# 6.2 Young Americans, Part I.

About 77% of young adults think they can achieve the American dream. Determine if the following statements are true or false, and explain your reasoning.

1. The distribution of sample proportions of young Americans who think they can achieve the American dream in samples of size 20 is left skewed.

* **True**
* Success failure condition is not satisfied as
  + Success 20 \* 0.77 = 15.4
  + Failure 20 \* (1-0.77) = 4.6

We expect to see at least 10 successes and 10 failures in our sample for the distribution to be normal.

* For most sample we expect the p = 0.77 which is bound by 100% suggesting that it would take left skewed shape.

1. The distribution of sample proportions of young Americans who think they can achieve the American dream in random samples of size 40 is approximately normal since n >= 30.

* **False**
* In case of proportion we need to factor in the size of success and failure, both have to be > 10,
* n >= 30 does not apply in case of proportions.
* Here Success failure condition is satisfied as
  + 1. Success 40 \* 0.77 = 30.8
    2. Failure 40 \* (1-0.77) = 9.2

We expect to see at least 10 successes and 10 failures in our sample for the distribution to be normal. In this case failures are < 10, so distribution could not be normal

1. A random sample of 60 young Americans where 85% think they can achieve the American dream would be considered unusual.

**False**

SEp = =SQRT((0.77\*(1-0.77)/60)) = 0.054329243

Z = (0.85 – 0.77)/0.054329243 = 1.472503491

1.472503491 SEs away from the mean, which would not be considered unusual, if is greater than 2 SE away then it is considered unusual.

1. A random sample of 120 young Americans where 85% think they can achieve the American dream would be considered unusual.

**True**

SEp = SQRT((0.77\*(1-0.77)/120)) = 0.038416576

Z = (0.85 - 0.77)/ 0.038416576 = 2.08

2.0824344 SEs away from the mean, which would be considered unusual, if is greater than 2 SE away, then it is considered unusual.

# 6.8 Elderly drivers.

In January 2011, The Marist Poll published a report stating that 66% of adults nationally think licensed drivers should be required to retake their road test once they reach 65 years of age. It was also reported that interviews were conducted on 1,018 American adults, and that the margin of error was 3% using a 95% confidence level.

1. Verify the margin of error reported by The Marist Poll.

SEp = SQRT(0.66\*(1-0.66)/1018) = 0.01484696

For 95% confidence level margin of error ME = 1.96 \* 0.01484696 = 0.029100042 = 2.9100042% which is close to 3%

1. Based on a 95% confidence interval, does the poll provide convincing evidence that more than 70% of the population think that licensed drivers should be required to retake their road test once they turn 65?

Point estimate +- z\*SE

66 + (1.96 \* 1.484696) = 68.91000416

66 - (1.96 \* 1.484696) = 63.08999584

For the confidence level of 95% values ranges from **63.09 and 68.91%**, **70% does not fall in that bracket**. So the **poll does not provide** any convincing evidence that more than 70% of the population think that licensed drivers should be required to retake their road test once they turn 65.

# 6.18 Is college worth it? Part II.

Exercise 6.16 presents the results of a poll where 48% of 331 Americans who decide to not go to college do so because they cannot afford it.

1. Calculate a 90% confidence interval for the proportion of Americans who decide to not go to college because they cannot afford it, and interpret the interval in context.

SEp = SQRT(0.48\*(1 - 0.48)/331) = 0.027460491 = 2.7460491

Point estimate +- z\*SE

48 + (1.64\*2.7460491) = 52.50352052

48 - (1.64\*2.7460491) = 43.49647948

For 90% confidence interval the proportion of Americans who decide to not go to college because they cannot afford it lies between **43.49647948 % and 52.50352052 %**

1. Suppose we wanted the margin of error for the 90% confidence level to be about 1.5%. How large of a survey would you recommend?

ME = (90 confidence interval z score) \* SE

(1.645) \* SQRT(0.48\*(1 - 0.48)/n) = 0.015

SQRT(0.48\*(1 - 0.48)/n) = 0.015/1.645

SQRT(0.48\*(1 - 0.48)/n) = 0.009118541

(0.48\*(1 - 0.48)/n) = 0.009118541^2

n = 3002

**Survey size should be at least of 3002**.

# 6.26 The Daily Show.

A 2010 Pew Research foundation poll indicates that among 1,099 college graduates, 33% watch The Daily Show. Meanwhile, 22% of the 1,110 people with a high school degree but no college degree in the poll watch The Daily Show. A 95% confidence interval for (pcollege grad - pHS or less), where p is the proportion of those who watch The Daily Show, is (0.07, 0.15). Based on this information, determine if the following statements are true or false, and explain your reasoning if you identify the statement as false.

