(1) **Comparing Two Populations 1**

A study of the differences in cognitive function between normal individuals and patients diagnosed with schizophrenia was published in the American Journal of Psychiatry (Apr. 2010). The total time (in minutes) a subject spent on the Trail Making Test (a standard psychological test) was used as a measure of cognitive function. The researchers theorize that the mean time on the Trail Making Test for schizophrenics will be larger than the corresponding mean for normal subjects. The data for independent random samples of 41 schizophrenics and 49 normal individuals yielded the following results:

	Schizophrenia	Normal
Sample size	41	49
Mean time	104.23	62.24
Standard deviation	62.24	16.34

(a) Define the parameter of interest to the researchers.

(b) Set up the null and alternative hypothesis for testing the researchers' theory.

(c) The researchers conducted the test, part (b), and reported a p-value of .001. What conclusions can you draw from this result? (Use $\alpha = 0.01$)

(d) Find a 99% confidence interval for the target parameter. Interpret the result. Does your conclusion agree with that of the previous part?

a) We have large sample sizes $n_1 = 41 \geq 30$ and $n_2 = 49 \geq 30$, hence we have approximately $\overline{X_1} \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $\overline{X_2} \sim \mathcal{N}(\mu_2, \sigma_2^2)$. From this we obtain $\overline{X_1} - \overline{X_2} \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

The target parameter is $\mu_1 - \mu_2$. We take the idea from slide 32 of lecture 11.

b) We would choose $H_0: \mu_1 - \mu_0 = 0$ vs. $H_1: \mu_1 - \mu_0 > 0$

The p-value is

$$\mathbb{P}\left(\overline{X_1} - \overline{X_2} \geq \overline{x}_1 - \overline{x}_2\right) = 1 - \mathbb{P}\left(\frac{\overline{X_1} - \overline{X_2}}{s} < \frac{\overline{x}_1 - \overline{x}_2}{s}\right) = 1 - \phi\left(\frac{\overline{x}_1 - \overline{x}_2}{s}\right) \approx 10^{-5} ?\, ?$$

c) Since the p-value is very small, we reject H_0. If we decide to use a confidence level of $\alpha = 1/100$, then we would reject H_0.

$$s := \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

d) We would like to find $d \in \mathbb{R}$, such that

$$1 - \alpha = \mathbb{P}\left(\overline{x}_1 - \overline{x}_2 - d \leq \overline{X_1} - \overline{X_2} < \overline{x}_1 - \overline{x}_2 + d\right) = \mathbb{P}\left(-d \leq \overline{X_1} - \overline{X_2} - (\overline{x}_1 - \overline{x}_2) < d\right)$$

$$= \mathbb{P}\left(\frac{-d}{s} \leq \frac{\overline{X_1} - \overline{X_2} - (\overline{x}_1 - \overline{x}_2)}{s} < \frac{d}{s}\right) = 1 - 2\phi\left(\frac{-d}{s}\right) \Leftrightarrow \frac{-d}{s} = \Phi^{-1}\left(\frac{\alpha}{2}\right) \Leftrightarrow d = -\Phi^{-1}\left(\frac{\alpha}{2}\right)s$$

The interval is given by $[\overline{x}_1 - \overline{x}_2 - d, \overline{x}_1 - \overline{x}_2 + d) \approx [16.24, 67.74)$

The interval does not contain 0, hence we are 99% confident, that the hypothesis H_0 can not be rejected.

(2) **Comparing Two Populations 2**

Suppose you wish to compare a new method of teaching reading to slow learners with the current standard method. You decide to base your comparison on the results of a reading test given at the end of a learning period of six months. Of a random sample of 22 slow learners, 10 are taught by the new method and 12 are taught by the standard method. All 22 children are taught by qualified instructors under similar conditions for the designated six-month period. The results of the reading test at the end of this period are given below.

New Method: 80, 76, 70, 80, 66, 85, 79, 71, 81, 76.

Standard Method: 79, 73, 72, 62, 76, 68, 70, 86, 75, 68, 73, 66.

(a) Use the data in the table to estimate the true mean difference between the test scores for the new method and the standard method. Use a 95% confidence interval.

(b) Interpret the interval you found in the previous part.

(c) What assumptions must be made in order that the estimate be valid? Are they reasonably satisfied?

a) Since the sample sizes $n_1 = 10 < 30$ and $n_2 = 12 < 30$ are small, we use the t-test from slides 33/34 of lecture 11.

