

Solutions to the sample problems for the final exam:

1. a. The expected number is $50(.133) = 6.65$.

b. The most likely number is... well, here I just did some experimentation, trying different values for x in $\text{binompdf}(50,.133,x)$. But I knew to start looking around $x = 6$, based on the calculation in part a, and indeed, the mostly likely outcome is 6 Asians, which happens with probability 16.49%.

c. So, OK. There are several different ways of setting this up; I will do $\text{binomcdf}(50,.133,10) - \text{binomcdf}(50,.133,4) = .7510$. I'm not sure we've done a problem quite like this before, actually, but hopefully what I did makes sense.

d. $=1 - \text{binomcdf}(50,.133,14) = .0016$. Very small.

2. a. Since we are going to want to use a Student's t-model with 4 degrees of freedom, we need to assume that the cost of each car crashing is independent from the others – if we assume the 5 cars were chosen randomly, we can assume independence. We also need to assume that the costs of the cars are distributed normally. With such a small sample this would ordinarily be quite a bold assumption. In the real world, we might be justified in making this assumption if we have had years of experience recording the costs of repairs.

b. But we're told to make that assumption, so we go ahead! We calculate the sample mean to be \$540 and the standard deviation to be \$299. I do a T-Interval and input this data... and now we can say with 95% confidence that the average damage will lie between about \$169 and \$911. OK, that's a big interval. But what do you expect from a sample size of 5?

3. [You will not have to be this wordy on the final exam!] Let p designate the true proportion of Americans who believe in intelligent life on other planets. We have taken a random sample of $n=950$ Americans and computed a sample proportion, $\hat{p} = .54$. If we think of the sample proportion as a random variable, the distribution of that variable will be Normal with mean p and standard deviation $\sqrt{\frac{pq}{950}}$, if certain conditions are true. We need to check that the number of “successes” and “failures” are greater than 10 – in this context, a success is a person who believes in intelligent life on other planets, and a failure is a person who doesn't. We don't know the true proportion, but the actual recorded number of successes is $.54(950) = 513$, and the number of failures is 437, which are well above 10, so we consider this condition to be met.

We also need to check that our sample size is less than 10% of all Americans, because technically speaking, since we are drawing without replacement, once we poll one person about their beliefs, the probability that the next person will believe in life on other planets will change, but only very very slightly. But if we draw too many people, these little tiny errors will add up and mess with our results. That's why we try to stay under 10% of the

population – which we definitely are. Meeting that condition doesn't ensure independence, though; we also need to look at the methodology of the sample. Since the subjects were chosen randomly, I think independence is assured.

Now that our conditions are met, we can do a 1-proportion Z-interval on our calculator. I get (.508, .572), meaning that we can say with 95% confidence that the true proportion of America that believes in intelligent life on other planets is between 50.8% and 57.2%.

(Note that if I wasn't using that function on my calculator, there would be a lot more calculations I'd have to show by hand, including the computation of the standard error:

$$\sqrt{\frac{(.54)(.46)}{950}} = .016, \text{ and something that looked like this : } .54 \pm 1.96(.016).$$

4. The probability of getting exactly two 2s is $\text{binompdf}(12, 1/6, 2) = .29609$. On the other hand, the probability of getting a mean die roll of 3 or less is... well, first we have to figure out the sampling distribution. The sampling distribution has mean 3.5 (that's the expected value of one die roll). The std. dev of one die roll is... (actually, that itself would be a good final exam question– “what is the *exact* value of the std. dev, no decimal

approximations allowed!) $\sqrt{\frac{35}{12}} \approx 1.7078$. So the sampling distribution is approximately

Normal with mean 3.5 and standard deviation $\frac{\sqrt{35/12}}{\sqrt{12}} = \frac{\sqrt{35}}{12} \approx .493$. Thus, the

probability of getting a sample with mean 3 or less is $\text{normalcdf}(-E99, 3, 3.5, .493) = .15524$. Hence it is more likely that you will get exactly 2 twos, than that you will get a mean die roll of 3 or lower.

