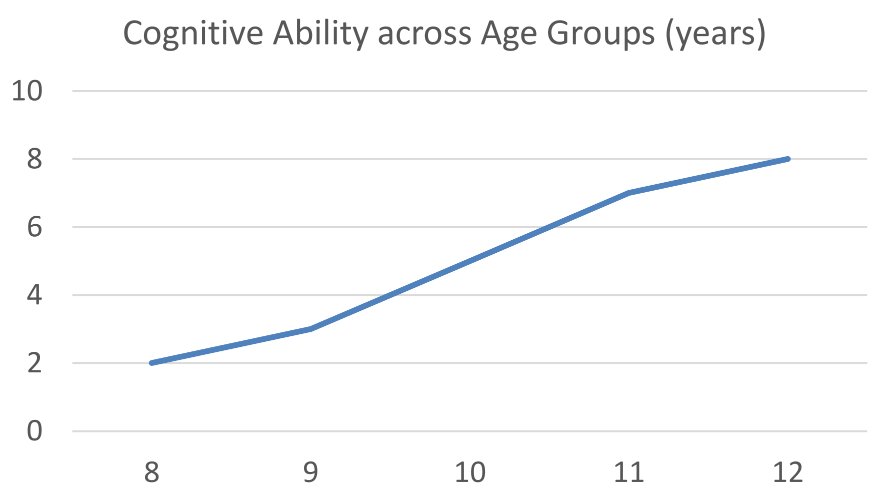
Testing Knowledge (ANOVA):

- True/False: Factor is another term for independent or treatment variable (Ans: True)
- An experiment has one independent variable with five levels. A one-way ANOVA reveals a significant F-ratio. Can you tell which level of IV is different from which other level? (Ans: No)
- True/False: A post hoc test can be undertaken when an ANOVA's F ratio is not significant

Testing Knowledge (ANOVA):

- True/False: Factor is another term for independent or treatment variable (Ans: True)
- An experiment has one independent variable with five levels. A one-way ANOVA reveals a significant F-ratio. Can you tell which level of IV is different from which other level? (Ans: No)
- True/False: A post hoc test can be undertaken when an ANOVA's F ratio is not significant. (Ans: False)

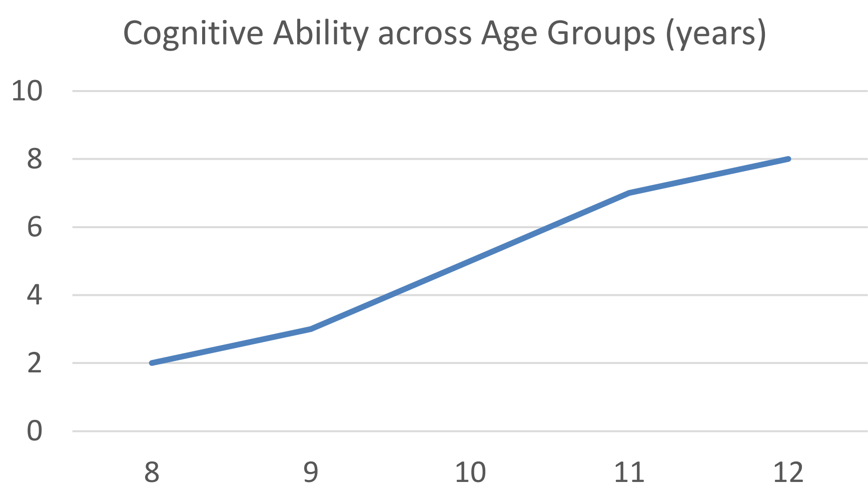
Contrast Analysis



Source	SS	df	MS	F	p
Age levels	260	4	65	1.03	0.40
Within-groups	2835	45	63		

- 50 children are tested for their cognitive ability across 5 age groups (N=10)
- IV? DV?
- $F(4,45)=1.03$, $p=0.4$)
- There is no significant main effect of age.
- But... there is a linear trend.
- Such data are better analyzed with contrast analysis.

Contrast Analysis



Age	8	9	10	11	12	∑
Cog. Ability	2	3	5	7	8	25
T(nX)	20	30	50	70	80	250
λ	-2	-1	0	+1	+2	0
Tλ	-40	-30	0	70	160	160

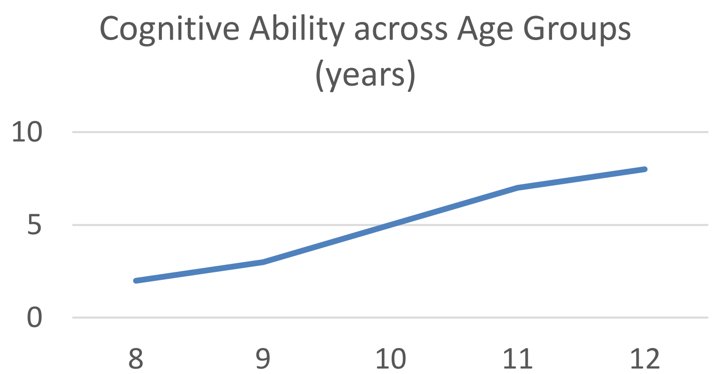
- $MS_{contrast} = SS_{contrast}$
  $= L^2/(n\sum\lambda^2)$

- $L = \sum(T\lambda)$
  $= T_1\lambda_1 + T_2\lambda_2 + T_3\lambda_3 \ldots T_k\lambda_k$

  where,
  k = number of conditions
  n = number of observations within each level
  T = nX
  λ = contrast weights linked to a hypothesis, must sum to 0.