(a) At the 5% significance level, the data provide convincing evidence of a difference between the proportions of college graduates and those with a high school degree or less who watch The Daily Show.

**True**

Ho – No difference between the proportions of college graduates and those with a high school degree or less who watch The Daily Show

Ha– There is difference between the proportions of college graduates and those with a high school degree or less who watch The Daily Show

Point estimate = p1 - p2 = 0.33 -0.22 = 0.11

n1 = 1099, n2 = 1100



SEp1 –p2 = 0.018899283

Z = point estimate - null value / SE

Z = (0.11 – 0)/ 0.018899283 = 5.82

For Z ≥ 3.50, the probability is greater than or equal to 0.9998

Using the normal model for this test statistic, we identify the right tail area as 0.0002 which is p-value and it is less than 0.05 so we reject the null hypothesis and conclude that data provide convincing evidence of a difference between the proportions of college graduates and those with a high school degree or less who watch The Daily Show.

(b) We are 95% confident that 7% less to 15% more college graduates watch The Daily Show than those with a high school degree or less.

**False**

* We are 95% confidence that difference between the proportions of college graduates and those with a high school degree or less who watch The Daily Show is between 7% and 15%

(c) 95% of random samples of 1,099 college graduates and 1,110 people with a high school degree or less will yield differences in sample proportions between 7% and 15%.

**True 🡺**

* 95% confidence interval indicates that there is 95% probability for the proportion difference to fall in that range

(d) A 90% confidence interval for (pcollege grad - pHS or less) would be wider.

**False**

Point estimate = p1 - p2 = 0.33 -0.22 = 0.11

n1 = 1099, n2 = 1100



SEp1 –p2 = 0.018899283

Z =1.65

Point estimate +- z\*SE

* 0.11 + (1.65\*0.018899283) =14%
* 0.11 - (1.65\*0.018899283) = 7%

(e) A 95% confidence interval for (pHS or less - pcollege grad) is (-0.15,-0.07).

**True**

Point estimate = p1 - p2 = 0.22 -0.33 = -0.11

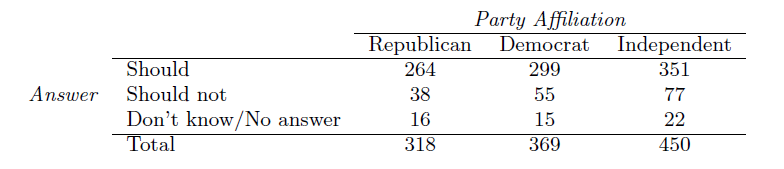
SEp1 –p2 = 0.018899283

Point estimate +- z\*SE

* -0.11 + (1.65\*0.018899283) = -7%
* -0.11 - (1.65\*0.018899283) = -15%

# 6.32 Full body scan, Part I.

A news article reports that Americans have differing views on two potentially inconvenient and invasive practices that airports could implement to uncover potential terrorist attacks." This news piece was based on a survey conducted among a random sample of 1,137 adults nationwide, interviewed by telephone November 7-10, 2010, where one of the questions on the survey was \Some airports are now using `full-body' digital x-ray machines to electronically screen passengers in airport security lines. Do you think these new x-ray machines should or should not be used at airports?" Below is a summary of responses based on party affiliation.



(a) Conduct an appropriate hypothesis test evaluating whether there is a difference in the proportion of Republicans and Democrats who think the full-body scans should be applied in airports. Assume that all relevant conditions are met.

