AP Stats

Review for the AP Exam - Chapters 12 & 13

Name: Key

Hour: ______

(2012 #4)

1. A survey organization conducted telephone interviews in December 2008 in which 1,009 randomly selected adults in the United States responded to the following question.

At the present time, do you think television commercials are an effective way to promote a new product?

Of the 1,009 adults surveyed, 676 responded "yes." In December 2007, 622 of 1,020 randomly selected adults in the United States had responded "yes" to the same question. Do the data provide convincing evidence that the proportion of adults in the United States who would respond "yes" to the question changed from December 2007 to December 2008?

① $H_0: p_{08} = p_{07}$ $H_a: p_{08} \neq p_{07}$

② Two-sample z-test for the differences in the proportions of all US adults who would answer "yes" to the question.

③ SRS is stated. Samples are taken in different years so independence can be assumed.

$$\hat{p}_c = \frac{676+622}{1009+1020} = .640$$

$n_{08}\hat{p}_c = 1009(.640) = 645.76$

$n_{08}(1-\hat{p}_c) = 1009(.360) = 363.24$

$n_{07}\hat{p}_c = 1020(.640) = 652.8$

$n_{07}(1-\hat{p}_c) = 1020(.360) = 326.4$

All are ≥ 5, we may continue with normal approximation.

④ $$z = \frac{\left(\frac{676}{1009} - \frac{622}{1020}\right) - 0}{\sqrt{\frac{.64(.36)}{1009} + \frac{(.64)(.36)}{1020}}} \approx 2.823$$

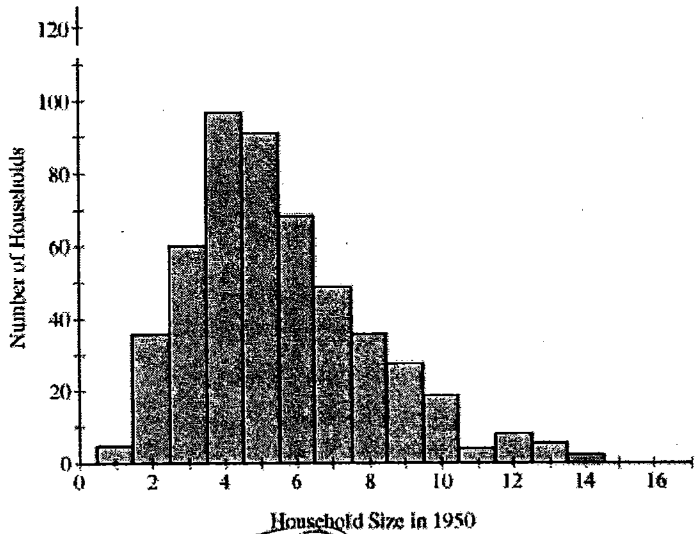
p-value $= 2P(z > 2.823)$

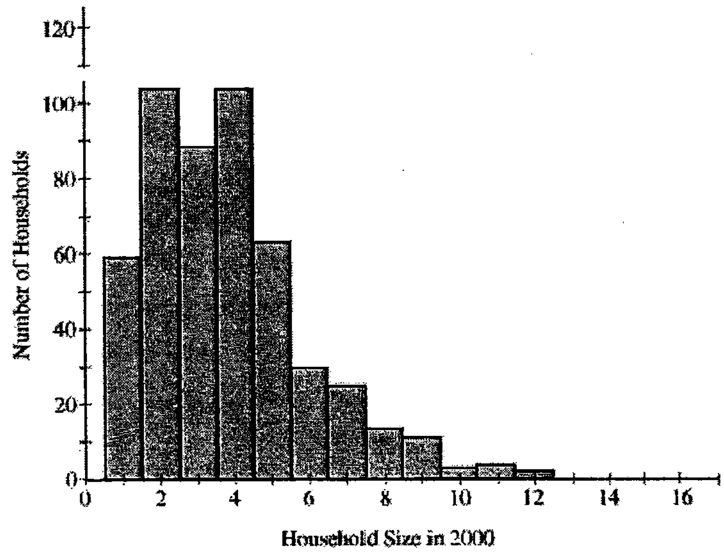
$= .005$

⑤ Since $p = .005 < .05$, we reject H_0. There is convincing evidence that the proportion of US adults who would respond "yes" to the question changed from 2007 to 2008.