- E.g., $L = 160$
  $\sum\lambda^2 = (-2)^2+(-1)^2+(0)^2+(1)^2+(2)^2$
  $= 10$
  $MS_{contrast} = (160^2)/(10)(10)$
  $= 256$
  $F_{contrast} = MS_{contrast} / MS_{within}$
  $= 256/63$
  $= 4.06$
  df = 1 (for contrast)
  $F_{critical} = 4.055$

Contrast Analysis



Report:
The mean performance scores appear to observe a linear trend, where ability increased systematically with age. A one-way analysis of variance (ANOVA) with age level as the independent variable did not reach significance, $F(4,45)=1.03$, p=ns. A planned contrast where performance was predicted to increase with age was significant, indicating that cognitive ability improved with age, $F(1,45)=4.06$, $p<0.05$.

- $MS_{contrast} = SS_{contrast}$
  $= L^2/(n\sum\lambda^2)$
- $L = \sum (T\lambda)$
  $= T_1\lambda_1 + T_2\lambda_2 + T_3\lambda_3 \ldots T_k\lambda_k$

  where,
  k = number of conditions
  n = number of observations within each level
  $T = nX$
  λ = contrast weights linked to a hypothesis, must sum to 0.
- E.g., $L = 160$
  $\sum\lambda^2 = (-2)^2+(-1)^2+(0)^2+(1)^2+(2)^2$
  $= 10$
  $MS_{contrast} = (160^2)/(10)(10)$
  $= 256$
  $F_{contrast} = MS_{contrast} / MS_{within}$
  $= 256/63$
  $= 4.06$
  df = 1 (for contrast)
  $F_{critical} = 4.055$

Two-way Analysis of Variance

2. Two-way ANOVA
   - determines differences between means of two or more independent variables

You investigate if game types have an influence on subjective anxiety. Your independent variables are game-type and age, with 2 levels each. Your dependent variable is a self-reported score on a questionnaire for anxiety. You have 4 experimental conditions and run 10 participants per condition. This is a between-group 2 x 2 factorial design.

Two-way Analysis of Variance

You investigate if game types have an influence on subjective anxiety. Your independent variables are game-type and age, with 2 levels each. Your dependent variable is a self-reported score on a questionnaire for anxiety. You have 4 experimental conditions and run 10 participants per condition. This is a between-group 2 x 2 factorial design for Age and Game Type.

		Factor B – Game Type	
		Puzzle	Shooter
Age	young	*Young people who played Tetris*	*Young people who played Doom*
	old	*Old people who played Tetris*	*Old people who played Doom*

Two-way Analysis of Variance

You investigate if game types have an influence on subjective anxiety. Your independent variables are game-type and age, with 2 levels each. Your dependent variable is a self-reported score on a questionnaire for anxiety. You have 4 experimental conditions and run 10 participants per condition. This is a between-group 2 x 3 factorial design for Age and Game Type.

		Factor B – Game Type		
		Puzzle	Shooter	RPG
Age	young	Young people who played Tetris	Young people who played Doom	Young people who played Final Fantasy
	old	Old people who played Tetris	Old people who played Doom	Old people who played Final Fantasy

Two-way Analysis of Variance

How many test conditions are there altogether?

How many participants were tested?

		Factor B – Game Type		
		Puzzle	Shooter	RPG
Age	young	*Young people who played Tetris*	*Young people who played Doom*	*Young people who played Final Fantasy*
	old	*Old people who played Tetris*	*Old people who played Doom*	*Old people who played Final Fantasy*

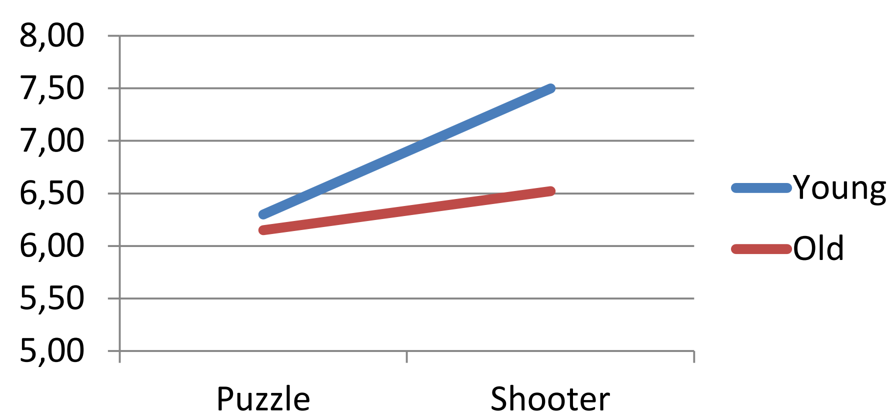
Two-way Analysis of Variance

		Game-Type: Puzzle	Game-Type: Shooter	Mean_i.
Age	21-30	5.70	7.30	
		5.50	7.90	
		6.90	7.10	
		8.00	8.10	
		6.00	8.50	
		6.70	7.70	
		7.10	7.40	
		5.90	7.80	
		6.50	6.20	
		4.70	7.00	
	Mean_1j	*6.30*	*7.50*	***6.90***
	51-60	5.40	5.10	
		5.30	7.80	
		7.30	5.90	
		7.10	9.10	
		5.90	5.20	
		6.30	6.10	
		6.20	7.10	
		6.90	5.30	
		6.30	6.20	
		4.80	7.40	
	Mean_2j	*6.15*	*6.52*	***6.34***
	Mean_.j	***6.23***	***7.01***	**6.62**

H0:

Main effects:
- there is no significant age difference
- there is no significant game difference

