Review for Exam 2

1 Time in the shower

The distribution of the amount of time spent in the shower (in minutes) of all Americans is right-skewed with mean of 8 minutes and a standard deviation of 10 minutes. A random sample of 100 US residents will be collected and the *median* shower time will be computed. We wish to find the probability that this median is less than 7 minutes. Which of the following is true?

(a*)** With this information, we cannot find the approximate probability that the median of this sample will be less than 7 minutes.

(b) Because the sample size is large, we can use the Normal model to find the approximate probability that the median shower time in the sample will be less than 7 minutes.

(c) The sample size is not large enough to apply the normal model to find this probabilty. We should take a larger sample size.

2 p577, #36

In July of 2004, the Gallup Poll asked 1005 US adults if they actively try to avoid carbohydrates in their diet. That number increased to 27% from 20% in a similar poll in 2002. Is this a statistically significant increase? Explain.

We run this via a hypothesis test.

- Let p_1 represent the true proportion in 2004 and p_2 represent the true proportion in 2002.

$$H_0 : p_1 - p_2 = 0$$
$$H_A : p_1 - p_2 \neq 0$$

- Check assumptions/conditions: (1) $np > 10$ and $n(1-p) > 10$ for each group, (2) random sample, and (3) less than 10% of the population.
- Draw a picture.
- Compute our test statistic (pooled). Note that $\hat{p}_c = 0.235$ and $n_1 = 1005$ and n_2 is assumed also to be 1005.

$$\begin{aligned} SE &= \sqrt{\frac{0.235 * 0.765}{1005} + \frac{0.235 * 0.765}{1005}} = 0.019 \\ Z &= \frac{(0.27 - 0.20) - 0}{0.019} = 3.70 \end{aligned}$$

  One tail = 1 - 0.9999 = 0.0001 $\rightarrow$ p-value $= 2 * 0.0001 = 0.0002$.

- Because p-value < 0.05, we reject the null hypothesis. That is, we do find strong evidence that there was in fact an increase in the proportion of people who actively try to avoid carbohydrates.

3 Alcohol consumption

The population distribution of the number of alcoholic beverages consumed each week by those 65 years old and older is right-skewed, with a mean of 1.2 drinks and standard deviation 5 drinks. Which of the following is most likely true?
(a) We have enough information to find the probability that a single randomly selected person from this population will drink more than 3 drinks per week.
(b) We have enough information to find the probability that a random sample of 6 people from this population will drink on average 3 or more drinks per week.
(c*)** We have enough information to estimate the probability the average number of drinks of a random sample of 100 people from this population will be more than 3 drinks per week.

4 p553, #9

In August 2004, *Time* magazine reported the results of a random telephone poll commissioned by the Spike network. Of the 1302 men who responded, only 39 said that their most important measure of success was their work.
(a) Estimate the percentage of all American males who measure success primarily by their work. Use a 98% confidence interval. Don't forget to check the conditions first.
Our estimate is $\hat{p} = 39/1302 = 0.03$.

- (1) $n\hat{p}$ and $n(1-\hat{p})$ are each greater than 10. (2) Random sample. (3) Less than 10% of the population.
- Setup the confidence interval formula:

$$estimate \pm Z * SE$$

We determine Z from the normal probability table: 2.33. We identify the standard error as

$$SE = \sqrt{\frac{0.03 * 0.97}{1302}} = 0.0047$$

Our interval:

$$0.03 \pm 2.33 * 0.0047 \rightarrow (0.019, 0.041)$$

(b) Suppose we want to conduct a hypothesis test to test whether this data shows the fraction of men who measure success by their work is below the 5% mark. What does your confidence interval indicate? Explain.

Because our interval for p does not contain 0.05, we would reject the notion that $p = 0.05$.

5 Clean air vehicles

A census conducted by UCLA in the past found that 10% of all cars with long term UCLA parking permits were clean air vehicles. A student environmentalist group believes that today that proportion is even larger. To test this, they perform a SRS of 200 cars with long-term parking permits. Which of the following is the best pair of hypotheses to test their conjecture?
(a) Null: the population proportion is greater than 0.10. Alternative: the population proportion is less than 0.10.
(b) Null: the sample proportion is 0.10. Alternative: the sample proportion is greater than 0.10.
(c*)** Null: the population proportion is 0.10. Alternative: the population proportion is greater than 0.10.
(d) Null: the sample proportion is 0.10. Alternative: the sample proportion is not equal to 0.10.

