Inferential STATS

sample statistic	population parameter	description
n	N	number of members of sample or population
$\bar{x}$ "x-bar"	μ "mu" or μ_x	mean
M or Med	(none)	median
s (TIs say Sx)	σ "sigma" or σ_x	standard deviation For variance, apply a squared symbol (s^2 or σ^2).
r	ρ "rho"	coefficient of linear correlation
$\hat{p}$ "p-hat"	p	proportion
z t χ^2	(n/a)	calculated test statistic

STATS II

Regression: Basic idea

Regression line:
The line that minimizes the sum of squared vertical distances from each point to the line.

ANOVA: Basic idea

Are differences between groups due a real effect, or just random?

- Study the variance between and within the groups

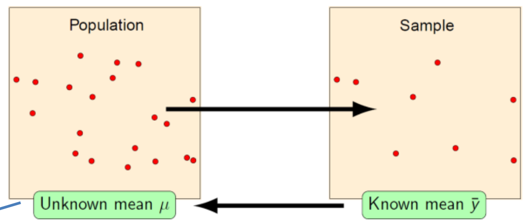
Confidence intervals (CIs)

- An $x\%$ CI contains an (unknown) population parameter with $x\%$ certainty.
- When repeating the study many times, about $x\%$ of the CIs will contain the parameter.

Hypothesis testing

- The probability of the current sample result is so small, under the null hypothesis, that it is unlikely that the population parameter has a certain value (defined under H_0).

Inference: To derive as a conclusion from facts or premises.



The sample mean $\bar{y}$ can be used to:

- Estimate μ
- Make probabilistic statements about μ:
  - "The 95% CI for μ is $(-0.4, 1.3)$"
  - "With 99% certainty we can say that $\mu > 0$"

knowledge about the sampling distribution needed!!!

1. Collect a sample. Compute the sample mean: $\bar{y}_1$.
2. Collect a sample. Compute the sample mean: $\bar{y}_2$.

Yn

The sampling distribution of the mean

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{y_1 + y_2 + \cdots + y_n}{n}$$

probability distribution of a statistic in the sample: y => estimate μ

Sampling distribution of $\bar{y}$:

- Mean $:= \mu_{\bar{y}} = \mu$
- SD $:= \sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$: Standard Error (SE) of the mean
- It is normally distributed if the population of y values is also normally distributed (regardless of the sample size n):

$$y_i \sim N(\mu, \sigma) \Longrightarrow \bar{y} \sim N(\mu_{\bar{y}}, \sigma_{\bar{y}}) = N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- The shape of the distribution of $\bar{x}$ depends on the shape of the population distribution
  - if a population has the $N(\mu,\sigma)$ distribution, then the sample mean $\bar{x}$ of n independent observations has the $N(\mu,\sigma/\sqrt{n})$ distribution
  - when n is large, the sampling distribution of $\bar{x}$ is approximately Normal, $N(\mu,\sigma/\sqrt{n})$, for any population with mean μ and finite standard deviation σ
    - = ***central limit theorem***
- Any linear function or linear combination of independent Normal random variables is also Normally distributed

IF NOT NORMAL

Central limit theorem:

- For large sample sizes, the distribution of $\bar{x}$ is close to a Normal distribution
  - No matter the shape of the population distribution, as long as the population has a finite standard deviation σ.
- When n is large, the sampling distribution of the sample mean $\bar{x}$ is approximately Normal:

$$\boxed{\bar{x} \text{ is approximately } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

Sampling distributions help quantifying which values of the statistic are most/least probable. This allows associating probabilities to sample values: Significance tests: p-values & Confidence intervals: The lower and upper boundaries

Significance tests

- Regression
  - t-tests: Parameters
  - F-tests: Model fit
  - Wald tests (χ^2): Parameters Logistic regression
- Correlation
  - t-test: Special case ($H_0 : \rho = 0$)
  - z-test: Fisher's Z transformation
- Analysis of variance (ANOVA)
  - t-tests: Contrasts, multiple comparisons
  - F-tests: Model fit
- Non-parametric tests:
  - Using exact sampling distributions
  - Using approximating distributions (e.g., Z, χ^2)

- Null hypothesis (H_0): Statement quantifying a value for the population parameter of interest <= Against
- Alternative hypothesis (H_a): Statement contradicting the null hypothesis (smaller, larger, different) <= In favour

$$\text{test statistic} = \frac{\text{(estimate statistic)} - \text{(hypothesized value} | H_0 \text{ is true)}}{\text{SE}_{\text{statistic}}}$$

From the sample

From the population

<=

One-sample z test

- In the population: $y \sim N(\mu, \sigma)$, σ known
- Sampling distribution: $\overline{y} \sim N(\mu, \sigma/\sqrt{n})$

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

The p-value is the probability of getting an outcome as extreme or more extreme than the actually observed outcome from the sample, given that H_0 were true.

- The smaller the p-value, the stronger the evidence against the null hypothesis, that is, the more unlikely the null hypothesis is.
- What is 'small'? Compare p with the significance level α (e.g., $\alpha = 5\%$)

Significance tests: One-sample t test

- $y \sim N(\mu, \sigma)$. Both μ and σ unknown
- $H_0 : \mu = \mu_0$
- Estimate σ with $s = \sqrt{\frac{1}{n-1}\sum_i (y_i - \overline{y})^2}$
- Test statistic:

$$t = \frac{\overline{y} - \mu_0}{s/\sqrt{n}} \sim t(n-1)$$

The t distribution is used to compute the p-value.

Significance tests: Two-sample t test

- $y_1 \sim N(\mu_1, \sigma_1)$, $y_2 \sim N(\mu_2, \sigma_2)$. All μs and σs unknown
- $H_0 : \mu_1 = \mu_2$ or, equivalently, $H_0 : \mu_1 - \mu_2 = 0$
- Test statistic
  - If we can assume that $\sigma_1 = \sigma_2$ (ideal):

$$t = \frac{(\overline{y}_1 - \overline{y}_2) - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2),$$

where s_p = pooled SD $= \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}}$.

  - Otherwise:

$$t = \frac{(\overline{y}_1 - \overline{y}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(k)$$

(k approximated by software)

The t distribution is used to compute the p-value.

- Two populations:
  - Boys ($n_1 = 115$, mean $= 26.39$)
  - Girls ($n_2 = 125$, mean $= 26.98$)
- Pooled SD $= 13.30$
- y = POSTLET, knowledge of the alphabet

Independent Samples Test

		t-test for Equality of Means						
							95% Confidence Interval of the Difference	
		t	df	Sig. (2-tailed)	Mean Diff.	SE. Diff.	Lower	Upper
POSTLET	Equal variances assumed	$\boxed{-.340}$	$\boxed{238}$	$\boxed{.734}$	-.5847	1.72	-3.9694	2.8000
	Equal variances							

$t = \frac{26.39 - 26.98}{13.30 \times \sqrt{1/115 + 1/125}}$

df = 115+125-2

Don't reject H_0

95% confidence interval (95% CI):
95% CI = sample mean ± 95%marge

› What does 95% mean?
- suppose we draw 100 samples from a population
- for each sample: 95% chance that population mean is within the CI
  - CI is different per sample, due to different sample means
- So, population mean: appr. 95x within (for each sample different) CI and appr. 5x outside CI

CI's

A specific CI either contains or does not contain the parameter.

$$\text{CI} = \text{estimate} \pm \underbrace{\text{margin of error}}_{\text{Critical value from sampling distribution} \times \text{SE}}$$

From the sample

From the population

Known σ: z confidence interval

- In the population: $y \sim N(\mu, \sigma)$, σ known

$$\text{CI} = \overline{y} \pm z^* \times \boxed{\frac{\sigma}{\sqrt{n}}}$$

SE of the mean

z^* = critical value from $N(0,1)$

Unknown σ: t confidence interval

- In the population: $y \sim N(\mu, \sigma)$. Both μ and σ unknown
- Estimate σ with $s = \sqrt{\frac{1}{n-1}\sum_i (y_i - \overline{y})^2}$

$$\text{CI} = \overline{y} \pm t^* \times \frac{s}{\sqrt{n}}$$

t^* = critical value from $t(n-1)$

Comparing two means

- Assume equal variances:
  $y_1 \sim N(\mu_1, \sigma)$, $y_2 \sim N(\mu_2, \sigma)$.
  μ's and σ unknown
- Sample sizes: n_1, n_2
- Recall the test statistic:

$$t = \frac{(\overline{y}_1 - \overline{y}_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2), s_p = \text{pooled SD}$$