Interaction effect:
- there is no significant interaction of age and game

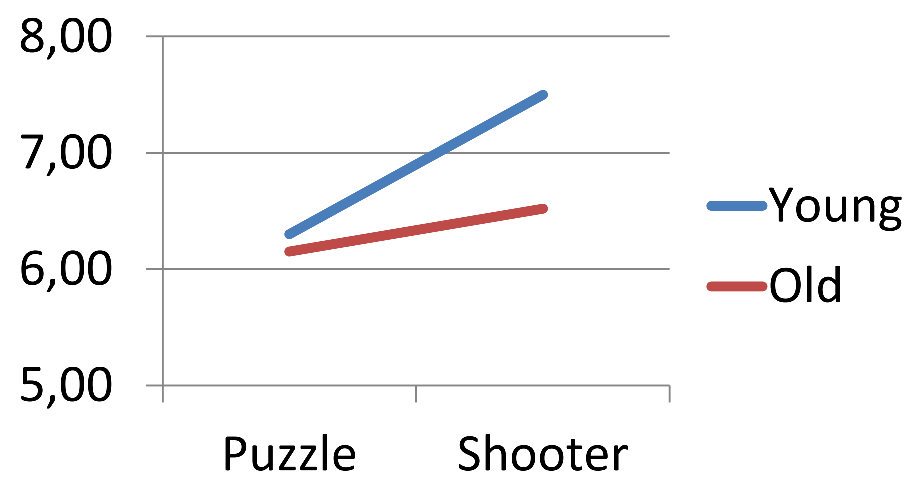


Two-way Analysis of Variance

Age		Game-Type: Puzzle	Shooter	Mean_i.
	21-30	5.70	7.30	
		5.50	7.90	
		6.90	7.10	
		8.00	8.10	
		6.00	8.50	
		6.70	7.70	
		7.10	7.40	
		5.90	7.80	
		6.50	6.20	
		4.70	7.00	
	Mean_1j	*6.30*	*7.50*	***6.90***
	51-60	5.40	5.10	
		5.30	7.80	
		7.30	5.90	
		7.10	9.10	
		5.90	5.20	
		6.30	6.10	
		6.20	7.10	
		6.90	5.30	
		6.30	6.20	
		4.80	7.40	
	Mean_2j	*6.15*	*6.52*	***6.34***
	Mean_.j	***6.23***	***7.01***	**6.62**

Sum-of-Squares:

- $SS_{Total} = SS_{cells} + SS_{error}$
- $SS_{cells} = SS_{age} + SS_{game} + SS_{age*game}$



Two-way Analysis of Variance

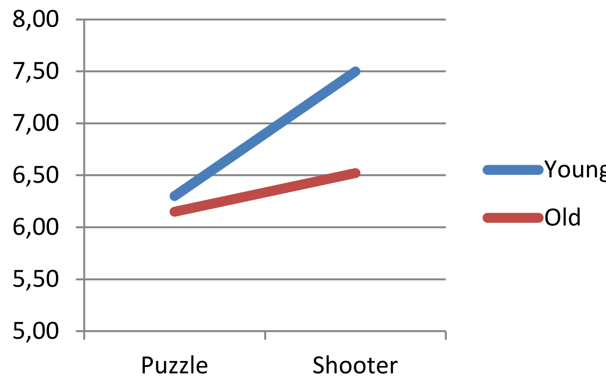
Source of Variation	**SS**	**df**	**MS**	**F**
Age	3.19	max(i)-1	SS_{Age}/df_{Age}	MS_{Age}/MS_{Error}
Game	6.16	max(j)-1	SS_{Game}/df_{Game}	MS_{Game}/MS_{Error}
Age x Game	1.72	df_{Age} x df_{Game}	$SS_{AgeGame}/df_{AgeGame}$	$MS_{AgeGame}/MS_{Error}$
Error	33.90	max(i) x max(j) x (n-1)	SS_{Error}/df_{Error}	
Total	44.98	N-1		

Statistical assumptions of two-way ANOVA

- A two-way ANOVA can analyzed data that is either an interval or ratio scale (i.e., means can be computed from data).
- The data within each sample are either (a) randomly or (b) independently sampled.
- The parent populations from which data samples are drawn:
  - *are normal.*
  - *have equal variances*
- The ANOVA design should be:
  - fully factorial (i.e., every combination is represented by distinct expt. conditions)
  - each cell should have the same number of participants.

Reporting Two-way Analysis of Variance

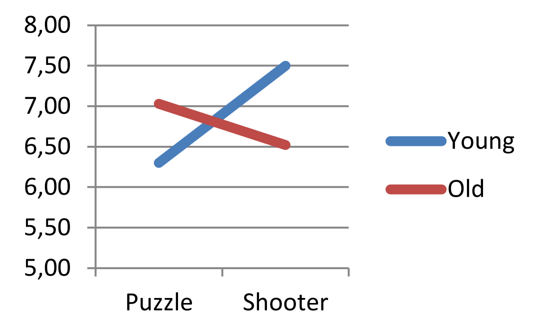
Source of Variation	SS	df	MS	F	η_p^2
Age	3.19	1.0	3.19	3.39	0.07
Game	6.16	1.0	6.16	6.54*	0.12
Age x Game	1.72	1.0	1.72	1.83	0.04
Error	33.90	36.0	0.94		
Total	44.98	39.0			



There is a significant main effect of Game-Type ($F(1,36)=6.54$, $p=0.01$, $\eta_p^2=0.12$). The effect size for this was moderate ($f=0.37$). Shooter games result in more stress than puzzle games. There is no significant main effect of Age ($F(1,36)=3.39$, $p=0.07$, $\eta_p^2=0.07$) and no significant interaction of Game-Type and Age ($F(1,36)=1.83$, $p=0.18$, $\eta_p^2=0.04$).

Reporting Two-way Analysis of Variance with significant interaction

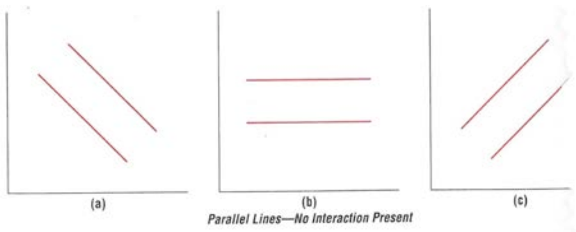
Source of Variation	SS	df	MS	F	η_p^2
Age	0.16	1.00	0.16	0.14	0.07
Game	1.19	1.00	1.19	1.04	0.12
Age x Game	7.31	1.00	7.31	6.41	0.04
Error	41.04	36.00	1.14		
Total	49.70	39.00			



There is no significant main effects of either Game-Type ($F(1,36)=0.14$, $p=0.71$, $\eta_p^2=0.00$). or Age ($F(1,36)=1.04$, $p=0.31$, $\eta_p^2=0.02$). There was a significant interaction of Game-Type and Age ($F(1,36)=6.41$, $p=0.02$, $\eta_p^2=0.13$). The effect size for this interaction was moderate ($f=0.38$). Young people are more stressed when playing shooter games than puzzle games. The opposite pattern is observed in old people.

