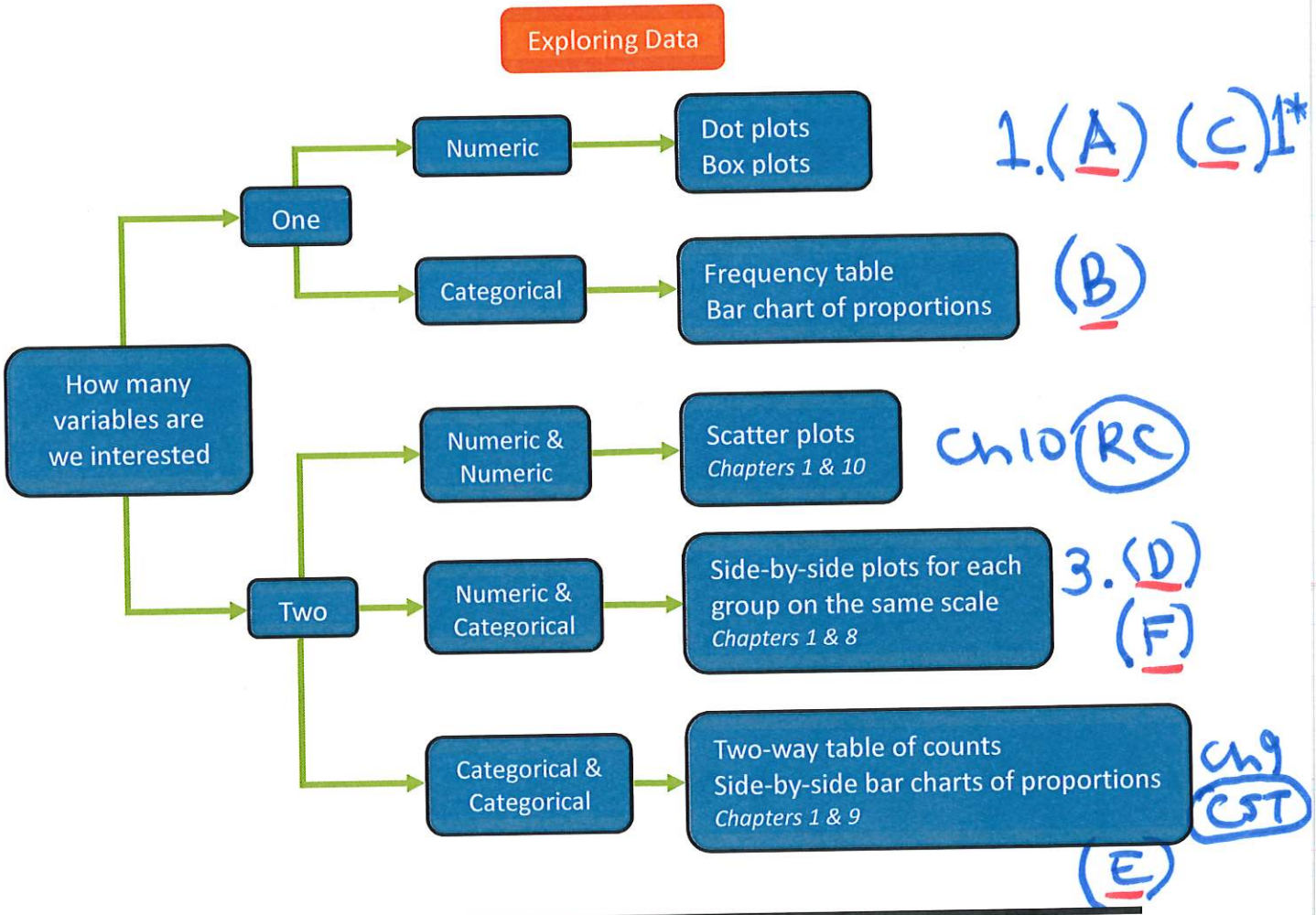
From page 24 of the notes...

