University of Texas at Austin, Department of Mathematics
M358K - Applied Statistics

The Problem Bank for In-Term III

1.1. **The normal approximation to the binomial.**

Problem 1.1. (5 points) The probability that a randomly chosen player clears a jump in *Geometry Dash* is 0.4. If 100 independent players attempt the jump, what is the approximate probability that strictly fewer than 30 will be successful?

(a) About 0.0052
(b) About 0.0162
(c) About 0.0250
(d) About 0.0274
(e) None of the above.

Solution: (b)
Let X be the random variable representing the number of players who successfully perform the jump. Then

$$X \sim Binomial(n = 100, p = 0.4).$$

The expected value of X is $\mathbb{E}[X] = (100)(0.4) = 40$. The variance of X is $Var[X] = 40(0.6) = 24$, so that its standard deviation equals $SD[X] = \sqrt{24} = 4.899$. Using the normal approximation to the binomial, we obtain

$$\mathbb{P}[X < 30] = \mathbb{P}[X \leq 29.5] = \mathbb{P}\left[\frac{X - 40}{4.899} \leq \frac{29.5 - 40}{4.899}\right] \approx \Phi(-2.14) = 0.0162.$$

1.2. **A single proportion: Confidence intervals.**

Problem 1.2. (5 points) In a random sample of 1000 homes in a particular large city, it is found that 228 have fireplaces. Find the 99%-confidence interval for the proportion of homes in this city that have fireplaces. *Note: Round your margin of error to four places after the decimal point.*

(a) 0.228 ± 0.0143
(b) 0.228 ± 0.0342
(c) 0.228 ± 0.0256
(d) 0.228 ± 0.0412
(e) None of the above.

Solution: (b)
The point estimate is $\hat{p} = 228/1000 = 0.228$. The critical value z^* associated with the 99% confidence level is $z^* = 2.576$. So, the confidence interval is

$$0.228 \pm 2.576\sqrt{\frac{0.228(1 - 0.228)}{1000}} = 0.228 \pm 0.0342$$

1.3. **A single proportion: Hypothesis testing.**

Problem 1.3. (5 points) An up-and-coming chef claims that his soufflé success rate is at least 80%. To test his claim, the celebrated *Culinary Guild*, asks him to make a total of 81 soufflés. Out of this sample, 21 soufflés drop and 60 are successes. What is the $p-$value the resident statistician will report to the trustees of the *Culinary Guild*?

a.: About 0.0537.
b.: About 0.0764.
c.: About 0.0885.
d.: About 0.0918.
e.: None of the above

Solution: d.
The null and the alternative hypotheses are

$$H_0 : p = 0.80 \quad vs. \quad H_a : p < 0.80.$$

The observed proportion of successes is $\hat{p} = 60/81 = 20/27 = 0.74$. The observed value of the $z-$statistic is (under the null hypothesis), in our usual notation,

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{20}{27} - 0.8}{\sqrt{\frac{0.8(0.2)}{81}}} \approx -1.3333.$$

According to the standard normal tables, the $p-$value is 0.0918

1.4. **Two proportions: Confidence intervals.**

Problem 1.4. (5 points) A survey of 1000 students in Austin found that 274 chose Pikachu as their favorite Pokemon. In another survey of 760 students in Pittsburgh, 240 chose Pikachu. Find the 95%-confidence interval for the difference in the two population proportions. *Note: Round your point estimates and the margin of error to four places after the decimal point.*

(a) $(-0.0849, 0.0013)$
(b) $(-0.0344, 0.0213)$
(c) $(-0.0951, -0.0033)$
(d) $(0.0149, 0.0913)$
(e) None of the above.

Solution: (a)
Let the proportion of Pikachu fans in Austin be p_A and let the proportion of Pikachu fans in Pittsburgh be p_P. Then, the point estimates of the two proportions equal

$$\hat{p}_A = \frac{274}{1000} = 0.274 \quad \text{and} \quad \hat{p}_P = \frac{240}{760} = 0.3158.$$

The critical value associated with the 95%-confidence level is 1.96. The standard error equals

$$\sqrt{\frac{0.274(1-0.274)}{1000} + \frac{0.3158(1-0.3158)}{760}} = 0.022.$$

So, the confidence interval is $(0.274 - 0.3158) \pm 1.96(0.022)$, i.e., $(-0.0849, 0.0013)$

1.5. **Two proportions: Hypothesis testing.**

Problem 1.5. (5 points) We want to compare the proportions of people from town A (pop. 25000) and people from town B (pop. 50000) whose favorite music genre is country music. Two polls are conducted. In town A, 120 out of the total of 200 people like country music best. In town B, 240 out of the total of 500 people like country music best. You conduct a hypothesis test for whether there is a difference in population proportions. Which $p-$value would you report? *Note: All your estimates should be rounded to four decimal places after the decimal point.*

(a) 0.0019
(b) 0.0042
(c) 0.0192
(d) 0.0431
(e) None of the above.