(2012 #3)

2. Independent random samples of 500 households were taken from a large metropolitan area in the United States for the years 1950 and 2000. Histograms of household size (number of people in a household) for the years are shown below.





CUSS

(a) Compare the distributions of household size in the metropolitan area for the years 1950 and 2000.

The median for 1950 is 5 people and the median for 2000 was 3 or 4, so the average household size was larger in 1950 than 2000. Neither distribution has unusual features. The range in 1950 of 13 was also larger than the range of 11 people in 2000. Both household distributions are skewed right.

(b) A researcher wants to use these data to construct a confidence interval to estimate the change in mean household size in the metropolitan area from the year 1950 to the year 2000. State the conditions for using a two-sample *t*-procedure, and explain whether the conditions for inference are met.

① The data must come from two independent random samples. This condition is met since it was stated these were random samples from 1950 and 2000.

② The population sizes are 10x the sample sizes. $N_1 \geq 10(500) = 5000$ $N_2 = 10(500) = 5000$ since the number of households in a large metropolitan area would be easily ≥ 5000 in both 1950 and 2000, this condition is satisfied.

③ The sample sizes must both be large enough to use a normal approximation. Since n_1 and $n_2 = 500 \geq 30$, the central limit theory applies and this condition is satisfied.

(2009B #6)

3. Two treatments, A and B, showed promise for treating a potentially fatal disease. A randomized experiment was conducted to determine whether there is a significant difference in the survival rate between patients who receive treatment A and those who receive treatment B. Of 154 patients who received treatment A, 38 survived for at least 15 years, whereas 16 of the 164 patients who received treatment B survived at least 15 years.

(a) Treatment A can be administered only as a pill, and treatment B can be administered only as an injection. Can this randomized experiment be performed as a double-blind experiment? Why or why not?

In order to be double-blind neither the patients or physicians or researchers can know who is receiving the treatment and who is receiving the placebo. To do this a neutral party will know the treatments given while the physician or researcher in direct contact with the patient will give one group treatment A and a placebo injection, while the second group will receive treatment B and a placebo pill.

(b) The conditions for inference have been met. Construct and interpret a 95 percent confidence interval for the difference between the proportion of the population who would survive at least 15 years if given treatment A and the proportion of the population who would survive at least 15 years if given treatment B.

(1) Two-sample CI for the difference in the proportions of the population who would survive more than 15 years. (Difference = Treatment A - Treatment B)

(2) Conditions for inference is stated as met.

(3) $\left(\frac{38}{154} - \frac{16}{164}\right) \pm 1.96\sqrt{\frac{\left(\frac{38}{154}\right)\left(1-\frac{38}{154}\right)}{154} + \frac{\left(\frac{16}{164}\right)\left(1-\frac{16}{164}\right)}{164}}$

$= (.067, .231)$

(4) Based on these samples, I am 95% confident that the true difference in the proportion of the population who would survive at least 15 years if given Treatment A and the proportion of the population who would survive at least 15 years if given treatment B is between .067 and .231

In many of these types of studies, physicians are interested in the ratio of survival probabilities, $\frac{p_A}{p_B}$, where p_A represents the true 15-year survival rate for all patients who receive treatment A and p_B represents the true 15-year survival rate for all patients who receive treatment B. This ratio is usually referred to as the relative risk of the two treatments.

For example, a relative risk of 1 indicates the survival rates for patients receiving the two treatments are equal, whereas a relative risk of 1.5 indicates that the survival rate for patients receiving treatment A is 50 percent higher than the survival rate for patients receiving treatment B. An estimator of the relative risk is the ratio of estimated probabilities, $\frac{\hat{p}_A}{\hat{p}_B}$.