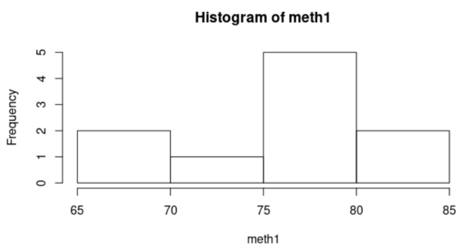
The variances are $s_1^2 = 34.04$ and $s_2^2 = 40.24$, hence we rather don't assume their equality. The 95% confidence interval is given by $[-1.36, 9.49)$

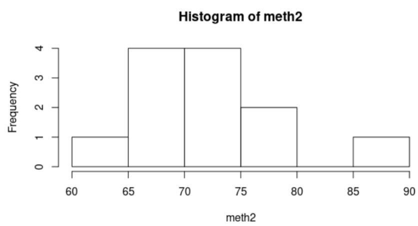
```
> meth1 <- c(80, 76, 70, 80, 66, 85, 79, 71, 81, 76)
> meth2 <- c(79, 73, 72, 62, 76, 68, 70, 86, 75, 68, 73, 66)
>
> t.test(meth1, meth2)

        Welch Two Sample t-test

data:  meth1 and meth2
t = 1.5643, df = 19.769, p-value = 0.1336
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.360091  9.493424
sample estimates:
mean of x mean of y
 76.40000  72.33333
```

b) Since the interval contains 0, we fail to reject the hypothesis $H_0: \mu_1 - \mu_2 = 0$, when compared against $H_1: \mu_1 - \mu_2 \neq 0$ if we choose $\alpha = 1/20$

We can interpret the confidence interval as follows. If we take k samples like the one we just analysed, then we expect that the true parameter (which we estimated here by the mean) lies within the confidence interval 0.95 * k times.

c) We assumed that the samples were independent and that they are normally distributed. Looking at their histograms, it seems fairly reasonable to assume normal distribution.





(3) **Missing Information**

An investigation of ethnic differences in reports of pain perception was presented at the annual meeting of the American Psychosomatic Society (Mar. 2001). A sample of 55 blacks and 159 whites participated in the study. Subjects rated (on a 13-point scale) the intensity and unpleasantness of pain felt when a bag of ice was placed on their foreheads for two minutes. (Higher ratings correspond to higher pain intensity.) A summary of the results is provided in the following table.

	Blacks	Whites
Sample Size	55	159
Mean pain intensity	8.2	6.9

(a) Why is it dangerous to draw a statistical inference from the summarized data? Explain.

(b) What values of the missing sample standard deviations would lead you to conclude (at $\alpha = 0$ .05) that blacks, on average, have a higher pain intensity rating than whites?

(c) What values of the missing sample standard deviations would lead you to an inconclusive decision (at $\alpha = 0.05$) regarding whether blacks or whites have a higher mean intensity rating?

a) It is dangerous, because we were not provided the standard deviation of the sample. If the standard deviation was high, then it might be difficult to draw a statistical inference.

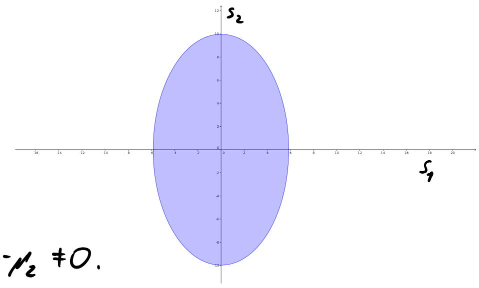
b) Since the sample sizes $n_1 = 55 \geq 30$ and $n_2 = 159 \geq 30$ are large, we choose the test on slide 37 of lecture 11. We test $H_0: \mu_1 - \mu_2 = 0$ vs. $H_1: \mu_1 - \mu_2 > 0$

$$\alpha \geq \mathbb{P}(\bar{X}_1 - \bar{X}_2 \geq \bar{x}_1 - \bar{x}_2) = 1 - \mathbb{P}\left(\frac{\bar{X}_1 - \bar{X}_2}{s} < \frac{\bar{x}_1 - \bar{x}_2}{s}\right) = 1 - \Phi\left(\frac{\bar{x}_1 - \bar{x}_2}{s}\right)$$

$$\Leftrightarrow \Phi\left(\frac{\bar{x}_1 - \bar{x}_2}{s}\right) \geq 1 - \alpha \Leftrightarrow \frac{\bar{x}_1 - \bar{x}_2}{s} \geq \Phi^{-1}(1-\alpha) \Leftrightarrow s \leq \frac{\bar{x}_1 - \bar{x}_2}{\Phi^{-1}(1-\alpha)} \approx 0.7903$$

where $s := \sqrt{s_1^2/n_1 + s_2^2/n_2}$

Since s_1 and s_2 are positive, only these values in the ellipse are possible for s_1 and s_2.



c) Here it seems more reasonable to test $H_0: \mu_1 - \mu_2 = 0$ vs. $H_1: \mu_1 - \mu_2 \neq 0$.