One lingering question: was a Normal model appropriate to use to approximate the sampling distribution? Well, recall what the conditions are (see top of page 418) – we need that the data values are sampled randomly, and that the data values are independent. Both are definitely satisfied here – the data values are the very definition of random, and the data values are definitely independent of each other (how will the dice affect each other?)

5a. The proportion in question is given by $\text{normalcdf}(137, E99, 100, 15) = .0068$.

b. So you could easily do some hypothesis testing here, with the null hypothesis being that the average Korean score is 100, and the alternative hypothesis being $\mu > 100$. The sampling distribution is Normal with mean 100 and SD $\frac{15}{\sqrt{12}}$ -- we don't need to use the standard error here since we're given the SD of the actual distribution of scores – and

then we can find a p-value by computing $normalcdf\left(126.3, E99,100, \frac{15}{\sqrt{12}}\right)$ = a really really small number.

But that would be totally missing the point – the sample here is extremely far from being a random sample of Koreans! It is a sample of Kim’s friends in college! We would certainly expect that people in college would, in general, score better on these tests than an overall population. We might also expect that Kim might gravitate towards choosing friends who are as ‘smart’ as he. In other words, this sample might be extremely non-representative of Koreans; his conclusion is unsupported by the facts.

(Moreover, his conclusion includes the phrase ‘smarter’, and you should know that a great deal of controversy has been raging in academia for decades about what IQ scores really measure. So from that perspective, that’s another reason to take issue with Kim’s conclusion.)

6. My null hypothesis is that there is no difference between using aspirin once a day, and not using aspirin once a day, in preventing heart attacks. If we let p_1 be the true proportion of people who have heart attacks if they’re using aspirin, and p_2 be the true proportion of people who have heart attacks if they’re not steadily using aspirin, then my null hypothesis can be written as $p_1 = p_2$, or $p_1 - p_2 = 0$. The alternative hypothesis is that $p_1 < p_2$, or $p_1 - p_2 < 0$.

I need to check that the number of success and failure in each group is greater than 10, which is true. I need to check that the sample sizes are less than 10% of the total population – since the population in question seems to be ‘all people everywhere’ this is met. While the volunteers were selected randomly for one treatment or the other, I’m not sure that a group of volunteer doctors is necessarily representative of all people everywhere, but that isn’t going to stop me from doing the analysis. Within each group, it’s fair to assume that the results are independent; it’s also fair to assume that the results from one group are independent of the other, especially since the use of a placebo indicated the participants were blinded.

So the conditions are met! I’m using a 2-proportion Z-test. I get a p-value of... really incredibly small. The evidence is overwhelming that aspirin once a day reduces the risk of heart disease (at least among volunteer doctors!)

(Historical note: the experiment was originally designed to last longer than 5 years, but when these results were calculated, the experimenters called it off and contacted all subjects receiving the placebo, and advised them to start taking aspirin once a day.)

7. I think it’s an experiment – the subjects were randomly placed into one group or the other. A placebo treatment was used (the spice tea without the herb was a placebo). The experiment was clearly double-blind; neither the participants nor the evaluating psychologist knew who was in which group.

8. a. The event that Heather's average was below 3 is much *less* likely, thanks to the Law of Large Numbers. See the illustrations and discussion on pages 413-414. Notice how the variation goes down as the number of rolls goes up.

b. I think that specifically what we're looking at here is two sampling distributions for the mean, both based on the distribution of the numbers on a fair die. The average die roll is 3.5, and the standard deviation is 1.7078. Jon's distribution is approximately Normal

with mean 3.5 and standard deviation $\frac{1.7078}{\sqrt{50}} = .242$; Heather's is approximately Normal

with mean 3.5 and standard deviation $\frac{1.7078}{\sqrt{200}} = .121$. The probability of Jon getting an

average value below 3 is therefore approximately

$\text{normalcdf}(-999, 3, 3.5, .242) = 0.019$. The probability for Heather is

$\text{normalcdf}(-999, 3, 3.5, .121) = .000017$. That's significantly smaller!