6 Russians and democracy

A survey of a random sample of Russians found that 52% of the sample approve of the change from communism to a multi-party system. (Pew Foundation, November 2009). A 95% confidence interval is calculated to be (49% to 55%). Which of the following is the best interpretation?
(a) There is a 95% chance that our sample proportion is between 49% and 55%.
(b) We are very confident that if the survey were to be repeated, between 49% and 55% of the new sample would approve of the change to a multi-party system.
(c) There is a 95% chance that the proportion of all Russians who approve of the change to a multi-party system is between 49% and 55% .
(d*)** We are very confident that the proportion of all Russians who approve of the change to a multi-party system is between 49% and 55%.

7 p529, #24

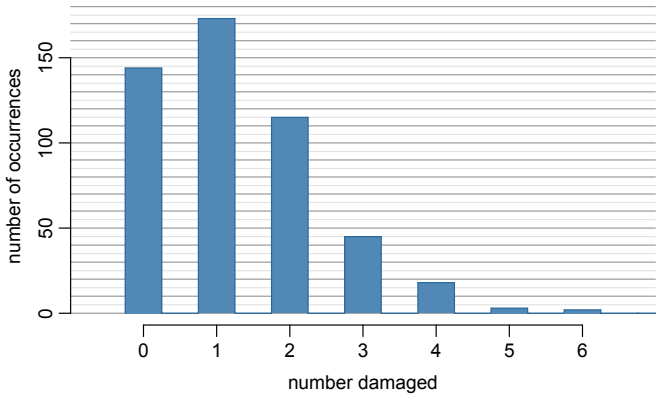
An appliance manufacturer stocks washers and dryers, and sometimes they become damaged in handling. The company's goal is to keep the number of damaged appliances under 2%. Suppose an inspector comes by one day to check and see if they are meeting their target, and he finds 5 of 60 randomly selected appliances are damaged.
(a) Is this strong evidence that the warehouse is failing to meet the company goal? Test an appropriate hypothesis test, however, check your assumptions and conditions first. (Only proceed if the assumptions/conditions are okay.)

- Let p represent the true proportion of appliances in the warehouse that are damaged.

$$H_0 : p = 0.02$$
$$H_A : p > 0.02$$

- Problem with the assumptions/conditions: $np = 1.2$ is much less than 10. Do not proceed with the test (with the normal model for $\hat{p}$).

(b) In part (a), you should have found that one of our assumptions failed. We can still run a test using simulation. Below are results for the simulated number of damaged appliances in 60 observations if the null hypothesis were true and $p = 0.02$ (500 simulations). Use this graph to estimate the p-value and complete the test.



Our test statistics is the number of damaged appliances. We find the p-value to be about $5/500 = 0.01$. Since the p-value is less than 0.05, we reject H_0 and conclude that the true proportion of damaged appliances is in fact greater than 0.02.

8 Cloud seeding

Meteorologists are studying the effects of cloud "seeding" on rain production. The theory is that by dropping ice crystals from an airplane into clouds, the clouds are more likely to produce rain. They conclude that the cloud seeding does not work, using a significance level of 5%. Which of the following can we conclude?

(a) The researchers would have reached the same conclusion had they used a significance level of 10%.
(b*)** The researchers would have reached the same conclusion had they used a significance level of 1%.
(c) The probability that we have reached the wrong conclusion about the effectiveness of cloud seeding is at least 5%.

9 Hypothesis test of a mean

In a hypothesis test, $H_0 : \mu = 45$ and $H_A : \mu < 45$. A sample of 100 people finds a sample average of 47. The p-value is:

(a*)** greater than 0.05
(b) less than 0.05
(c) not appropropriate to calculate for this test.

