Quick stats review for PSS

Wei Ji Ma

Feb 10, 2020

These notes describe the concepts from statistics that I consider most relevant for understanding psychology research papers.

Contents

1 Mean and SEM

Let's say we have a data set. Our starting assumption that each measurement in our data set is drawn from a larger population. This is almost always the case. For example, if we survey people by phone on a election topic, we are not necessary interested in these specific individuals but in the larger, unobserved population of all registered voters. The terminology "population" comes from demographic studies. In psychology, we are similarly not interested in the specific people we measure but in making more general statements. This is also why data sets are often called *samples*: the measurements could have been sampled from a larger population. The larger population is not directly accessible and we have to draw conclusions about its properties solely based on our sample. This process is called *statistical inference* and the associated quantities are called *inferential statistics*. You can think of statistical inference as sophisticated guesswork.

At first sight, statistical inference seems like a tall order – after all, how can you say something about values you never observed? What comes to the rescue is the assumption that the sample is representative of the population. That is, we assume that each measurement is an independent draw from the population. This assumption allows us to make statements about the population as a whole.

1.1 Population mean and variance

Consider the case that we are interested in a continuous variable x. We further assume that there exists a very large *population* of values, which we will denote by $(x_1, x_2, \ldots, x_N)$. Each measurement in a sample is considered an independent draw from the population. We define the *population mean* or *population average* as

$$\mu \equiv \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{1}$$

which is shorthand for

$$\mu \equiv \frac{1}{N}(x_1 + x_2 + x_3 + \cdots + x_N).$$

(The symbol $\sum$ is called the sum symbol, and $\sum_{i=1}^{N} x_i$ is read as "the sum over i from 1 to N of x_i".) The *population variance* is the average of the squared distance between each data point and the population mean:

$$\sigma^2 \equiv \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2. \tag{2}$$

Population mean and population variance are both fixed but unknown numbers. Population variance is a measure of the spread of the values around the mean. The *population standard deviation*, σ, is the square root of the population variance.

1.2 A small sample

In practice, we never measure from the whole population. Instead we have a small sample $x_1, \ldots, x_n$ consisting of n data points, where n is much smaller than N. These points can be represented as points on a line, as in Figure 1.

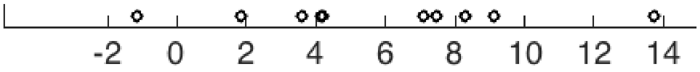

Figure 1: Example of a sample of continuous data.

1.3 Estimating the population mean

Given our small sample of n observations, how can we make a good guess of the population mean μ? Naturally, the *sample mean*, denoted by $\bar{x}$, makes sense:

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

This means that the sample mean – which can also be regarded as a purely descriptive statistic – as an inferential statistic (an estimator). An estimator is denoted by a "hat" (ˆ) above the symbol for the population parameter.

1.4 Estimating the population variance

We now turn to estimating the population variance. Our first try is to consider the sample variance as an estimator of the population variance. The sample variance is defined as

$$\hat{\sigma}^2 = \mathrm{var}_n(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{3}$$

How good an estimator of the population variance is this? It turns out that the sample variance var_n is a *biased* estimator of the true variance; this means that when calculated many times on different samples from the same population, it does on average not return the correct answer. To solve this problem, we introduce the *unbiased* sample variance, where we divide by $n-1$ (the sum is still over all n measurements):

$$\hat{\sigma}^2_{\mathrm{unbiased}} = \mathrm{var}_{n-1}(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{4}$$

The *sample standard deviation* is the square root of the sample variance.

1.5 Standard error of the mean

The *standard error of the mean* (SEM) is the amount of variation in the sample mean that is expected if one were to repeat the experiment infinitely often. The more such variation, the less "reliable" our guess of the population mean is. Stated yet differently, the SEM measures how far the sample mean tends to be away from the population mean. The equation is:

$$\mathrm{SEM} = \frac{\hat{\sigma}_{\mathrm{unbiased}}}{\sqrt{n}}.$$

This says that the SEM is equal to the sample standard deviation divided by the number of measurements. Because of the approximation, the equation should not be used if sample size n is small. You typically want a sample of at least 5 observations.