- Confidence interval:

$$\text{CI} = (\overline{y}_1 - \overline{y}_2) \pm t^* \times s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

t^* = critical value from $t(n_1 + n_2 - 2)$

Example: 'Sesame Street' Data

- Two populations:
  - Boys ($n_1 = 115$, mean = 26.39)
  - Girls ($n_2 = 125$, mean = 26.98)
- Pooled SD = 13.30
- y = POSTLET, knowledge of the alphabet

Independent Samples Test

		t-test for Equality of Means						
							95% Confidence Interval of the Difference	
		t	df	Sig. (2-tailed)	Mean Diff.	SE. Diff.	Lower	Upper
POSTLET	Equal variances assumed	-.340	238	.734	-.5847	1.72	-3.9694	2.8000
	Equal variances							

Significance test

Diff $\pm t^*(238) \times \text{SE}_{\text{Diff}}$

The margin of error of a CI decreases (i.e., the CI becomes smaller) if:

- The confidence level decreases
- The sample size increases
- The SD decreases

Possible correct interpretations of a 95% CI = (a, b):

- We say that we are 95% confident that the unknown parameter lies between a and b.
- We arrived at these numbers by a method that gives correct results in 95% of the time.
- In the long run, 95% of all samples lead to an interval that covers the unknown parameter.

A two-sided test with significance level α rejects the hypothesis

$$H_0 : \mu = \mu_0$$

if and only if the value μ_0 lies outside the $(1-\alpha)$% CI for μ.

Example

Suppose the 95% CI for the difference between the means of two groups is

$$95\%\ \text{CI} = (-3.97, 2.80).$$

It can be concluded that the null hypothesis $H_0 : \mu_1 = \mu_2$ cannot be rejected for $\alpha = 5\%$ because 0 $(= \mu_1 - \mu_2)$ is contained in the CI.

Nonparametric Tests

Parametric tests (e.g., t- and z-tests) depend on model assumptions, most notably

$$y \sim N(\mu, \sigma)$$

If normality violated CLT can be used, if n is large. => If n is small, then: remove outliers/transform data; Use resampling techniques, other model; use nonparametric test!!!

VS

Nonparametric tests are distribution free => no assumption about distribution, but:

Almost all parametric tests have a nonparametric counterpart:

- t-test for two independent samples
  - Wilcoxon rank sum test (today)
- t-test for two paired samples
  - Sign test (Stats I)
  - Wilcoxon signed rank test (read in book, section 15.2)
- ANOVA's F-test (Lecture 4)
  - Kruskal-Wallis test (Lecture 6)

Rank Scores

Scores	-2	4	8	1	0	-3	5	5
Ordered scores	-3	-2	0	1	4	5	5	8
Ordered ranks	1	2	3	4	5	6.5	6.5	8
Ranks	2	5	8	4	3	1	6.5	6.5

Used in nonparametric tests instead of the raw scores, but less information/precise
=> Distance ignored (only order) + less power, if normal (but less strict) => Too conservative: Failing to reject false H_0 too often

WILCOXON (Mann-Whitney) rank sum test:

H_0: Both distributions are the same.

H_a: One distribution has systematically larger/smaller values.

Example: Independent samples from two groups

Group 1	6	10	8	13
Group 2	11	9	15	

Test:

H_0: Both distributions are the same.
H_a: Group 1 has systematically smaller values than Group 2.

Procedure:

1 – Merge scores in one set	6	10	8	13	11	9	15
2 – Order scores	6	8	9	10	11	13	15
3 – Assign ranks	1	2	3	4	5	6	7

Ranks

	GROEP	N	Mean Rank	Sum of Ranks
SCORE	controle	4	3.25	$\boxed{13.00}$
	experiment	3	5.00	15.00
	Total	7		

$W_1 = 13$

Test Statistics[b]

	SCORE
Mann-Whitney U	3.000
Wilcoxon W	13.000
Z	-1.061
Asymp. Sig. (2-tailed)	.289
Exact Sig. [2*(1-tailed Sig.)]	.400[a]
Exact Sig. (2-tailed)	.400
Exact Sig. (1-tailed)	$\boxed{.200}$
Point Probability	.086

a. Not corrected for ties.
b. Grouping Variable: GROEP

$z = -1.061$, approx. $p = .289/2 = .144$

Exact $p = .200$

Do not reject H_0: There is no statistical evidence that scores in Group 1 are systematic smaller than in Group 2.

Test statistic: Use the sum of rank scores

$W_1 = 13$

($W_2 = 15$ would also do)

Idea: If the scores in group 1 are systematically smaller than those in group 2 ($= H_a$), then so is the sum of rank scores.

Test: Reject H_0 is W_1 is too small

Sampling distribution of W: Either

- Exact. Ideal. Given by software for small sample sizes ($n_1 + n_2$)
- Approximation. Use the normal distribution.

Normal approximation for W:

$$z = \frac{W - \mu_W}{\sigma_W} = \frac{W - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}}$$

For the running example:

Population 1	6	10	8	13
Population 2	11	9	15	

- $n_1 = 4,\ n_2 = 3$
- $W_1 = 13$
- Therefore, $z = \frac{13-4\times8/2}{\sqrt{(4\times3\times8)/12}} = -1.061$
- p-value $= P(Z < -1.061) = .144$: Do not reject H_0

Suppose a variable X can take the values 1, 2, 3, or 4.
The probabilities associated with each outcome are described by the following table:

Outcome	1	2	3	4
Probability	0.1	0.3	0.4	0.2

The probability that *X* is equal to 2 or 3 is the sum of the two probabilities: *P(X = 2 or X = 3) = P(X = 2) + P(X = 3)* = 0.3 + 0.4 = 0.7. Similarly, the probability that *X* is greater than 1 is equal to 1 - *P(X = 1)* = 1 - 0.1 = 0.9, by the complement rule.

Use continuity correction whenever using a continuous distribution for probability calculations of discrete variables.

For the Wilcoxon rank sum test:

$$z = \frac{W - \frac{n_1(n_1+n_2+1)}{2} \pm 0.5}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}}$$

- $+0.5$: Left one-sided H_a, or two-sided H_a with $W < \mu_W$
- -0.5: Right one-sided H_a, or two-sided H_a with $W > \mu_W$

SPSS does not provided continuity correction.

Error & Power (Significance)

	Reality: H_0 is true	Reality: H_0 is false
Test: Don't reject H_0	☺	Type II error (β)
Test: Reject H_0	Type I error (α)	☺

$\alpha = P(\text{rejecting } H_0 | H_0 \text{ is true})$
$\beta = P(\text{not rejecting } H_0 | H_0 \text{ is false})$

Not fixed.
Power: $P(\text{reject } H_0 | H_0 \text{ is false}) = 1 - \beta$.

Example: z-test

- Population: $X \sim N(\mu, \sigma = 1)$
- Test at $\alpha = 5\%$:
  H_0: $\mu = 6.0$
  H_a: $\mu > 6.0$
- Results from a sample: $n = 25$, $\bar{x} = 6.4$.
- Test statistic:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{6.4 - 6.0}{1/5} = 2.0$$

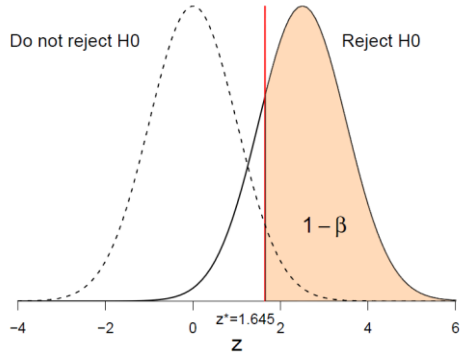
- Critical value: $P(Z > z^*) = .05 \implies z^* = 1.645$.

$z = 2.0 > 1.645 = z^*$

1-0.05=0.95
=> ~1.645

Test result: Reject H_0.

Now suppose that μ is actually is equal to 6.5.



The actual distribution has shifted to the right.
The probability of rejecting H_0 is much larger than α.

