Math 20340: Statistics for Life Sciences
Fall 2008
Mid-semester Exam 2
Solutions

1. Answer these questions using the standard normal table included at the end of this exam. In the first two parts, z is a standard normal.

1. Find a z_0 such that $P(z > z_0) = .8$.
   Solution: $z_0 = -.84$
2. What is the probability that z is between -2 and 1?
   Solution: $.8413 - .0228 = .8185$
3. Which is more likely — a normal random variable taking a value that is within 2/3 of a standard deviation of its mean, or taking a value that differs (above or below) from its mean by at least 2/3 of a standard deviation? Justify.
   Solution: Normal random variable taking a value that is within 2/3 of a standard deviation of its mean is same as standard normal taking value between $-.66...$ and $+.66...$, which is $.7486 - .2514 = .4972$. So probability of taking a value that differs (above or below) from its mean by at least 2/3 of a standard deviation is $1 - .4972 = .5028$. This latter probability is greater, so differing by at least 2/3 a standard deviation is more likely.

2.

1. A recent poll of 484 randomly selected Kuwaitis concluded that 56% of Kuwaitis would prefer that Barack Obama be elected US president. The poll notes that "one can say with 95% confidence that the margin of sampling error is $\pm 5\%$". Explain what this means.
   Solution: An acceptable answer is that "there is a 95% probability that the poll is correct, that is, that the true support for Obama among the whole population is within 5% of 56%". A better answer is that a statistical process is being used to produce the point estimate, that, if the sampling is done truly randomly, would result in the true value being within the margin of error of the estimated value, 95% *of the times that the process is performed.*

   An unacceptable answer is that "95% of the time that 484 people are polled, the support among them for Obama is within 5% of 56%" (indeed, all but 271 people in he entire population may support McCain; but the pollster has unluckily included all those 271 in his random sample). The correct statement here is that "95% of the time that 484 people are polled, the support among them for Obama is within 5% of p, where p is the true proportion".

   Another unacceptable answer is that "the pollster is 95% confident that ...", since this doesn't explain what "95% confident" means.

2. 14 national polls have been published in the last three days, each trying to estimate the proportion of the voting population that supports Barack Obama. Each poll quotes a 95% confidence margin of error. Say that a poll is GOOD if the true proportion lies within the margin of error of the estimated proportion, and BAD otherwise. What is the probability that at least one of the 14 polls published in the last three days is BAD?

   Solution: By definition of margin of error, the probability of one poll being GOOD is .95. Assuming the polls are independent, the probability of all 14 being good is $.95^{14} = .048...$, so the probability of at least one being BAD is $1 - .048... = .5123....$

3. Experience has shown me that when I write a seven question exam, the amount of time it takes a student to complete it is a normal random variable with mean 42 minutes and standard deviation 4 minutes. I want 90% of the class to finish the exam on time. How many minutes should I allow for the exam?

Solution: Let x be the amount of time it takes a random student to finish. x is normally distributed with mean 42, standard deviation 4. I want to find a time x_t such that $P(x < x_t) = .9$ (so 90% of class finish on time). Standardizing, this is same as $P(z < (x_t - 42)/4) = .9$, so $(x_t - 42)/4 = 1.28$ or $x_t = 47.12$. So I should allow 47.12 minutes for the exam.

4. I recently traveled to the planet Zog and succeeded in spotting 100 Zogians. I was interested in estimating the average number of legs that a Zogian has. 20 of the Zogians I spotted had no legs, 10 had one leg, 20 had two legs and 50 had three legs. Assume that the 100 Zogians I spotted form a random sample from the planet's population.

1. What is the sample mean $\bar{x}$ of my observations?

   Solution: $\bar{x} = \frac{20*0+10*1+20*2+50*3}{100} = \frac{200}{100} = 2.$

2. What is the sample standard deviation s of my observations?

   Solution: $s = \sqrt{\frac{20(0-2)^2+10(1-2)^2+20*(2-2)^2+50*(3-2)^2}{99}} = \sqrt{frac14099} = 1.189....$

3. Give a point estimator for the average number of legs of a Zogian, with a 95% confidence margin of error.

   Solution: Point estimator is $\bar{x} = 2$; 95% confidence margin of error is $\pm 1.96 * SE \approx 1.96 * \frac{s}{\sqrt{100}} = 1.96 * \frac{1.189...}{10} = .233....$

4. It seems clear (if my sample is representative) that the number of legs a Zogian has is not distributed normally. Why then is it ok to use a normal distribution to approximate the distribution of $\bar{x}$?

   Solution: Central limit theorem tells us that whatever the population distribution, the distribution of the sample mean is approximately normal if sample size is large enough (and $n = 100$ certainly is large enough).

5. (20 pts) The average income of Notre Dame alumnus (measured in thousands of dollars) is 80, with standard deviation 11. Thirty Domers find themselves sitting randomly together on game day, and pass the time during a media time out by computing their average annual income.

How likely is it that they get an answer greater than 85?

Solution: Let $\bar{x}$ be the answer they get. For random samples of size 30, $\bar{x}$ is distributed approximately normally with mean 80 and SE $11/\sqrt{30} = 2.008....$ So the answer "85" is greater than the mean by $5/2.008... = 2.48...$ standard deviations. The probability of a normal random variable taking a value that exceeds its mean by this much is $1 - .9936 = .0064$.

6. I have 36 friends in England. On the first Saturday of each month, I chat online with each of them, one after another. Each of the 36 chats last 10 minutes on average, with a standard deviation of 3 minutes. What's the probability that the total time I spend chatting online exceeds $5\frac{1}{2}$ hours?

Solution: The total time x I am on the phone is the sum of 36 samples from a distribution with mean 10 and standard deviation 3. By the Central Limit Theorem, this is distributed approximately normally with mean $36 * 10 = 360$ and standard deviation $3 * \sqrt{36} = 18$. I'm interested in the probability that this random variable takes on a value at least 330. $P(x > 300) = P(z > (330 - 360)/18) = P(z > -1.66...) = 1 - P(z \leq -1.66...) = 1 - .0475 = .9525$.

7. A curious professor wants to determine whether there is a difference between the proportion of female students who favor partial credit exams over multiple-choice exams and the proportion of male students who do. He surveys 100 randomly chosen students, 60 female and 40 male, and finds that 35 of the females and 20 of the males favor partial credit exams.

Construct a 90% confidence interval for the difference between the proportion of females and males who favor partial credit exams.

Solution: $n_W = 60$, $\hat{p_W} = 35/60 = .583...$; $n_M = 40$, $\hat{p_M} = 20/40 = .5$.

Point estimator for $p_W - p_M$ is $\hat{p_W} - \hat{p_M} = .083...$

Estimate for standard error is $\sqrt{\frac{\hat{p_W}\hat{q_W}}{n_W} + \frac{\hat{p_M}\hat{q_M}}{n_M}} = .10149...$

90% confidence interval: $.083... \pm 1.645 * .10149... = (-.0836..., .25...)$.