9. Our sample mean, \bar{y} , is 15.9, the standard deviation s is .35, and of course our sample size n is 49. Our null hypothesis is that the true mean weight of the shipment is 16 ounces as claimed ($H_0 : \mu = 16$); the alternative hypothesis is that the true mean is smaller than 16 ounces ($H_A : \mu < 16$).

To use a T-test (which I would love to do), we need to check that the boxes are independent (that seems OK, they were randomly selected) and that the weights of those boxes are normally distributed. We don't actually have any information about the distribution of weights in our sample, but 49 is a pretty big sample size, so I say, let's just go ahead.

We do our T-test and get a p-value of .0256. That's small enough for me to reject the null hypothesis; I think this shipment is underweight, on average, and that the grocery company should send it back.

10a. 1%. (You have a 1/10 chance of being one of the 10 winners, and then you have a 1/10 chance of being the big winner from the group of 10.)

b. 10%.

c. To answer this we need to understand our probability model, which is as follows:

-3 : 90%

+7 : 9%

+72 : 1%.

Based on this, we can then easily calculate that the expected value is \$-1.35, and the SD is 7.907.

11.a. The expected number is 28.5.

b. Do $\text{binomcdf}(100, .285, 20) = .0349$.

c. We already know the probability of getting 20 or fewer successes. The probability of getting 40 or more is $1 - \text{binomcdf}(100, .285, 39)$. These are two disjoint events, so the

probability of one or the other happening is the sum of the probabilities. The sum of these two probabilities is .0413.

d. The sampling distribution of the mean weight of our shrovlians is Normal with mean 13.2 and SD 0.12. So we calculate $\text{normalcdf}(-E99, 13, 13.2, 0.12)$, which is about 4.78%.

e. We know your sampling distribution is $N(13.2, 0.12)$. Your friend's has the same mean but a slightly different sampling distribution : $N(13.2, 0.085)$. (Their SD is smaller because their sample size is larger. If we let X be a random variable representing the average weight of your sample, and Y be a random variable representing the average weight of your friend's sample, then what we're really interested in is $X-Y$, which is also going to be Normal, with mean 0 and $SD = 0.147$ or so. Now we want the probability that $X-Y > 0$. You can do a normalcdf for this, but you don't need to; the probability is 50%.

f. Now we have to look at $X+Y$, which is normal with mean 26.4 and SD 0.147. We do $\text{normalcdf}(26.7, 34583458738457345, 26.4, 0.147) = .0206$ so that's about 2.06%. Not very likely.

12. a. If you believe that a bag of M&M's is constructed at random, that each of the 40 candies in that bag were selected at random, then these are Bernoulli trials – there are two outcomes (success = yellow or red; failure = neither), there is a constant chance of success for each trial (.4), and the number candies in the bag is less than 10% of all M&M's. On the final exam, that answer would be perfectly acceptable. However, I do not believe a bag of M&M's is constructed at random – I bet there's some quality control mechanism to make sure that a certain number of each color of candy gets into each bag, messing up the randomness. So I bet these aren't quite truly Bernoulli trials!

b. No – we're not recording success and failures, we're recording a *number*.

c. No – it sounds like the probability of getting a question right changes from question to question.

d. *Technically* the trials aren't independent. To see why, suppose 42,000,000 out of 300,000,000 Americans say they never have time to relax. Then when we survey our first respondent, there is a $42000000/300000000$ chance that they say this. Suppose they do say this and are counted as a "success." Then the probability of our next respondent saying the same thing is $41999999/299999999$. So they're not independent – BUT those probabilities are so incredibly close that we can treat them as independent anyway. We do this all the time. Every time we calculate a confidence interval or do some sort of hypothesis test regarding proportions, we are doing this. So I'd say it's ok to treat these as Bernoulli trials.