One, Two and Three or More Means; Using SPSS

We will now use SPSS to consider:

- One sample (A)
- Two independent samples (D) 3.
- Paired data comparisons (C) 1*
- More than 2 samples (F)

(special case → 1)



1. (A) (C) 1*

(B)

Ch10 (RC)

3. (D) (F)

Ch9 (CST) (E)

Code		Form of analysis
1	A	One sample *t*-test on a mean → 1 num var, 1 sample
2	B	One sample *t*-test on a proportion → 1 cate var, 1 sample
1*	C	One sample *t*-test on a mean of differences → 2 num vars, 1 sample
3	D	Two sample t-test on a difference between two means → 1 num var, 1 cate var (2 levels), 2 samples
4	E	*t*-test on a difference between two proportions → 2 cate vars (either 1 or 2 samples)
F	F	One-way analysis of variance *F*-test → 1 num var, 1 cate var (3+ levels) → 3+ samples

Ch9 Chi-sq test for indep

Ch10 Regression & correlation (simple linear regression)

Checking the Normality assumption

✓ You should make sure data don't show separation into clusters or have a multi-modal nature and then apply the 15-40 rule as follows:

Sample Size Guidelines – "15 – 40 Guide"

Small (total $n \leq 15$ or so)	Medium ($15 <$ total $n < 40$)	Large (total $n \geq 40$ or so)
no outliers	no outliers	no gross outliers
at most, slight skewness	not strongly skewed	data may be strongly skewed

✓ The one sample/paired *t*-test and CIs are reasonably robust against non-Normality, but sensitive to outliers in small-medium samples. The two sample *t*-test & CIs and the *F*-test & Tukey pairwise CIs are very robust against non-Normality. The *F*-test is reasonably robust with respect to the standard deviations assumption, but the Tukey pairwise CIs are not.

Q. Which **one** of the following statements about the validity of confidence intervals of the form sample mean ± *t* standard errors is **false**?

(1) It is critical that the sample is random.
(2) It is critical that the distribution being sampled is Normal.
(3) It is critical that the observations come from the same distribution.
(4) Outliers and clusters of data can invalidate confidence intervals.
(5) It is critical that observations are independent.

R. Which **one** of the following statements about *t*-tests is **false**?

(1) *t*-tests may not be valid if there are outliers present and the sample is not large.
(2) *t*-tests may not be valid when the data show clustering.
(3) In general, *t*-tests are not robust against the Normality assumption.
(4) *t*-tests will generally work well for any large sample.
(5) *t*-tests may not be valid if the data are clearly skewed and the sample is not large.

S. Before conducting formal tests, one should look at plots of the data. Which **one** of the following statements is **false**?

(1) Plots may highlight strange or interesting features of the data which cannot be seen in a formal test.
(2) Summaries of the important features of the data can often be obtained from looking at plots.
(3) Plots are used to check the validity of the assumptions for the formal tests.
(4) Inferences, i.e. conclusions about the population, drawn from plots do not need to be verified by formal tests.
(5) Additional points of interest are often suggested by plots.

3. 2-sample *t*-tests

- ✓ Two independent samples
- ✓ Parameter = $\mu_1 - \mu_2$
- ✓ **Hypotheses:** H_0: $\mu_1 - \mu_2 = 0$ vs H_1: $\mu_1 - \mu_2 \neq 0$