- Under the wrong hypothesis $H_0 : \mu = 6.0$, the test rejects H_0 if

Right value = Solve sample mean (x)

$$\begin{aligned} z = \frac{\bar{x} - 6.0}{1/5} &> 1.645 \\ \bar{x} - 6.0 &> 0.329 \\ \bar{x} &> 6.329. \end{aligned}$$

- So, what is the correct probability of rejecting H_0 (i.e., under $H_a : \mu = 6.5$)?

$$\begin{aligned} P_{H_a}(\bar{x} > 6.329) &= P_{H_a}\left(\frac{\bar{x} - 6.5}{1/5} > \frac{6.329 - 6.5}{1/5}\right) \\ &= P_{H_a}(Z > -.855) \\ &= .804. \end{aligned}$$

1-((0,1977+0,1949)/2)
=> ~0,804

Conclusion: There is an 80% probability of correctly rejecting H_0 if $H_a : \mu = 6.5$ were true.

- Usually, the alternative hypothesis covers a range of values, e.g. $H_a : \mu > 6.0$.
- However, only one such value (at most), such as $\mu = 6.5$, may be true.
- The power depends on this unknown true value.

Effect size (ES)

- How large is the difference between the means of two groups?
- How strong is the association between two variables?

Different from stat. signif.

An effect size is simply an objective and standardized measure of the magnitude of observed effect (see Field, 2005a; 2005b). The fact that the measure is standardized just means that we can compare effect sizes across different studies that have measured different variables, or have used different scales of measurement. So, an effect size based on the Beck depression inventory could be compared to an effect size based on levels of serotonin in blood.

Example:

- Is $\bar{x} = 8$ a relevant finding if the population is $N(6, 1)$?
- Is $\bar{x} = 8$ a relevant finding if the population is $N(6, 5)$?
- Pop1: $d = \frac{8-6}{1} = 2$.
  Sample mean is 2 SD above μ.
- Pop2: $d = \frac{8-6}{5} = .4$.
  Sample mean is only .4 SD above μ.

For example, for Cohen's d:

- Small: $d = .20$
- Medium: $d = .50$
- Large: $d = .80$

Size of effect
<= Cohen's d
Social Psy. =>

- < 0.1 = trivial effect
- 0.1 - 0.3 = small effect
- 0.3 - 0.5 = moderate effect
- > 0.5 = large difference effect

Two important types: Of ES

- Based on distances between means (e.g., Cohen's d).
  - One sample randomly drawn from $N(\mu, \sigma)$, μ and σ known:

$$d = \frac{\bar{y} - \mu}{\sigma}$$

  - Two sample means:

$$d = \frac{\bar{y}_1 - \bar{y}_2}{s_p}$$

  (s_p = pooled SD)
- Based on proportions of explained variance.
  - Regression (and ANOVA): R^2 (*later lectures*)
  - ANOVA: ω^2 (*later lectures*)

The power of a test depends on:

1. Sample size n.
2. Significance level α.
3. Effect size.
4. (Nature of the test).

Power calculations for relationship!!!

- Increase sample size.
- Increase α.
  $\longrightarrow$ Risk: More Type I errors.
- Increase the ES.
- Decrease error variance (noise), e.g. by
  - Making groups more homogeneous.
  - Add additional variables to the model.

A: Three things:

- Significance level α.
- Sample size n.
- ES (distribution under H_a).

Formulas contain α, $1-\beta$, ES, and n. Plug in any three values to compute the fourth. The math is complicated to perform by hand.

How to calculate Power:

- Power tables (similar to tables with critical values).
- Use special programs such as G*Power.

priori analysis => Required N (Compute sample size n as a function of the required power level, pre-specified α level, and population ES to be detected. => n low = power low / n high = waste of R.)

posteriori (retrospective) analysis => Achieved Power

- Based on observed ES in the sample, which is only an estimate of true ES.
- Performed after test; retrospective power offers no additional information for explaining nonsignificant results.
- Limited usability; it can be a *post mortem* analysis.
- Useful as an a priori analysis for next experiment.
- Better to calculate CIs and ESs.

Power table for the two-sample t test:
Sample size per group needed given the power and d, for two-tailed $\alpha = 5\%$ (one-tailed $\alpha = 2.5\%$):

Power	ES = Cohen's d .20	.30	.40	.50	.60	.70	.80	1.0	1.2	1.4
.60	246	110	62	40	28	21	16	11	8	6
.70	310	138	78	50	35	26	20	13	10	7
.80	393	175	99	64	45	33	26	17	12	9
.90	526	234	132	85	59	44	34	22	16	12

Q: [a priori] How many participants are needed so that there is an 80% probability of finding a difference between two groups of one SD (i.e., $d = 1$), for $\alpha = 5\%$?
A: 17 participants per group, that is, $17 \times 2 = 34$ in total.

T-test => ANOVA

2 groups comparing mean => Two-sample T-test (special kind of ANOVA)

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \underbrace{t(n_1 + n_2 - 2)}_{\text{Under } H_0}$$

$H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$

- μ_1, μ_2, σ: unknown
- Same σ assumed
  (If not: Slightly different t-test needed, recall Lec. 2)
- Sample sizes: n_1, n_2

If more than two groups => ANOVA – F-test (Example Sesame Street)

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

versus

H_a : not all μs are equal

F-Test

- For two groups we have $t^2 = F$

$$\underbrace{t \sim t(n_1 + n_2 - 2)}_{t\text{-test}} \longrightarrow \underbrace{F \sim F(1, n_1 + n_2 - 2)}_{F\text{-test (from ANOVA)}}$$

ANOVA: Generalizes the idea of comparing the between and within types of variance to more than two groups.

VS.

T-Test

Look at t^2 (and assume $n_1 = n_2 = n$ for simplicity):

$$t^2 = \frac{\frac{n}{2}(\bar{y}_1 - \bar{y}_2)^2}{s_p^2} = F \sim F(1, n_1 + n_2 - 2)$$

difference between means → estimator of the variance between groups

pooled variance of both groups → estimator of σ^2, the variance within groups

The t statistic is therefore a measure of the ratio

$$\frac{\text{between groups variation}}{\text{within groups variation}}$$

Differences between the means of both groups are assessed by the t-test by comparing the two types of variances.

ANOVA I

Principle: Study differences in the means of I independent groups
Procedure: Compare the between and the within group variances using the F-test

Variation between groups => Plot of means (Do means differ? => Use CIs)
Variation between and within groups => Use Boxplot
=>**CIs:** Small overlap: **sig.** differences Vs. Large overlap: non sig. di_erences

Which group(s) differ?

Sample size: $n = 240$

Independent variable (factor):
VIEWCAT = number of times that Sesame Street is viewed
($1 =$ few, . . . , $4 =$ a lot)

Dependent variable:
POSTNUMB = knowledge of numbers (scores 0–54)

VIEWCAT	Mean	SD	n
1	21.65	11.39	54
2	27.50	11.53	60
3	33.03	12.38	64
4	36.77	11.08	62
Total	30.05	12.85	240

Split the total variance in two parts:

- A part that can be explained by differences between groups.
- A part that remains unexplained within groups.

$$\underbrace{\sum_{ij}(y_{ij} - \bar{y})^2}_{\text{SST}} = \underbrace{\sum_{ij}(\bar{y}_i - \bar{y})^2}_{\text{SSG}} + \underbrace{\sum_{ij}(y_{ij} - \bar{y}_i)^2}_{\text{SSE}}$$

Total SS

Group SS

between

Error SS

within

$$\text{SST} = \text{SSG} + \text{SSE}$$

- i indexes the groups
- j indexes the persons in a group
- $y_{ij} =$ observation of person j in group i
- $\bar{y}_i =$ mean of the DV y in group i (i.e., over all persons in group i)
- $\bar{y} =$ overall, or grand, mean of y (i.e., over all persons in all groups)

Test statistic:

$$F = \frac{\text{MSG}}{\text{MSE}} = \frac{\text{SSG}/\text{df}_G}{\text{SSE}/\text{df}_E}$$

If H_0 holds: $F \approx 1$

1. $\text{MSE} = s_p^2$ always estimates σ^2, the common group variance.
2. Under H_0, MSG also estimates σ^2.
3. Hence, under H_0, the F ratio is ≈ 1.

If H_0 does not hold: $F > 1$

Convert SS's in variances: Divide by degrees of freedom (df)

Mean Squares (MS)

$\text{Var} = \text{SST}/\text{DFT}$

	Total	Group	Error
SS	SST	SSG	SSE
df	$\text{df}_T = n - 1$	$\text{df}_G = I - 1$	$\text{df}_E = n - I$
MS	$\text{MST} = \text{SST}/\text{df}_T$	$\text{MSG} = \text{SSG}/\text{df}_G$	$\text{MSE} = \text{SSE}/\text{df}_E$

Variance in y!

