

STAT 35000 Exam Jam Solutions

Contents

1	First Day	2
	Question 1: PDFs, CDFs, and Finding $E(X)$, $V(X)$	2
	Question 2: Bayesian Inference	3
	Question 3: Binomial to Normal Approximation	4
	Question 4: Sampling Distributions	5
	Question 5: Confidence Intervals	6
	Question 6: One Sample Hypothesis Testing, Power	7
	Question 7: Two Sample Hypothesis Testing	9
	Question 8: Paired Sample Hypothesis Testing	10
	Question 9: One Sample Proportion Hypothesis Testing, Power	11
	Question 10: Simple Linear Regression	13
2	Day 2	14
	Question 1: Unions, Intersections, and Bayes Theorem	14
	Question 2: Normal Distribution	15
	Question 3: Binomial to Normal Approximation	16
	Question 4: Confidence Intervals	17
	Question 5: One Sample Hypothesis Testing	18
	Question 6: One Sample Hypothesis Testing, Power	19
	Question 7: Two Sample Proportion Hypothesis Testing	21
	Question 8: Simple Linear Regression	22
	Question 9: Anova	23

1 First Day

1. A continuous random variable has a pdf given below

$$f(x) = \begin{cases} ke^{-x/100}, & 0 < x < \infty \\ 0, & \text{elsewhere} \end{cases}$$

- Find value k such that $f(x)$ is a valid p.d.f.
- Write out the c.d.f., or $F(x)$, of X
- Find $P(0 < X < 80)$.
- Find $E(X)$ and $V(X)$.
- Suppose we got a new variable $Y = -2X + 50$. Find $E(Y)$ and $V(Y)$.

Solution

(a) $\int_0^{\infty} ke^{-x/100} dx \rightarrow k = \frac{1}{100}$

(b)

$$F(x) = \begin{cases} 1 - e^{-x/100}, & 0 \leq x < \infty \\ 0, & x < 0 \end{cases}$$

(c) $F(80) = 1 - e^{-80/100} = 0.55067$

(d) Notice that this is an exponential distribution. $E(X) = 100$ and $V(X) = 100^2$. If you did not catch that, you would have to solve $E(X) = \int_0^{\infty} x \frac{1}{100} e^{-x/100} dx$ and then $E(X^2) = \int_0^{\infty} x^2 \frac{1}{100} e^{-x/100} dx$. From there, we could find $V(X) = E(X^2) - (E(X))^2$

(e) Remember the rule

$$E(aX + b) = aE(X) + b$$

$$V(aX + b) = a^2V(X)$$

So,

$$E(Y) = E(-2X + 50) = (-2)E(X) + 50 = (-2)(100) + 50 = -150$$

$$V(Y) = V(-2X + 50) = (-2)^2V(X) = (4)(100^2) = 40000$$

2. Suppose that there are two drawers that contain different number of marbles. The first drawer contains 5 white marbles and 5 blue marbles. The second drawer contains 9 white marbles and 3 blue marbles. Mike, randomly opens one of the drawers, but he is twice as likely to open the second drawer opposed to the first drawer. Then, without looking, he randomly picks up one marble. Mike does not treat either drawer any different nor the marble.

- (a) Create a tree diagram with the corresponding probabilities (The tree should have two layers).
- (b) Suppose that we had a following hypothesis,

$$H_0 : \text{Marble was picked from the 1st drawer}$$

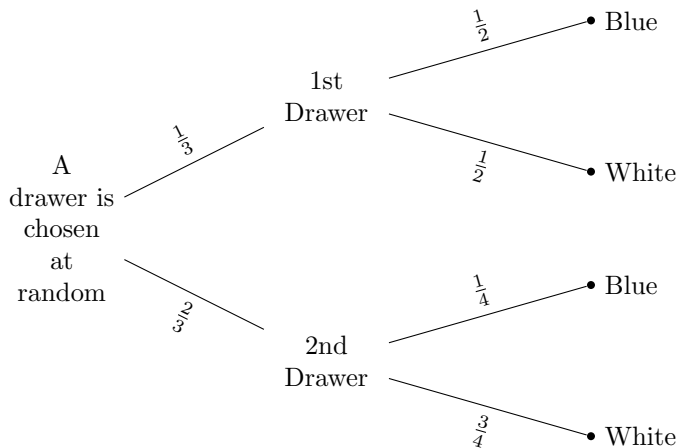
$$H_a : \text{Marble was picked from the 2nd drawer}$$

What is the probability for each hypothesis to happen?