Exercise: Consider the data set $\{-0.1, 0.2, 0.3, 0.5, 1.0\}$. Calculate the sample mean and the SEM. In Excel, if the data are in cells A1 to A5, use COUNT(A1:A5) to get the number of data points, n. Use AVERAGE(A1:A5) to get the sample mean, and STDEV.S(A1:A5) to get the sample standard deviation. Combining these commands, you can calculate the SEM.

The SEM gives a sense of a spread of the sample mean if the experiment were repeated many times (many data sets). The SEM decreases when the sample size grows larger: as you have more data, you have better knowledge about the population mean. Keep in mind, however, that the sample variance does not change systematically as n increases, because it estimates the population variance, which is an intrinsic property of the population.

The SEM is numerically representing using the symbol ±. For example, one could write "The mean height was 174.3 cm ± 1.1 cm." to indicate that the sample mean was 174.3 cm and the SEM was 1.1 cm. The SEM is graphically represented as an error bar around the sample mean. In fact, most error bars that you will see in graphs in papers will be SEMs.

2 Hypothesis testing

Often in psychology, as in any empirical science, we want to draw a binary conclusion: Is mean performance significantly higher than chance? Does therapy reduce symptoms? Does attentional load affect performance? In other words: does a "treatment" have an "effect"? (This terminology comes from medicine but is widely used outside of it.) Drawing such binary conclusions is called hypothesis testing. The hypothesis that there is no effect, and any apparent difference between groups or conditions in the data is just due to chance, is called the *null hypothesis*. If the observed data (as summarized using a *test statistic*) are very unlikely if we assume the null hypothesis, then we *reject the null hypothesis*. Any standard hypothesis test follows a set of:

Step 1. Identify the hypotheses. The null hypothesis is the "boring" one, that there is no effect, or no difference between conditions. The alternative hypothesis is there is an effect or difference between conditions.

Step 2. Define a *test statistic*, which is a suitably chosen summary of the data (just like the sample mean or the sample variance were summaries of the data). The test statistic is designed with two criteria in mind: to distinguish between the hypotheses (it will tend to be higher if the alternative hypothesis is true than if the null hypothesis is true), and to make Step 3 easy. In practice, we don't have to worry about defining the test statistic, since statisticians have done a lot of work to come up with suitable statistics. What we do need to do in this step is calculate the value of the test statistic for the data.

Step 3. Calculate the theoretical distribution of the test statistic *if the null hypothesis were true* ("under the null hypothesis"). In most cases, this distribution will only depend on the number of measurements (the sample size). *The data themselves are not used in this step!*

Step 4. Using the distribution of the test statistic from Step 3, calculate the probability that under the null hypothesis, the test statistic takes a value greater than the observed value of the test statistic from Step 2. This probability is called the p-value and it is a measure of how extreme the observed data are under the null hypothesis. If the p-value is small enough (smaller than a criterion value called the *significance level*), the observed data are considered sufficiently unlikely under the null hypothesis, and we *reject* the null hypothesis.

All standard hypothesis tests, no matter how fancy, follow this same recipe (an exception is *Bayesian hypothesis testing*, which we will not cover in PSS).

2.1 Gaussian assumption

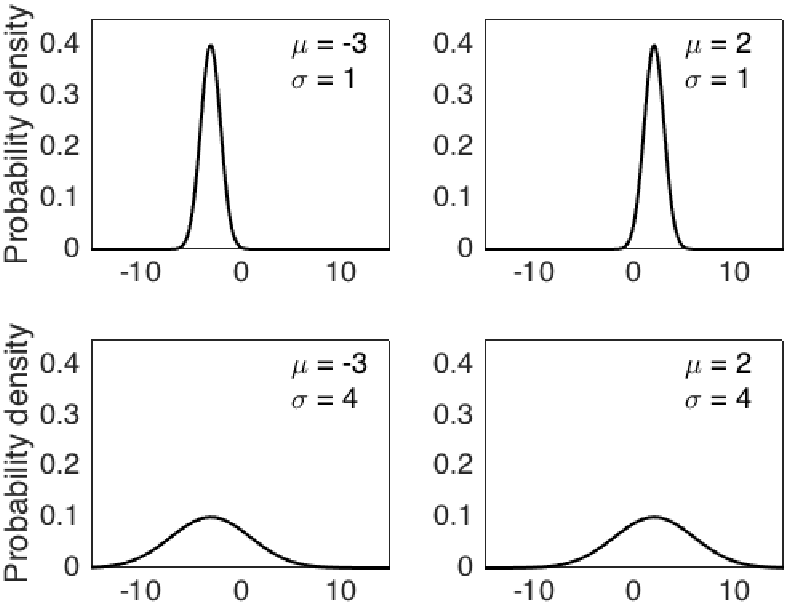

Figure 2: Example Gaussian distributions.