Solution: (b)
Let p_A be the probability that a randomly chosen person from town A prefers country music and let p_B be the probability that a randomly chosen person from town B prefers country music. We are testing

$$H_0 : p_A = p_B \quad vs. \quad H_a : p_A \neq p_B.$$

The point estimates for the two proportions are $\hat{p}_A = \frac{120}{200} = 0.60$ and $\hat{p}_B = \frac{240}{500} = 0.48$. The pooled estimate of the proportion is

$$\hat{p} = \frac{120 + 240}{200 + 500} = 0.5143$$

The observed value of the $z-$statistic, under the null, is

$$z = \frac{0.60 - 0.48}{\sqrt{(0.5143)(1 - 0.5143)\left(\frac{1}{200} + \frac{1}{500}\right)}} = 2.87$$

So, the $p-$value is $2\Phi(-2.87) = 2(0.0021) = 0.0042$.

1.6. **The chi-squared distribution.**

Problem 1.6. (5 points) Let the random sample $X_1, \dots, X_{10}$ be drawn from a normal distribution with mean 0 and variance 4. Define

$$Y = \sum_{i=1}^{10} X_i^2.$$

Find the constant q such that

$$\mathbb{P}[Y \leq q] = 0.975.$$

(a) 78.32
(b) 79.04
(c) 80.48
(d) 81.92
(e) None of the above.

Solution: (d)
The random variable $\frac{Y}{4}$ has the χ^2-distribution with 10 degrees of freedom. In the χ^2-tables, we find that $\chi^2_{0.025}(df = 10) = 20.48$. Since

$$\mathbb{P}[Y \le q] = \mathbb{P}\left[\frac{Y}{4} \le \frac{q}{4}\right],$$

we conclude that $\frac{q}{4} = 20.48$. Hence, $q = 81.92$.

1.7. **The goodness-of-fit test.**

Problem 1.7. (5 points) You want to test whether a six-sided die is fair. Here are the observed counts in 120 rolls of the die:

Side	1	2	3	4	5	6
Count	20	22	17	19	24	18

What is your decision?

(a) Reject at the 0.01 significance level.
(b) Fail to reject at the 0.01 significance level; reject at the 0.025 significance level.
(c) Fail to reject at the 0.025 significance level; reject at the 0.05 significance level.
(d) Fail to reject at the 0.05 significance level; reject at the 0.10 significance level.
(e) Fail to reject at the 0.10 significance level.

Solution: (e)
Let $p_i, i = 1, \dots, 6$ be the probability that the die falls on i. We are testing

$$\begin{aligned} H_0 :&\quad p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6} \\ &\quad vs. \\ H_a :&\quad \text{At least one of the probabilities } p_i \text{ is different from } \frac{1}{6}. \end{aligned}$$

The observed value of the χ^2-statistic is

$$q^2 = \frac{1}{20}((20-20)^2 + (22-20)^2 + (17-20)^2 + (19-20)^2 + (24-20)^2 + (18-20)^2) = 1.7.$$

With $6 - 1 = 5$ degrees of freedom, using the χ^2-table, we see that the $p-$value is more than 0.30.

1.8. **The chi-squared test of independence.**

Problem 1.8. (5 points) *Source: "Probability and Statistics for Engineers and Scientists" by Walpole, Myers, Myers, and Ye.*
A study was conducted to determine whether there is an association between being for or against tax reform and income level. The results are displayed in the following table:

	Low	Medium	High	Total
For	182	213	203	598
Against	154	138	110	402
Total	336	351	313	1000

Your goal is to test whether being for or against the tax reform is independent from income level. The observed value of the relevant test statistic is 7.85. What is your decision for this hypothesis test?

(a) Reject at the 0.01 significance level.

(b) Fail to reject at the 0.01 significance level; reject at the 0.025 significance level.

(c) Fail to reject at the 0.025 significance level; reject at the 0.05 significance level.

(d) Fail to reject at the 0.05 significance level; reject at the 0.10 significance level.

(e) Fail to reject at the 0.10 significance level.

Solution: (b)
The distribution of the test statistic is approximately χ^2 with $(3-1)(2-1) = 2$ degrees of freedom. Consulting the χ^2-table, we see that the given observed value of the test statistic is between the critical values $\chi^2_{0.01}(df = 2)$ and $\chi^2_{0.025}(df = 2)$. So, the $p-$value is between 0.01 and 0.025.

