Answer Key to Second Midterm Examination, Math 225, Spring, 2001

Problem 1: TRUE or FALSE:

> A histogram with well-chosen bins shows more details about the shape of a distribution than a boxplot.

Solution: True.

Problem 2: TRUE or FALSE:

> For $p < 0.5$, a binomial distribution is skewed to the left.

Solution: False. If $p < 0.5$, the distribution is skewed to the right. The skewness will be stronger when p is far from 0.5 and when n is small.

Problem 3: TRUE or FALSE:

> There are $5!/(2!3!)$ ways to order three items out of five.

Solution: False. There are $5!/(2!3!)$ ways to choose three items from five without regard to order and $5!/2!$ ways to do it with regard to order.

Problem 4: TRUE or FALSE:

> A 95% confidence interval for μ will be wider than a 99% confidence interval for μ when based on the same data.

Solution: False. It would be more narrow.

Problem 5: TRUE or FALSE:

> A 95% confidence interval for μ with σ known has a margin of error of 2.4. A sample size twice as large is necessary to decrease the margin of error to 1.2.

Solution: False. The sample size would need to be four times as large.

Problem 6: TRUE or FALSE:

> A 95% confidence interval for the mean number of children of 40-year-old American women cannot be $(1.7, 1.9)$ because the number of children a woman has is a whole number.

Solution: False. The mean of a distribution of whole numbers might not be a whole number.

Problem 7: TRUE or FALSE:

> If a hypothesis test is significant at the 5% level, the p-value must be smaller than 0.05.

Solution: True (if you are not picky about how things are defined if the p-value is **exactly** 0.05).

Problem 8: TRUE or FALSE:

> A p-value of 0.02 means that there is only a two percent chance that the null hypothesis is correct.

Solution: False. The p-value is a conditional probability calculated given the null hypothesis is true. It is not the probability of the null hypothesis.

Problem 9: TRUE or FALSE:

> If a hypothesis test is significant at the 1% level, it is also significant at the 5% level.

Solution: True. If a test is significant at the 1% level, the p-value is smaller than 0.01. It is, therefore, also smaller than 0.05.

Problem 10: TRUE or FALSE:

> The one-way ANOVA procedure assumes that the population standard deviations are the same for all populations in the study.

Solution: True.

Problem 11:
A fair coin is tossed 40 times. Let X be the number of heads.

(a) X has a binomial distribution with $n = 40$ and $p = 0.5$.

(b) X will almost certainly equal 20 because $\mu = np = 20$.

(c) $P(X > 25) = \sum_{k=26}^{40} {}_{40}C_k (0.5)^{40}$.

(d) The parameter n is large enough for the normal approximation to be accurate.

(e) $P(X > 25)$ is approximately the same as the probability a standard normal random variable is greater than $z = 25.5$.

Solution: (a), (c), and (d) are correct. Notice that $(0.5)^k(0.5)^{40-k} = (0.5)^{40}$. Also, $np = n(1-p) = 20 > 5$. (b) is false because binomial random variables are not almost certainly equal to their mean. (e) is false. It would be correct if 'standard normal random variable' were changed to 'normal random variable' with mean 20 and standard deviation $\sqrt{40(0.5)(0.5)}$.

Problem 12:
In a sample of five thirty-year-old men, the measured thyroid suppressing hormone levels are 0.4, 1.7, 4.5, 5.7, and 51.2.

(a) The mean is greater than the median.

(b) There are no outliers.

(c) The sample is skewed to the left.

(d) In a boxplot, the height of the middle box is 4.0.

(e) The standard deviation is larger than 2.0.

Solution: The data has a large outlier and so is skewed to the right with the mean larger than the median. A typical deviation from the mean is much larger than 2.0. (a) and (e) are correct. (d) was marked correct in either case because there is no single standard way to calculate a quartile. (b) and (c) are incorrect.

Problem 13:
Elizabeth predicts the suit of the top card of each of twenty well-shuffled standard decks. If she does not have ESP, she has a one in four chance of correctly predicting the suit of each card. In the experiment, Elizabeth correctly predicts the suits of nine cards. If X is a binomial random variable with $n = 20$ and $p = 0.25$, $P(X = 9) = 0.027$, $P(X \geq 9) = 0.041$, and $P(X > 9) = 0.014$. We are looking for evidence of ESP which allows a higher chance of correctly predicting the suits.

