Review for the final examination

Examples

1. Match each item in column A with the correct item in column B.

A	B
$.01 < p\text{-value} < .025$	(a) Do not reject H_0 at $\alpha = .1$
$p\text{-value} = .06$	(b) Reject H_0 at $\alpha = .1$, but not at .05
$p\text{-value} > .25$	(c) Reject H_0 at $\alpha = .05$, but not at .01
$p\text{-value} = .007$	(d) Reject H_0 at $\alpha = .01$

2. A research worker wants to determine the average time it takes a machine to rotate the tires of a car, and she wants to be able to assert with 95% confidence that the mean of her sample is off by at most .5 minutes. If $\sigma = 1.6$ minutes, how large a sample will she have to take?

3. A public health survey is to be designed to estimate the proportion p of a population having a defective version of football helmets. How many persons should be examined if the public health commissioner wishes to be 98% certain that the error of estimation is below .05 when:

   (a) there is no knowledge about the value of p?

   (b) p is known to be about .3?

4. A random sample of 75 SUNY Stony Brook students is given a verbal aptitude test. Here are the raw scores:

42	42	43	43	46	47	49	50	50	50	52	52	52	53	54	54	54	54	56	56
56	57	57	57	57	57	58	58	58	58	59	59	59	59	60	60	61	61	61	62
62	63	63	63	64	64	65	66	66	66	66	66	66	67	68	69	70	70	71	71
71	71	72	73	74	74	75	75	77	78	78	80	80	85	90					

   (Hint: $\overline{x} = 62.0$, $s = 10.3$)

   (a) Construct a 95% confidence interval for the population mean score.

   (b) Construct a 99% confidence interval for the proportion who scored 60 or above on the exam.

5. A random sample of 60 seventh grade students in Suffolk County is given an exam on reading comprehension. Below are their raw scores.

68	73	74	77	78	83	84	84	84	85	86	87	87	88	88	88
90	90	91	92	92	93	93	94	94	95	96	97	98	98	98	98
98	99	99	100	100	102	102	102	104	104	104	107	107	108	108	109
109	109	111	111	113	113	114	114	114	115	115	133				

   Nationally, the population mean score on this exam is 100 points. Furthermore, nationally, 25% of the students score 110 points or higher on this exam. (Hint: $\overline{x} = 97.45$, $s = 12.48$)

   (a) Analyze these data if the goal is to show the population mean score of Suffolk County students is different from the nationwide population mean score. Use $\alpha = .05$.

   (b) Analyze these data if the goal is to show the population proportion of Suffolk County students who score 110 points or higher is smaller than the nationwide population proportion of students who score 100 points or higher. Use $\alpha = .05$.

6. All 10 year-old children are given an examination on mathematical achievement. For the entire country, the (population) mean score is 60. A random sample of 10 year-olds is selected from Utah. We do not know the population mean or standard deviation for Utah. Below is a listing of the data set with summary statistics. (Hint: $n = 20$, $\overline{x} = 56.70$ and $s = 8.11$)

45 46 48 48 50 50 51 52 53 54 56 61 62 62 63 63 64 65 67 74

   (a) Construct a 90% confidence interval for the population mean score in Utah.
   (b) Construct a 90% confidence interval for σ in Utah.
   (c) I want to show that the population mean score in Utah is below the national mean. What does an analysis of the data show? Use $\alpha = .1$.

7. All 10 year-old children are given an examination on mathematical achievement. For the entire country, the (population) mean score is 60. Random samples of 10 year-olds are selected from two states - Iowa and Ohio. We do not know the population means or standard deviations for these two states. Below is a listing of the data sets with summary statistics.

For Iowa ($n = 50$, $\overline{x} = 59.32$ and $s = 6.80$):

43	47	48	49	49	52	54	54	54	55	55	55	55	55	56	56	56	56
56	57	58	58	58	59	59	59	59	59	59	59	61	61	61	62	62	62
62	63	64	64	65	66	67	68	69	71	71	72	73	73				

For Ohio ($n = 75$, $\overline{x} = 62.15$ and $s = 6.55$):

50	50	52	52	52	52	53	53	53	54	54	55	55	55	55	56	56	56
57	57	57	57	58	58	59	59	60	60	60	60	60	60	61	61	61	62
62	62	62	63	64	64	64	64	64	64	65	65	65	65	66	66	67	67
67	67	67	68	68	68	69	69	69	69	69	70	71	71	71	72	72	72
73	74	76															

Compare the population means in Iowa and Ohio using $\alpha = .1$ by constructing an appropriate confidence interval.

8. Independent random samples are selected from each of two populations. The summary statistics are:

$$n_1 = 4,\ \overline{x} = 13.8,\ s_1^2 = 11.0;\quad n_2 = 7,\ \overline{y} = 10.0,\ s_2^2 = 8.0.$$

Conduct a hypothesis test if the goal is to show that the first population has a larger mean than the second population. Use $\alpha = .05$.

9. A dealer of a certain Japanese car claims that it is easier to change spark plugs on a 4-cylinder model that he sells than on a comparable US model. To test the claim, 8 mechanics were assigned to change spark plugs on each of the two models. Here are the times (in minutes) they required to change the plugs on the two models.

	Mechanic number							
	1	2	3	4	5	6	7	8
US model	19	24	19	16	16	23	19	26
Japanese model	15	21	22	13	18	21	20	25
Difference	4	3	-3	3	-2	2	-1	1

Hint: For the US model data, the sample mean is 20.25 minutes, and the sample standard deviation is 3.69 minutes. For the Japanese model data, those numbers are 19.38 and 3.89. For the differences, those numbers are 0.88 and 2.59.