Interpreting Interactions (x-axis: A; lines: B)

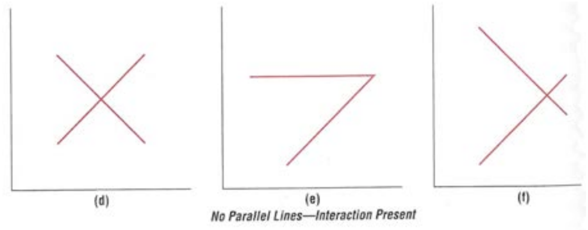
<u>No interaction:</u>
defined by parallel lines



<u>Interpretations (based on statistical analysis)</u>
a) Main effects (A and B)
b) Main effect of B, no main effect of A
c) Main effects (A and B)

Interpreting Interactions (x-axis: A; lines: B)

Significant Interaction:
defined by non-parallel lines

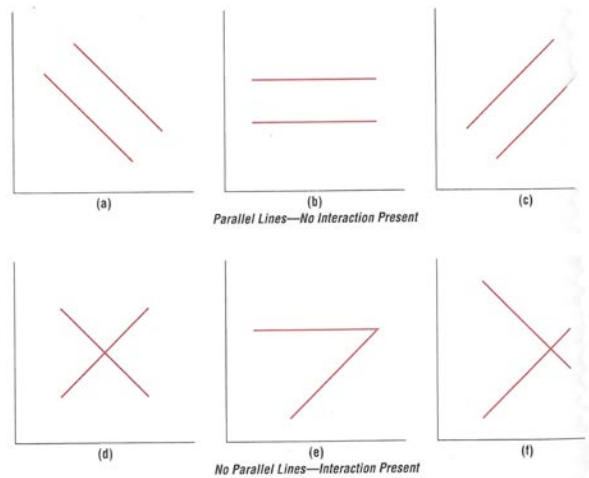


Interpretations (likely results of statistical analysis)
a) No main effect of A and B
b) Main effect of B, no main effect of A
c) Main effect of A, no main effect of B

If there is a significant interaction, interpret the interaction and ignore the main effects.

Interpreting Interactions (x-axis: A; lines: B)

If there is a significant interaction, interpret the interaction and ignore the main effects.

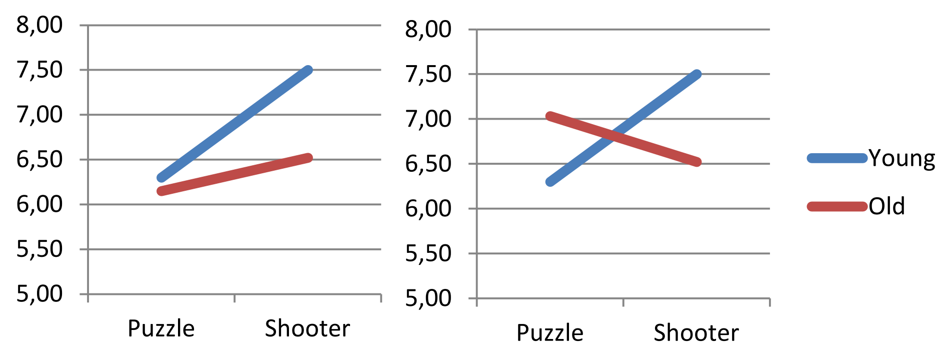


<u>No interaction:</u>
defined by parallel lines

<u>Interaction:</u>
defined by non-parallel lines

<u>Examples</u>
Only interpret an interaction if the analysis reports a significant interaction.

Left figure looks like a significant interaction even though it only has a main effect of Game



Statistical assumptions of two-way ANOVA

- A two-way ANOVA can analyzed data that is either an interval or ratio scale (i.e., means can be computed from data).
- The data within each sample are either (a) randomly or (b) independently sampled.
- The parent populations from which data samples are drawn:
  - *are normal.*
  - *have equal variances*
- The ANOVA design should be:
  - fully factorial (i.e., every combination is represented by distinct expt. conditions)
  - each cell should have the same number of participants.
- Each factor of a two-way ANOVA can have more than 2 levels
  - e.g., a 3 x 2 full factorial design for Game-Type (puzzle, shooter, role-playing) and Display-Resolution (high, low)

Repeated measures ANOVA

- repeated measures ANOVAs can be performed where all the levels of an IV is experienced by all the participants
  - **one-way; only one factor**
  - *multi-factorial; all factors are repeated measures*
  - *mixed-design; some factors are repeated measures*

  Reasons
  - fewer participants are required
  - reduce inter-subject variability; each participant serve as his/her own control group
  - allow designs with carry-over effects (e.g., effect of meditation over time)
- $SS_{total} = SS_{between} + SS_{subjects} + SS_{within}$
  - unlike a one-way between-groups ANOVA, we can now account for variance between participants (i.e., $SS_{subjects}$)

Repeated measures one-way ANOVA

You are interested in understanding whether the anxiety experienced in playing a shooter game changes with time. You test six participants. Each participant plays the same shooter game, once per week and they report their anxiety to you after playing the game. You measure them three times.

What is the IV?
What is the DV?
What is the null hypothesis, H0?
What is the alternative hypothesis, H_{alt}?