(a) The null hypothesis is $p > 0.25$.

(b) The p-value is 0.027.

(c) The p-value is 0.041.

(d) The p-value is 0.014.

(e) Elizabeth has been shown statistically to have ESP.

Solution: (c) is the only correct solution. The null hypothesis is the assumption $H_0 : p = 0.25$. The p-value is the probability of obtaining a result at least as extreme as that observed assuming the null hypothesis is true. This is the binomial probability $P(X \geq 9) = 0.041$ when $n = 20$ and $p = 0.25$. A hypothesis test asks if chance alone can explain the difference between what is observed and what is expected. The p-value of 0.041 implies some evidence against chance alone, but it is hardly conclusive. Also, there are possible explanations other than ESP for Elizabeth guessing correctly fairly often. Perhaps the cards are marked, or she can see a reflection in a mirror or someone's glasses, or the cards are not as well-shuffled as claimed. Or maybe she really did just get lucky — twenty is not a large sample and there is considerable initial doubt about the truth of the alternative hypothesis.

Problem 14:
A 95% confidence interval for the mean number of upsets in the first round of a basketball tournament is 7.8 ± 1.3 based on a sample of sixteen such tournaments. Think of this sample as a random sample from a larger hypothetical population of possible tournaments where the mean number of first round upsets is μ.

(a) The number of first round upsets was between 6.5 and 9.1 in 95% of the sample.

(b) In the future, we expect 95% of all tournaments to have 7, 8, or 9 first round upsets.

(c) The sample mean is 7.8.

(d) The sample standard deviation is 1.3.

(e) A two sided hypothesis test with $H_0 : \mu = 10$ would be rejected at the $\alpha = 0.05$ level.

Solution: (c) and (e) are correct. A confidence interval for μ is not the middle 95% of the sample. It is also not the middle 95% of the population. We center confidence intervals at the estimate, so the sample mean is 7.8. The margin of error, not the standard deviation, is 1.3. The value 10 is not in the 95% confidence interval, so a two-sided hypothesis test with $H_0 : \mu = 10$ would be significant at the 5% level and have a p-value less than 0.05.

Problem 15:
$\bar{X}$ is the sample mean of n numbers randomly drawn from a population with mean $\mu = 100$ and standard deviation $\sigma = 12$.

(a) The distribution of $\bar{X}$ is approximately normal for any value of n.

(b) If $n = 25$, the distribution of $\bar{X}$ is approximately normal for any population.

(c) If the population is normal, the distribution of $\bar{X}$ is exactly normal for any n.

(d) If the population is fairly symmetric and $n = 16$, the 95th percentile of the distribution of $\bar{x}$ is about $100 + (1.645)(3)$.

(e) If the population is normal and $n = 16$, $P(97 < \bar{X} < 103) \approx 0.68$.

Solution: By the Central Limit Theorem, it will be approximately normal for sufficiently large n, but the necessary size of n depends on the particular population. (a) is incorrect because we do not know if the population is normal. (b) is incorrect because $n = 25$ will not work for all populations. (c) is correct because the sampling distribution of $\bar{X}$ will be normal for any n if the population is normal. (d) is correct because the standard error is $12/\sqrt{16} = 3$ and the 95th percentile is 1.645 standard errors above the mean for any normal distribution. (e) is correct because approximately 68% of the area of any normal curve is within one standard deviation of the mean.

Problem 16:
In an experiment on the effectiveness of nasal strips to improve the ability to process oxygen, sixteen college athletes ride an exercise cycle, once with a nasal strip and once without, with two days of rest in between. The response variable is the volume of oxygen processed per second while at maximum exertion. The test statistic is $t = 1.72$. The numerator is the mean with the nasal strip minus the mean without.

(a) This is a matched pair experiment.

(b) There are fifteen degrees of freedom.

(c) Assuming equal population variances, there are thirty degrees of freedom.