We assume (and this is a strong assumption!) that each data point is independently drawn from a Gaussian (or normal) distribution with unknown population mean μ and unknown population standard deviation σ::

$$p(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}. \tag{5}$$

Example Gaussian distributions are shown in Figure 2 .

2.2 T-test

We have a sample $x_1, \ldots, x_n$. We assume that each measurement is independently drawn from a normal distribution with the same unknown mean μ and unknown standard deviation σ.

Step 1 is to identify the hypotheses. The null hypothesis, denote by $\mathcal{H}_0$, is that the data follow a normal distribution with mean $\mu = \mu_0$. The alternative hypothesis, denoted by $\mathcal{H}_1$, is that the data follow a normal distribution with mean *different from* μ_0. Note that we don't specify different in what way. This is typical of all hypothesis tests.

Exercise: Use the same data as before, and $\mu_0 = 0$. Why would we even consider the possibility that $\mu < \mu_0$? After all, the sample mean is clearly positive.

Step 2. The t-statistic is defined as

$$t = \frac{\bar{x} - \mu_0}{\mathrm{SEM}(x)}. \tag{6}$$

The interpretation of t is "how many SEMs the sample mean is away from the null mean μ_0". It is akin to a signal-to-noise ratio. Note that the data come in in two places: in $\bar{x}$ and in $\mathrm{SEM}(x)$.

Exercise: Using the data from the previous exercise, calculate t. In Excel, the commands given above, along with specifying μ_0, should give you everything you need to answer this.

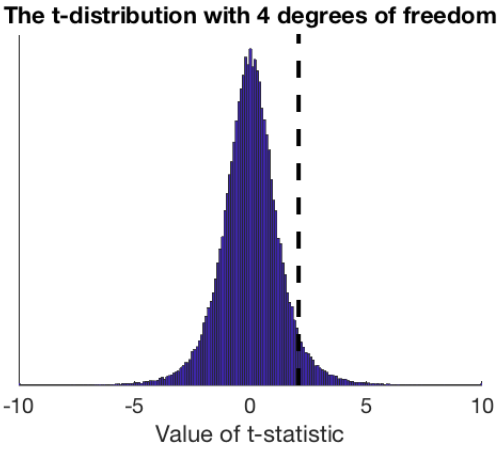

Figure 3: (Simulated) distribution of the t-statistic under the null hypothesis, when $n = 5$. The dashed line indicates the t value of the actual sample that we are using in the exercises.

Step 3. Our goal is to quantify how extreme the value of t found in Step 2 is under the null hypothesis. In Step 3, we work towards this goal by computing the probability distribution of t under the null hypothesis. The distribution of $t^{(\mathrm{null})}$ under the null hypothesis can be derived exactly. This calculation, in which the normal assumption plays an essential role, is the theoretical basis of the t-test. The distribution is called a *Student t distribution* with $n - 1$ "degrees of freedom". It has a mean of 0 and is symmetric around 0.

Step 4. We calculate the theoretical probability that the test statistic under the null hypothesis $t^{(\mathrm{null})}$ takes a value that is in absolute value equal to or greater than the value of t in the data (from Step 2).

Exercise: What is the p-value from the t-test in the previous exercise? In Excel, this is calculated using TDIST(t, $n - 1$, 2), where the 2 stands for two-tailed distribution.