$s_p^2 =$ pooled variance!

Reject H0 is F is too large (i.e., F >> 1) => Largeness of F = sampling distribution $F \sim F(I-1, n-I)$

Source	SS	df	MS	F
Group	$\sum_{ij}(\bar{y}_i - \bar{y})^2$	$I - 1$	SSG/df_G	MSG/MSE
Error	$\sum_{ij}(y_{ij} - \bar{y}_i)^2$	$n - I$	SSE/df_E	
Total	$\sum_{ij}(y_{ij} - \bar{y})^2$	$n - 1$		

Example: Sesame Street

ANOVA

POSTNUMB

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7574.205	3	2524.735	18.699	.000
Within Groups	31864.091	236	135.017		
Total	39438.296	239			

Reject $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

Rejecting H_0 implies:

- Significant difference between groups
- At least one group is different from the others.

Which group(s) differ?
Perform statistical inference (next week's lecture)

- Planned comparisons: Contrasts
- Post hoc comparisons: Multiple comparisons

Homoscedastic, if all random variables in the sequence have the same finite variance

Two ways to compute CIs for group means:

- Based on the pooled SD, s_p (ideal when homoscedasticity is met)

$\sqrt{MSE}$

$$\text{CI for group } i = \bar{y}_i \pm t^*_{n-I} \frac{s_p}{\sqrt{n_i}}$$

n = total sample size

- Based on the groups SD, s_i (when homoscedasticity is violated)

$$\text{CI for group } i = \bar{y}_i \pm t^*_{n_i-1} \frac{s_i}{\sqrt{n_i}}$$

n_i = group sample size

Example: Sesame Street

VIEWCAT	Mean	SD	n
1	21.65	11.39	54
2	27.50	11.53	60
3	33.03	12.38	64
4	36.77	11.08	62
Total	30.05	12.85	240

$$s_p = \sqrt{\frac{(54-1)\times 11.39^2 + \cdots + (62-1)\times 11.08^2}{54+60+64+62-4}} = 11.62$$

- Based on s_p (ideal):

$$95\%\ \text{CI}_{\text{group 2}} = 27.50 \pm \underbrace{1.970}_{t^*_{236}} \times \frac{11.62}{\sqrt{60}} = (24.54, 30.46)$$

- Based on s_i:

T-table (CL)

$$95\%\ \text{CI}_{\text{group 2}} = 27.50 \pm \underbrace{2.001}_{t^*_{59}} \times \frac{11.53}{\sqrt{60}} = (24.52, 30.48)$$

Why not multiple t-test? => u1/u2/u3/u4 => 4-groups = 6 t-tests (Too large overall error rate, or experiment-wise error rate) => Therefore probability of at least one Type I Error high (increases with number of tests) => By chance capitalization

$$\begin{aligned}\text{overall error rate} &= \text{probability of at least one false rejection}\\ &= 1 - (\text{probability of no false rejection})\\ &\approx 1-(1-.05)^6\\ &= .265.\end{aligned}$$

=> α=5% * 6 tests => Formula for pairs = k(k-1)/2

Conclusion: There is a 26% probability of at least one false rejection $\longrightarrow$ overrejecting

=> To avoid capitalization make use of inference procedures = Planned comparisons: Contrasts // Post hoc comparisons: Multiple comparisons

Contrast (ψ)

= describe systems of procedures that can be used to draw conclusions from datasets arising from systems affected by random variation, such as observational errors, random sampling, or random experimentation

Contrast = Hypotheses constructed prior to data collection => written as linear combinations of group means => G1=T1 => G2=T2 => G3=Control => larger scores = better effect (does it work)

$$\psi = \sum_i a_i\mu_i = a_1\mu_1 + a_2\mu_2 + \cdots + a_I\mu_I$$

Coefficients of the contrast

=> ai's defined by researcher, but unknown μ's

Solution: Estimate from sample (μ's $\longrightarrow$ $\bar{y}$'s)

$$\text{sample contrast} = c = \sum_i a_1\bar{y}_1 + a_2\bar{y}_2 + \cdots + a_I\bar{y}_I$$

=> Contrast: T1 more effective than T2? = Contrast $1 = 1\mu_1 + (-1)\mu_2 + 0\mu_3$
Coefficients of contrast 1: $1, -1, 0$ => (-1) => $H_{01}: \mu_1 - \mu_2 = 0$

=> Contrast: Combined T1&T2 effective? = Contrast $2 = .5\mu_1 + .5\mu_2 + (-1)\mu_3$
Coefficients of contrast 2: $.5, .5, -1$ => (0.5) => $H_{02}: \frac{\mu_1+\mu_2}{2} - \mu_3 = 0$

Inference using contrasts:

- Standard error of c: $SE_c = s_p\sqrt{\sum_i \frac{a_i^2}{n_i}}$
- Statistical testing: $H_0: \psi = 0$ (H_a one- or two-sided)

$$t = \frac{c}{SE_c} \sim \underbrace{t(n-I)}_{\text{Under}H_0}$$

- Confidence interval:

$$c \pm t^* SE_c$$

t^* =critical value from $t(n-I)$

A note on the coefficients of contrasts.

$\psi = 1\mu_1 + (-1)\mu_2 + 0\mu_3 \longrightarrow$ coefficients: $1, -1, 0$

- A multiple of these coefficients could be used, e.g.
  - ✓ $(2, -2, 0)$, $(-.5, .5, 0)$, ...

  This does not affect the t test (t statistic & p-value the same).
- However, this does affect the computation of CIs!

  SPSS 'avoids' this problem:
  - 'One-way' ANOVA does not provide CI's anyway.
  - 'GLM' has pre-built contrasts that avoid this problem.

GLM=`repeated' contrast compares the mean of each level (except the last) to the mean of the subsequent level

Example of contrasts (3 contrasts are testes):

"The more children watch Sesame Street, the better they know numbers."

- IV = 'VIEWCAT' (Group 1 = few, ..., Group 4 = a lot)
- DV = Knowledge of numbers (scores 0-54)

Contrast	Sample contrast	Coefficients
$\psi_1 = \mu_1 - \mu_2$	$c_1 = \bar{y}_1 - \bar{y}_2$	$(1, -1, 0, 0)$
$\psi_2 = \mu_2 - \mu_3$	$c_2 = \bar{y}_2 - \bar{y}_3$	$(0, 1, -1, 0)$
$\psi_3 = \mu_3 - \mu_4$	$c_3 = \bar{y}_3 - \bar{y}_4$	$(0, 0, 1, -1)$

SPSS 'One-way ANOVA':

Descriptives

postnumb

	N	Mean
1.00	54	21.6481
2.00	60	27.5000
3.00	64	33.0313
4.00	62	36.7742
Total	240	30.0542

Contrast Tests

	Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
postnumb	1	-5.8519	2.17959	-2.685	236	.008
	2	-5.5313	2.08805	-2.649	236	.009
	3	-3.7429	2.07059	-1.808	236	.072

E.g., for contrast 2 ($c_2 = \bar{y}_2 - \bar{y}_3$):

- $c_2 = 27.5000 - 33.0313 = -5.5313$
- $SE_{c_2} = \underbrace{11.6197}_{s_p = \sqrt{MSE}} \sqrt{1/60 + 1/64} = 2.0880$
- $t = c_2 / SE_{c_2} = -5.5313/2.0880 = -2.649$

Reject H_0 : $\mu_2 - \mu_3 = 0$

SPSS 'GLM':

Contrast Results (K Matrix)

VIEWCAT Repeated Contrast			Dependent Variable POSTNUMB
Level 1 vs. Level 2	Contrast Estimate		-5.852
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		-5.852
	Std. Error		2.180
	Sig.		.008
	95% Confidence Interval for Difference	Lower Bound	-10.146
		Upper Bound	-1.558
Level 2 vs. Level 3	Contrast Estimate		-5.531
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		-5.531
	Std. Error		2.088
	Sig.		.009
	95% Confidence Interval for Difference	Lower Bound	-9.645
		Upper Bound	-1.418
Level 3 vs. Level 4	Contrast Estimate		-3.743
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		-3.743
	Std. Error		2.071
	Sig.		.072
	95% Confidence Interval for Difference	Lower Bound	-7.822
		Upper Bound	.336

Same results as before:

- $c_2 = -5.5313$
- $SE_{c_2} = 2.0880$
- $p = .009$

Confidence interval:

$$c \pm t^* SE_c$$

t^* = critical value from $t(n - I)$

$$-5.531 - \underbrace{1.97}_{t^*} \times 2.088 = -9.645$$

$$-5.531 + 1.97 \times 2.088 = -1.418$$

Post-Hoc multiple comparison: Tests (CIs) for differences between all pairs of averages.=> Only to be used if prior to the analysis no specific hypotheses can be (or have been) defined.