13.a. The maximum it can do is 120 hit points of damage. The wizard would have to be *very* lucky to get this result!

b. The expected value is $3.5 * 20 = 70$ points. The SD is about 7.6375. (I used the SD of one die roll, calculated from an earlier problem; squared that to get the variance; added that to itself 20 times, one for each die; took the square root of that.)

c. OK. We don't *quite* know about the sampling distribution of the sum of the dice. But we do know about the sampling distribution of \bar{y} , the *average* of all the rolls. We know

that it will be distributed Normally with mean 3.5 and SD 1.7078 divided by the square root of 20, so that's about .3819. Now, a \bar{y} greater than or equal to 4.25 will kill the dragon. (4.25 times 20 = 80.) So, what is the probability of us getting that? Well... we can do `normalcdf(4.25, big number, 3.5, .3819)` and we get about a 2.5% chance. This spell is unlikely to kill the dragon (unfortunately for the wizard.)

14. We do a t -interval here and get (298.49, 321.51).

We had to assume that we had a random sample (seems likely) and that the population was nearly Normal.

What our interval means: We are 95% confident that the hot dogs have an average sodium content of between 298.49 mg and 321.51 mg of sodium.

What the confidence means: 95% of all such samples of 40 hot dogs will generate intervals containing the true average sodium content.

15. a. Yes, roughly; the larger sample size means a smaller margin of error in our confidence interval, so everything else being equal, we'll be a little more accurate in our interval.

b. We are 95% confident that the actual average sodium content of these hot dogs is between 309.73 mg and 326.27 mg.

c. 70% of 465 mg is 325.5 mg. 325.5 mg is in our interval! We don't have any sense of the actual distribution of hot dog sodium levels – all the values in our interval are equally likely to be the true average sodium content. This means we cannot legitimately claim that these hot dogs are “reduced sodium”. It is true that 325.5 is very close to the endpoint of one of our intervals, but *this does not matter* – it is possible that these hot dogs do not deserve the title.

16a. Mean number of games won: 81. Median: 82.5. (The mean is lower than the median because a few teams were really bad, skewing the average. It shouldn't be surprising that the average number of games won is 81 – in fact mathematically it can be shown that the average *must* be 81, since 81 is exactly half of all 162 games, and in every game one team has to win and one has to lose.) SD: 9.262.

b. Mean number of runs scored: 776.833. Median: 769.

c. Mean number of runs allowed: 776.833. (It's the same as the mean runs allowed! And this should make sense too, because every run a team scores is a run that some *other* team allows...). The median: 776.5

d. So, make a scatterplot. It looks pretty chaotic; not a lot of pattern; although there is a hint of a positive trend. Which is what you might expect. But the correlation, R , is only .61 – that's not extraordinarily high. There is some sort of weak relationship here, maybe, but it's definitely not strong. The equation of the regression line is $wins = 0.0818 RS + 17.438$.

e. OK, so I'd expect it to be negative, because in a very general sense, a team that doesn't score a lot of runs isn't going to win very many games. The scatterplot sort of bears this

out: there's a roughly negative sort of trend. But like the last scatterplot, there's also a lot of noise and cloudiness; not a very strong relationship here. R is about $-.655$, which is still not extraordinarily high. Not strong evidence of a linear relationship.

f. It's easy to have your calculator do this for you, remember? If you have RS in L2 and RA in L3, you can put your cursor on the title of L4 and press enter, then do $L4 = L2 - L3$ and your calculator will automatically fill in L4 with the values of $RS - RA$ for each team. Now make a scatterplot and you should be struck by how much *more* linear this looks than the previous scatterplots! Which indicates to me that it's much more accurate to predict the quality of a team by not looking at pure runs scored, or runs allowed, but rather in looking at the number of runs by which they outscored their opponents. The correlation is $.897$, which is significantly higher than in the previous parts of this problem.

g. And the regression line is $wins = 0.0853(RS - RA) + 81$.

h. What stands out about the regression line to me is the y -intercept of the line – 81 games. Our line is predicting that a team that scores exactly as many runs as it allows will win half its games! (Remember that one season is 162 games.) A very sensible prediction, if you ask me. The slope is harder to interpret, but here we go: for each extra run a team scores more than its competition over the course of a season, that team is predicted to win $.0853$ more games. That doesn't *sound* like a lot, but it works out that if you score 100 more runs than your opponents, our model translates that into 8 or 9 more won games, which is a significant margin in baseball!

Whew. That was fun, huh?