2.3 Reporting the result of a t-test

Suppose you analyze a data set with $n = 10$ measurements. You do a t-test and find $t = 2.90$ and $p = 0.018$. In a paper, you would write this result up as "The mean is significantly different from μ_0 ($t(9) = 2.90$, $p = 0.018$)." or "We reject the null hypothesis that the mean is equal to μ_0 ($t(9) = 2.90$, $p = 0.018$)." It is implied that this is a one-sample t-test, and that the notion of significance is based on a significance level of $\alpha = 0.05$. The "(9)" corresponds to the number of degrees of freedom, which is $n - 1$. Some remarks:

- It is incorrect to state "The mean is different from μ_0." without "significantly".
- It is also incorrect to state "The null hypothesis is false." Hypothesis testing is not about truth.
- Even if the sample mean $\bar{x}$ is greater (or smaller) then μ_0, it is incorrect (but quite common) to state "The mean is significantly greater (or smaller) than μ_0". Even when the *sample* mean is greater than μ_0, it is still possible that the *population* mean is smaller than μ_0, or conversely. There is no way around this.
- If you had not found significance, you would have written, for example, "The mean is not significantly different from μ_0 ($t(9) = 1.83$, $p = 0.10$)." or "We fail to reject the null hypothesis that the mean is equal to μ_0 ($t(9) = 1.83$, $p = 0.10$)." It is incorrect to state "The mean is equal to μ_0." or "The null hypothesis is true."
- You typically don't report more than three significant digits in the t-value (e.g. 2.90 is ok, 2.90139 is overkill) and no more than two significant digits in the p-value (e.g. $p = 0.018$ is ok, $p = 0.0175623$ is overkill).
- In many papers, people write $p < 0.05$, $p < 0.01$, or $p < 0.001$ for significant results, instead of giving the exact p-value. This habit is a remnant from days when p-values had to be looked up in tables, in which only a limited set of values were specified. Nowadays, we can easily calculate the exact value. However, if the p-value is very low, then it often does not matter how low exactly. Nevertheless, it is good practice to specify values between 0.001 and 0.05 instead of simply writing "$p < 0.05$".
- In a plot, it is common to indicate $p < 0.05$ with a single star (*), $p < 0.01$ with two stars (**), and $p < 0.001$ with three stars (***).

2.4 Remarks on the t-test

- The t-test we have discussed so far is the *two-tailed* or *two-sided* t-test. There are also *left-tailed* and *right-tailed* t-tests. In these tests, the alternative hypothesis is $\mu > \mu_0$ or $\mu < \mu_0$, respectively, without consideration of the other direction. Thus, these tests are only appropriate if you are absolutely certain that you can ignore the other direction. This is rarely the case, so we will not discuss these tests any further.
- The t-test assumes that the measurements follow a normal distribution. If they don't, you can still apply the t-test but the resulting p-value might not be accurate. If you are not sure, it might be better to use a *non-parametric test.*

- A major limitation of standard hypothesis testing is that you will have 5% false alarms all the time (or whatever your significance level is) - the probability that there is no effect but the data are in the tail of the distribution.

- In standard hypothesis testing, the alternative distribution never gets exactly specified. In Step 4, we only calculate a probability under the null hypothesis. Another approach is *Bayesian hypothesis testing*, in which the alternative and the null are treated on the same footing. A topic that you can delve into yourself if you are interested!

3 Unpaired data

We now consider hypotheses involving the comparison of continuous data from two groups (samples). We consider two scenarios. In the first scenario, the measurements in the two groups are unrelated (e.g. they come from altogether different individuals) - we will call such samples *independent* or *unpaired*. The two samples could have different sizes. In the second scenario, each measurement in one group corresponds to a measurement in the other group. One example is when the same individuals are tested at two different time points. We will call these samples *paired*. The two samples necessarily have the same size. Different tests apply in each scenario.

In the first scenario, each group contains different individuals. For example, the individuals in two groups all complete a cognitive test, and their scores are as in Table 1. Keep in mind that within each vector, the order of the measurements is arbitrary: it is *not* the case that the first measurement in Group 1 corresponds to the first one in Group 2, etc. Very often, one of the groups is an experimental group or treatment group, whereas the other group is the control group. Individuals in the control group do not undergo a particular experimental manipulation or treatment. The question of interest is whether the groups differ from each other.

Group 1 scores	**Group 2 scores**
3.8	3.1
2.7	4.1
3.0	2.8
2.9	3.7
3.3	3.5

Table 1: Example of two independent samples.