(d) A hypothesis test with a one-sided alternative that the nasal strip increases the measured variable, is significant at the 5% level.

(e) The p-value is between 0.05 and 0.10.

Solution: This is a matched pair experiment and (a) is correct because the two measurements on the same individual can be thought of as having been sampled together as a pair. The sample size is 16 pairs of observations, so (b) is correct. There were not two independent samples, so (c) is incorrect. The p-value is the area to the right of 1.72 under a t distribution with 15 degrees of freedom. From the t table, this area is between 0.05 and 0.10. Thus, (d) is incorrect and (e) is correct.

Problem 17:
A physical therapist compares five new treatments with a control treatment.

(a) It is good statistical practice to make five separate 95% pairwise confidence intervals between the new treatments and the control.

(b) With Bonferroni's method to adjust for multiple comparisons, five separate 99% confidence intervals would have at least a 95% chance of all containing the true differences.

(c) Scheffe's method requires stating the desired comparisons before analyzing the data.

(d) The multiplier for the margin of error using Scheffe's method comes from an F distribution.

(e) The multiplier for the margin of error using Bonferroni's method comes from a t distribution.

Solution: (a) is incorrect because you should adjust for multiple comparisons. (b) is correct. (c) is incorrect because Scheffe's method allows you to make any comparisons after the fact. (d) and (e) are both correct. (Your test has 17(e) graded incorrectly. Increase your score by 1.)

Problem 18:
An investigator tests several treatments for acne. The response is percentage improvement. Values range from 46 to 72. There are no outliers or strong skewness. An ANOVA table is here.

Source	Sum of Squares	DF	Mean Square	F	p-value
Among	2133.66	2	1066.83	262.12	0.0000
Within	130.30	32	4.07		
Total	2263.96	34			

(a) There are two different treatments in the study.

(b) There were 35 subjects in the study.

(c) This is a two-way analysis of variance.

(d) The area to the right of 262.12 under an F distribution with 34 degrees of freedom is essentially 0.

(e) Chance alone can explain the differences in response to the treatments.

Solution: There are two degrees of freedom among treatments, so there are three treatments and (a) is incorrect. (b) is correct. (c) is incorrect because there is only a single explanatory variable. (d) is incorrect because there is no F distribution with 34 degrees of freedom. The p-value is the area to the right of 262.12 under an F distribution with 2 and 32 degrees of freedom. (e) is incorrect because the very small p-value means chance cannot explain the difference alone.

Problem 19: (10 points)
Birth weights of babies are normally distributed with a mean of 120 ounces and a standard deviation of 22 ounces. What is the probability that exactly two of four randomly sampled babies weigh more than 132 ounces?

Solution: The probability that a single birth weight is larger than 132 is the area to the right of (132-120)/22=0.55 under a standard normal curve. This area is 0.2912. The probability that exactly two of four babies weigh more than 132 ounces is a binomial probability with $n = 4$ and $p = 0.2912$. $P(X = 2) = {}_4C_2(0.2912)^2(1 - 0.2912)^2 = 0.2556$.

Problem 20: (15 points)
In a study on the effects of carbon monoxide exposure on patients with coronary artery disease, men were selected from three different medical centers. Twenty-one men came from Johns Hopkins Medical Center, sixteen came from Los Amigos Medical Center, and twenty-three came from St. Louis University Medical Center. The response variable is the one second forced expiratory volume (FEV) in liters prior to treatment. Side-by-side boxplots indicate fairly symmetric distributions with similarly sized spread in each sample. The three sample means are 2.63, 3.03, and 2.88, respectively. Is there evidence that the three populations of subjects have different mean FEV values? A partial ANOVA table is below.

Source	Sum of Squares	Degrees of Freedom	Mean Square	F-statistic	P-value
Among	1.58				0.052
Within	14.48				
Total					

(a) Complete the ANOVA table.

(b) What is the pooled estimate of the standard deviation of the chance error?

(c) Circle all correct statements.

