4.3 The Neyman-Pearson paradigm

We wish to check whether a coin is fair. We've performed an experiment by tossing the coin 10 times. We get 7 heads.

- Write the null hypothesis for your example. Is the null hypothesis simple of composite? Write the test statistic, denote it by X - what is its distribution under the null hypothesis?
- Say your alternative hypothesis is $H_A : p > 0.5$. Is this a simple or composite hypothesis? What kind of values speak in favour of the alternative hypothesis? Is the alternative hypothesis one-sided or two-sided?
- In our case, $X = 7$. What is the probability of seing this event on our sample under the null hypothesis?
- Say that the rejection region consists of the value $\{10\}$. What is the level of significance α in this case? What is the level of significance if the rejection region equals $\{6, 7, 8, 9, 10\}$?
- Determine the rejection region so that $\alpha = 0.05$. Can you reject the null hypothesis at this level?
- What is the power of the test for this α, if we assume that the true value of parameter equals $p = 0.6$ or $p = 0.7$? What is the value of the type II error?
- Assume that your alternative hypothesis equals $H_A : p \neq 0.5$. Is this a simple or a composite hypothesis? Is it one or two-sided?
- What is the rejection region if $\alpha \leq 0.05$? What is the exact significance level for this rejection region?
- Calculate the power of the test in this example.

Understanding the ideas with R:

- Repeat an experiment with 10 coins 1000 times. Check the probabilities of chosen rejection regions.
- Change the probability of heads and check the power of the test.

p\k	0	1	2	3	4	5	6	7	8	9	10
0.5	0.001	0.011	0.055	0.172	0.377	0.623	0.828	0.945	0.989	0.999	1
0.6	0.000	0.002	0.012	0.055	0.166	0.367	0.618	0.833	0.954	0.994	1
0.7	0.000	0.000	0.002	0.011	0.047	0.150	0.350	0.617	0.851	0.972	1

Table 1: The cumulative probabilities for the binomial distributions with $n = 10$ ($P(X \leq k|p)$)

- Increase the sample size (for example take 20 coins) and check how the power changes.

4.4 Test power

The available literature claims that the mean haemoglobin value of an athlete that stays at least 14 days at the altitude above 1500m increases for 2 g/l, but no change in variance is expected. At normal altitudes, the values are approximately normally distributed $X \sim N(\mu_1, 5^2)$, where μ_1 is the athlete's individual mean.
An athlete often trains on altitude, but in shorter time intervals. He wishes to know whether his mean haemoglobin values nevertheless increases. He made 12 (independent) measurements during the season, 8 of these were during the altitude training and 4 otherwise. The goal of this exercise is to determine the power of his test for significance level $\alpha = 0.05$.

- What is the null and what the alternative hypothesis?
- Propose a test statistic. Calculate its distribution under the null hypothesis.
- Calculate the power of the test, i.e. the probability that he can reject the null hypothesis if his mean indeed increases for 2 g/l?
- How would the power change if he had an equal number of measurements in each of the phases?
- How does the power of the test depend on the variance of individual's measurements? How does it depend on the true differences in the population?

4.5 Generalized likelihood ratio test

An assumption that is used in the Athlete Biological Passport is that haemoglobin varies equally in all athletes. We wish to test this assumption on a sample of k athletes. Let the values of i-th athlete be normally distributed ($i = 1, \ldots, k$), i.e. $X_{ij} \sim N(\mu_i, \sigma_i^2)$, where $j = 1, \ldots, n_i$ denote the individual's measurements. Assume that all measurements are independent.

- Write the null and alternative hypothesis
- Consider first the case of only one athlete with n measurements. How would we estimate his parameters μ and σ^2 with the method of maximum likelihood?
- Show that under the alternative hypothesis (for k athletes), the parameter estimates equal

$$
\begin{aligned}
\hat{\mu}_i &= \frac{1}{n_i} \sum x_{ij} \\
\hat{\sigma}_i^2 &= \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu}_i)^2
\end{aligned}
$$

- How do we estimate means under the null hypothesis?
- What is the variance estimate under the null hypothesis?
- How would you test the null hypothesis with the generalized likelihood ratio test?