1.9. **t-procedures: Single mean: Confidence intervals.**

Problem 1.9. (5 points) *Source: Ramachandran-Tsokos.*
A Dobson unit is the most basic measure used in ozone research. The unit is named after G.M.B. Dobson, one of the first scientists to investigate atmospheric ozone (between 1920 and 1960). He designed the Dobson spectrometer – the standard instrument used to measure ozone from the ground. The total ozone levels were measured in Dobson units at 12 randomly selected locations of earth on a particular day. You look at the $q-q$ plot and convince yourself that you can assume the normal distribution for the total ozone levels. The sample average is 285.7 and the sample standard deviation is 43.9. Construct the 95%-confidence interval for the mean total ozone level.

a.: 285.7 ± 20.8468
b.: 285.7 ± 27.8929
c.: 285.7 ± 12.6728
d.: 285.7 ± 43.9
e.: None of the above

Solution: b.
The sample mean has the $t-$distribution with $12 - 1 = 11$ degrees of freedom. So, the critical value associated with the 0.025 upper-tail probability equals $t^* = 2.201$. The standard error we obtain from our sample equals $43.9/\sqrt{12}$. So, the confidence interval we seek is

$$285.7 \pm 2.201\left(\frac{43.9}{\sqrt{12}}\right) = 285.7 \pm 27.8929. \tag{1.1}$$

1.10. **t-procedures: Single mean: Hypothesis testing.**

Problem 1.10. (5 points) Bags of tinsel are arriving at your holiday store. They are declared to contain 25lbs of tinsel each, but you suspect that the bags you are receiving are actually lighter than declared. You decide to test your hypothesis. A sample of 25 bags is collected and weighed. The sample average is 24.8lbs and the sample standard deviation is 1.6lbs. Assuming the normal distribution of the tinsel weight per bag, what are the bounds on the $p-$value you obtain for your sample?

a.: Less than 0.025.
b.: More than 0.025, but less than 0.05.
c.: More than 0.05, but less than 0.15.
d.: More than 0.15, but less than 0.25.
e.: More than 0.25.

Solution: e.
The null and the alternative hypotheses are

$$H_0 : \mu = 25 \quad vs. \quad H_a : \mu < 25.$$

The observed value of the $t-$statistic is (under the null hypothesis), in our usual notation,

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{24.8 - 25}{1.6/\sqrt{25}} = -0.625.$$

The number of degrees of freedom for our $t-$distribution is $25 - 1 = 24$. The $p-$value is more than 0.25.

1.11. **t-procedures: Paired data: Confidence intervals.**

Problem 1.11. (5 points) A weight-loss program is being evaluated. The percentage of body fat is measured for 10 adults before the program and after the program. The average body fat loss is recorded as 0.447 while the sample standard deviation turned out to be 1.7296. What is the 90%-confidence interval for the mean of the difference in body fat (in percentages)? Assume that the distribution of body fat is normal. *Note: Please, round your margin of error to four places after the decimal point.*

(a) 0.447 ± 0.0126
(b) 0.447 ± 0.4494
(c) 0.447 ± 1.0026
(d) 0.0447 ± 1.7296
(e) None of the above.

Solution: (c)
The point estimate for the mean difference is given at 0.447. The critical value of the $t-$distribution with $10 - 1 = 9$ degrees of freedom corresponding to the 90% confidence level is $t^* = 1.833$. So, the confidence interval is of the form

$$0.447 \pm 1.833 \left(\frac{1.7296}{\sqrt{10}} \right) = 0.447 \pm 1.0026.$$

1.12. **t-procedures: Paired data: Hypothesis testing.**

Problem 1.12. (5 points) *Source: "Probability and Statistical Inference" by Hogg, Tanis, and Zimmerman.*
A vendor of milk sells dry milk to a company which makes baby formula. They are both interested in the fat content of milk since that helps the babies' brains grow. Both the vendor and the company take an observation from each batch and measure the fat content in percent. There is a total of 10 batches assessed. The average of the company's test results was 0.905. The average of the vendor's test results was 0.778. The sample standard deviation of the differences is 0.2719. Assume that the fat content in normally distributed. You are testing to see whether there is a difference between the company's and the vendor's mean fat content in milk. What can you say about the $p-$value for this hypothesis test?

(a) It's below 0.05.
(b) It's between 0.05 and 0.10.
(c) It's between 0.10 and 0.20.
(d) It's between 0.20 and 0.30.
(e) It's over 0.30.

Solution: (c)

The observed value of the $t-$statistic, under the null hypothesis, equals

$$t = \frac{0.905 - 0.778}{\frac{0.2719}{\sqrt{10}}} = 1.477.$$

The number of degrees of freedom is $10 - 1 = 9$. So, the upper-tail probability associated with the observed value of the $t-$statistic lies between 0.05 and 0.10. Since we are testing for a **difference**, this is a two-tailed test. So, the $p-$value is between 0.10 and 0.20.