Repeated measures one-way ANOVA

H0: $\mu_1 = \mu_2 = \mu_{...} = \mu_j$;
H_{alt}: At least one mean is different from another mean

$SS_{total} = SS_{between} + SS_{subjects} + SS_{within}$

$$SS_{total} = \sum X_{ij}^2 - \frac{(\sum X_{ij})^2}{N}$$

- N=number of observations
- i=participant-index
- j=IV-index
- df = N-1

$$SS_{between} = \frac{(\sum X_j)^2}{n} - \frac{(\sum X_{ij})^2}{N}$$

- n = number of participants
- df = k-1
- k = number of IV levels

$$SS_{subjects} = \frac{(\sum X_i)^2}{k} - \frac{(\sum X_{ij})^2}{N}$$

- df =n-1

$$SS_{within} = \sum X_{ij}^2 - \frac{(\sum X_i)^2}{k} - \frac{(\sum X_j)^2}{n} + \frac{(\sum X_{ij})^2}{N}$$

(see Excel worksheet for example)

Repeated measures one-way ANOVA

	Week1	Week2	Week3
Mean	9.25	5.75	5.08

Source-Table	SS	df	MS	F	p	sig
Between	60.11	2.00	30.06	23.62	0.00	TRUE
Subjects	5.24	5.00	1.05	0.82		
Within	12.72	10.00	1.27			
Total	78.07	17.00				

HSD=1.79

There is a significant main effect of Time ($F(2,10)=23.6$, $p=0.00$, $\eta^2=0.00$). The effect size for this is large ($f=1.83$). Post hoc comparisons of means based on Tukey's HSD test indicated that participants reported high anxiety in the first week, compared to either of the subsequent weeks.

(see Excel worksheet for example)

ANCOVA

- One last twist
- Account for a-priori factors

- Covariate – „An IV with known effect”
  - Must be continuous
  - „predictor”
  - „risk factor”
- Use to account for „obvious” things

Example

My dog has a skin condition. I'd like to study this. I think it's a meat allergy.

Let's take 60 dogs and give them 50g, 100g and 150g meat protein a day for 4 weeks (between groups). We then measure the inflammation levels in the blood.

IV=?

DV=?

Example

My dog has a skin condition. I'd like to study this. I think it's a meat allergy.

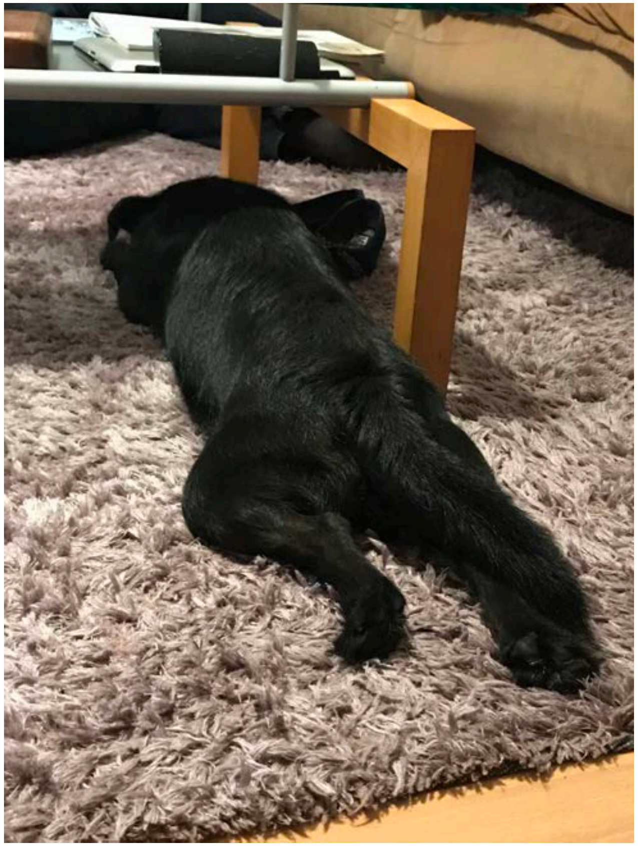
Let's take 60 dogs and give them 50g, 100g and 150g meat protein a day for 4 weeks (between groups). We then measure the inflammation levels in the blood.

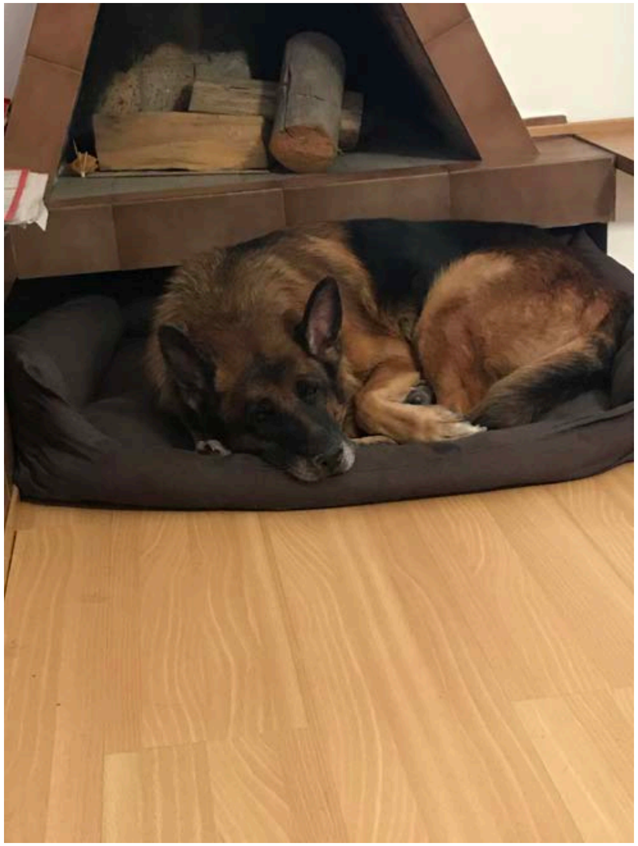
Inflammation ~ protein consumed

Wait!

https://www.eliteprodogtraining.com/wp-content/uploads/2015/10/dogs.jpg

Wait!