* Ho: There is no difference in the proportion of Republicans and Democrats who think the full-body scans should be applied in airports.
* Ha: There is difference in the proportion of Republicans and Democrats who think the full-body scans should be applied in airports.

p1 = 264/318 = 0.830188679

p2 = 299/369 = 0.810298103

p1 - p2 = 299/369 = 0.019891



SE = SQRT(((0.830188679\*(1-0.830188679))/318) + ((0.810298103 \* (1-0.810298103))/369))

SE = 0.029324

Z = (Point estimate – NULL VALUE)/SE)

Z = 0.678318101

P value for this is = (1 - 0.7517) = 0.2483

Since p > 0.05 the, we fail to reject NULL hypothesis and conclude that there is no difference in the proportion of Republicans and Democrats who think the full-body scans should be applied in airports. In other words there is no evidence there is an association between party and body scan opinion.

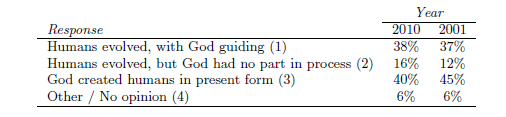
(b) The conclusion of the test in part (a) may be incorrect, meaning a testing error was made. If an error was made, was it a Type 1 or a Type 2 Error? Explain.

**Type 2 error**

* Rejecting Alternative hypothesis even though alternative is true. We may have incorrectly concluded there is no difference in the proportion of Republicans and Democrats who think the full-body scans should be applied in airports.
* In other words Type 2 error 🡺 Failing to reject NULL hypothesis (Ho) even though alternative hypothesis (Ha) is true.

# 6.42 Evolution vs. creationism.

A Gallup Poll released in December 2010 asked 1019 adults living in the Continental U.S. about their belief in the origin of humans. These results, along with results from a more comprehensive poll from 2001 (that we will assume to be exactly accurate), are summarized in the table below:



(a) Calculate the actual number of respondents in 2010 that fall in each response category.

* Humans evolved, with God guiding = 1019 \* 0.38 = **387**
* Humans evolved, but God had no part in process = 1019 \* 0.16 = **163**
* God created humans in present form = 1019 \* 0.40 = **408**
* Other / No opinion = 1019 \* 0.06 = **61**

(b) State hypotheses for the following research question: have beliefs on the origin of human life changed since 2001?

* Ho
  + Proportions of 4 response type is same in 2001 and 2010
  + p1 = 0.37, p2 = 0.12, p3 = 0.45, p4 = 0.6
* Ha
  + In 2010 at least one proportion is different from that of 2001.

(c) Calculate the expected number of respondents in each category under the condition that the null hypothesis from part (b) is true.

* Humans evolved, with God guiding = 1019 \* 0.37 = **377**
* Humans evolved, but God had no part in process = 1019 \* 0.12 = **122**
* God created humans in present form = 1019 \* 0.45 = **459**
* Other / No opinion = 1019 \* 0.06 = **61**

(d) Conduct a chi-square test and state your conclusion. (Reminder: Verify conditions.)

Condition Verification

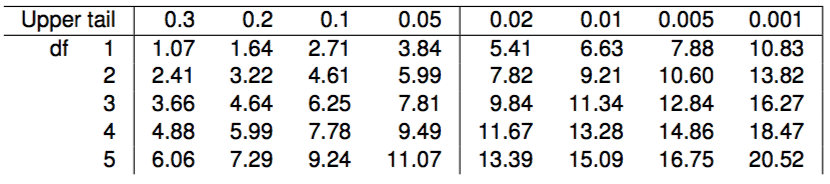
* Independence – We can safely assume that all participant are independent of each other. There is no reason for one person’s choice to influence the choice of other person
* Sample Size – Each value is greater than 5 (n > 5) for each of the choices
* df > 1 – Degrees of freedom must be greater than 1.
  + df = (rows \* cols) – 1 = (4 \* 1) – 1 = 3

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Actual Count In 2010** | **Expected Count in 2010 (Same as that of 2001)** | **X^2df** |
| Humans evolved, with God guiding | 387 | 377 | 0.265251989 |
| Humans evolved, but God had no part in process | 163 | 122 | 13.77868852 |
| God created humans in present form | 408 | 459 | 5.666666667 |
| Other / No opinion | 61 | 61 | 0 |
| **TOTAL** | **1019** | **1019** | **19.71060718** |

X2df = 19.71

df = (rows \* cols) – 1 = (4 \* 1) – 1 = 3

So based on the table X2df = 19.71 and df = 3, p-value is less than 0.001

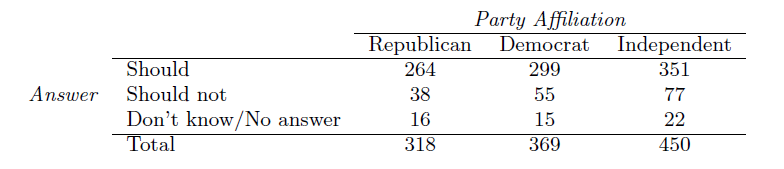


Since p-value is less than 0.001 which is less than significance level of 0.05 we can safely reject NULL hypothesis.

