Biostatistics

A sample of questions for the final exam

1. Consider the following two plots. Write down these probabilities:

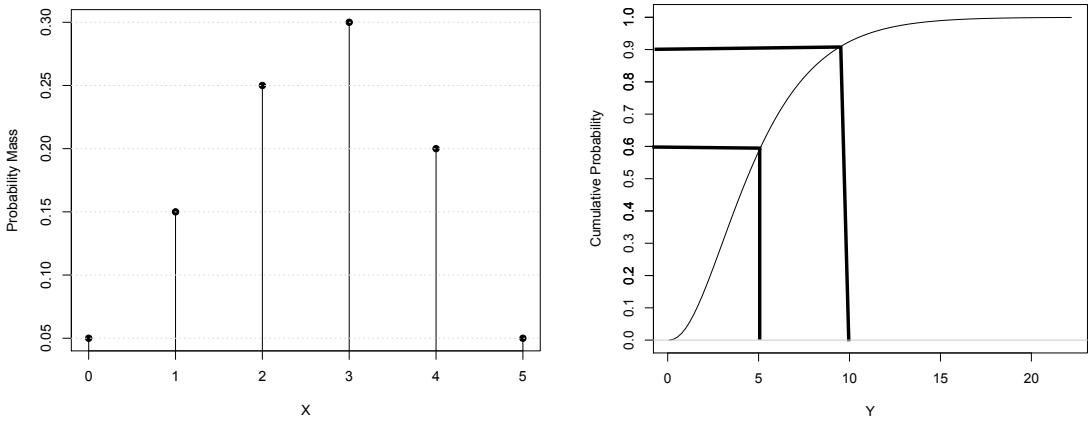
   a- $P(X \leq 2) = 0.05 + 0.15 + 0.25 = 0.45$

   b- $P(1 \leq X < 3) = 0.15 + 0.25 = 0.40$

   c- $P(5 \leq Y < 10) = 0.9 - 0.6 = 0.3$

   d- $P(Y > 10) = 1 - 0.9 = 0.1$

   e- $P(Y = 10) = 0$



2. We assume that blood pressure, X, in the whole population has a $N(\mu, \sigma^2)$ distribution. To estimate μ and σ, we randomly selected 9 people and measured their blood pressure. The sample mean and sample standard deviation are $\bar{x} = 110$ and $S = 6$.

   Knowing that based on the $t(8)$ distribution, 0.8-quantile = 0.9, 0.85-quantile = 1.1, 0.9-quantile = 1.4, and 0.95-quantile = 1.9, answer the following questions (*20 points*).

   a- Write down the point estimate for μ along with its standard error.

   b- Write down the margin of error at 0.8 confidence level.

   c- Write down the 80% confidence interval estimation for μ based on this sample.

   d- Assume that we want to test the null hypothesis $H_0 : \mu = 120$ using a one sample t test, what is the t score?

   a)
   $\hat{\mu} = \bar{x} = 110$ and $SE = \frac{S}{\sqrt{n}} = 6/3 = 2$
   b)
   To find the margin of error, we first need to find $t_{0.8}$ critical value. This is the 0.9-quantile based on the $t(8)$ distribution, which is 1.4. The margin of error is $1.4 \times 2 = 2.8$.

c)
The $80\% CI$ is $[110 - 2.8, 110 + 2.8]$.
d)
t-score $= \frac{\bar{x}-\mu_0}{SE} = \frac{110-120}{2} = -5$.

3. A drug company is testing a new flue vaccine which is believed to be more effective than the one that is currently used. Of 100 people who used this new vaccine, 10 people eventually got the flu (the vaccine was not effective for them). What is the 95% confidence interval for the proportion of people who would still get the flue if we use this vaccine (instead of the old one) for the whole population (*10 points*).