Only to be used after ANOVA rejected H_0.
Use t tests (two-sided), but:

✓ Caution: Chance capitalisation

Tests (CIs) must therefore be adjusted.

CIs for multiple comparisons:

$$CI_{ij} = (\bar{y}_i - \bar{y}_j) \pm t^{**} s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

Based on all groups!

- LSD (least significant differences).
  - Number of tests is not controlled (only each individual test is improved).
  - Poor control of overall error rate when # tests increases.

$$t^{**} = t^{*}_{(1-\alpha/2)}, \text{ with } df = df_{\text{DFE}}$$

Bonferroni.

- Adjusts Type I error level per test to α/k (k = number of tests) so that overall error rate $\leq \alpha$.
- **Motivation** – The Bonferroni inequality:

$$P(\text{at least one Type I error}) \leq$$
$$\leq \sum_{i=1}^{k} \underbrace{P(\text{Type I error in test } i)}_{\alpha/k} = k \times \frac{\alpha}{k} = \alpha$$

$$t^{**} = t^{*}_{(1-(\alpha/k)/2)}, \text{ with } df = df_{\text{DFE}}$$

Some guidelines for assumed equal variances (Field, 2009):

- Equal sample sizes: Tukey, Bonferroni (more conservative).
- Similar sample sizes: Gabriel.
- Very different samples sizes: Hochberg's GT2.
- When $I = 3$ LSD performs the best.

$p_{\text{Bonf.}} = 6 \times p_{\text{LSD}}$

Multiple Comparisons

Dependent Variable: POSTNUMB

Procedure	(I) VIEWCAT	(J) VIEWCAT	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
LSD	1.00	2.00	-5.8519*	2.17959	.008	-10.1458	-1.5579
		3.00	-11.3831*	2.14708	.000	-15.6130	-7.1532
		4.00	-15.1260*	2.16287	.000	-19.3870	-10.8650
	2.00	1.00	5.8519*	2.17959	.008	1.5579	10.1458
		3.00	-5.5312*	2.08805	.009	-9.6448	-1.4177
		4.00	-9.2742*	2.10428	.000	-13.4198	-5.1286
		1.00	11.3831*	2.14708	.000	7.1532	15.6130
		2.00	5.5312*	2.08805	.009	1.4177	9.6448
		4.00	-3.7429	2.07059	.072	-7.8221	.3363
	4.00	1.00	15.1260*	2.16287	.000	10.8650	19.3870
		2.00	9.2742*	2.10428	.000	5.1286	13.4198
		3.00	3.7429	2.07059	.072	-.3363	7.8221
Bonferroni	1.00	2.00	-5.8519*	2.17959	.047	-11.6510	-.0527
		3.00	-11.3831*	2.14708	.000	-17.0958	-5.6704
		4.00	-15.1260*	2.16287	.000	-20.8808	-9.3713
	2.00	1.00	5.8519*	2.17959	.047	.0527	11.6510
		3.00	-5.5312	2.08805	.052	-11.0869	.0244
		4.00	-9.2742*	2.10428	.000	-14.8730	-3.6754
	3.00	1.00	11.3831*	2.14708	.000	5.6704	17.0958
		2.00	5.5312	2.08805	.052	-.0244	11.0869
		4.00	-3.7429	2.07059	.432	-9.2521	1.7662
	4.00	1.00	15.1260*	2.16287	.000	9.3713	20.8808
		2.00	9.2742*	2.10428	.000	3.6754	14.8730
		3.00	3.7429	2.07059	.432	-1.7662	9.2521

Procedure

- $t^{*}_{1-\alpha/2}(236) = 1.970$
- $s_p = \sqrt{MSE} = 11.620$
- $n_1 = 54, \; n_2 = 60$

CI:
$$-5.8519 \pm 1.970 \times 11.620 \sqrt{\frac{1}{54} + \frac{1}{60}}$$

- $t^{*}_{1-(\alpha/6)/2}(236) = 2.661$

CI:
$$-5.8519 \pm 2.661 \times 11.620 \sqrt{\frac{1}{54} + \frac{1}{60}}$$

	N	Mean
1.00	54	21.6481
2.00	60	27.5000
3.00	64	33.0313
4.00	62	36.7742
Total	240	30.0542

Advantages of contrasts over post hoc:

- Fewer tests:
  - ✓ Smaller overall error rate
  - ✓ Less chance capitalisation
  - ✓ Fewer Type I errors.
- More power:
  - ✓ Better probability to reject H_0 when it is not true (fewer Type II errors).

Disadvantage of contrasts:

- Researchers don't always know beforehand which comparisons should be made.

Test Statistics[a,b]

	effect
Chi-Square	9.336
df	2
Asymp. Sig.	.009
Exact Sig.	.003
Point Probability	.000

Approximation using the χ^2-distribution

Exact distribution

If ANOVA assumptions are severely violated => Use **Kruskal-Wallis procedure**

Replace all scores by their rank scores.
Perform 'regular' ANOVA on the ranks.

Hypotheses:

H_0 The distribution is the same in all groups.

H_a Scores are systematically larger in some groups than others.

Test statistic:

(When there are no ties. Slight adjustment is needed with ties; SPSS does this automatically.)

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{I} \frac{R_i^2}{n_i} - 3(n+1) = \frac{12}{n(n+1)} SSG_{\text{ranks}}$$

Sampling distribution:

- Exact: When n_i's are small and no ties exist.
- Approximation: When n_i's are not too small (say, ≥ 5) then use this approximation:

$$H \sim \chi^2(I-1)$$

Effectiveness of information ($I = 3$, $n = 16$)

Flyer Score	Flyer Rank	Meeting Score	Meeting Rank	Video Score	Video Rank
9	5.5	9	5.5	17	13
10	7	13	8	15	10.5
8	3.5	18	14	16	12
6	2	8	3.5	14	9
5	1	15	10.5	19	15
				20	16
	$R_1 = 19$		$R_2 = 41.5$		$R_3 = 75.5$

$$H = \frac{12}{16 \times 17} \left(\frac{19^2}{5} + \frac{41.5^2}{5} + \frac{75.5^2}{6} \right) - 3 \times 17 = 9.30$$

Conclusion: Reject H_0, that is, there are systematic differences between groups.

ANOVA

Large F => Reject Ho (there are differences) <= Where and largeness?

ANOVA

SCORE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	79.500	3	26.500		.075
Within Groups	107.500	12	8.958		
Total	187.000	15			

Which of the following claims, based on this ANOVA table, is true?

- a The F-value is 0.34.
  No: $F = 26.500/8.958 = 2.96$.
- b There is a significant difference between group means (based on $\alpha = 5\%$).
  No: $p = .075 > .05$.
- c The F-test uses an F-distribution with 15 df's.
  No: df's $= (3, 12)$
- d The total variance on the variable score is smaller than $26.500 + 8.958 = 35.458$.
  Yes: Var $=$ SST/DFT $= 187.000/15 = 12.467 < 35.458$.

- IV $=$ Therapy (four treatment groups)
- DV $=$ Agoraphobia score

In a study about the efficiency of different therapies to treat agoraphobia a one-way ANOVA is performed. There are 4 treatment groups, all with equal sample sizes. The ANOVA table is given above. For Group 2, the average score is 23.3 (SD $=$ 3.48). What is the 99% CI for this mean, based on the pooled sd?

- a $\boxed{(18.7, 27.9)}$
- c $(19.5, 27.1)$.
- b $(17.3, 29.3)$.
- d $(20.8, 25.8)$.

$$99\%\ \text{CI} = 23.3 \pm t^*_{(1-.01/2);df=12} \times \frac{s_p}{\sqrt{n}} = 23.3 \pm 3.055 \times \sqrt{\frac{8.958}{4}}$$

In a one-way ANOVA with 4 groups, the researcher wants to test the following null hypotheses:

1. H_0 : The mean of groups 1 and 2 is equal to that of groups 3 and 4.
2. H_0 : The mean in group 3 is equal to that of group 4.