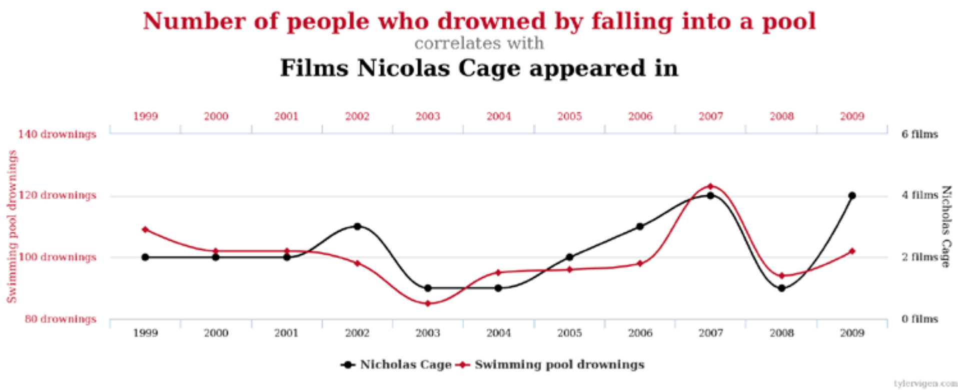
Example

My dog has a skin condition. I'd like to study this. I think it's a meat allergy.

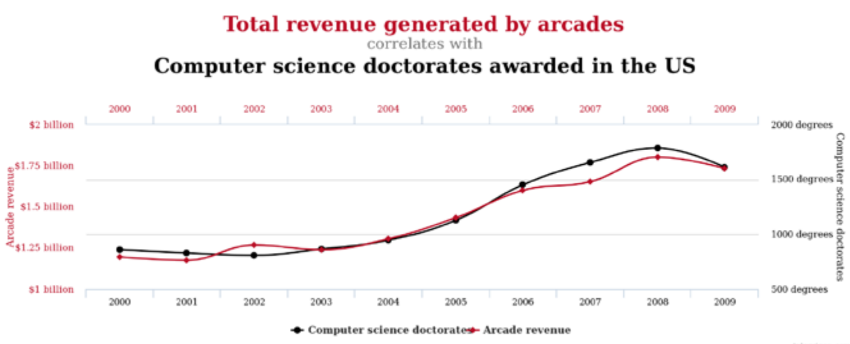
Let's take 60 dogs and give them 50g, 100g and 150g meat protein a day for 4 weeks (between groups). We then measure the inflammation levels in the blood.

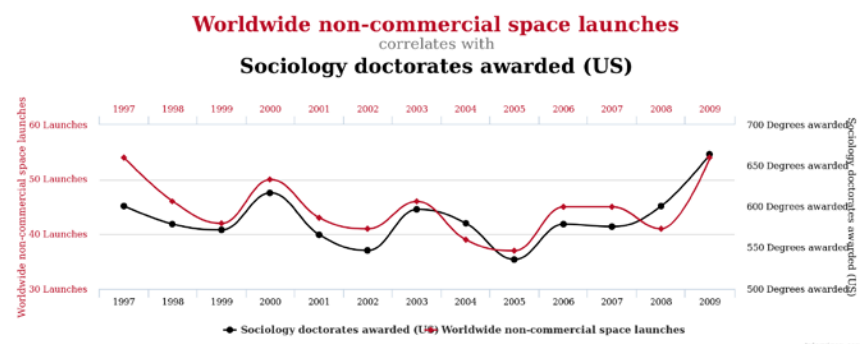
Inflammation ~ protein consumed + body weight

Covariate!