- (c) Given that Mike opened the 2nd drawer, what is the probability that the marble is white?
- (d) Now, lets say that the marble happens to be blue. What is the updated belief for the two hypothesis? In other words, what is the probability for each hypothesis to happen now? (Use Bayes rule)

Solution

(a) Solution is displayed below,



(b) There is $\frac{1}{3}$ chance for Null hypothesis and $\frac{2}{3}$ chance for Alternative hypothesis to happen.

(c) $P(W|2^{nd}) = P\left(\frac{W \cap 2^{nd}}{2^{nd}}\right) = \frac{(\frac{2}{3})(\frac{3}{4})}{(\frac{2}{3})} = 0.75$

(d) $P(H_0|B) = \frac{(\frac{1}{3})(\frac{1}{2})}{(\frac{1}{3})(\frac{1}{2}) + (\frac{2}{3})(\frac{1}{4})} = 0.5$

$P(H_1|B) = \frac{(\frac{2}{3})(\frac{1}{4})}{(\frac{1}{3})(\frac{1}{2}) + (\frac{2}{3})(\frac{1}{4})} = 0.5$

3. Chris loves eating snickers candy. He rolls a fair six sided die and if the die lands on 5, he eats the candy. Let X denote the amount of snicker bars that Chris eats from a bag of 50.
- Name the exact distribution of X and its parameters.
 - Find $P(X \geq 4)$.
 - Name the approximate distribution for X and its parameters. Explain why approximation is valid.
 - Find the approximate probability that Chris eats 10 or less snickers from the bag of 50, $P(X \leq 10)$.

Solution

- Let X denote the number of candy that Chris eats from the bag of 50. Then $X \sim \text{Binomial}(p = \frac{1}{6}, n = 50)$
- $P(X \geq 4) = 1 - P(X \leq 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) = .976$
Using a calculator: $1 - \text{binomcdf}(50, \frac{1}{6}, 3)$.
- X can be approximated with a normal distribution because n is large ($np > 15$ and $n(1-p) > 15$).
 $X \sim \text{Normal}(np = 8.333, npq = 6.944)$
- Because we are using Binomial to Normal Approximation, we will use 0.5 continuity correction, thus our answer will be $P(X \leq 9.5) = P(Z \leq \frac{9.5 - 8.333}{\sqrt{6.944}}) = P(Z \leq .4427) = 0.671$
Using a calculator: $\text{normalcdf}(-9^{10}, 9.5, 8.333, \sqrt{(6.944)})$

4. Suppose X has the following distribution

x	0	1	3
$p(x)$.1	.5	.4

Suppose we take a random sample of $n=2$.

- Calculate $P(\bar{X}_2 = 2)$.
- Find the distribution of \bar{X}_2
- Find $E[\bar{X}]$ and $V[\bar{X}]$ for $n = 2$.

Solution

- It is useful to create a table first for all of the possible averages.

\bar{x}	0	1	3
0	0	.5	1.5
1	.5	1	2
3	1.5	2	3

$$\text{So, } P(\bar{X} = 2) = (.5)(.4) + (.5)(.4) = 0.2 + 0.2 = 0.4$$

- Distribution function is shown below

\bar{x}	0	.5	1	1.5	2	3
$p(\bar{x})$.01	.1	.25	.08	.4	.16

- $$E[\bar{X}] = E[X] = 1.7$$

$$V[\bar{X}] = \frac{V[X]}{n} = \frac{E[X^2] - (E[X])^2}{n} = \frac{1.21}{2} = 0.605$$

5. In certain chemical processes it is very important that a particular solution have a pH of exactly 8.20. A method for determining the pH for solutions of this type is known to give measurements that are normally distributed with a mean equal to the actual pH and with a standard deviation of 0.1. Suppose 6 independent measurements yielded the following pH values.

