Math 207
Introduction to Statistics

Exam IV Review Exercises

Use `R` as needed. Please work your answer out on scratch paper first and keep your work neat.

1. A box contains three tickets with 0 and one with 1. A random sample of 100 tickets is made.

   a) What is the chance that 30 or more of the selected tickets will be 1?

   b) What is the chance that 20% or less of the tickets will be 1?

2. A die is rolled n times and the total number of spots counted. For each of the following cases, determine if the result can be explained as a chance variation. Calculate the test statistic z and the P-value (i.e., the chance of getting a test statistic as extreme or more extreme than the one given).

   a) $n = 10$, sum $= 40$

   b) $n = 100$, sum $= 375$

   c) $n = 225$, sum $= 750$

3. A simple random sample of 100 registered voters in a Winchester is made. Of this sample, 45% say that they plan to vote for the Coffee party candidate.

a) Find a 95% confidence interval on the city-wide percentage of Coffee party voters.

b) True or false? There is about a 95% chance that the true percentage of Coffee party voters is in the range you found in a).

c) True or false? If you took a second sample of 100 voters, there is about a 95% chance that its average will be in the range you found in a).

d) True or false? There is about a 95% chance that percentage of doctors in Winchester who plan to vote for the Coffee party candidate is in the range you found in a).

e) To woo doners, the the Coffee party claims that its recently unknown candidate already has 51% of the vote. Run a hypothesis test using this as the null hypothesis. (That is, assume that the percent of Coffee party voters really is 51% but the observed percentage was 45%. Find z and P.)

f) Repeat e) based on a 55% claim.

g) Repeat e) based on a 60% claim.

4. True or false?

a) If $P = 1\%$, then the data supports the alternative hypothesis more than the null?

b) If $P = 1\%$, then there is only a 1% chance that the null hypothesis is true.

c) If the null hypothesis is true and $P = 1\%$, then there is only a 1% chance of getting a test statistic as extreme or more extreme than the observed value.

d) The observed significance level is equal to the P-value.

e) The standard error for the average of a random sample decreases with sample size.

f) The observed significance level depends on the data.

g) If $P = 95\%$, the null hypothesis must be true.

h) If $P = 95\%$, the null hypothesis is plausible.

i) The standard error for a sum increases with sample size.

j) The standard error increases with the standard deviation of the data.

k) The null hypothesis says that the observed difference is simply due to chance.

5. A construction crew measures the length of a site by adding 9 length measurements. From prior experiments, they know that the SD of their tape measure is about 2". They conclude that the sum of their measurements will be off by about how much? 18", 6", 2" or 2/3".

6. Would taking the average of 16 measurements decrease the error by a factor of about 16, 4, 1/4 or 1/16?

7. Four hundred draws are made at random from the box

$$\left| \text{60,000 } \boxed{0}\text{s} \quad \text{20,000 } \boxed{1}\text{s} \right|$$

Note that $\sqrt{(6/8) \times (2/8)} \,/\, \sqrt{400} \approx 0.02$. True or false?

a) The expected value for the percentage of 1s among the draws is exactly 25%.

- The expected value for the percentage of 1s among the draws is around 25%, give or take 2% or so.

b) The percentage of 1s among the draws will be around 25%, give or take 2% or so.

c) The percentage of 1s among the draws will be exactly 25%.

d) The percentage of 1s in the box is exactly 25%.

e) The percentage of 1s in the box is around 25%, give or take 2% or so.

8. A die is rolled n times and the percent of sixes is counted. For each of the following cases, determine if the result can be explained as a chance variation. Calculate the test statistic z and the P-value (i.e., the chance of getting a test statistic as extreme or more extreme than the one given).

a) $n = 10$, percent $= 20\%$

b) $n = 100$, sum $= 20\%$

c) $n = 225$, sum $= 20\%$

9. (Hypothetical) The British Imperial Yard is sent to Paris for calibration against The Meter. The length is determined 100 times. This sequence of measurements averages out to 91.4402 cm, and the SD is 800 microns (a *micron* is a millionth of a meter).

a) Is a single reading off by around 80 microns, or 800 microns?

b) Is the average of all 100 readings off by around 80 microns, or 800 microns?

c) Find a 95%-confidence interval for the exact length of the Imperial Yard.

Solutions

1. a) The average of the box is $(3 \times 0 + 1 \times 1)/4 = 1/4 = 0.25$. The SD of the box is $\sqrt{0.25 \times 0.75} = 0.43$. The expected value of the sum is $100 \times (1/4) = 25$. The standard error for the sum is $\sqrt{100} \times 0.43 = 4.3$. If $x = 30$, then $z = (30 - 25)/4.3 = 1.16$. The chance of getting 30 or more 1s is $1 - \text{pnorm}(1.16) = 12.3\%$.

b) The expected value for the percent of 1s equals the average of the box: 25%. The standard error for the percent is $0.43/\sqrt{100} = 4.3\%$. If $x = 20\%$, then $z = (20 - 25)/4.3 = -1.16$. The chance of getting 20% or fewer 1s is $\text{pnorm}(-1.16) = 12.3\%$.