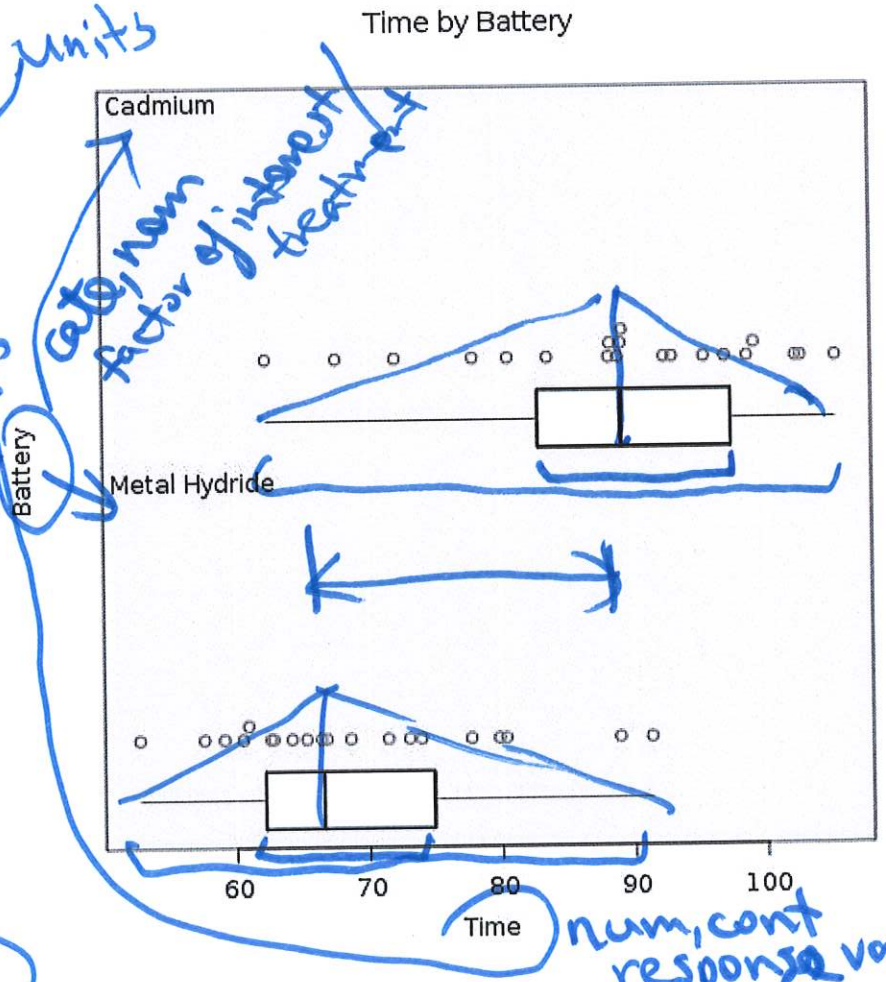
All four datasets used on pages 17 to 26 are available here:

www.tinyURL.com/stats-HTS

as is a hand-out on how to generate the output in SPSS.

- ✓ **Assumptions** for 2-sample *t*-tests
  1. Observations *within* the samples are **independent** – *CRITICAL!*
  2. The two samples/groups are **independent**, i.e., observations *between* samples/groups are independent of each other – *CRITICAL!*
  3. **The Normality Assumption**: The population or underlying distributions are Normal. No clusters or multi-modes allowed.

Example: Which cell phone battery is better? A random sample of 40 cellphones of the same make and model were chosen. Half of the cellphones were randomly selected to have a nickel-cadmium battery put in them and the rest had a nickel-metal hydride battery. The talk time (in minutes) before the batteries needed to be recharged was recorded.



Cadmium cellphone batteries last longer on average than metal hydride cellphone batteries. The talk times for both battery types have similar variability. The talk times for cadmium cellphone batteries are slightly / moderately / strongly negatively (left) skewed / reasonably symmetric / positively (right) skewed while the talk times for metal hydride cellphone batteries are slightly / moderately / strongly negatively (left) skewed / reasonably symmetric / positively (right) skewed.

1 Let μ_C be the underlying mean talk time for cadmium cellphone batteries and μ_M be the underlying mean talk time for metal hydride cellphone batteries, i.e. $\mu_C - \mu_M$ is the difference in underlying mean talk time between the 2 types of battery.

