Statistics Exercises - Solutions

Covariance, correlation, the t-test

1. Consider the following data:

x	0	2	4	5	7	8	10
y	4	3	1	1	2	1	0

Calculate the correlation coefficient between variables x and y.
First we calculate the means $E(x) = 5.14$ and $E(y) = 1.71$. Then, we calculate the unbiased estimators for the variance $Var(x) = 12.14$ and $Var(y) = 1.90$. Using the formula for the unbiased estimator for the covariance, we get $Cov(x, y) = -4.12$. Our correlation coefficient then is $Corr(x, y) = -0.86$.

2. Consider the following data:

x	0	2	4	6	8
y	0	1	2	3	4

Calculate the correlation coefficient between variables x and y. What does this mean?

We first calculate the means $E(x) = 4$ and $E(y) = 2$. Then we calculate the unbiased estimators for the variance: $Var(x) = 10$ and $Var(y) = 2.5$. The covariance can then be calculated as $Cov(x, y) = 5$. Thus our correlation coefficient is 1. This means that the data = "fully" linearly correlated, or that there is a function $y = ax + b$ that translated the x data into the y data.

3. We want to check if on average a cup of butter contains 250 grams or less. A sample test of 25 cups reveals an average of 248.2 grams and a standard deviation of 2.5 grams. Check whether the smaller mean is a significant difference for a significance threshold of 5%.
Or null hypothesis is $H_0 : \mu = 250$. We test this against $H_1 : \mu < 250$. Therefore our statistic is $T = \frac{\bar{X}-250}{S}\sqrt{n}$. From the sample we get a t-value of $T = \frac{248.2-250}{2.5}\sqrt{25} = -3.6$. We reject the null hypothesis for too small values of T. Therefore $P(T \leq t_\alpha; H_0) = \alpha = 0.05$. Using the table we find $P(T(24) \geq 1.71) = P(T(24) \leq -1.71) = 0.05$, therefore $t_\alpha = -1.71$. Since our t-value is smaller than t_α, we reject the null hypothesis. Therefore the cups' weights are on average lower than expected.

4. We want to find out if persons of 40 years old are on average heavier than persons of 30 years old. We have taken two independent sample sets, given as follows:

30y weight	77	65	73	58	63	49	51	82	103	69					
40y weight	102	73	56	55	83	72	88	70	81	85	44	71	62	78	75

Give an analysis of this situation and use the t-test to find out whether our hypothesis is valid for a significance threshold of 10%.
Our null hypothesis is $H_0 : \mu_X = \mu_Y$ which we want to check agains $H_1 : \mu_X < \mu_Y$. This is a two-sample unpaired test, so our statistic is $T = \frac{\bar{X}-\bar{Y}}{S\sqrt{\frac{1}{n}+\frac{1}{m}}}$. From the sample sets we get $\bar{X} = 69$, $\bar{Y} = 73$, $\hat{\sigma}_X^2 = 255.8$ and $\hat{\sigma}_Y^2 = 213.7$. Therefore $S = \sqrt{\frac{9\hat{\sigma}_X^2+14\hat{\sigma}_Y^2}{23}} = 15.2$. Filling this in for the statistic T we get $T = -0.65$.

We reject the null hypothesis for too small values of T. Then, $P(T(23) \leq t_\alpha; H_0) = P(T(23) \geq -t_\alpha; H_0) = 0.10$. By looking in the table we find $-t_\alpha = 1.32$, thus $t_\alpha = -1.32$.

Our t-value is not smaller than t_α, we cannot reject the null hypothesis. Therefore we cannot prove that people of age 40 are on average heavier than people of age 30.

5. We want to check if a marketing strategy for playing more on-line games has worked. For 8 subjects, we have measured the number of hours they spend per week on on-line games before and after the marketing strategy. This data is given as follows:

hours played before	2	3	1	4	6	2	12	8
hours played after	3	2	0	4	7	4	10	10

Find out whether the marketing strategy works with a significance threshold of 5%.

This is a paired two-sample t-test. Our null hypothesis is $H_0 : \mu_Z = 0$ where $Z = X - Y$. We are going to test this null hypothesis against $H_1 : \mu_Z < 0$ since we expect Y to be larger than X. The statistic is then $T = \frac{\bar{Z}}{\hat{\sigma}_Z}\sqrt{n}$. We calculate $\bar{Z} = -0.25$ and $S_Z = 1.49$. Our t-value then is $T = -0.47$.

We reject our null hypothesis for too small values of T. Then, $P(T(7) \leq t_\alpha; H_0) = P(T(7) \geq -t_\alpha) = 0.05$. By looking in the table we find $-t_\alpha = 1.89$, so $t_\alpha = -1.89$. Since our t-value is not smaller than t_α we cannot reject the null hypothesis. Therefore, we cannot prove that the marketing strategy actually worked.