10 Sulphinpyrazone and deaths after heart attacks

Researchers believe the drug sulphinpyrazone will reduce deaths in heart attack patients. Below is a summary table for the results of a double-blind experiment to test its efficacy.

		outcome		
		lived	died	Total
trmt	drug	692	41	733
	placebo	682	60	742

Setup a hypothesis test to see whether the drug provides significant evidence of improvement. Check all assumptions and conditions.
For convenience, the pooled standard error for this interval is provided:

$$SE = \sqrt{\frac{0.068 * 0.932}{742} + \frac{0.068 * 0.932}{733}} = 0.013$$

where $\hat{p}_1 = 60/742 = 0.081$, $\hat{p}_2 = 41/733 = 0.056$, and $\hat{p}_c = 0.068$.

SOLUTION.

- Let p_1 represent the proportion who died in the control group and p_2 represent the proportion who died in the drug group.

$$H_0 : p_1 - p_2 = 0$$
$$H_A : p_1 - p_2 > 0$$

- Check assumptions/conditions: (1) $n\hat{p}_c$ and $n(1 - \hat{p}_c)$ are each greater than 10 for both groups. (2) The data was randomized. (3) Less than 10% of the population.
- Draw a picture.
- Compute our test statistic (pooled).

$$Z = \frac{(0.081 - 0.056) - 0}{0.013} = 1.92$$

  One tail = 1 - 0.9726 = 0.0274 → p-value = 0.0274 since it is a one-sided test.

- Because p-value < 0.05, we reject the null hypothesis. That is, we find strong evidence that the drug is effective in reducing the number of deaths in heart attack victims.

11 Female faculty

A sociologists wants to test whether a greater proportion of women are now working as faculty in science departments at research universities than in the year 2000. To test this, she takes a random sample of faculty from research universities in the U.S.. She calculates a test statistic of $z = 1.0$. The p-value is about

(a*)** 0.16 **(b)** 0.32 **(c)** 0.68 **(d)** 0.84

12 Census income

A random sample of 500 census forms from the year 2000 in the southern Californian metropolitan area results in a 95% confidence interval of \$25,590 to \$34,273 for the mean annual income of people older than 22. Which of the following is true?

(a) If we took many random samples of size 500, about 95% of them would produce a mean between \$25,590 and \$34,273.
(b*)** It is possible that the mean income is actually as large as \$35,000, but this is unlikely and would be surprising.
(c) 95% of the population has an income between \$25,590 and \$34,273.
(d) If another random sample of 500 forms were taken, there's a 95% chance that the mean of the new sample would be between \$25,590 and \$34,273.

13 p574, #9

Researchers at the National Cancer Institute released the results of a study that investigated the effect of weed-killing herbicides on house pets. They examined 827 dogs from homes where an herbicide was used on a regular basis, diagnosing malignant lymphoma in 473 of them. Of the 130 dogs from homes where no herbicides were used, only 19 were found to have lymphoma. The standard error for the difference between the two proportions is given by

$$SE_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{0.572 * 0.428}{827} + \frac{0.146 * 0.854}{130}} = 0.036$$

where $\hat{p}_1 = 0.572$ represents the proportion of dogs with malignant lymphoma from the herbicide group and $\hat{p}_2 = 0.146$ represents the sample proportion with malignant lymphoma from the second group.

(a) Construct a 95% confidence interval for the difference.

The assumptions are verified (np and $n(1-p)$, random sample okay, less than 10% of population). We let p_1 represent the proportion of dogs from homes that regularly used herbicide that got cancer, and p_2 is the proportion of dogs who got cancer in the homes that did not regularly use herbicide.

$$\begin{aligned} estimate &\pm Z * SE \\ \hat{p}_1 - \hat{p}_2 &\pm 1.96 * 0.036 \\ 0.426 &\pm 1.96 * 0.036 \rightarrow (0.355, 0.497) \end{aligned}$$

(b) State an appropriate conclusion.

We are 95% confident that using herbicide regularly is associated with a 35.5% to 49.7% higher rate of malignant lymphoma in dogs.
(c) Suppose you increased the confidence level to 99%. Would your interval (i***) get wider, (ii) shrink, (iii) stay the same, or (iv) none of the above?
(d) If you increased the sample size of each group and held the confidence level constant, would your confidence interval (i) get wider, (ii***) shrink, (iii) stay the same, or (iv) none of the above?