Statistics 403 Problem Set 4

Due in lab on Friday, October 9th

1. Suppose we plan to carry out a two-sample comparison based on data $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, where $m = n = 10$, and it is known that $\mathrm{SD}(X) = 3$ and $\mathrm{SD}(Y) = 5$. What is our power to detect an effect $\mu_x - \mu_y = 0.5$?

   Solution: The two pieces to the power are $P(T > 2)$ and $P(T < -2)$. To calculate $P(T > 2)$ we need

$$\begin{aligned} 2 - \delta\sqrt{(m+n)/(\sigma_x^2/q_x + \sigma_y^2/q_y)} &= 2 - 0.5\sqrt{20/(9/(1/2) + 25/(1/2)))} \\ &= 1.73 \end{aligned}$$

   so $P(T > 2) = P(Z > 1.73) = 0.04$.

   To calculate $P(Z < -2)$ we need

$$\begin{aligned} -2 - \delta\sqrt{(m+n)/(\sigma_x^2/q_x + \sigma_y^2/q_y)} &= -2 - 0.5\sqrt{20/(9/(1/2) + 25/(1/2)))} \\ &= -2.27 \end{aligned}$$

   so $P(T < -2.27) = P(Z < -2.27) = 0.01$. Thus the overall power is around 0.05.

2. Suppose we plan to carry out a two-sample comparison based on data $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, where $m = 10$, and it is known that $\mathrm{SD}(X) = 3$ and $\mathrm{SD}(Y) = 5$. Our interest is in the power to detect an effect $\mu_x - \mu_y = 0.5$. Explain why the power does not come close to 1, regardless of how large n becomes. What power do we get for large values of n? (Hint: it's easier if you calculate the power using the expression for T on slide 60, not the equivalent expression on slide 63).

   Solution:

   Let T be the test statistic. The power is $P(T > 2) + P(T < -2)$. Focus first on $P(T > 2)$:

$$P\left(\frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_x^2/m + \sigma_y^2/n}} > 2\right) = P(Z > 2 - \delta/\sqrt{\sigma_x^2/m + \sigma_y^2/n}).$$

   As n grows, the term σ_y^2/n gets small, so we can ignore it. Thus for very large n, the power will be

$$P\left(Z > 2 - \delta/\sqrt{\sigma_x^2/m}\right) + P\left(Z < -2 - \delta/\sqrt{\sigma_x^2/m}\right).$$

As long as m, δ, and σ_x^2 are fixed, the power will not be close to 1. This means that in a two sample comparison, it is not enough to increase one of the two sample sizes. To get very high power, you need to increase the sizes of both groups being compared.

3. For what value of x will the following distribution have the least possible variance?

x	-1	x	1
$P(X = x)$	0.3	0.2	0.5

Solution: The expected value is $EX = 0.2(1 + x)$. The variance is

$$0.3(-1 - 0.2(1 + x))^2 + 0.2(x - 0.2(1 + x))^2 + 0.5(1 - 0.2(1 + x))^2$$

which equals

$$0.3(-1.2 - 0.2x)^2 + 0.2(0.8x - 0.2)^2 + 0.5(0.8 - 0.2x)^2$$

which simplifies to

$$0.16x^2 - 0.08x + 0.76$$

which has derivative

$$0.32x - 0.08.$$

Setting the derivative to zero and solving for x yields $x = 1/4$.

4. Is it possible to find values for c and d such that the following joint probability table satisfies $E(Y|X = 0) = E(Y|X = 1)$? If so, give the values. If not, state that it is impossible.

		Y		
		-1	0	1
X	0	0.1	c	0.4
	1	0	0.1	d

Solution: The conditional probability table is

		Y		
		-1	0	1
X	0	0.1/(0.5+c)	c/(0.5+c)	0.4/(0.5+c)
	1	0	0.1/(0.1+d)	d/(0.1+d)

Thus,

$$E(Y|X = 0) = 0.3/(0.5 + c)$$

$$E(Y|X = 1) = d/(0.1 + d).$$

Since $c + d + 0.6 = 1$, it also follows that $d = 0.4 - c$, thus

$$E(Y|X = 1) = (0.4 - c)/(0.5 - c).$$

Setting $E(Y|X = 0) = E(Y|X = 1)$ gives $c^2 - 0.2c - 0.05 = 0$, which has roots

$$\frac{0.2 \pm \sqrt{0.2^2 + 4 \cdot 0.05}}{2}$$

The solution with a negative sign is negative, which can't be a probability, so we take the solution with a positive sign. Thise gives $c = 0.35$ and $d = 0.05$.

5. Suppose that in a certain population of people, the average blood pressure is 170. All of these people are then treated with a drug intended to lower blood pressure. However the drug has side effects and 30% of the people stop taking the drug. Among the people who continue taking the drug ("treatment compliers"), the average blood pressure is 140. The average blood pressure for those who stop taking the drug rapidly returns to 170.

   (a) What is the overall average blood pressure during the period that the treatment is offered? Hint: use the double expectation theorem.

   Solution: The double expectation theorem says that we should average the averages of the compliers and non-compliers, using the fractions of people who comply and who do not comply as weights. Thus we would get

   $$140 \cdot 0.7 + 170 \cdot 0.3 = 149.$$

   (b) Now suppose that the standard deviation of blood pressure among the treatment compliers is 5 and the standard deviation of blood pressure among those who stop taking the drug is 10. What is the overall (marginal) standard deviation of blood pressure during the period that the treatment is offered? Hint: use the law of total variation.

Solution: According to the law of total variation we must consider two things. First, we need the mean of the variances within the two groups. These variances are 25 and 100, so the mean of the conditional variances is

$$E\text{var}(\text{BP}|\text{compliance}) = 0.3 \cdot 100 + 0.7 \cdot 25 = 47.5.$$

The other piece of the overall variance is the variance of the conditional means. The conditi

$$\text{var}E(\text{BP}|\text{compliance}) = 0.7 \cdot (140 - 149)^2 + 0.3 \cdot (170 - 149)^2 = 189.$$

Thus the overall variance of blood pressure is $47.5 + 189 = 236.5$, so the overall standard deviation of blood pressure is $\sqrt{236.5} = 15.4$.