2 H_0: $\mu_C - \mu_M = 0$ vs 3 H_1: $\mu_C - \mu_M \neq 0$

T-Test

Steps 4, 5, 6, 8

Group Statistics

	Battery	N	Mean	Std. Deviation	Std. Error Mean
Time	Cadmium	20	88.7050	11.76066	2.62976
	Metal Hydride	20	69.1900	10.30528	2.30433

t_0 = est − hyp val / std err = $(\bar{x}_C - \bar{x}_M - 0) / se(\bar{x}_C - \bar{x}_M)$ = 19.515 / 3.49651

Ch 7 df = 19

Step 4

$se(\bar{x}_C - \bar{x}_M)$

Independent Samples Test

		Levene's Test for Equality of Variances: F	Sig.	t-test for Equality of Means: t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference: Lower	Upper
Time	Equal variances assumed	.079	.780	5.581	38	.000	19.51500	3.49651	12.43668	26.59332
	Equal variances not assumed			5.581	37.356	.000	19.51500	3.49651	12.43267	26.59733

p-val

$\bar{x}_C - \bar{x}_M$

Step 8

est ± t* × se(est)

Step 6

7. We have no / weak / some / strong / very strong evidence (*P-value* = .000) against there being no difference in the underlying mean talk time between the two types of cellphone batteries, that is, we have no / weak / some / strong / very strong evidence of a difference in the underlying mean talk time between the two types of cellphone battery.

9. With 95% confidence, we estimate that, the underlying mean talk time for a cadmium cellphone battery is, on average, somewhere between 12 and 27 minutes less / more than that of a metal hydride battery.

There are no doubts about the validity of the *t*-test. Because a random sample of 40 cellphones of the same make and model was taken, there are no concerns about the independence assumptions. In this two independent samples situation, $n_1 + n_2 = 40$ and, with respect to the Normality assumption, *t*-procedures work well for this number of observations despite the talk times being somewhat skewed.

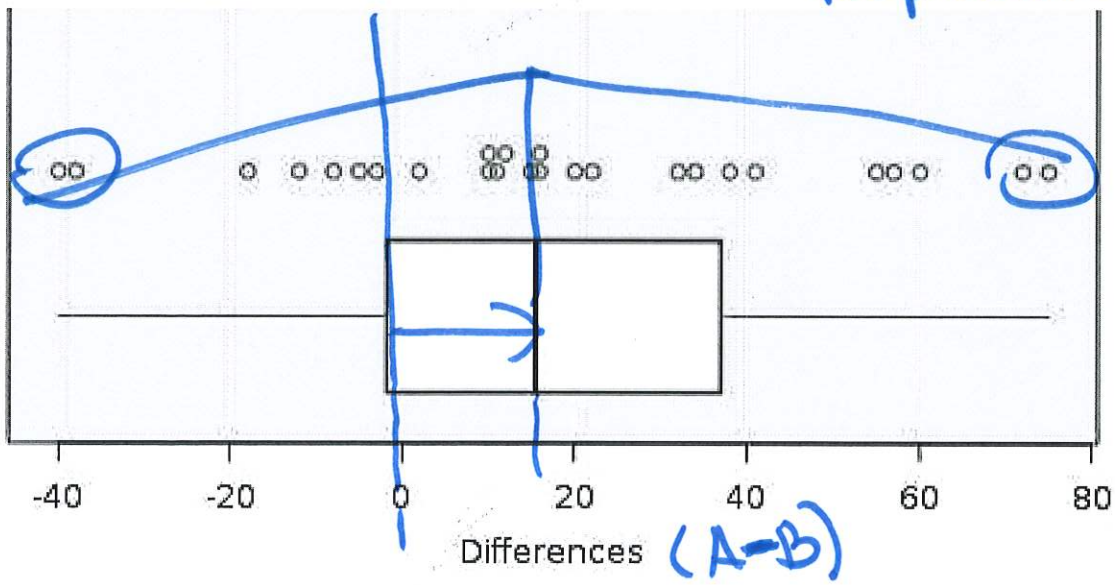
Paired data: use a paired data *t*-test (c) [Special case of one sample/single mean]

- ✓ 1 sample (group) of data
- ✓ Two measurements taken on each experimental unit
- ✓ With related or paired data we analyse the differences and use a 1-sample *t*-test, i.e. we treat the differences as a single sample.
- ✓ Parameter = μ_{diff}
- ✓ **Hypotheses**: H_0: $\mu_{diff} = 0$
  vs H_1: $\mu_{diff} \neq 0$
- ✓ **Assumptions** for the paired-*t*-test:
  1. Observations (differences) *within* the sample are **independent** – *CRITICAL!*
  2. **The Normality Assumption**: The population or underlying distribution of the differences is Normal. No clusters or multi-modes allowed.