   (a) Compute a 95% confidence interval for the difference in population means.
   (b) Now assume that the same data had been obtained by 16 different mechanics, with 8 selected at random to work on the Japanese model and the remaining 8 assigned to the US model. Now compute a 95% confidence interval for the difference in population means.

Answers to the Examples

1. (c), (b), (a), (d)

2.

$$n = \left(\frac{z_{\alpha/2}\sigma}{d}\right)^2 = \left(\frac{z_{.025}\sigma}{d}\right)^2 = \left(\frac{1.96 \cdot 1.6}{.5}\right)^2 = 39.3 \Longrightarrow n = 40$$

3.

$$n = \frac{1}{4}\left(\frac{z_{\alpha/2}}{d}\right)^2 = \frac{1}{4}\left(\frac{2.33}{.05}\right)^2 = 542.89 \Longrightarrow n = 543$$

$$n = (.3)(.7)\left(\frac{z_{\alpha/2}}{d}\right)^2 = (.3)(.7)\left(\frac{2.33}{.05}\right)^2 = 456.03 \Longrightarrow n = 457$$

4. (a)

$$\overline{x} \pm z_{.025}\frac{s}{\sqrt{n}} = 62.0 \pm 1.96\frac{10.3}{\sqrt{75}} = 62.0 \pm 2.3 = (59.7,\ 64.3)$$

(b) $\hat{p} = 41/75 = .547$. A 99% CI is

$$\hat{p} \pm z_{.005}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .547 \pm 2.58\sqrt{\frac{(.547)(.453)}{75}} = .547 \pm .148 = (.399,\ .695).$$

5. (a) H_0: $\mu = 100$ versus H_1: $\mu \neq 100$
The rejection region is $|Z| > z_{.025} = 1.96$.

$$z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{97.45 - 100}{12.48/\sqrt{60}} = \frac{-2.55}{1.611} = -1.58$$

Since $|z| = 1.58 < 1.96$, do not reject H_0.
p-value $= P(|Z| > 1.58) = 2P(Z < -1.58) = 2(.057) = .114$

(b) H_0: $p \geq .25$ versus H_1: $p < .25$
The rejection region is $Z < -z_{.05} = -1.645$. Since $\hat{p} = 10/60 = .1667$,

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{.1667 - .25}{\sqrt{(.25)(.75)/60}} = -1.49 > -1.645$$

Do not reject H_0.
p-value $= P(Z < -1.49) = .068$

6. df $= n - 1 = 19$

(a)

$$\overline{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = \overline{x} \pm t_{.05}\frac{s}{\sqrt{n}} = 56.7 \pm 1.729\frac{8.11}{\sqrt{20}} = 56.7 \pm 3.14 = (53.56,\ 59.84)$$

(b)

$$\left(s\sqrt{\frac{n-1}{\chi^2_{.05}}},\ s\sqrt{\frac{n-1}{\chi^2_{.95}}}\right) = \left(8.11\sqrt{\frac{19}{30.14}},\ 8.11\sqrt{\frac{19}{10.12}}\right) = (6.44,\ 11.11)$$

(c) H_0: $\mu \geq 60$ versus H_1: $\mu < 60$
Rejection region: $T \leq -t_{.1} = -1.328$

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{56.7 - 60}{8.11/\sqrt{20}} = -1.820 < -1.328 \Longrightarrow \text{Reject } H_0$$

$1.729 = t_{.05} < 1.820 < t_{.025} = 2.093 \Longrightarrow .025 < p\text{-value} < .05$

7. A 90% confidence interval for the difference of the means is

$$\begin{aligned}\overline{x} - \overline{y} \pm z_{.025}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} &= 59.32 - 62.15 \pm 1.645\sqrt{\frac{(6.80)^2}{50} + \frac{(6.55)^2}{75}} \\ &= -2.83 \pm 2.01 = (-4.84,\ -.82)\end{aligned}$$

Since 0 is not included in this confidence interval, reject H_0. There is a significant difference between the two means.

8. H_0: $\mu_1 - \mu_2 = 0$ versus H_1: $\mu_1 - \mu_2 > 0$
df $= n_1 + n_2 - 2 = 9$. Rejection region: $T \geq t_\alpha = t_{.05} = 1.833$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{3(11.0) + 6(8.0)}{4 + 7 - 2} = 9 \Longrightarrow s_p = 3$$

$$t = \frac{\overline{x} - \overline{y}}{s_p\sqrt{(1/n_1) + (1/n_2)}} = \frac{13.8 - 10.0}{3\sqrt{(1/4) + (1/7)}} = 2.02 > 1.833 \Longrightarrow \text{Reject } H_0$$

$1.833 = t_{.05} < 2.02 < 2.262 = t_{.025} \Longrightarrow .025 < p\text{-value} < .05$

9. (a) df $= n - 1 = 7$

$$\overline{d} \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}} = \overline{d} \pm t_{.025}\frac{s_d}{\sqrt{n}} = .88 \pm 2.365\frac{2.59}{\sqrt{8}} = .88 \pm 2.17 = (-1.29,\ 3.05)$$

(b) df $= n_1 + n_2 - 2 = 8 + 8 - 2 = 14$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{7(3.69)^2 + 7(3.89)^2}{8 + 8 - 2} = 14.37 \Longrightarrow s_p = 3.79$$

A 95% confidence interval is

$$\overline{x} - \overline{y} \pm t_{\alpha/2}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \overline{x} - \overline{y} \pm t_{.025}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= 20.25 - 19.38 \pm 2.145(3.79)\sqrt{\frac{1}{8} + \frac{1}{8}} = .87 \pm 4.06 = (-3.19,\ 4.93)$$