What are the contrast coefficients the researcher has to put in to the SPSS-procedure OneWay in order to test these contrasts?

- a 1st contrast: $(.5, -.5, .5, -.5)$. 2nd contrast: $(0, 0, -1, 1)$.
- c 1st contrast: $(.5, -.5, .5, -.5)$. 2nd contrast: $(0, 0, 1, 1)$.
- b $\boxed{\text{1st contrast: } (.5, .5, -.5, -.5).\ \text{2nd contrast: } (0, 0, -1, 1).}$
- d 1st contrast: $(.5, .5, -.5, -.5)$. 2nd contrast: $(0, 0, 1, 1)$.

1st contrast $H_0 : \frac{\mu_1+\mu_2}{2} = \frac{\mu_3+\mu_4}{2} \longrightarrow \psi_1 = .5\mu_2 + .5\mu_2 - .5\mu_3 - .5\mu_4 = 0$

2nt contrast $H_0 : \mu_3 = \mu_4 \longrightarrow \psi_2 = 0\mu_1 + 0\mu_2 + 1\mu_3 - 1\mu_4 = 0$

ANOVA II – Two way

Two way ANOVA $\longrightarrow$ Two factors ($=$categorical IVs)

- Group membership is defined by 2 factors, say A and B
- A has I categories, B has J categories
  - ✓ In total there are $I \times J$ groups.
- In each (i, j) group:

$$y_{ij} \sim N(\mu_{ij}, \sigma)$$

normality assumption — homoscedasticity assumption

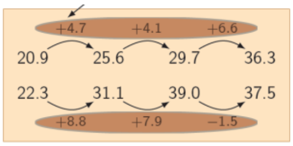
- Test effects
  - ✓ Main effects: Factor A, Factor B
  - ✓ Interaction effect: Between factors A and B

VIEWCAT * SETTING

Dependent Variable: POSTNUMB

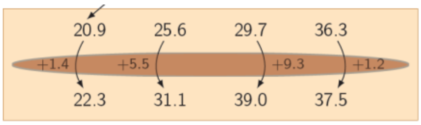
VIEWCAT	SETTING	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1.00	1.00	20.880	2.278	16.392	25.368
	2.00	22.310	2.115	18.143	26.478
2.00	1.00	25.564	1.824	21.971	29.158
	2.00	31.095	2.486	26.198	35.992
3.00	1.00	29.659	1.779	26.154	33.163
	2.00	39.043	2.375	34.364	43.723
4.00	1.00	36.289	1.848	32.649	39.930
	2.00	37.542	2.325	32.961	42.122

Interaction effect $A \times B$:
Differences between the $2 \times 4 = 8$ group means



- Interaction effect:
  Are there differences between differences in means?

Differences can also be looked at across factor 'SETTING'.



- DV: POSTNUMB $=$ knowledge of numbers (scores 0–54)
- IVs:
  - ✓ VIEWCAT $=$ viewing frequency; 4 groups (1=few,. . . ,4=a lot)
  - ✓ SETTING $=$ viewing location; 2 groups (1 $=$ home; 2 $=$ school)

POSTNUMB * VIEWCAT

POSTNUMB

VIEWCAT	Mean	Std. Deviation	N
1.00	21.6481	11.39374	54
2.00	27.5000	11.52668	60
3.00	33.0313	12.38403	64
4.00	36.7742	11.07563	62
Total	30.0542	12.84577	240

POSTNUMB * SETTING

POSTNUMB

SETTING	Mean	Std. Deviation	N
1.00	28.7692	12.67175	143
2.00	31.9485	12.93121	97
Total	30.0542	12.84577	240

Main effect A:
Differences between A's marginal means

Main effect B:
Differences between B's marginal means

Use a cross table to display the means.

SETTING	VIEWCAT 1	2	3	4	
1 = home	20.9 ($n = 25$)	25.6 ($n = 39$)	29.7 ($n = 41$)	36.3 ($n = 38$)	28.8 ($n = 143$)
2 = school	22.3 ($n = 29$)	31.1 ($n = 21$)	39.0 ($n = 23$)	37.5 ($n = 24$)	31.9 ($n = 97$)
	21.6 ($n = 54$)	27.5 ($n = 60$)	33.0 ($n = 64$)	36.8 ($n = 62$)	30.1 ($n = 240$)

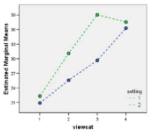
- Main effect VIEWCAT:
  Are these four marginal means significantly different from each other?

  These are means weighted by sample size:
  - ✓ $21.6 = (25 \times 20.9 + 29 \times 22.3)/54$
  - ✓ $27.5 = (39 \times 25.6 + 21 \times 31.1)/60$

- Interaction effect:
  Are there differences between differences in means?

Main effect SETTING:
Are these two marginal means significantly different from each other?

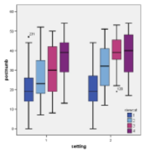
These are means weighted by sample size:
- ✓ $28.8 = (25 \times 20.9 + \cdots + 38 \times 36.3)/143$
- ✓ $31.9 = (29 \times 22.3 + \cdots + 24 \times 37.5)/97$



Visualisation by Means plot => Model effect displayed =>

Interaction effect:

"Is the increase of POSTNUMB scores across the levels of VIEWCAT processed at the same rate for the conditions SETTING = 1 ('home') and SETTING = 2 ('school')?"



Visualization of group variances by boxplots => Model assumption displayed

Normality: Do the boxes look roughly symmetric?
(Obs: This is a necessary but not a sufficient condition!)
Homoscedasticity: Do the boxes have approx. the same height?

Main effect A

H_0 : There is no main effect for factor A.
H_a : There is a main effect for factor A.

Main effect B

H_0 : There is no main effect for factor B.
H_a : There is a main effect for factor B.

Interaction effect

H_0 : There is no interaction effect $A \times B$.
H_a : There is an interaction effect $A \times B$.

=> ANOVA testing signif. of 3 effects

$$\text{SST} = \underbrace{\text{SSA} + \text{SSB} + \text{SSAB}}_{\text{explained}} + \underbrace{\text{SSE}}_{\text{error}}$$

=> Factor A with I levels; Factor B with J levels

Source	SS	df	MS	F
Factor A	SSA	$I-1$	SSA/df$_A$	MSA/MSE
Factor B	SSB	$J-1$	SSB/df$_B$	MSB/MSE
Factor $A \times B$	SSAB	$(I-1)(J-1)$	SSAB/df$_{AB}$	MSAB/MSE
Error	SSE	$n-IJ$	SSE/df$_E$	
Total	SST	$n-1$	Var(y)	

Relation of SS's one- and two wav ANOVA:

- The SST's coincide, of course.
- If all n_{ij} are equal: $\text{SSA}_{\text{one way}} = \text{SSA}_{\text{two way}}$.
- Otherwise: SSA and SSE depend on the so-called types of SS
  - ✓ In Stats II we will always use Type III SS (SPSS' default).
  - ✓ More on this topic: Statistics III.
- In general: The SSE is reduced when factor B is added.

GLM Output: Tests of Between-Subjects Effects

Dependent Variable: postnumb

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
viewcat	7811,239	3	2603,746	20,070	,000
setting	1092,095	1	1092,095	8,418	,004
viewcat * setting	647,633	3	215,878	1,664	,176
Error	30098,196	232	129,734		
Corrected Total	39438,296	239			

a. R Squared = ,237 (Adjusted R Squared = ,214)

Effects being tested

Main effects: Both significant.
Interaction effect: Not significant.