Example: A market research company is interested in which of two similar electric shavers, model A or model B, is preferred by consumers. 26 men who daily use an electric shaver, but not one of the models of interest are randomly selected to participate in the study. Half the men were randomly allocated to use model A one morning followed by model B the next morning whilst the order was reversed for the remaining men. After every shave, each man completed a questionnaire rating his satisfaction with the shaver. Satisfaction was measured as a score based on the answers to the questionnaire and is given in a range from 1 to 100. (Larger scores indicate greater satisfaction).

units

response var (diffs)



The difference in the satisfaction score is centred above / ~~below~~ zero suggesting that the average satisfaction score for model A is higher / ~~lower~~ than that for model B. The difference in average satisfaction score is ~~slightly~~ / ~~moderately~~ / ~~strongly negatively (left) skewed~~ / reasonably symmetric / ~~positively (right) skewed~~.

1. Let μ_{diff} be the underlying mean difference in the average satisfaction score for the model A electric shaver and the average satisfaction score for the model B electric shaver. (Note: Diff = model A – model B)

2. H_0: $\mu_{diff} = 0$ vs 3. H_1: $\mu_{diff} \neq 0$

$t_0 = \frac{\bar{x}_{diff} - 0}{se(\bar{x}_{diff})} = \frac{18.231}{5.955}$

T-Test

Step 4,5,6,8

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean ($\bar{x}_{diff}$)	Std. Deviation	Std. Error Mean ($se(\bar{x}_{diff})$)	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Model A - Model B	18.231	30.362	5.955	5.967	30.494	3.062	25	.005

Step 4 Step 5 Step 3 Step 6 p-val

7. We have ~~no~~ / ~~weak~~ / ~~some~~ / strong / ~~very strong~~ evidence (*P-value* = .005) that there is a shaver effect on the average satisfaction score for the model A electric shaver and the average satisfaction score for the model B electric shaver.

9. With 95% confidence, we estimate that, the average satisfaction score for the model A electric shaver is, on average, somewhere between 6.0 and 30.5 units ~~less~~ / greater than that for the average satisfaction score for the model B electric shaver.

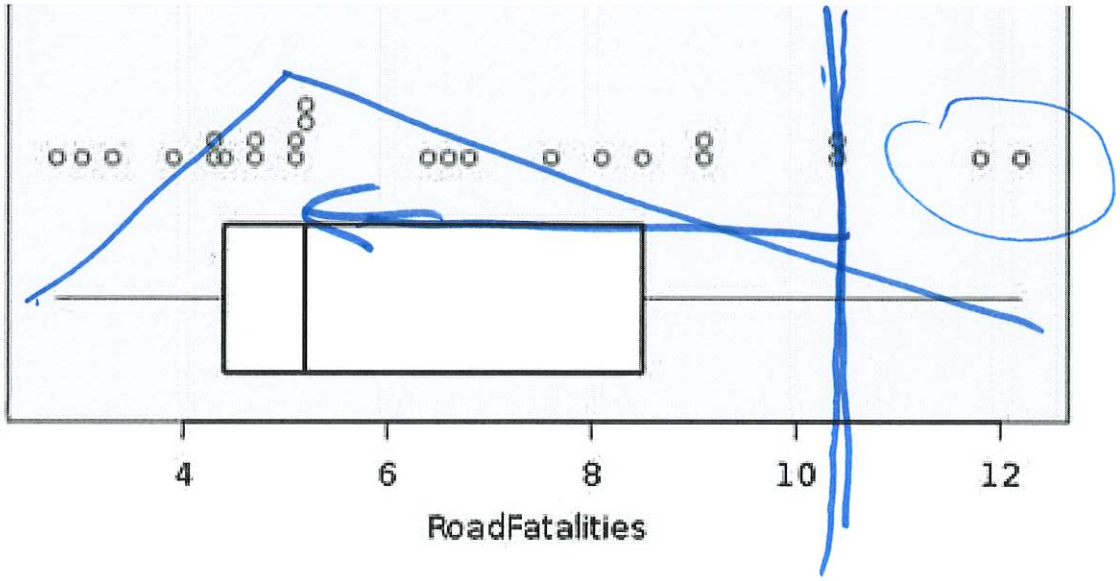
There are no doubts about the validity of the *t*-test. Because a random sample of 26 men was taken, there are no concerns about the independence assumption. In this paired data situation, $n = 26$ and, with respect to the Normality assumption, *t*-procedures work well for this number of observations as the differences are reasonably symmetric.