2. Represent the die as a box with six tickets numbered 1 to 6. The average of the box is 3.5. The SD is 1.7.

a) If $n = 10$, the expected value for the sum is 35 and the SE for sum is $\sqrt{10} * 1.7 = 5.4$. If $x = 40$, then $z = (40 - 35)/5.4 = 0.93$. The P-value is $1 - \text{pnorm}(0.93) = 17.6\%$.

b) If $n = 100$, the expected value for the sum is 350 and the SE for sum is $\sqrt{100} * 1.7 = 17$. If $x = 375$, then $z = (375 - 350)/17 = 1.47$. The P-value is $1 - \text{pnorm}(1.47) = 7.1\%$.

c) If $n = 225$, the expected value for the sum is 787.5 and the SE for sum is $\sqrt{225} * 1.7 = 25.5$. If $x = 750$, then $z = (750 - 787.5)/25.5 = -1.47$. The P-value is $\text{pnorm}(-1.47) = 7.1\%$.

3. Represent the situation as a box of 0s and 1s. Each Coffee party voter is a 1. Voters for other candidates are 0s. We don't know the percent of 1s or 0s. Among the random sample of 100 voters, 45% are 1s.

a) If we assume that the percent of 1s in the box is 45%, then we can estimate the SD of the box as $\sqrt{0.45 \times 0.55} \approx 0.5$. The SE for the percent of 1s is $0.5/sqrt100 = 5\%$. So, a 95% confidence interval on the percent of 1s is $45\% \pm 2 \times 5\% = 45\% \pm 10\%$.

b) True. Actually, prior to taking the sample, there is a 95% chance that the true value of the percent of 1s will be in the confidence interval that we compute.

c) False. The new interval should cover the true value even if the previous one doesn't the new sample average.

d) False. Doctors in Winchester form a different population. Their voting habits may or may not coincide with those of the population as a whole.

e) If the average of the box is really 51%, then the SD is $\sqrt{0.51 \times 0.49} \approx 0.5$ and an SE of $0.5/\sqrt{100} = 5\%$. That gives a z-score of $(45 - 51)/5 = -1.2$. The P-value is $\text{pnorm}(-1.2) = 11.5\%$. This gives weak support for the null hypothesis: the difference between observed (45%) and expected (51%) is due to chance.

f) If the average of the box is really 55%, then we still get about 0.5 for the SD and 5% for the SE. The z-score is $(45 - 55)/5 = -2$ and the P-value is $\text{pnorm}(-2) = 2.3\%$. This supports the alternative hypothesis: the difference between observed (45%) and expected (55%) is not due to chance.

g) If the average of the box is really 60%, then we get about 0.49 for the SD and 4.9% for the SE. The z-score is $(45 - 60)/4.9 = -3.1$ and the P-value is $\text{pnorm}(-3.1) \approx 0.1\%$. This supports the alternative hypothesis: the difference between observed (45%) and expected (60%) is not due to chance.

4. a) T; b) F; c) T; d) T; e) T; f) T; g) F; h) T; i) T; j) T; k) T

5. Each measurement is like drawing from a box of numbered errors. The SD of the errors is 2". So, the SE for the sum of 9 measurements is $\sqrt{9} \times 2 = 6$".

6. $\text{SE}_{\text{AV}} = \text{SD}_{\text{box}}/\sqrt{n}$ Since $n = 16$, the error decreases by a factor of $\sqrt{16} = 4$.

7. a) T; b) F; c) T; d) F; e) T; f) F

8. Model the situation as a box with five 0s in it and one 1. The 1 represents rolling a six. For this box, the average is 0.167 and the SD is $\sqrt{0.167 \times (1 - 0.167)} \approx 0.37$. The expected value for the percent of 1s is 16.7%.

a) If $n = 10$, the standard error for the percent of 1s is $16.7/\sqrt{10} = 5.3\%$. So, the z-score for the given observation is $(20 - 16.7)/5.3 = 0.62$. The P-value is $1 - \text{pnorm}(0.62) \approx 27\%$. This supports the null hypothesis: the difference between the observed value (20%) and expected value (16.7%) is due to chance.

b) If $n = 100$, the standard error for the percent of 1s is $16.7/\sqrt{100} = 1.67\%$. So, the z-score for the given observation is $(20 - 16.7)/1.67 \approx 2$. The P-value is $1 - \text{pnorm}(2) \approx 2.3\%$. This supports the alternative hypothesis: the difference between the observed value (20%) and expected value (16.7%) is not due to chance.

c) If $n = 225$, the standard error for the percent of 1s is $16.7/\sqrt{225} = 1.1\%$. So, the z-score for the given observation is $(20 - 16.7)/1.1 = 3$. The P-value is $1 - \text{pnorm}(3) \approx 0.1\%$. This supports the alternative hypothesis: the difference between the observed value (20%) and expected value (16.7%) is not due to chance.

9. a) 800 microns is the SD of the measurement process. b) The SE for the average is $800/\sqrt{100} = 80$ microns. c) 91.4402 cm $\pm$ 160 microns.