Rejecting H_0 implies:

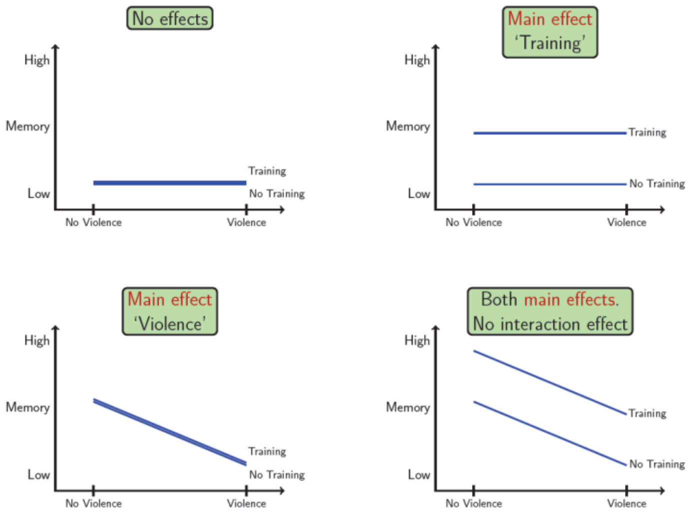
- Significant difference between groups.
- At least one groups is different from the others.

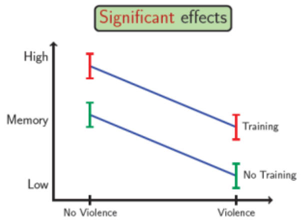
Further investigation is required:

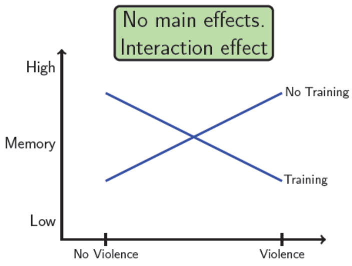
- Visually — use plots
- Perform statistical inference
  - Planned comparisons: Contrasts (more details in Stats III)
  - Post hoc comparisons: Multiple comparisons (only for main effects)

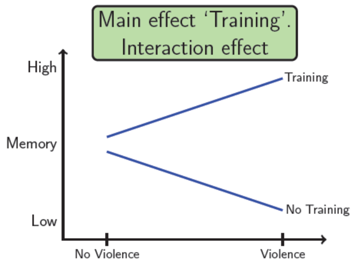
To figure out, which group/factor!!

CIs (based on s_p) help looking for differences

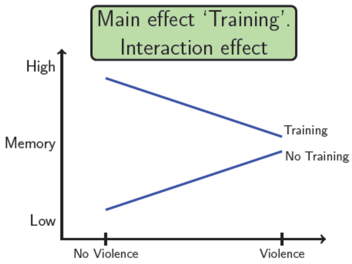


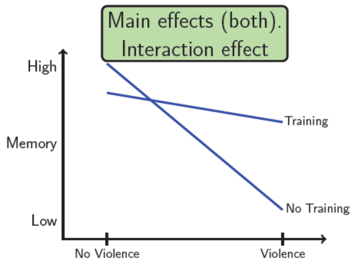






Take within-groups variance into account (sampling variation).





$$\boxed{\omega^2 < \eta^2 < \eta_p^2}$$

Effect Size (ANOVA)

Small effect: $\omega^2 < .06$
Medium effect: $.06 \leq \omega^2 < .14$
Large effect: $.14 \leq \omega^2$

There are several effect size indices for ANOVA models.

- Eta squared (η^2)
  Proportion of the total sample variance explained by the effect.

$$\eta^2 = \frac{SS_{\text{effect}}}{SST}$$

✓ $\eta^2 = R^2$ in one way ANOVA.
✓ **Advantage:** Effects are additive for balanced designs (same n per group).

$$\sum_{\text{All effects}} SS_{\text{effect}} = SSM$$

✓ **Disadvantage:**
  - η^2 depends on the number and size of the remaining effects.
  - η^2 does not estimate the proportion of variance accounted for in the population.

- Partial eta squared (η_p^2)
  Proportion of the (effect + error) sample variance explained by the effect.

$$\eta_p^2 = \frac{SS_{\text{effect}}}{SS_{\text{effect}} + SSE}$$

✓ $\eta_p^2 = R^2$ in one way ANOVA.
✓ **Advantage:** η_p^2 does not depend so much on the remaining effects.
✓ **Disadvantage:**
  - Effects are not additive for balanced designs.
  - η_p^2 does not estimate the proportion of variance accounted for in the population.

η_p^2 is the only effect size index reported by SPSS.

- Omega squared (ω^2)
  Estimate of the proportion of the variance explained by the effect in the population (no repeated measures, balanced design).

$$\omega^2 = \frac{SS_{\text{effect}} - df_{\text{effect}} \times MSE}{MSE + SST}$$

✓ **Advantage:** It no longer overestimates the population effects, as η^2 and η_p^2 did.
✓ **Disadvantage:** Effects are not additive.

Omega squared is an estimate of the dependent variance accounted for by the independent variable in the population for a fixed effects model

Dependent Variable: postnumb

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
viewcat	7811,239	3	2603,746	20,070	,000	,206
setting	1092,095	1	1092,095	8,418	,004	,035
viewcat * setting	647,633	3	215,878	1,664	,176	,021
Error	30098,196	232	129,734			
Corrected Total	39438,296	239				

a. R Squared = ,237 (Adjusted R Squared = ,214)

E.g., for 'viewcat':

- $\eta^2 = \frac{SS_{\text{effect}}}{SST} = \frac{7811.239}{39438.296} = .198$
- $\eta_p^2 = \frac{SS_{\text{effect}}}{SS_{\text{effect}}+SSE} = \frac{7811.239}{7811.239+30098.196} = .206$
- $\omega^2 = \frac{SS_{\text{effect}} - df_{\text{effect}} \times MSE}{MSE+SST} = \frac{7811.239 - 3 \times 129.734}{129.734 + 39438.296} = .188$

Report finding:

The effect of 'viewcat' on 'postnumb' is statistically significant, $F(3, 232) = 20.07$, $p < .001$, $\omega^2 = .19$.

- ANOVA also possible with more than two factors.
- The underlying principles are unchanged.
- The 'interaction' concept is extended.
  E.g., with three factors A, B, and C:
  ✓ Second-order effects: $A \times B$, $A \times C$, $B \times C$
  ✓ Third-order effects: $A \times B \times C$.
- Higher-order effects (say, above 3rd order) are hard to interpret. Avoid if possible.

Regression analysis (Info about RS by linear model)

Predictors (x-variables): May be
✓ Continuous (as in Lectures 7 through 10)
✓ Categorical (code variables; Lecture 12)
✓ Some continuous, some categorical (Statistics III)

Dependent variable (y-variable): Continuous.

Simple linear regression (SLR):
Only one continuous predictor is used.

- SLR tries to model a linear relationship between x and y.
- Therefore, there is a strong connection between regression and correlation.

$$r = \frac{cov(x,y)}{sd(x)sd(y)} = \frac{1}{n-1} \sum_i \left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

r = x and y are continuous

Careful: r is sensitive to outliers.

WHY?

Step 1: Recall the two-sample t test.

- Two populations: $y_1 \sim N(\mu_1, \sigma)$, $y_2 \sim N(\mu_2, \sigma)$. Parameters μ_1, μ_2, and σ unknown. Same σ assumed.
- Take one sample from each population; sample sizes n_1 and n_2.

$$\boxed{H_0 : \mu_1 = \mu_2 \text{ versus } H_a : \mu_1 \neq \mu_2}$$

Step 2: Recall ANOVA.

- Two or more populations: $y_i \sim N(\mu_i, \sigma)$, $i = 1, \ldots, I$. Parameters $\mu_1, \ldots, \mu_I$, and σ unknown. Same σ assumed.
- Take one sample from each population; sample sizes n_i.

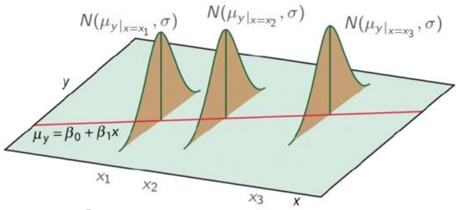
$$\boxed{H_0 : \mu_1 = \mu_2 = \ldots = \mu_I \text{ versus } H_a : \text{Not all } \mu\text{'s are equal}}$$

Step 3: Let's stretch a bit more the ANOVA's model assumptions.

- Admit that, for each value of x (say x_*), there is an associated subpopulation of y values:

$$y|_{x=x_*} \sim N(\mu_{y|_{x=x_*}}, \sigma).$$

All $\mu_{y|_{x=x_*}}$'s and σ are unknown. Same σ assumed.
- Also, admit that the $\mu_{y|_{x=x_*}}$'s are linearly related with x.



The SLR assumes a linear relationship between x and μ_y in the population.