(c) Using the data from the randomized experiment described above, compute the estimate of the relative risk.

$\frac{38}{154} \Big/ \frac{16}{164} = \boxed{2.529}$

The sampling distribution of $\frac{\hat{p}_A}{\hat{p}_B}$ is skewed. However, when both sample sizes n_A and n_B are relatively large, the distribution of $\ln\left(\frac{\hat{p}_A}{\hat{p}_B}\right)$ — the natural logarithm of relative risk — is approximately normal with a mean of $\ln\left(\frac{p_A}{p_B}\right)$ and a standard deviation of $\sqrt{\frac{1-p_A}{n_A p_A}+\frac{1-p_B}{n_B p_B}}$, where p_A and p_B can be estimated by using $\hat{p}_A$ and $\hat{p}_B$.

When a 95 percent confidence interval for $\ln\left(\frac{p_A}{p_B}\right)$ is known, an approximate 95 percent confidence interval for $\frac{p_A}{p_B}$ — the relative risk of the two treatments — can be constructed by applying the inverse of the natural logarithm to the endpoints of the confidence interval $\ln\left(\frac{p_A}{p_B}\right)$.

(d) The conditions for inference are met for the data in the experiment above, and a 95 percent confidence interval for $\ln\left(\frac{p_A}{p_B}\right)$ is (0.3868, 1.4690). Construct and interpret a 95 percent confidence interval for the relative risk, $\frac{p_A}{p_B}$, of the two treatments.

$e^{.368} = 1.472$ $e^{1.4690} = 4.345$ $(1.475, 4.345)$

Based on these samples, we are 95% confident that the relative risk, $\frac{p_A}{p_B}$, of the two treatments is between 1.472 and 4.345

(e) What is an advantage of using the interval in part (d) over using the interval in part (b)?

Instead of focusing on only the difference in survival rates in part b, we can see how much better treatment A is. The chances of survival being 1.472 to 4.345 times better with treatment A is powerful.

(2004 #6)

4. A pharmaceutical company has developed a new drug to reduce cholesterol. A regulatory agency will recommend the new drug for use if there is convincing evidence that the mean reduction in cholesterol level after one month of use is more than 20 milligrams/deciliter (mg/dl), because a mean reduction of this magnitude would be greater than the mean reduction for the current most widely used drug.

The pharmaceutical company collected data by giving the new drug to a random sample of 50 people from the population of people with high cholesterol. The reduction in cholesterol level after one month of use was recorded for each individual in the sample, resulting in a sample mean reduction and standard deviation of 24 mg/dl and 15 mg/dl, respectively.

(a) The regulatory agency decides to use an interval estimate for the population mean reduction in cholesterol level for the new drug. Provide a 95 percent confidence interval. Be sure to interpret this interval.

① One-sample t CI for μ = the mean reduction in cholesteral level for all people with high cholesterol who take the new drug.

② SRS is stated
$N \geq 10n = 10(50) = 500$ There are more than 500 people w/ high cholesterol, so we can assume independence. $n = 50 > 30$ so the CLT says we may procede with normal approximation

③ $24 \pm 2.010\left(\frac{15}{\sqrt{50}}\right) = (19.737, 28.263)$ $df = 50 - 1 = 49$

④ Based on this sample, we are 95% confident that the mean reduction in cholesterol level for all people with high cholesterol that take this drug is between 19.737 and 28.263 mg/dl.

(b) Because the 95 percent confidence interval includes 20, the regulatory agency is not convinced that the new drug is better than the current best-seller. The pharmaceutical company tested the following hypotheses.

$$H_0: \mu = 20 \text{ versus } H_a: \mu > 20,$$

where μ represents the population mean reduction in cholesterol level for the new drug.

The test procedure resulted in a t-value of 1.89 and a p-value of 0.033. Because the p-value was less than 0.05, the company believes that there is convincing evidence that the mean reduction in cholesterol level for the new drug is more than 20. Explain why the confidence interval and the hypothesis test led to different conclusions.

