XIV. Tests for Correlation Coefficients

Scenario. This is a continuation of the prior section.

Data and parameters. You will be given

sample size	$\boldsymbol{n}$
correlation coefficient	$\boldsymbol{r}$
significance level	$\boldsymbol{\alpha}$

Research Objective. The goal will be to decide whether or not $\boldsymbol{r}$ describes a real relationship between the two variables – i.e., whether or not $\boldsymbol{r}$ is statistically significant. If $\boldsymbol{r}$ describes a real relationship between the two variables, then $\boldsymbol{r} \neq 0$. Note that $\boldsymbol{r}$ can be statistically significant but not be a useful predictor.

Solution Template

Step 1. Make a list of the know parameters:

sample size	$\boldsymbol{n}$
correlation coefficient	$\boldsymbol{r}$
significance level	$\boldsymbol{\alpha}$

Step 2. Identify the *null* and *alternative* hypotheses:

$$\begin{array}{rl} \text{Null Hypothesis} & \boldsymbol{H_0} : \boldsymbol{r} = 0 \\ \text{Alternative Hypothesis} & \boldsymbol{H_A} : \boldsymbol{r} \neq 0 \quad \boldsymbol{or} \\ & \boldsymbol{H_A} : \boldsymbol{r} > 0 \quad \boldsymbol{or} \\ & \boldsymbol{H_A} : \boldsymbol{r} < 0 \end{array}$$

Step 3. Compute the test statistic

$$\text{test statistic} = \frac{\boldsymbol{r} \times \sqrt{\boldsymbol{n} - 2}}{\sqrt{1 - \boldsymbol{r}^2}}$$

Step 4. Find the cutoffs using the table in Appendix B.

Step 5. Plot the cut-off, the rejection region for the null hypothesis and the test statistic in exactly the same manner as for any other hypothesis test.

End of Solution Template

Remark 1. This test statistic is actually a *Student's t* with $(\boldsymbol{n}-2)$ *degrees of freedom.* For larger values of $\boldsymbol{n}$ – say $\boldsymbol{n} > 60$ – this test statistic is approximately normally distributed.

Remark 2. Recall that you can make two types of error:

	$\boldsymbol{H}_0$ true	$\boldsymbol{H}_A$ true
accept $\boldsymbol{H}_0$	OK	Type II error
reject $\boldsymbol{H}_0$	Type I error	OK

With hypothesis tests you can only control Type I error, the error of rejecting $\boldsymbol{H}_0$ when $\boldsymbol{H}_0$ is in fact true.

Example. In a sample of 282 subjects it is found that there is a correlation of -0.13 between income and weight. Is this significant at the $\boldsymbol{\alpha} = 5\%$ level?

Solution.

Step 1. The parameters are

sample size	$\boldsymbol{n} = 282$
correlation coefficient	$\boldsymbol{r} = -0.13$
significance level	$\boldsymbol{\alpha}5\%$

Step 2. **Step 2.** Identify the *null* and *alternative* hypotheses:

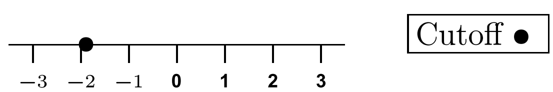
$$\text{Null Hypothesis} \qquad \boldsymbol{H}_0 : \boldsymbol{r} = 0$$
$$\text{Alternative Hypothesis} \qquad \boldsymbol{H}_A : \boldsymbol{r} \neq 0$$

Step 3. Next we compute the test statistic:

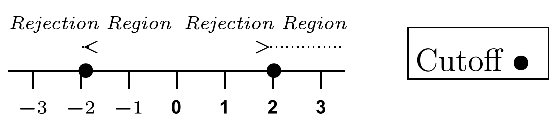
$$\begin{aligned}
\text{test statistic} &= \frac{\boldsymbol{r} \times \sqrt{\boldsymbol{n}-2}}{\sqrt{1-\boldsymbol{r}^2}} \\
&= \frac{-0.13 \times \sqrt{282-2}}{\sqrt{1-(-0.13)^2}} \\
&= \frac{-0.13 \times 16.73}{\sqrt{0.9831}} \\
&= \frac{-2.17}{.99} \\
&= -2.20
\end{aligned}$$

Step 4. Find the cutoffs from the table in Appendix B: cutoffs $= \pm 1.96$.

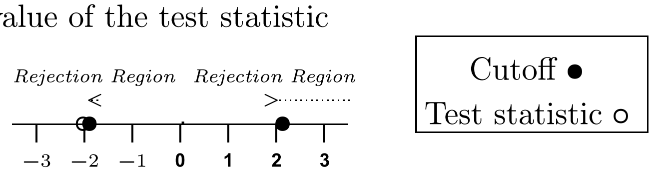
Step 5. To help make the decision whether or not to reject the null hypothesis, *first draw a number line* and plot the cutoffs from step 4.



Next find the rejection regions.



Finally, plot the value of the test statistic



Since the test statistic falls in the (left) rejection region, we reject $\boldsymbol{H}_0$ and accept $\boldsymbol{H}_{\boldsymbol{A}}$. ▮

Questions. Would you want to use weight to predict income? Based on the above data, do you think that employers discriminate against overweight employees? What other variables might you need to control to conclude that age and weight are correlated?

Problems

1. In a sample of size 13, x and y measurements on each subject are found to be correlated at $r = -0.254$. Is this significant at the 5% level?

2. In a sample of size 867, x and y measurements on each subject are found to be correlated at $r = -0.254$. Is this significant at the 5% level?

3. In a sample of size 974, x and y measurements on each subject are found to be correlated at $r = 0.38$. Is this significant at the 5% level?

4. In a sample of size 231, x and y measurements on each subject are found to be correlated at $r = 0.11$. Is this significant at the 5% level?