3.1 T-test for two independent samples

Step 1: Data and hypotheses. We assume that both samples have the same size, n. (There is an extension for unequal sample sizes, but the t statistic is less intuitive in that case, so we will not discuss that here. If you are interested, check out the Wikipedia page on the t-test.) We denote the samples by $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. We assume that in each sample, the measurements are drawn from a normal distribution, and that both distributions have the same standard deviation. That only leaves the question of whether the means of the normal distributions (the two population means) are the same or not. Thus, our hypotheses are:

H_0: The means of the two populations are equal.
H_1: The means of the two populations are different.

Step 2: Test statistic. The t-statistic is

$$t = \frac{\bar{y} - \bar{x}}{\sqrt{\mathrm{SEM}(x)^2 + \mathrm{SEM}(y)^2}}.$$

Thus, t is larger when the sample means are farther apart, but also when the sample standard deviations are smaller.

Exercise: Compute t for the data in Table 1.

Step 3: Distribution of the test statistic under the null hypothesis. Under H_0, t follows a t-distribution, but now with $2n-2$ degrees of freedom, double the number for the one-sample t-test.

Step 4: p-value. The p-value of the two-tailed test can be calculated using Step 3. To report the result of a two-sample t-test, write "There was a/no significant difference in [dependent variable] between Group 1 and Group 2 (two-sample t-test, $t(2n-2) = ...$, $p = ...$)."

Exercise: Continuing from the previous exercise, compute the p-value. At $\alpha = 0.05$, is there a significant different in the scores between the groups? In Excel, if the data from Group 1 are in A1:A5, and the data from Group 2 are in B1:B5, the command T.TEST(A1:A5, B1:B5, 2, 2) immediately returns the p-value. The first "2" refers to a two-tailed test, the second "2" to unpaired data.

4 Paired data

We now consider the comparison of two paired samples, $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. There is now a direct correspondence between x_1 and y_1, between x_2 and y_2, etc. A typical example of paired data is when the same individuals are tested before and after some treatment. That means that there are two numbers for each individual. For an example, see Table 2.

4.1 Plotting paired data

A tempting way to plot paired samples x and y is like in Fig. 4 but with means and SEMs calculated separately for x and y, as in Fig. 4A. Looking at this, you might think that the difference between x and y is not significant because the error bars overlap so much. However, this is misleading, because at the level of individuals you would see what is shown in Fig. 4B. Thus, there is a small but very consistent increase from x to y. Another more informative way to plot paired data is a scatterplot (Fig. 4C). To characterize the difference, it is best to first subtract the two samples on an individual basis: we call these the *pairwise differences* (Table 3B). We then calculate mean and SEM *of the pairwise differences.* Then we find an entirely different plot (Fig. 4D). From this plot, it does seem that the difference between y and x is significant. What happened in Fig. 4A is that the small but systematic change between x and y got "washed out" by the variation among individuals. Since we are interested in the former, and not the latter, it's not the right way to plot.

4.2 Paired two-sample t-test

Like in the paired t-test, we assume that in each sample, the measurements are drawn from a normal distribution, and that both distributions have the same standard deviation. That only leaves

Individual	**Condition**	**Score**
1	1	7.7
1	2	9.1
2	1	18.4
2	2	18.2
3	1	16.9
3	2	17.0
4	1	11.1
4	2	11.3
5	1	6.0
5	2	6.5
6	1	10.2
6	2	10.9
7	1	11.4
7	2	12.4
8	1	13.1
8	2	15.1
9	1	11.0
9	2	11.5
10	1	13.8
10	2	16.0

Table 2: Example of two paired samples. We use "condition" instead of "group" to make clear that the same individual participates in both conditions.

Individual	**Pairwise difference**
1	1.4
2	-0.2
3	0.1
4	0.2
5	0.5
6	0.7
7	1.0
8	2.0
9	0.5
10	2.2

Table 3: Pairwise differences based on the data from Table 2.

the question of whether the means of the normal distributions (the two population means) are the same or not. Thus, our hypotheses are:

H_0: The means of the two populations are equal.
H_1: The means of the two populations are different.