8.10	8.45	8.60
8.25	8.30	8.25

- Calculate the MOE.
- Construct a 95% Confidence Interval for the average pH level.
- What does this confidence interval represent?
- What conclusion can be made at the $\alpha = 0.05$ level of significance?

Solution

- First, we can get $Z_{\alpha/2}$ using a calculator: $invNorm(1 - \alpha/2, 0, 1) = 1.96$
 $MOE = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = (1.96) \frac{.1}{\sqrt{6}} = .08$
- $(L, U) = \bar{X} \pm MOE = 8.325 \pm .08 = (8.245, 8.405)$
- With 95% confidence, we expect the true proportion of pH level to fall into the given interval.
- Since pH of 8.20 given above does not fall in the interval calculated in part 2, at 95% confidence level we can reject the null that the pH level is 8.20.

6. It is advertised that certain machinery on average lasts 12.7 years. Paul wants to test and see whether the average life of the machinery is actually less than 12.7 years. He tests 20 machines and finds that the sample mean was 11.8. Assume the lifetime of the machinery is normally distributed with standard deviation of 1.5.
- State the null and alternative hypothesis
 - Name the parameters and describe what they represent.
 - What method would you use?
 - 1-sample T
 - 1-sample Z
 - 2-sample T
 - 2-sample Z
 - Calculate the p-value.
 - Assuming $\alpha = .01$, state the conclusion and interpret the results in the context of the problem.
 - Suppose that the true average time a machine lasts is actually 12. For the test above, what is the power to successfully detect the alternative?
 - If we want the above power to be 95%, at least how many machines you need to have in your sample?

Solution

- $H_0 : \mu = 12.7$
 $H_a : \mu < 12.7$
- $\mu =$ average life a certain machinery will last
- 1-sample Z because we are told it is normally distributed with $\sigma = 1.5$
- First, we need to calculate t-stat. $z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{11.8 - 12.7}{1.5/\sqrt{20}} = -2.68$. We are looking for
 $P(z < -2.68) = .0037$
 Using Calculator: $normalcdf(-9^{10}, 11, 8, 12.7, \frac{1.5}{\sqrt{20}})$
- Because $p - val < \alpha$, we have sufficient evidence to reject the null hypothesis and state that on average, the machinery lasts less than 12.7 years.
-

$$\begin{aligned}
 \text{Power} &= 1 - P(\text{Type II Error}) \\
 &= 1 - P(\text{Don't Reject } H_0 | \mu = 12) \\
 &= 1 - P\left(\frac{x - 12.7}{1.5/\sqrt{20}} > -2.33 | \mu = 12\right) \\
 &= 1 - P(x > 11.92 | \mu = 12) \\
 &= P(x \leq 11.92 | \mu = 12) \\
 &= \text{Using Calculator: } normalcdf(-9^{10}, 11.92, 12, 1.5/\sqrt{20}) \\
 \text{Power} &= 0.4057
 \end{aligned}$$

(g)

$$0.95 = 1 - P(\text{Type II Error})$$

$$0.05 = P(\text{Don't Reject } H_0 | \mu = 12)$$

$$0.05 = P\left(\frac{x - 12.7}{1.5/\sqrt{n}} > -2.33 | \mu = 12\right)$$

$$0.05 = P(x > (-2.33)(1.5)/(\sqrt{n}) + 12.7 | \mu = 12)$$

$$0.05 = P\left(z > \frac{(-2.33)(1.5)/(\sqrt{n}) + 12.7 - 12}{1.5/\sqrt{n}}\right)$$

$$1.645 = -2.33 + (.7)(\sqrt{n})/1.5$$

$$n = 72.55$$

$$n = 73(\text{Remember to always round up for sample size})$$

7. Mrs. McBroom teaches two different independent classes. Class A has 25 students and Class B has 16 students. Both classes took the same standardize exam. Class A average was 80 with standard deviation of 10 and Class B average was 85 with standard deviation of 7. Mrs. McBroom wants to know if there is significant difference between the scores in two of her classes (use $\alpha = .05$).
- State the null and alternative hypothesis
 - What method would you use?
 - 1-sample T
 - 1-sample Z
 - 2-sample T
 - 2-sample Z
 - Find the test statistic and
 - Find the rejection region.
 - Using the rejection region, state the conclusion and interpret the results in the context of the problem.