1. 1-sample: use a One Sample *t*-test (A)

- ✓ 1 sample (group) of data
- ✓ Parameter = μ
- ✓ **Hypotheses:** H_0: $\mu = \mu_0$ vs H_1: $\mu \neq \mu_0$ — hyp. val. is some # from story
- ✓ **Assumptions** for 1-sample *t*-tests
  1. Observations *within* the sample are **independent** – *CRITICAL!*
  2. **The Normality Assumption**: The population or underlying distribution is Normal. No clusters or multi-modes allowed.

Example: Of interest is whether the most recent road fatalities per capita per year for 25 randomly selected countries has changed from the historical average of 10.5 per 100,000 inhabitants per year (for the same 25 countries in the mid-eighties). The data was collected by the World Health Organization (WHO).

n — units — hyp val



The annual road fatalities per 100,000 inhabitants is centred at about 5. Road fatalities ranges from roughly 3 to 12 per 100,000 inhabitants and is `slightly / strongly negatively (left) skewed / reasonably symmetric / positively (right) skewed.`

1. Let μ be the underlying mean annual road fatalities per 100,000 inhabitants

2. H_0: $\mu = 10.5$ vs 3. H_1: $\mu \neq 10.5$ [3. $\mu < 10.5$]

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Road fatalities per 100,000 inhabitants per year	25	6.5180	2.76567	.55313

Step 4 $\bar{x}$ $se(\bar{x})$ Step 5

$t_0 = \frac{6.5180 - 10.5}{.55313}$

One-Sample Test

	Test Value = 10.5					
					95% Confidence Interval of the Difference	
	t_0	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Road fatalities per 100,000 inhabitants per year	-7.199	24	.000	-3.98200	-5.1236	-2.8404

pval

Step 5 $n-1$ Step 6 $\bar{x} - 10.5$

(5.3764, 7.6596) ✓

7. We have ~~no~~ / ~~weak~~ / ~~some~~ / ~~strong~~ / very strong evidence (*P-value* = 0.000) that the underlying mean annual road fatalities per 100,000 inhabitants has changed compared to the historical average of 10.5 per 100,000 inhabitants per year.

9. With 95% confidence, we estimate that the underlying mean annual road fatalities is somewhere between 5.4 and 7.7 per 100,000 inhabitants.

There are no doubts about the validity of the *t*-test. Because a random sample of 25 countries was taken, there are no concerns about the independence assumption. In this one-sample situation, $n = 25$ and, with respect to the Normality assumption, *t*-procedures work well for this number of observations as the data is slightly positively (right) skewed.

T, U, V, W then break! 2.40-3pm...

T. Thirty observations of the relative return of over-the-counter stocks bought in the week of the 9th to the 13th of May, 1994 are given below.

-0.2940	-0.1092	-0.1053	-0.0707	-0.0563
-0.0541	-0.0423	-0.0398	-0.0396	-0.0390
-0.0381	-0.0323	-0.0221	-0.0169	-0.0139
-0.0081	0.0038	0.0057	0.0156	0.0172
0.0182	0.0192	0.0423	0.0459	0.0476
0.0667	0.0714	0.0780	0.1176	0.1224

Note: $\bar{x} = -0.01030$ and $s = 0.0786$

Investors would like to know how the market performed. One measure of market performance is the mean relative return for the week.

