11th problem set for Probability and Statistics

Hypothesis Testing and Linear regression

1. We have one measurement $X \sim N(\mu, 1)$. We want to test the hypothesis H_0: $\mu = 5$ at a significance level $\alpha = 5\%$.

   (a) What critical region will we choose – the set of measurements where we reject the hypothesis?

   (b) Instead of one measurement, we take n measurements (naturally independent). What will be the critical region for $\bar{X}_n$?

   (c) If in fact $\mu = 4$ and we have $n = 10$ measurements, what is the probability that we do not reject the hypothesis?

   (d) Let X still have mean μ and variance 1, but not necessarily be normal. What changes?

   (e) What if we know nothing about the variance of X?

2. What if we sample from two populations? (Both with normal distribution.)

   (a) What can we test?

   (b) Create an appropriate model.

   (c) Think of two different situations that need to be approached differently. (Hint: in one situation, it is necessary for both populations to have the same size.)

3. You want to determine whether or not more than 3/4 of citizens of country X would vote for candidate Y for President in 2025. In a random poll sampling of $n = 137$ citizens of X, we collected responses $x_1, \ldots, x_n$ (each is 1 or 0, if they would vote for Y or not). We observe 131 "yes" responses: $\sum_{i=1}^{n} x_i = 131$. Perform a hypothesis test and state your conclusion.

4. According to the manufacturer's promise, their machine makes errors in at most 3% of cases. In 600 trials, there were 28 errors. Assess the manufacturer's promise (as the null hypothesis) at a significance level of 5%.

   (a) Model the number of errors exactly using the binomial distribution.

   (b) Model the number of errors approximately using the normal distribution (with appropriate μ, σ^2).

5. A genetics engineer was attempting to cross a tiger and a cheetah. She predicted a phenotypic outcome of the traits she was observing to be in the following ratio 4 stripes only: 3 spots only: 9 both stripes and spots. When the cross was performed and she counted the individuals she found 50 with stripes only, 41 with spots only and 85 with both. According to the chi-square goodness of fit test at significance level of 5%, did she get the predicted outcome?

6. A company manufactures an electronic device to be used in a very wide temperature range. The company knows that increased temperature shortens the life time of the device, and a study is therefore performed in which the life time is determined as a function of temperature. Fit the line of regression, for the following data:

Temperature in Celsius (t)	10	20	30	40	50	60
Life time in hours (y)	420	360	275	175	115	35