Solution

(a) $H_0 : \mu_A = \mu_B$
 $H_a : \mu_A \neq \mu_B$

- (b) 2-sample T test because $n_1, n_2 < 30$ and σ is not given.

(c)
$$t = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{S_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{(80 - 85) - 0}{8.966 \sqrt{\frac{1}{25} + \frac{1}{16}}} = -1.742$$

Where,

$$S_p = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}} = \sqrt{\frac{(10^2)(25 - 1) + (7^2)(16 - 1)}{25 + 16 - 2}} = 8.966$$

Using Calculator: $2 - SampTTest(\bar{x}_1 = 80, s_1 = 10, n_1 = 25, \bar{x}_2 = 85, s_2 = 7, n_2 = 16, \mu_1 \neq \mu_2, Pooled = Yes) \rightarrow -1.742$

- (d) Rejection region is $|t| > t_{\alpha/2, df}$ or $|t| > 2.021$ (If you use a table, it will not be an exact t value)
 Using Calculator: $invT(1 - 0.025, 39) = 2.023$
- (e) Since $|-1.742| < 2.021$ we fail to reject the null hypothesis. We conclude that there is not enough evidence at $\alpha = 0.05$ to say that the average score in the two classes is different.

8. A group of students wanted to see whether their new filter invention could filtrate Zinc out of the drinking water. Five pairs of data were taken recording the zinc concentration in water before and after the filtration. The data is shown below.

	Zinc %					Mean	S.D.
Before	10	12	9.5	16	13.5	12.2	2.66
After	5	8	6	8.5	9.5	7.4	1.85
After-Before	-5	-4	-3.5	-7.5	-4	-4.8	1.6

Is there any difference in Zinc Concentration Before and After the filtration?

- State the Null and Alternative Hypothesis.
- What do the parameter(s) represent?
- What method would you use?
 - 1-sample T
 - 1-sample Z
 - 2-sample T
 - 2-sample Z
- Calculate test statistic.
- State your conclusion and interpret the results in the context for the researcher using $\alpha = 0.01$

Solution

- $H_0 : \mu_d = 0$
 $H_a : \mu_d \neq 0$
- μ_d = the average difference before and after the filtration
- 1-Sample T test because it is a paired test with $n < 30$ and σ unknown.
- $$t = \frac{\bar{x} - \mu_d}{s/\sqrt{n}} = \frac{-4.8 - 0}{1.6/\sqrt{5}} = -6.708$$
 Using Calculator: $T - Test(\mu_0 = 0, \bar{x} = -4.8, s = 1.6, n = 5, \mu \neq \mu_0) \rightarrow -6.708$
- We found that $|t| = 6.708$ and the critical value with $\alpha = 0.01$ is $t_{\alpha/2} = 4.604$ (Using Calculator: $invT(1 - .005, 4) = 4.604$). Since $6.708 > 4.604$, we can reject the null hypothesis and state that the filter does work and it eliminates zinc from water at 99% confidence.

9. The CEO of a large company says that about 60 percent of its 10000 customers are satisfied with their services. To test this claim, a local news reporter tests 100 customers. From that sample, 64 said they were satisfied with the service. Assume $\alpha = 0.02$.
- State the Null and Alternative Hypothesis
 - Calculate test statistic.
 - State the conclusion of your test.
 - Suppose that we later found out that the true percent satisfaction rate was actually 55%. Find the power of the test.
 - If we wanted to have a power of 95%, what is the minimal sample size we would need?