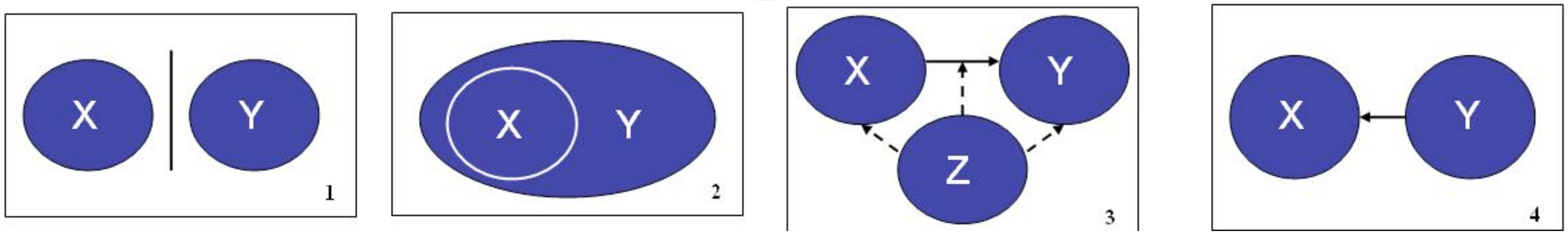


http://www.tylervigen.com/spurious-correlations

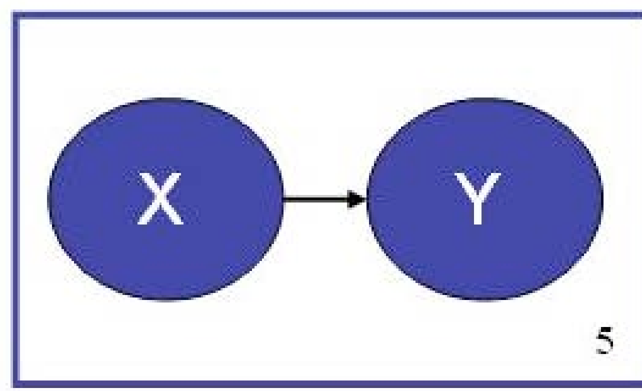




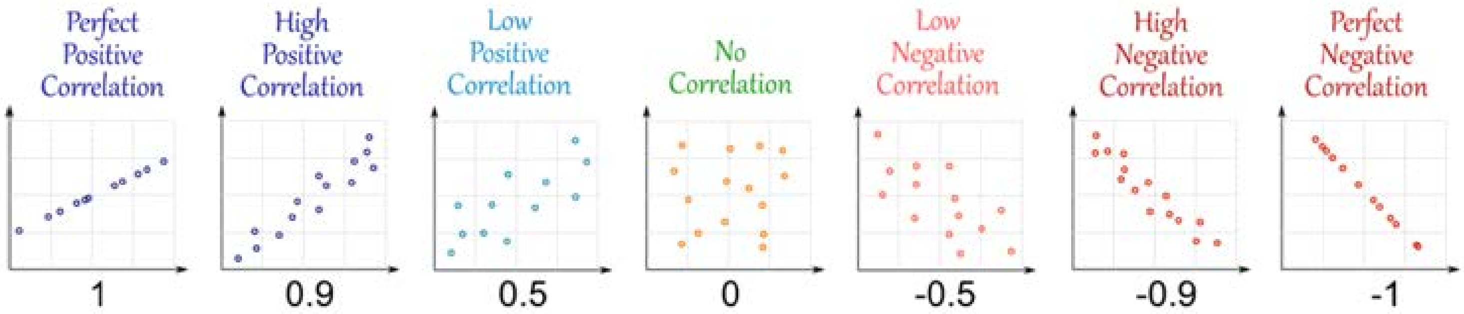
Why is correlation not (necessarily) causation?



1. Coincidence
2. X is an subset of Y
3. Moderator- or Mediator-effect?
   a hidden third variable underlies the relationship.
4. Opposite causation
5. **Causal relationship**



Possible relationships



https://www.mathsisfun.com/data/correlation.html

- **positive correlation:** As the value of X increases/decreases, the corresponding value of Y also increases/decreases.
- **negative correlation:** As the value of X increases/decreases, the corresponding value of Y also decreases/increases.
- **zero correlation:** there is no pattern or predictive relationship between variables X and Y

Correlation

- does not *necessarily* imply causation
- the extent to which two variables are associated (or co-vary)
- this covariance could be *positive* or *negative*
- assumes that the data is from a bivariate normal population
- assumes that the data is based on a *interval* or *ratio* scale

Pearson's Product-Moment Correlation Coefficient (*r*)

$$r = \frac{\sum z_x z_y}{N}$$

Pearson *r (rho):*

- a statistic
- quantifies the extent to which two variables *X* and *Y* are associated,
- whether the direction of the association is positive, negative, or zero

ID	Pace of Life	Incidence of Heart Disease	P-mean	H-mean	P-mean*H-mean	(P-mean)^2	(H-mean)^2
1	27.67	24.00	4.83	1.16	20.25	23.31	17.59
2	25.33	29.00	2.49	6.16	22.87	6.19	84.54
...	...	...	...	...	...	...	...
36	15.00	16.00	-7.84	-6.84	29.84	61.50	14.48
MEAN	22.84	19.81		SUM	200.89	318.08	951.64

1. calculate means for variables *X* and *Y*
2. calculate $(X_i - \bar{X})$ and $Y_i - \bar{Y}$
3. calculate $(X_i - \bar{X})(Yi - \bar{Y})$, $(Xi - \bar{X})^2$, and $(Y_i - \bar{Y})^2$
4. sum up $(X_i - \bar{X})(Y_i - \bar{Y})$, $(Xi - \bar{X})^2$, and $(Yi - \bar{Y})^2$
5. calculate $\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$

Magnitude of *r*

Range of *r*	**Descriptive label for r**
± .80 to 1.00	Very strong
± .60 to .79	Strong
± .40 to .59	Moderate
± .20 to .39	Weak
± .00 to .19	Very weak

Correlations can be described in terms of the association between the two variables.

For example, we can say that there is a weak positive correlation (r=0.37) between "pace of life" and "incidence of heart disease".

When is *r* significant?

H0: $\rho=0$

H_A: $\rho>\neq0$ or $\rho>0$ or $\rho<0$

$$t = \frac{observed - expected}{standard\ error}$$

$$standard\ error = \sqrt{\frac{1-r^2}{n-2}}$$

$$t = \frac{r-\rho}{\sqrt{\frac{1-r^2}{n-2}}} = r\sqrt{\frac{n-2}{1-r^2}}$$

To calculate significance:

1. calculate t-value
2. calculate df (i.e., n-2)
3. find critical t-value (2-tailed)
4. reject H0, if $t>t_{crit}$

Alternative calculation:

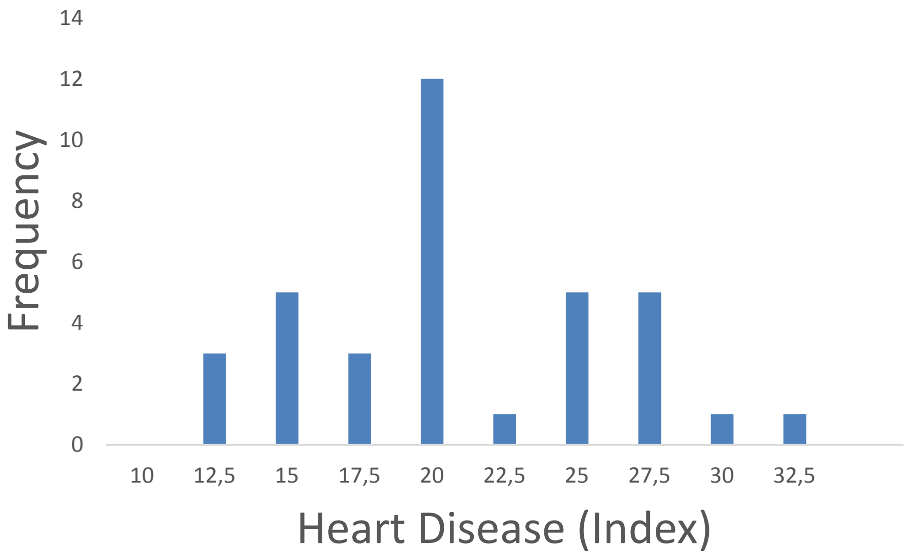
1. calculate df (i.e., n-2)
2. find critical *r* value
3. reject H0, if $r>r_{crit}$

