Observational cross sectional sample survey (hypothetical data)

Sex and salaries[1]

Background:	There is concern about sex equity in salaries (wage gap)[2]
Population:	Software engineers from a large software company
Hypothesis:	Mean annual salary for male and female differ
H0:	There is no difference in mean annual salary between sexes.
Sampling Design:	A stratified (38 male and 38 female) sample of total 76 employees was drawn randomly from 1220 engineers of promotion level A
Recorded:	Actual annual salary (salary) , Sex (sex) and Job experience (exp) in years.

Exercise

Part 1

1. Open **Salary Sex JobExperience ND.sav**, label, label value
2. Start with **Boxplots** (salary vs. sex ; salary vs. job experience; job experience vs. sex)

---

3. Choose a simple one factorial model: **salary sex**
4. Formulate the model and the null hypothesis
5. Test the null hypothesis

---

6. Repeat 3-5 for two simple one factorial models: **salary exp** and **exp sex**

7. Choose an appropriate model to **adjust for exp**
8. ANOVA / ANCOVA
9. Formulate the model and the null hypothesis
10. Test the null hypotheses
11. What conclusions can be made concerning the population

Part 2

12. Repeat the analysis with data from **Salary Sex JobExperience CD.sav**

---

[1] Modified example from Raabe-Hesketh S (2008). Multilevel and Longitudinal Modeling Using Stata. Stata Press, pp. 20-25.
[2] Background: e.g. Brown E et al. (2007). Sex and salaries at the University of Manitoba. http://www.cerforum.org/conferences/200705/papers/BrownPrenticeTroutt.pdf.

Experimental study (hypothetical data)

Undisturbed sleeping-duration (USD) under Baldrian (Valerian)

Baldrian 2 groups

Observation Valerian (Baldrian) has been used as a medicinal anti-anxiety herb and sleep aid since the days of the Romans

Model It is unclear which of the numerous compounds is the true "active" – but the combination of compounds appears to work together in the brain in a manner similar to the action of prescription tranquilizers such as Valium and Halcion.

Hypothesis The undisturbed sleep duration (USD) will change under valerian

Null hypothesis There will be no change in the undisturbed sleep period
Mean USD under treatment = Mean USD in control (no treatment)

Sampling
Population All patients with medium sleeping disorder in a clinic
Sampling Design Simple random sample (**n=40**) from population

Experimental design **Two treatment groups:** with and without treatment.
Randomized trial: Assign randomly each 20 patients to one of the two groups.
Duration: One week
Variable: Change of USD = (after one week USD) – (before trial USD)

Statistical test ANOVA (t-test)

Baldrian 3 groups same as above, but:

Sampling
Population All patients with medium sleeping disorder in a clinic
Sampling Design Simple random sample (**n=60**) from population

Experimental design **Three treatment groups**: with and without treatment and placebo
Randomized trial: Assign randomly each 20 patients to one of the three groups
Duration: One week
Variable: Change of USD = (after one week USD) – (before trial USD)

Statistical test ANOVA

Variable	Variable Label	Value Labels
group	Treatment Group	1 No 2 Yes (Baldrian) 3 Placebo
age	Age	
value	Change of USD (hours)	

Variable	Variable Label	Value Labels
ageclass	Age Class	1 <=60 2 61+

Baldrian (cont.)

Part 1

- Import data from **Baldrian2Age.xls**, label, label value
- Make new variable "ageclass"
- Save SPSS-File: **Baldrian2Age.sav**

Part 2

- Overview of data (Descriptive / Boxplots)
- Age could be a confounding variable
  Check whether Age is correlated with Treatment allocation
- Test the hypothesis (simple)
- Enlarge the model by including AgeCat and Interaction
- What is your conclusion?
  Relevance?

Same for Baldrian3Age.xls

Randomized experimental study (hypothetical data)

Weight reduction study

Background:	The long-term benefit of various treatments of patients with overweight Body mass index (BMI) ≥ 33[3] is unknown
Population:	Patients with BMI $\geq$ 33
Therapy:	Two treatments (diet, diet + psychotherapy) were tested against control.
H0:	There is no difference in long-term weight reduction
Sampling Design:	A stratified sample (15 male and 15 female) of total 30 patients with BMI $\geq$ 33 was randomly drawn from from a hospital's file.
Experimental Design:	Each 5 female and 5 male of this sample were randomly allocated to one of the three treatment
Recorded:	Treatment-group; sex; height (cm) at start; weight at start (kg); weight after one year (kg)

Exercise

Part 1

1. **Weight NI3 Start0.sav:** label, value label and measure
2. Add new variables WeightDelta, WeightDeltaPercent, BMIStart, BMIEnd, BMIDelta, BMIDeltaPercent
3. Save SPSS-File: **Weight NI3.sav**

Part 2

4. Descriptives (Tables / Boxplots)
5. Is BMIStart similar distributed in therapy groups ?

Part 3

6. Formulate a model
7. Test hypothesis
8. Discussion:
    a. Relevance?
    b. Generalization?

Part 4

9. Repeat part 2 and part 3 with data from **Weight I3 Start0.sav**

---

[3] BMI (Body mass index) : = kg/m²
≤ 18.5 underweight ; (18.5; 25] normal; (25;30] overweight; 30+ obesity

Weight reduction study (cont.)

Variable	Label
Group	Treatmentgroup (for 6 month)
Sex	Sex
Height	Height
Weight0	Weight Start
Weight1	Weight End (after 1 year)

Value		Label
Group	1,00	Control
	2,00	Diet
	3,00	Diet + Therapy
Sex	1,00	male
	2,00	female

New variables:

Variable	Variable Label
WeightD	Weight Delta = weight1- weight0
WeightDP	WeightDeltaPercent = weightD/weight0
BMI0	BMI Start
BMI1	BMI End
BMID	BMI Delta=BMI1-BMI0
BMIDP	BMI Delta Percent = BMID/BMI0