   A)
   Our point estimate is the sample proportion $10/100 = 0.1$. The standard error is $\sqrt{0.1 \times 0.9/100} = 0.03$. The $z_{0.95} \approx 2$ (the more exact value is 1.96). So the margin of error at 0.95 confidence level is $2 \times 0.03 = 0.06$. The 95% CI is $[0.1 - 0.06, 0.1 + 0.06] = [0.04, 0.16]$

4. What would be the most appropriate probability distribution for each of the following random variables (10 points):

   a- Whether a tumor is benign or malignant: Bernoulli

   b- Number of people with a malignant tumor out of 10 patients with tumor: Binomial

   c- Size of tumors: Normal, assuming we choose the unit such that the values are not very close to zero; otherwise, we should use log-transformation of size as a random variable with a normal distribution

5. We want to find out the proportion of people in the US who smoke cigarette. We interview 100 people and find that 20 of them smoke. Write down the point estimate for the population proportion and the corresponding standard error. Also write down the margin of error, and confidence interval at 0.8 confidence level. We know that based on the standard normal distribution, 0.8-quantile = 0.84, 0.85-quantile = 1.04, 0.9-quantile = 1.28, and 0.95-quantile = 1.65, and 0.975-quantile= 1.96 (20 points).

$$\begin{aligned} \hat{\theta} &= p = 0.2 \\ SE &= \sqrt{0.2 \times 0.8/100} = 0.04 \\ MOE &= 1.28 \times 0.04 \\ CI &= [0.2 - 1.28 \times 0.04, 0.2 + 1.28 \times 0.04] \end{aligned}$$

6. We are interested in the distribution of blood pressure for people who are overweight ($BMI > 25$). We assume the distribution is normal with mean μ and standard deviation σ. We take a sample of 25 people who are overweight, we measure their blood pressure, and find that the sample mean is 140 and the sample standard deviation is 10.

   a- Write down the point estimate for the population mean and the corresponding standard error, as well as the margin of error, and confidence interval at 0.9 confidence level. We know that based on the $t(24)$ distribution, 0.85-quantile = 1.06, 0.9-quantile = 1.32, and 0.95-quantile = 1.71 (15 points).

$$\begin{aligned} \hat{\mu} &= \bar{x} = 140 \\ SE &= S/\sqrt{n} = 10/5 = 2 \\ MOE &= 1.71 \times 2 \\ CI &= [140 - 1.71 \times 2, 140 + 1.71 \times 2] \end{aligned}$$

b- We had hypothesized that the average blood pressure for overweight people is above 130. Write down the null hypothesis and calculate the t-score (10 points):

$$\begin{aligned} H_0 &: \mu = 130 \\ t - score &= \frac{\bar{x} - \mu_0}{SE} = \frac{140 - 130}{2} = 5 \end{aligned}$$

7. An experiment was conducted by students at The Ohio State University in the fall of 1993 to explore the nature of the relationship between a person's heart rate and stepping up and down on steps of various heights. One of the objectives of this study was to see whether the hight of steps matters. Students were randomly assigned to two groups: Low-steps, where the height of steps was 5.75 inches (this group was coded as 0), and High-steps, where the height of steps was 11.5 inches (this group was coded as 1). Student performed the exercise (stepping up and down) for three minutes, after which their heart rates were measured. The investigators had hypothesized that the average heart rate would be different between the two groups. A two-sample t-test was used to analyze the data. The results are shown below. Answer the following questions:

   a- Explain why we use a two-sample t-test and not a z-test or a paired t-test (5 points): We use a two-sample t-test since we are comparing two populations (high and low steps) in terms of their means, and we don't know σ's. We can't use z-test since we don't know σ; we can't use paired $t-$test since the observations in the two groups are independent.

   b- Is this a one-sided or two-sided test? (5 points): Two-sided since we don't specify and direction for departure from H_0.

   c- Is the relationship between heart rate and the height of steps statistically significant at 0.01 level? Write down the t-score and p-value (5 points): It is significant since the $p-$value=0.002427 is below 0.01. The t-score is -3.371.

   d- Alternatively, the investigators could do the experiment as follows: ask each student to do the exercise with a low-step first, measure his/her heart rate, ask him/her to relax for a while, then repeat the exercise with a high-step, and measure his/her heart rate again. What type of hypothesis testing method you would recommend in this case? Why? (10 points) : In this case, we would use a paired t-test since the two sets of observations are paired (each subject has a pair of observations).

```
> t.test(HR~Height, alternative='two.sided', conf.level=.95, var.equal=FALSE, data=Dataset)

        Welch Two Sample t-test

data:  HR by Height
t = -3.371, df = 25.107, p-value = 0.002427
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -34.79382  -8.40618
sample estimates:
 mean in group Low mean in group High
             96.6              118.2
```