The paired two-sample t-test is conceptually very easy: it is identical to a one-sample t-test on the *pairwise differences* $y_i - x_i$, with $\mu_0 = 0$. Thus, the number of degrees of freedom is also $n - 1$. To report the result of a paired t-test, write "There is a/no significant difference in [dependent variable]

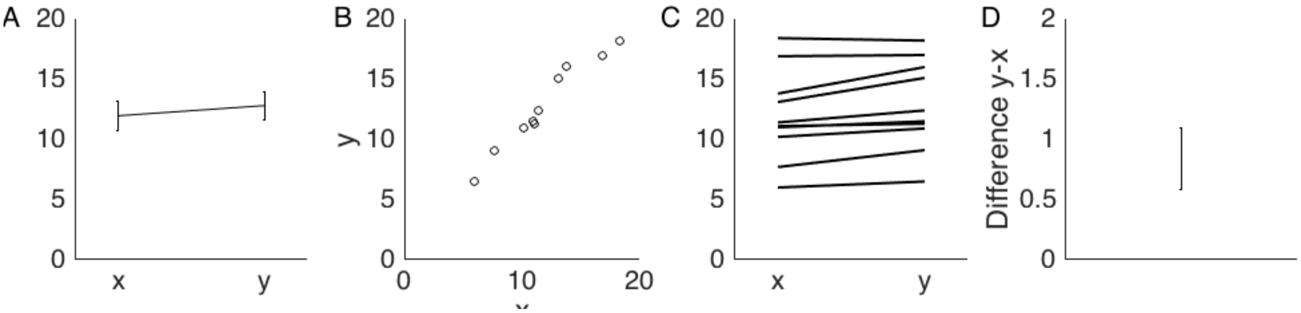

Figure 4: (A) Common but some what misleading way of plotting paired data. (B) More informative way (each line is an individual). (C) Scatterplot representation of the same data (D) Mean and SEM of the pairwise differences.

between Condition 1 and Condition 2 (paired t-test, $t(n-1) = ...$, $p = ...$)." It is useful here to mention that the t-test was paired. In Excel, if the data from Condition 1 are in A1:A10, and the data from Condition 2 are in B1:B10, the command T.TEST(A1:A10, B1:B10, 2, 1) immediately returns the p-value. The "2" refers to a two-tailed test, the "1" to paired data.

5 Correlations

So far, we have talked about univariate data, where we are measuring only a single variable. Very often, what is of interest is the relationship between two variables. Then we are talking about *bivariate data*.

5.1 Bivariate data

Bivariate data involve two variables, say x and y. They consist of pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where n is the number of measurements. Bivariate data come in two forms: two dependent variables, or one dependent and one independent variable. We will only discuss the former here. Bivariate data can be represented as a matrix with two columns, with each row being a measurement. We do not include a column for "measurement number" or "trial number". We also leave out "irrelevant" variables, i.e. variables that we either did not measure or that we believe are not relevant. An example is in Table 4.

We can plot the data in a *scatterplot*. Fig. 5 shows the data from Table 4. We now move towards numerically summarizing bivariate data. There are now two sample means:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{7}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{8}$$

Average room temperature (°C)	**Happiness rating**
15.2	4.59
19.8	5.00
13.9	4.76
14.2	4.55
16.1	5.08
20.2	5.10
16.8	4.89

Table 4: Example of bivariate data: happiness rating versus average room temperature in Asian cities. Example of two independent variables. Data approximated from Tu and Hsee (2016), *Consumer happiness derived from inherent preferences versus learned preferences*, Current Opinion in Psychology 10: 83-88.

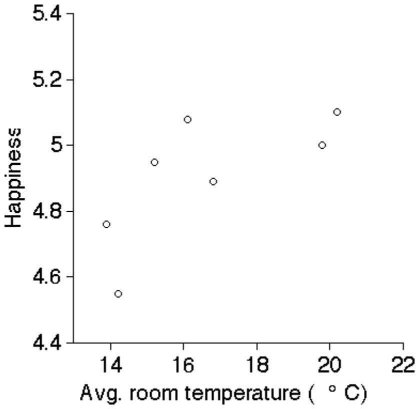

Figure 5: Scatterplot of the happiness data.