We want

$$\alpha = \mathbb{P}(|\bar{X}_1 - \bar{X}_2| \geq \bar{x}_1 - \bar{x}_2) = \mathbb{P}(\bar{X}_1 - \bar{X}_2 < -(\bar{x}_1 - \bar{x}_2)) + \mathbb{P}(\bar{X}_1 - \bar{X}_2 \geq \bar{x}_1 - \bar{x}_2) = 2\,\Phi\left(\frac{-(\bar{x}_1 - \bar{x}_2)}{s}\right)$$

$$\Leftrightarrow \frac{-(\bar{x}_1 - \bar{x}_2)}{s} = \Phi^{-1}\left(\frac{\alpha}{2}\right) \Leftrightarrow s = \frac{-(\bar{x}_1 - \bar{x}_2)}{\Phi^{-1}\left(\frac{\alpha}{2}\right)} \approx 0.663$$

(4) χ^2**-test for independence**

100 students from major mathematics of three Viennese universities were randomly chosen and asked which lecture, either a: calculus, b: algebra, or c: probability, they enjoyed most. The frequencies are given in the following table:

	Uni A	Uni B	Uni C	
calculus	10	5	5	20
algebra	10	20	10	40
probability	20	5	γ	$25+\gamma$
	40	30	$15+\gamma$	$85+\gamma$

Perform a χ^2-test to test whether the preference for a lecture is independent from the university, on a 5% significance level.

(a) Only use the following table which gives the 95%-quantile q of the χ^2-distribution with *df* degrees of freedom.

df	1	2	3	4	5	6	7	8	9
q	3.84	5.99	7.81	9.49	11.07	12.59	14.07	15.51	16.92

(b) Solve the previous exercise using `R`.

Let us choose $\gamma = 15$

a) we define the Matrix $O := \begin{pmatrix} 10 & 5 & 5 \\ 10 & 20 & 10 \\ 20 & 5 & \gamma \end{pmatrix}$

The total number of students that prefer calculus is 20 and which is $\frac{20}{85+\gamma}$ of all students. Since 40 of these students study at Uni A we expect $e_{11} := 40 \cdot \frac{20}{85+\gamma}$ of the students of Uni A to prefer calculus.

Similarly, we define $e_{ij} := \sum_{k=1}^{3} o_{kj} \cdot \frac{1}{85+c} \sum_{l=1}^{3} o_{il}$ and obtain

$$E = \frac{1}{85+\gamma} \begin{pmatrix} 20 \cdot 40 & 20 \cdot 30 & 20 \cdot (15+c) \\ 40^2 & 40 \cdot 30 & 40\,(15+c) \\ (25+\gamma)\,40 & (25+\gamma)\,30 & (25+\gamma)(15+\gamma) \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 40 & 30 & 30 \\ 80 & 60 & 60 \\ 80 & 60 & 60 \end{pmatrix} = \begin{pmatrix} 8 & 6 & 6 \\ 16 & 12 & 12 \\ 16 & 12 & 12 \end{pmatrix}$$

$$\sum_{i=1}^{3} \sum_{j=1}^{3} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \approx 14.58$$

This sum is approximately $\chi^2(m)$ distributed, where $m = 9-1 = 8$

The 95% quantile of the $\chi^2(8)$ distribution is 15.51 which is larger than 14.58, hence we fail to reject the hypothesis, that the preference does not depend on the university.

b) p-value = 0.068

(5) **Hypnosis**

Some researchers claim that susceptibility to hypnosis can be acquired or improved through training. To investigate this claim six subjects were rated on a scale of 1-20 according to their initial susceptibility to hypnosis and then given 4 weeks of training. Each subject was rated again after the training period. In the ratings below, higher numbers represent greater susceptibility to hypnosis.

Subject	Before	After
1	10	18
2	16	19
3	7	11
4	4	3
5	7	5
6	2	3

Specify and perform the appropriate hypothesis test. What potential issues exist with this analysis?

We can use the sign-test. If X_i is the susceptibility before the training and Y_i the susceptibility after the training, then $K := |\{i \in \{1,\dots,n\} \mid Y_i > X_i\}| \sim Bin(n, \theta)$ for some parameter θ. We test $H_0: \theta = 0.5$ vs. $H_1: \theta > 0.5$. In our case $n=6$.

In the sample above, we have 4 successes. The p-value is

$$\mathbb{P}(K \geq 4) = \sum_{i=4}^{6} \binom{6}{i} 0.5^i \cdot 0.5^{6-i} = \left(\tfrac{1}{2}\right)^6 \left(\binom{6}{4} + \binom{6}{5} + \binom{6}{6}\right) = \frac{1}{64}(15+6+1) = \frac{11}{32}$$

Hence we fail to reject the null Hypothesis. A problem with this test might be that it does not take the magnitude of the improvement in susceptibility into account.

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1