STAT100 Elementary Statistics and Probability

Exam 3, Friday, August 22, 2014 **Solution**

Instructions: Show all work clearly and in order, and circle your final answers. Justify your answers algebraically whenever possible. A correct answer with no justification may not receive full credit. The exam is worth a total of 110 points.

Calculator is allowed in this exam for basic calculation only. It is also allowed to bring the normal table, t-table, χ^2-table, formulas for (difference) of means, plus a piece of paper of formulas prepared by yourself.

1. Recorded here are the amounts of decrease in percent body fat for eight participants in an exercie program over three weeks.

$$2.8 \qquad 10.6 \qquad -1.2 \qquad 12.9 \qquad 15.1 \qquad -2.0 \qquad 7.6 \qquad 10.2.$$

(a) (8 points) Construct a 95% confidence interval for the population mean amount μ of decrease in percent body fat over the three-week program.

Solution: The estimator of μ is the sample mean $\bar{x} = 7$. **(2 points)**
The standard error is estimated by $S.E. = \frac{s}{\sqrt{n}} = \frac{6.435}{\sqrt{8}} = 2.275$. **(4 points)**
We use t distribution with $d.f. = 7$ for the cutoff. Look up the table and get $t_{.025} = 2.365$. **(6 points)**
Therefore, the 95% confidence interval is

$$[7 - 2.365 \cdot 2.275, 7 + 2.365 \cdot 2.275] = [1.620, 12.380]. \quad \textbf{(8 points)}$$

(b) (4 points) If you were to test $H_0 : \mu = 15$ vesus $H_1 : \mu \neq 15$ at $\alpha = .05$, what would you conclude from your result in part (a)?

Solution: This two-side hypotheses is related to the confidence interval in (a) as α matches. As 15 is outside the 95% confidence interval, the null hypothesis H_0 is rejected. **(4 points)**

(c) (10 points) Consider hypotheses $H_0 : \sigma = 8$ vesus $H_1 : \sigma \neq 8$. Should H_0 be rejected at $\alpha = .05$?

Solution: We perform a χ^2 test. The test statistic is

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{7 \cdot (6.435)^2}{8^2} = 4.529. \quad \textbf{(5 points)}$$

> We reject H_0 if $\chi^2 \geq \chi^2_{\alpha/2}$ or $\chi^2 \leq \chi^2_{1-\alpha/2}$. Look up the cutoffs $\chi^2_{.025} = 16.01$ and $\chi^2_{.975} = 1.69$ with $d.f. = 7$. **(8 points)**
> As 4.529<16.01 and 4.529>1.69, we retain H_0. **(10 points)**

2. Given the following 2 data sets from 2 populations.

$$n_1 = 250, \quad \bar{x} = 88.8, \quad s_1^2 = 7.5.$$

$$n_2 = 300, \quad \bar{y} = 91.7, \quad s_2^2 = 18.$$

(a) (8 points) Find the 94% confidence interval of the difference in population means.

> **Solution:** The estimator of $\mu_1 - \mu_2$ is $\bar{x} - \bar{y} = -2.9$. **(2 points)**
> The estimated standard error is $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = .3$. **(4 points)**
> As n_1, n_2 are big, we use Z-test. $\alpha = .06$. The cutoff is $z_{.03} = \Phi^{-1}(.97) = 1.88$. **(6 points)**
> Finally, the 94% confidence interval is
>
> $$[-2.9 - 1.88 \cdot .3, -2.9 + 1.88 \cdot .3] = [-3.464, -2.336]. \quad \textbf{(8 points)}$$

(b) (8 points) Consider hypotheses $H_0 : \mu_1 - \mu_2 = -2$ vesus $H_1 : \mu_1 - \mu_2 \neq -2$. Find the P-value, and conclude that if the data support the argument H_1 or not.