two sample variances:

$$\mathrm{var}_n(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{9}$$

$$\mathrm{var}_n(y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2, \tag{10}$$

and two sample standard deviations:

$$\mathrm{std}_n(x) = \sqrt{\mathrm{var}_n(x)} \tag{11}$$

$$\mathrm{std}_n(y) = \sqrt{\mathrm{var}_n(y)}. \tag{12}$$

5.2 Sample covariance

Since we have two variables, we can do one more thing that we could not do with a single variable: study their relationship. For example, Table 4 and Fig. 5 suggest that a higher room temperature is associated with higher happiness: we call this a *positive correlation*. Dog size and apartment size of dog owners might be positively correlated. Test performance and anxiety might be negatively correlated. If two quantities tend to go up and down together, they are positively correlated, whereas if one tends to goes down when the other goes up, then they are negatively correlated. We will now work towards a formal definition of correlation. We will need a new quantity called the *sample covariance* between samples x and y:

$$\text{cov}_n(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}). \tag{13}$$

We use a subscript n because, like for the variance, there is another version of the sample covariance where you divide by $n-1$ instead of by n. The covariance tells you how much two quantities *co-vary* (vary together). It can be an arbitrarily large positive number or an arbitrarily large negative number. Some properties: The covariance of a sample with itself is equal to its sample variance: $\text{cov}_n(x,x) = \text{var}_n(x)$. If y is constant (always takes the same value), then $\text{cov}_n(x,y) = 0$. The sample covariance is *symmetric*: $\text{cov}_n(y,x) = \text{cov}_n(x,y)$.

5.3 Sample correlation

The *sample Pearson correlation coefficient* (a term that is often shortened by leaving out any or all of the words "sample", "Pearson", or "coefficient"), is defined as

$$r(x,y) = \frac{\text{cov}_n(x,y)}{\text{std}_n(x)\text{std}_n(y)}, \tag{14}$$

Unlike the sample covariance, the sample correlation is always between -1 and 1. Moreover, it does not matter whether you normalize by n or by $n-1$ in the sample covariance and the sample standard deviations in Eq. (14), as long as you are consistent between both. What value of a correlation is considered strong depends on the type of data.

Exercise: Calculate the sample correlation of the data in Table 4. In Excel, if the x data are in cells A1 to A7 and the y data in B1 to B7, use CORREL(A1:A7, B1:B7) to calculate the Pearson correlation coefficient between x and y.

5.4 The significance of a correlation

How do we calculate whether a correlation is significant? Here, we examine the most common way: the t-test for correlations.

Step 1. The hypotheses are:

H_0: the population correlation coefficient equals 0.
H_1 : the population correlation coefficient is different from 0.

Step 2. The test statistic is the *t-statistic for a correlation*,

$$t = r\sqrt{\frac{n-2}{1-r^2}}. \tag{15}$$

Step 3. We now assume that (x,y) pairs are independently drawn from a two-dimensional Gaussian distribution. Then, under H_0, the t-statistic follows a t-distribution with $n-2$ degrees of freedom.

Step 4. The p-value can be calculated using the distribution from Step 2. You report "There is a/no significant correlation ($r = ...$, $t = ...$, $p = ...$)

> **Exercise:** Continue from the previous exercise. Calculate t and p.

Note that a correlation can be weak (r low) but highly significant (t large, p low), or conversely, strong (r high) but not significant (t low, p high). This is because t and p do not only depend on r but also on sample size n: larger sample size means larger t and lower p.

5.5 Caution with correlations

A few important caveats about correlations:

- Not every trend can be captured by a covariance or correlation. A correlation is good at quantifying the strength of *linear relations*, but more sophisticated measures are needed to quantify the strength of other relations, such as a "U-shaped" curve.

- Correlation does not imply causation. A famous example is that ice cream sales and murder rates are correlated with each other. This is not because one causes the other, but because they have a common cause, namely high temperatures. In neuroscience, the activity of two neurons could be correlated with each other not because one sends an axon to the other (anatomical connection), but because they receive common input from a third neuron.

- High correlations can occur by chance. There is no limit to the quantities one can get data from, and some will by chance be correlated. This is exploited cleverly by the website "Spurious correlations" (http://tylervigen.com/spurious-correlations). All of these are feature two dependent variables versus the independent variable of time.