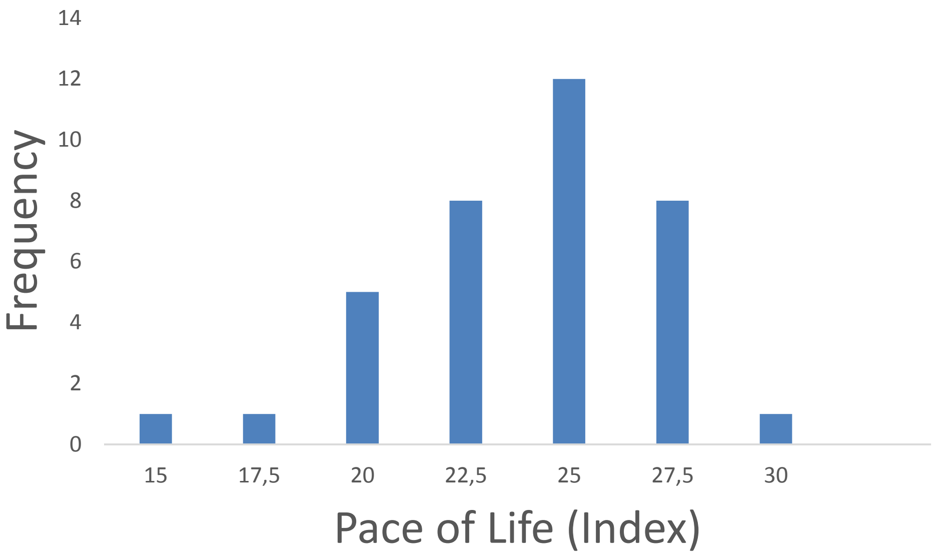
Report correlation:

"A significant weak correlation between "pace of life" and "incidence of heart disease" was found, r(34)=0.37, *p<0.05.* A fast pace of life tends to result in higher incidences of "heart disease".

Histogram

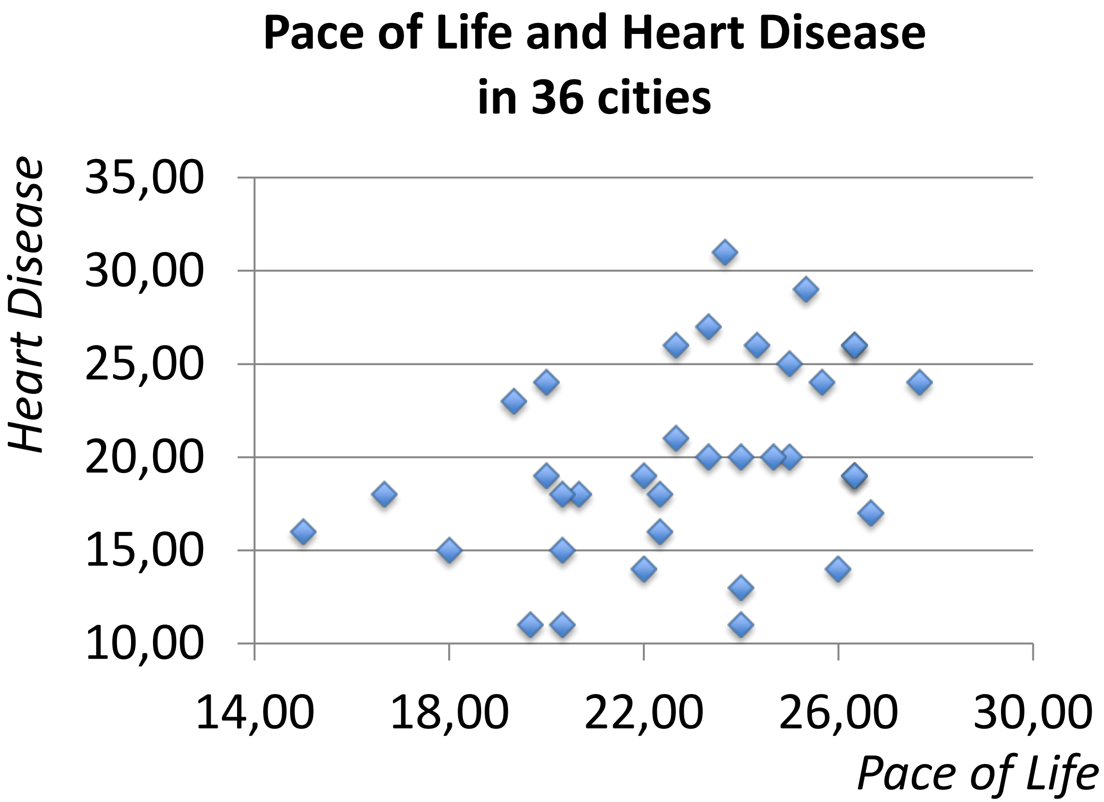
a diagram consisting of rectangles whose area is proportional to the frequency of a variable and whose width is equal to the class interval (e.g., 2.5).





Levine, R. (1990) The pace of life and coronary heart disease. *American Scientist*, 78, 450-459

Scatterplots



a graph in which the values of two variables are plotted along two axes, the pattern of the resulting points revealing any correlation present.

Levine, R. (1990) The pace of life and coronary heart disease. *American Scientist*, 78, 450-459