So, the population regression equation is:

$$\boxed{\mu_y = \beta_0 + \beta_1 x}$$

- μ_y: population mean y-score conditional on x.
- β_0: population intercept, i.e., the mean value of y when $x = 0$.
- β_1: population slope, i.e., the change rate of μ_y when x increases 1 unit.

Assumptions:

- Given x, the y values are normally distributed: $y|_x \sim N(\mu_{y|_x}, \sigma)$.
- The spread of the y values is the same for all subpopulations (i.e., same σ).

So, individual y scores spread around the mean μ_y according to the value of σ:

$$\boxed{y_i = \underbrace{\beta_0 + \beta_1 x_i}_{\mu_{y_i}} + \varepsilon_i} \longrightarrow \varepsilon_i \sim N(0, \sigma) \text{ (unrelated to } x\text{)}$$

The intercept β_0.

The slope β_1.

The SD σ.

Estimation of $\boxed{y_i = \underbrace{\beta_0 + \beta_1 x_i}_{\mu_{y_i}} + \varepsilon_i}$ by using data from sample => fact about a sample to estimate the truth about the whole population!!! => Computation of regression line!

Data = Model + Error

$$\underbrace{y_i = \beta_0 + \beta_1 x_i + \varepsilon_i}_{\text{Population}} \longrightarrow \underbrace{y_i = b_0 + b_1 x_i + e_i}_{\text{Sample}}$$

b_0: Sample estimate of β_0

b_1: Sample estimate of β_1

s: Sample estimate of σ

Computation from sample

Ordinary least squares (OLS) method

Find b_0, b_1 that minimize the sum of squared distances between the observations and the regression line:

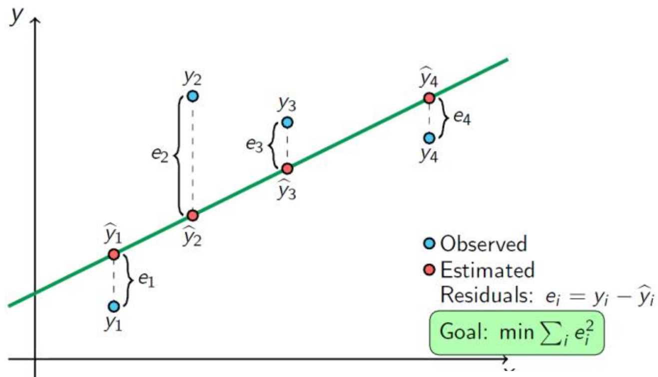
$$\min \sum_i e_i^2 = \min \sum_i (y_i - \hat{y}_i)^2 = \min \sum_i [y_i - (b_0 + b_1 x_i)].$$

Mathematical solution:

$$\boxed{b_1 = r_{xy} \frac{s_y}{s_x}} \qquad \boxed{b_0 = \overline{y} - b_1 \overline{x}}$$

Having b_0 and b_1 computed, s can be computed too:

$$\boxed{s^2 = \frac{\sum_i e_i^2}{n-2} = \frac{\sum_i (y_i - \hat{y}_i)^2}{n-2}}$$



Descriptive Statistics

	Mean	Std. Deviation	N
POSTLET	26.6958	13.2723	240
POSTNUMB	30.0542	12.8458	240

Correlations

		POSTLET	POSTNUMB
POSTLET	Pearson Correlation	1.000	.818
	Sig. (2-tailed)	.	.000
	N	240	240
POSTNUMB	Pearson Correlation	.818	1.000
	Sig. (2-tailed)	.000	.
	N	240	240

Coefficients[a]

Model		Unstandardized Coefficients		Stand. Coef.			95% Confidence Interval for B	
		B	S.E.	Beta	t	Sig.	Lower	Upper
1	(Constant)	1.290	1.258		1.025	.306	-1.189	3.768
	POSTNUMB	.845	.039	.818	21.953	.000	.769	.921

a. Dependent Variable: POSTLET

$$\boxed{b_1 = r_{xy} \frac{s_y}{s_x} = .818 \times \frac{13.27}{12.85} = .845}$$

$$\boxed{b_0 = \overline{y} - b_1 \overline{x} = 26.7 - .845 \times 30.05 = 1.290}$$

Model Summary

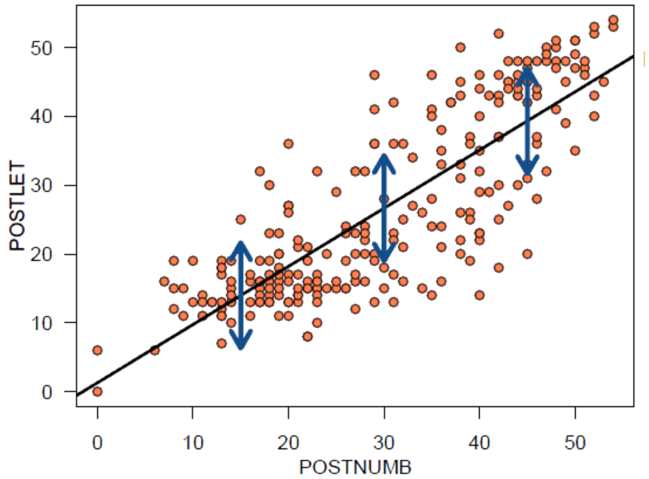
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.818[a]	.669	.668	7.6471

a. Predictors: (Constant), POSTNUMB

r(POSTNUMB, POSTLET); better, it is $r(y, \hat{y})$

percentage of explained variance → one measure of model fit

$$\boxed{s^2 = \frac{\sum_i e_i^2}{n-2} = 7.65^2} \rightarrow \text{estimate of } \sigma^2$$



$s \simeq 7.5 \longrightarrow$ length arrows $\simeq 15$

s is an estimate of the variability about the population regression line.

- The estimation of a regression model typically varies from sample to sample.
  (e.g., b_1 is 'just' one guess at the true value $\beta_1 \ldots$)
- Researchers often want more:

  What about the population regression model?

- The answer is inference.

In statistics and optimization, **statistical errors** and **residuals** are two closely related and easily confused measures of the deviation of an observed value of an element of a statistical sample from its "theoretical value". The **error** of an observed value is the deviation of the observed value from the (unobservable) *true* function value, while the **residual** of an observed value is the difference between the observed value and the *estimated* function value.

Inference in regression models depends on crucial assumptions:

- The residuals are normally distributed with equal SD σ:

$$\varepsilon_i \sim N(0, \sigma).$$

- The residuals are independent from x.

If these asumptions are met, it can be shown that the sampling distributions of b_0 and b_1 are also normal distributions:

$$b_0 \sim N(\beta_0, \sigma_{b_0}) \qquad b_1 \sim N(\beta_1, \sigma_{b_1})$$

Problem: σ_{b_0} and σ_{b_1} are unknown, because they depend on σ (the SD of the residuals in the population).
Solution: Use s (from the sample) instead of σ.

Consequences:

- $\sigma_{b_0} \xrightarrow[\sigma \to s]{} \mathrm{SE}_{b_0}$
- $\sigma_{b_1} \xrightarrow[\sigma \to s]{} \mathrm{SE}_{b_1}$

Given by SPSS

- Normal distributions are replaced by $t(n-2)$ (same df's as MSE!)

	β_0	β_1
CI	$b_0 \pm t^*_{n-2}\mathrm{SE}_{b_0}$	$b_1 \pm t^*_{n-2}\mathrm{SE}_{b_1}$
Test	$H_0: \beta_0 = 0$ vs $H_a: \beta_0 \neq 0$	$H_0: \beta_1 = 0$ vs $H_a: \beta_1 \neq 0$
	$t = \frac{b_0}{\mathrm{SE}_{b_0}} \sim t(n-2)$	$t = \frac{b_1}{\mathrm{SE}_{b_1}} \sim t(n-2)$

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients S.E.	Stand. Coef. Beta	t	Sig.	95% Confidence Interval for B Lower	95% Confidence Interval for B Upper
1	(Constant)	1.290	1.258		1.025	.306	-1.189	3.768
	POSTNUMB	.845	.039	.818	21.953	.000	.769	.921

a. Dependent Variable: POSTLET

SE_{b_0}

$t = 1.290/1.258$

SE_{b_1}

$t = .845/.039$

$.845 \pm 1.96 \times .039$

β_1 lies with 95% certainty between .769 and .921.
β_1 differs significantly from 0.