People beliefs on the origin of human life has changed in 2010 compared that one in the 2001

# 6.46 Full body scan, Part II.

The table below summarizes a data set we first encountered in Exercise 6.32 regarding views on full-body scans and political affiliation. The differences in each political group may be due to chance. Complete the following computations under the null hypothesis of independence between an individual's party affiliation and his support of full-body scans. It may be useful to first add on an extra column for row totals before proceeding with the computations.



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Party Affiliation | | | | | |  |
|  | **Republican** | **Expected Republican** | **Democrat** | **Expected Democrats** | **Independent** | **Expected Independent** | **TOTAL** |
| Should | 264 | 256 | 299 | 297 | 351 | 362 | **914** |
| Should not | 38 | 48 | 55 | 55 | 77 | 67 | **170** |
| Don't know/No answer | 16 | 15 | 15 | 17 | 22 | 21 | **53** |
| **Total** | **318** | **318** | **369** | **369** | **450** | **450** | **1137** |

(a) How many Republicans would you expect to not support the use of full-body scans?

**48**

(b) How many Democrats would you expect to support the use of full-body scans?

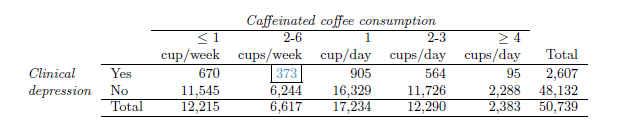
**297**

(c) How many Independents would you expect to not know or not answer?

**21**

# 6.48 Coffee and Depression.

Researchers conducted a study investigating the relationship between caffeinated coffee consumption and risk of depression in women. They collected data on 50,739 women free of depression symptoms at the start of the study in the year 1996, and these women were followed through 2006. The researchers used questionnaires to collect data on caffeinated coffee consumption, asked each individual about physician-diagnosed depression, and also asked about the use of antidepressants. The table below shows the distribution of incidences of depression by amount of caffeinated coffee consumption.



(a) What type of test is appropriate for evaluating if there is an association between coffee intake and depression?

* Chi-Square with two ways tables

(b) Write the hypotheses for the test you identified in part (a).

The hypotheses are:

* *H0*: Coffee intake and depression are independent. There is no relationship between coffee intake and depression.
* *HA*: Coffee intake and depression are dependent. There is a relationship between coffee intake and depression.

(c) Calculate the overall proportion of women who do and do not suffer from depression.

* Overall proportion of women who suffer from depression = 2607/50739 = 0.051380595 = **5.14%**
* Overall proportion of women who do not suffer from depression = 48132/50739 = 0.948619405 = **94.86%**

(d) Identify the expected count for the highlighted cell, and calculate the contribution of this cell to the test statistic, i.e. (Observed - Expected) ^2/Expected.

* Expected Count
  + (2607\*6617)/50739 = 340
* (Observed - Expected) ^2/Expected
  + (373-340) ^2/340 = 3.2029

(e) The test statistic is X^2 = 20:93. What is the p-value?

* df = (rows-1)\* (cols-1) = (2-1)\*(5-1) = 4
* p-value is less than 0.001 (p-value = 0.0003269507)

(f) What is the conclusion of the hypothesis test?

* Since the p-value (0.0003269507) is less than significance level of 0.05, we can **safely reject Null hypothesis**. There is a **relationship between coffee intake and depression.**

(g) One of the authors of this study was quoted on the NY Times as saying it was too early to recommend that women load up on extra coffee" based on just this study. Do you agree with this statement? Explain your reasoning.

* Tend to agree with statement, as the hypothesis test does prove out the relationship between coffee and depression.
* Look at the data, there is considerable amount of women who do not have clinical depression in spite of consuming coffee. There is seems to very weak association between coffee consumption and clinical depression