With a two-sided confidence interval, $H_a \neq 20$. In this case the p-value would be $2(.033) = .066$ and the agency would fail to reject and the two conclusions would not match.

(c) The company would like to determine a value L that would allow then to make the following statement.

We are 95 percent confident that the true mean reduction in cholesterol level is greater than L.

A statement of this form is called a one-sided confidence interval. The value of L can be found using the following formula.

$$L = \bar{x} - t^* \frac{s}{\sqrt{n}}$$

$\alpha = .05$ @ 95% CI

This has the same form as the lower endpoint of the confidence interval in part (a), but requires a different critical value, t^*. What value should be used for t^*?

$t^* = 1.677$ invT(.95, 49)

Recall that the sample mean reduction in cholesterol level and standard deviation are 24 mg/dl and 15 mg/dl, respectively. Compute the value of L.

$$L = 24 - 1.677\left(\frac{15}{\sqrt{50}}\right) = 20.443$$

(d) If the regulatory agency had used the one-sided confidence interval in part (c) rather than the interval constructed in part (a), would it have reached a different conclusion? Explain.

Yes. The interval is now > 20 using 95% one-sided t-test therefore the conclusion would have been there is convincing evidence that the mean reduction in cholesterol level is better for the new drug.

(2011 #4)

5. High cholesterol levels in people can be reduced by exercise, diet, and medication. Twenty middle-aged males with cholesterol readings between 220 and 240 milligrams per deciliter (mg/dL) of blood were randomly selected from the population of such male patients at a large local hospital. Ten of the 20 males were randomly assigned to group A, advised on appropriate exercise and diet, and also received a placebo. The other 10 males were assigned to group B, received the same advice on appropriate exercise and diet, but received a drug intended to reduce cholesterol instead of a placebo. After three months, posttreatment cholesterol readings were taken for all 20 males and compared to pretreatment cholesterol readings. The tables below give the reduction in cholesterol level (pretreatment reading minus posttreatment reading) for each male in the study.

Group A (placebo)

Reduction (in mg/dL)	2	19	8	4	12	8	17	7	24	1

Mean Reduction: 10.20 Standard Deviation of Reductions: 7.66

Group B (cholesterol drug)

Reduction (in mg/dL)	30	19	18	17	20	−4	23	10	9	22

Mean Reduction: 16.40 Standard Deviation of Reductions: 9.40

Do the data provide convincing evidence, at the $\alpha = 0.01$ level, that the cholesterol drug is effective in producing a reduction in mean cholesterol level beyond that produced by exercise and diet?

① $H_0: \mu_A - \mu_B = 0$ $H_a: \mu_A - \mu_B < 0$

② Two-Sample t-test for the true difference in mean cholesterol reductions for all such male patients at this hospital advised on appropriate exercise and diet. (Placebo(A) − Drug(B))

③ SRS is stated. $N_A > 10(10) = 100$ and $N_B = 10(10) = 100$. Since there are likely more than 200 men being advised on exercise and diet at the hospital, we assume independence. Boxplots are ~N [Show box plots]

④ $t = \dfrac{(10.20 - 16.40) - 0}{\sqrt{\left(\frac{7.66}{\sqrt{10}}\right)^2 + \left(\frac{9.40}{\sqrt{10}}\right)^2}} = -1.62$

df = 17.296

P = .062

⑤ Since .062 is $> \alpha = .01$ we fail to reject H_0. There is not convincing evidence that the cholesterol drug is effective in producing a reduction in cholesterol beyond that produced by exercise & diet.

(2013 #5)

6. Psychologists interested in the relationship between meditation and health conducted a study with a random sample of 28 men who live in a large retirement community. Of the men in the sample, 11 reported that they participate in daily meditation and 17 reported that they do not participate in daily meditation. The researchers wanted to perform a hypothesis test of

$$H_0 : p_m - p_c = 0$$
$$H_a : p_m - p_c < 0$$

where p_m is the proportion of men with high blood pressure among all the men in the retirement community who participate in daily meditation and p_c is the proportion of men with high blood pressure among all the men in the retirement community who do not participate in daily meditation.