> **Solution:** First, compute the test statistic
>
> $$z = \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{-2.9 - (-2)}{.3} = -3. \quad \textbf{(3 points)}$$
>
> Next, find the $P-$value
>
> $$P - \text{value} = P(|Z| > |z|) = 2\Phi(-3) = .0026. \quad \textbf{(6 points)}$$
>
> As the $P-$value is small, the data DOES support the argument H_1. **(8 points)**

3. To compare the crop yields from two strains of wheat, A and B, eight random experiments were contucted for each strain. The data of yield in pounds per plot is given below.

Strain A	23	39	19	43	33	29	28	42
Strain B	18	33	21	34	33	20	21	40

By calculation, we have for strain A, sample mean $\bar{x} = 32$ and sample variance $s_1^2 = 78$. Also, for strain B, sample mean $\bar{y} = 27.5$ and sample variance $s_2^2 = 70$. (You can use these values directly without verifying from the data.)

(a) (8 points) Test the null hypothesis that the mean yield of **strain A** is 30 pounds versus a two-sided alternative. Take $\alpha = .05$.

> **Solution:** Construct hypotheses: $H_0 : \mu_1 = 30$ vs $H_1 : \mu_1 \neq 30$. We use t-test, rejecting H_0 if $|T| > t_{\alpha/2}$ with $d.f. = 7$. **(2 points)**
>
> Compute the test statistic $T = \dfrac{\bar{X} - \mu_1}{S_1/\sqrt{n}} = \dfrac{32 - 30}{\sqrt{78}/\sqrt{8}} = .641$. **(4 points)**
>
> Look up the table to get the cutoff $t_{.025} = 2.365$ with $d.f. = 7$. **(6 points)**
>
> As $|.641| < 2.365$, we retain H_0. **(8 points)**

(b) (10 points) Obtain a 99% confidence interval for the difference of population mean yields between strain A and B.

> **Solution:** The estimator of $\mu_1 - \mu_2$ is $\bar{x} - \bar{y} = 4.5$. **(2 points)**
>
> As $s1/s2 = 1.056$ is between $1/2$ and 2, we pool the data for standard deviation
>
> $$s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = 8.602.$$
>
> The estimated standard error is $s_{pooled}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}} = 4.301$. **(6 points)**
>
> The cutoff is $t_{.005}$ is from t- distribution with $d.f. = n_1 + n_2 - 2 = 14$. Therefore, $t_{.025} = 2.977$. **(8 points)**
>
> The 99% confidence interval of $\mu_1 - \mu_2$ is
>
> $$[4.5 - 2.977 \cdot 4.301, 4.5 + 2.977 \cdot 4.301] = [-8.304, 17.304]. \quad \textbf{(10 points)}$$

(c) (12 points) Additional information shows each pair of data in the same column comes from experiments conducted at the same farm. Does the *paired* data substantiate the claim that strain A has a higher yield than strain B? Test with $\alpha = .05$ on the difference of the paired data.

> **Solution:** We first calculate the difference for each pair of data $D = X - Y$
>
> $$5 \quad 6 \quad -2 \quad 9 \quad 0 \quad 9 \quad 7 \quad 2,$$
>
> and its mean $\bar{D} = 4.5$ and variance $s_D = 4.106$. **(3 points)**
>
> We form hypothesis test on μ_D: $H_0 : \mu_D = 0$ vesus $H_1 : \mu_D > 0$ (indicating stain A has a higher yield.) **(5 points)** We use t-test, rejecting H_0 if $T > t_\alpha$ with $d.f. = 7$.