A 95% confidence interval for the mean relative return is [-0.040, 0.019]. Which **one** of the following statements is **false**?

(1) The P-value for testing $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$ is larger than 0.05.

(2) There is evidence, at the 5% level, to believe that the mean return is different from zero.

(3) A 99% confidence interval for the mean return would be wider than the 95% confidence interval.

(4) It is plausible that the mean relative return is zero.

(5) An estimate of the mean relative return is −0.01030.

U. Does too much sleep impair intellectual performance? Taub *et al.* (1971) examined this commonly held belief by comparing the performance of 12 subjects on the mornings following (1) two normal nights' sleep and (2) two nights of "extended sleep". In the morning they were given a number of tests of ability to think quickly and clearly. One test was for vigilance where the lower the score, the more vigilant the subject. The following data was collected:

Subject	1	2	3	4	5	6	7	8	9	10	11	12
Normal Sleep	8	9	14	4	12	11	3	26	3	11	10	1
Extended Sleep	8	9	15	2	21	16	9	38	10	11	16	41

To see if the data supports the view that too much sleep can be bad for you, we would test which of the following hypotheses?

(1) $H_0 : \bar{x}_1 - \bar{x}_2 = 0$ versus $H_1 : \bar{x}_1 - \bar{x}_2 < 0$

(2) $H_0 : \bar{x}_{Diff} = 0$ versus $H_1 : \bar{x}_{Diff} \neq 0$

(3) $H_0 : \mu_1 - \mu_2 = 0$ versus $H_1 : \mu_1 - \mu_2 \neq 0$

(4) $H_0 : p_1 - p_2 = 0$ versus $H_1 : p_1 - p_2 \neq 0$

(5) $H_0 : \mu_{diff} = 0$ versus $H_1 : \mu_{diff} \neq 0$

V. In order to study the harmful effects of DDT poisoning, the pesticide was fed to 6 randomly chosen rats out of a group of 12 rats. The other 6 unpoisoned rats comprised of the control group. The following data gives measurements of the amount of tremor detected in the bodies of each rat after the experiment. The more tremor, the more harmful.

Poisoned group:	12.207,	16.869,	25.050,	22.429,	8.456,	20.589
Control group:	11.074,	12.064,	9.351,	6.642,	9.686,	8.182

We wish to test

(1)	$H_0 : \mu_{diff} = 0$	versus	$H_1 : \mu_{diff} \neq 0$
(2)	$H_0 : \mu_1 - \mu_2 = 0$	versus	$H_1 : \mu_1 - \mu_2 \neq 0$
(3)	$H_0 : \bar{x}_1 - \bar{x}_2 = 0$	versus	$H_1 : \bar{x}_1 - \bar{x}_2 \neq 0$
(4)	$H_0 : p_1 - p_2 = 0$	versus	$H_1 : p_1 - p_2 \neq 0$
(5)	$H_0 : \bar{x}_{Diff} = 0$	versus	$H_1 : \bar{x}_{Diff} \neq 0$

Question W refers to the following information.

The heights (in cm) of the carapaces (shells) of a sample of 48 painted turtles were recorded. Shown below is a stem-and-leaf plot of this data set.

Units: 3|5 = 35cm

3 | 55778888999
4 | 0000111222344
4 | 55566789
5 | 0111113
5 | 567
6 | 01233
6 | 7

Figure: Stem-and-leaf plot of carapace height of painted turtles.

W. Based on this sample of size 48, a 95% confidence interval for the underlying mean carapace height of all painted turtles is 44.0cm to 48.8cm. The number of painted turtle carapace heights we would need to sample in order to halve the width of this interval is, approximately:

(1) 24
(2) 192
(3) 12
(4) 96
(5) 7

double precision! 48 ×4 → 192

double width
halve precision 48 ÷4 → 12