(a) If the study were to provide significant evidence against H_0 in favor of H_a, would it be reasonable for the psychologists to conclude that daily meditation causes a reduction in blood pressure for men in the retirement community? Explain why or why not.

No. You cannot conclude daily meditation causes a reduction in blood pressure because this was an observational study, not an experiment.

$\boxed{\text{OR}}$ Treatments were not randomly assigned.

$\boxed{\text{OR}}$ Confounding variables could occur if men who meditate also differ in some other way affecting blood pressure.

The psychologists found that of the 11 men in the study who participate in daily meditation, 0 had high blood pressure. Of the 17 men who do not participate in daily meditation, 8 had high blood pressure.

(b) Let $\hat{p}_m$ represent the proportion of men with high blood pressure among those in a random sample of 11 who meditate daily, and let $\hat{p}_c$ represent the proportion of men with high blood pressure among those in a random sample of 17 who do not meditate daily. Why is it not reasonable to use a normal approximation for the sampling distribution of $\hat{p}_m - \hat{p}_c$?

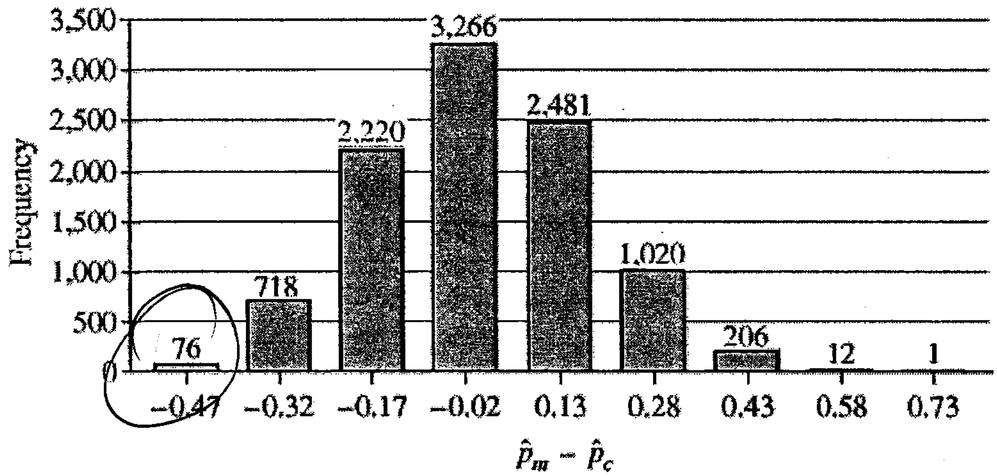
$$\hat{p}_c = \frac{0+8}{11+17} = \frac{8}{28} = 2.86$$

$$n_m \hat{p}_c = 11\left(\frac{8}{28}\right) = 3.143$$

$$n_c \hat{p}_c = 17\left(\frac{8}{28}\right) = 4.857$$

Since 3.143 and 4.857 are not greater than 5, we cannot use a normal approximation.

Although a normal approximation cannot be used, it is possible to simulate the distribution of $\hat{p}_m - \hat{p}_c$. Under the assumption that the null hypothesis is true, 10,000 values of $\hat{p}_m - \hat{p}_c$ were simulated. The histogram below shows the results of the simulation.



(c) Based on the results of the simulation, what can be concluded about the relationship between blood pressure and meditation among men in the retirement community?

$\hat{p}_m - \hat{p}_c = \frac{0}{11} - \frac{8}{17} = -.471$

Since the probability of obtaining a difference of $-.471$ or more is only $\frac{76}{10000} = .0076$, this is unlikely and we can conclude there is a relationship between blood pressure and meditation among men in this retirement community.