Compute the test statistic $T = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}} = \frac{4.5 - 0}{4.106/\sqrt{8}} = 3.100$. **(8 points)**

Look up the table to get the cutoff $t_{.05} = 1.895$ with $d.f. = 7$. **(10 points)**

As 3.100>1.895, we reject H_0. Therefore, the data substantiate the claim that strain A has a higher yield. **(12 points)**

4. A major clinical trial of a new vaccine for type-B hepatitis was conducted with a high-risk group of 1000 male volunteers. From this group, 600 men were given the vaccine and the other 400 a placebo. A follow-up of all these individuals yielded the data:

	Got hepatitis	Did not get hepatitis	Total
Vaccine	22	578	600
Placebo	58	342	400
Total	80	920	1000

To see if the vaccine is performing different from placebo, we form the hypotheses $H_0 : p_1 = p_2$ vesus $H_1 : p_1 \neq p_2$, where p_1, p_2 are proportions of people who get hepatitis by taking the vaccine or a placebo, respectively. Take $\alpha = .05$.

(a) (10 points) Use Z-test to determine if H_0 is rejected or not.

Solution: From the data, $\hat{p}_1 = \frac{22}{600} = .0367$, $\hat{p}_2 = \frac{58}{400} = .145$, $\hat{p} = \frac{80}{1000} = .08$.

The estimator of $p_1 - p_2$ is $\hat{p}_1 - \hat{p}_2 = -.1083$. **(2 points)**

The estimated standard error is $\sqrt{\hat{p}(1-\hat{p})}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = .0175$. **(5 points)**

The test statistic $z = \frac{-.1083}{.0175} = -6.19$. **(7 points)**

Two sided cutoff $z_{.025} = 1.96$. **(8 points)**

As $|-6.19| > 1.96$, we reject H_0. **(10 points)**

(b) (10 points) Use χ^2-test to determine if H_0 is rejected or not.

Solution: Compute the expected frequency. **(4 points)**

	Got hepatitis	Did not get hepatitis	Total
Vaccine	22	578	600
	(48)	(552)	
Placebo	58	342	400
	(32)	(368)	
Total	80	920	1000

Perform a χ^2 test **(8 points)**

	Got hepatitis	Did not get hepatitis	
Vaccine	14.083	1.225	
Placebo	21.125	1.837	
			$\chi^2 = 38.27$ $d.f. = 1$

The cutoff for χ^2 test with degree of freedom 1 is $\chi^2_{.05} = 3.84$.

As $38.27 > 3.84$, we reject H_0. **(10 points)**

5. The following table shows the final grades for STAT 100 summer course by sampling 200 students.

	A	B	C or lower	Total
Summer I	20	28	32	80
Summer II	42	38	40	120
Total	62	66	72	200

(a) (10 points) For the following model: 25% A, 35% B, 40% C or lower, test the goodness of fit with **Summer II** data and $\alpha = .05$. Does the data contradict the model?

Solution: Use a χ^2 test. **(6 points)**

	A	B	C or lower	Total
Observed O	42	38	40	120
Prob. under H_0	.25	.35	.4	1
Expeced freq. E	30	42	48	120
$(O-E)^2/E$	4.8	.381	1.333	$\chi^2 = 6.514$ $d.f. = 2$

The cutoff for χ^2 test with degree of freedom 2 is $\chi^2_{.05} = 5.99$. **(8 points)**

As $6.514 > 5.99$, we reject the null hypothesis. Hence, the data DOES contradict the model. **(10 points)**

(b) (12 points) Do the data provide strong evidence that the point distribution of Summer I and Summer II are different? Test with $\alpha = .05$.

Solution: Compute the expected frequency **(4 points)**

	A	B	C or lower	Total
Summer I	20 (24.8)	28 (26.4)	32 (28.8)	80
Summer II	42 (37.2)	38 (39.6)	40 (43.2)	120
Total	62	66	72	200

Perform a χ^2 test **(8 points)**

	Support	Indifferent	Opposed	Total
Union	.929	.097	.356	
Nonunion	.619	.065	.237	
				$\chi^2 = 2.303$ $d.f. = 2$

The cutoff for χ^2 test with degree of freedom 2 is $\chi^2_{.05} = 5.99$. **(10 points)**
As $2.303 < 5.99$, we retain H_0. Therefore, the data does NOT support that the point distributions are different. **(12 points)**