1. The three population mean FEVs are all equal because the p-value is greater than 0.05.
2. The null hypothesis that the population means are all equal is not rejected at the 5% level.
3. There is only mild evidence that the population mean FEVs at the three centers are unequal.
4. There is strong evidence that the population mean FEV at Johns Hopkins Medical Center is smaller than the population mean FEVs at the other two medical centers.

Solution: There are three populations, so there are two degrees of freedom among groups. There are a total of sixty patients, so there are $60 - 3 = 57$ degrees of freedom within groups and 59 degrees of freedom total. The rest of the table is below.

Source	Sum of Squares	Degrees of Freedom	Mean Square	F-statistic	P-value
Among	1.58	2	0.79	3.11	0.052
Within	14.48	57	0.254		
Total	16.06	59			

The pooled estimate of the standard deviation of a chance error is the square root of the mean square within, or $\sqrt{0.254} = 0.5$.

Statement 1 is false. Inability to reject the null hypothesis does not imply that the null hypothesis is true.

Statement 2 is true because the p-value is larger than 0.05.

Statement 3 is true because the p-value is smallish.

Statement 4 is false. If we cannot even say with much conviction that the means are unequal, we should not continue on to test which means are unequal.

Problem 21: (15 points)
Researchers who study bats are interested in the the distance bats must fly to feed. In an experiment, twenty-five male bats and twenty-five female bats are tagged and tracked by radio. For each bat, the average distance (in meters) of a feeding pass is measured. Summary data are shown here.

Sex	Sample Size	Sample Mean	Sample Standard Deviation
Female	25	205	100
Male	25	135	95

Furthermore, if the distances are sorted from smallest to largest separately within each sample and the corresponding pairs are matched, the mean difference (female minus male) is 70 meters with a standard deviation of 31 meters. One possible explanation for larger mean distances for females is that the males are more aggressive and drive the females away from the nearby feeding areas.

(a) It is most appropriate to analyze this problem with (matched pair / independent sample) methods.

(b) Construct a 95% confidence interval for either (1) the population mean paired difference in distances (female minus male) per feeding pass, or (2) the difference between the two population mean distances (female minus male) per feeding pass.

(c) Test either the hypothesis (1) the population mean paired difference is zero, or (2) the two population means are equal, versus a one-sided alternative. State hypotheses, calculate a test statistic, identify the degrees of freedom, and report a p-value.

(d) Circle all correct statements.

1. It is essentially certain that female bats fly farther than male bats on average.
2. There is fairly strong evidence that female bats fly farther than male bats on average.
3. There is fairly strong evidence that male bats force female bats away from nearby feeding areas.
4. There is fairly strong evidence that male bats and female bats fly the same distance on average.
5. Batgirl flies farther than Batman.

Solution: For matched pair techniques to be applicable, the data should be sampled in pairs. This is not the case. We have two populations of bats, male and female. If these had been brother/sister pairs of bats, matching would have been appropriate.

For two independent samples, we can choose to assume the population means are equal or not. If not, we need to estimate degrees of freedom which is best left to a computer. The sample standard deviations are quite close, so assuming the population standard deviations are equal is justified.

$$s_p = \sqrt{\frac{(25-1)100^2 + (25-1)95^2}{25+25-2}} = 97.5$$

This has to be between the two sample standard deviations as it is a type of averaging.

There are 48 degrees of freedom. Use $t^* = 2.014$ to be safe because our table has 45 df, but not 48 df.

A confidence interval is

$$(205-135) \pm 2.014 \times 97.5 \times \sqrt{\frac{1}{25} + \frac{1}{25}}$$

or

$$70 \pm 56$$

It is most appropriate to round a margin of error to two significant digits, and not more than the estimate itself.

For the one-sided hypothesis test, $H_0 : \mu_f = \mu_m$ and $H_a : \mu_f > \mu_m$. The test statistic is

$$t = \frac{(70 - 0)}{97.5\sqrt{\frac{1}{25} + \frac{1}{25}}} = 2.54$$

The p-value is between 0.005 and 0.01.

Statement 1 is too strong. Statement 2 is correct. Statement 3 is not correct. We can say that it is likely that chance alone cannot explain the difference, but we cannot conclude what is directly causing the difference without additional information. There are other possibilities. Statement 4 is incorrect.