Solution

(a)

$$H_0 : p_0 = 0.60$$

$$H_a : p_0 \neq 0.60$$

(b) Our sample proportion is $\hat{p} = \frac{64}{100}$. Then we can calculate our test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{.64 - .60}{\sqrt{(.6)(.4)/100}} = 0.82$$

Using Calculator: $1 - PropZTest(p_0 = 0.6, x = 64, n = 100, prop \neq p_0) \rightarrow 0.816$

(c) For a two sided test, assuming $\alpha = 0.02$, our critical value is $z_{cv} = 2.33$ (Using Calculator: $invNorm(1 - .01, 0, 1) = 2.33$). Because our test statistic (0.82) is less than our critical value (2.33), we have insufficient evidence to reject the null hypothesis and state that the true satisfaction rate is no different from 0.60.

(d)

$$\begin{aligned} Power &= 1 - P(\text{Type II Error}) \\ &= 1 - P(\text{Don't Reject } H_0 | p = .55) \\ &= 1 - P(-2.33 < \frac{\hat{p} - p_0}{\sqrt{(p_0)(1 - p_0)/n}} < 2.33 | p = 0.55) \\ &= 1 - P(-2.33 < \frac{\hat{p} - 0.60}{\sqrt{(0.60)(0.40)/100}} < 2.33 | p = 0.55) \\ &= 1 - P((-2.33)\sqrt{(0.60)(0.40)/100} + 0.60 < \hat{p} < (2.33)\sqrt{(0.60)(0.40)/100} + 0.60 | p = 0.55) \\ &= 1 - P(0.486 < \hat{p} < 0.714 | p = 0.55) \\ &= \text{Using Calculator: } normalcdf(0.486, 0.714, 0.55, \sqrt{\frac{(0.55)(0.45)}{100}}) = 0.90 \\ &= 1 - 0.90 \\ Power &= 0.10 \end{aligned}$$

(e)

$$\text{Power} = 1 - P(\text{ Type II Error })$$

$$0.95 = 1 - P(\text{ Don't Reject } H_0 | p = .55)$$

$$0.05 = P(-2.33 < \frac{\hat{p} - p_0}{\sqrt{(p_0)(1 - p_0)/n}} < 2.33 | p = 0.55)$$

$$0.05 = P(-2.33 < \frac{\hat{p} - 0.60}{\sqrt{(0.60)(0.40)/n}} < 2.33 | p = 0.55)$$

$$0.05 = P((-2.33)\sqrt{(0.60)(0.40)/n} + 0.60 < \hat{p} < (2.33)\sqrt{(0.60)(0.40)/n} + 0.60 | p = 0.55)$$

$$0.05 = P\left(\frac{(-2.33)\sqrt{(0.60)(0.40)/n} + 0.60 - 0.55}{\sqrt{(0.55)(0.45)/n}} < z < \frac{(2.33)\sqrt{(0.60)(0.40)/n} + 0.60 - 0.55}{\sqrt{(0.55)(0.45)/n}}\right)$$

$$0.05 = P(-2.29 + (0.10)\sqrt{n} < z < 2.29 + (0.10)\sqrt{n})$$

$$0.05 = P(z < 2.29 + (0.10)\sqrt{n})$$

$$-1.645 = 2.29 + (0.10)\sqrt{n}$$

$$n = 1548.4$$

$$n = 1549$$

10. Consider the following data.

x	y
1	6
2	15
3	12

Assume the form of the regression line is $y = mx + b + \epsilon$, where ϵ is a random error term whose mean is zero. Compute the following.

(a) \bar{x}, \bar{y}

(b) $S_{xy} = \sum_{i=1}^3 (x_i - \bar{x})(y_i - \bar{y})$

(c) $S_{xx} = \sum_{i=1}^3 (x_i - \bar{x})^2$

(d) The slope of the regression line $m = \frac{S_{xy}}{S_{xx}}$.

(e) Find the equation of the regression line using the slope from part (d) and the fact that the regression line must go through the point (\bar{x}, \bar{y}) .

(f) Estimate y when $x = 10$.

(g) What is the sum of the residuals? What is the sum of the residuals squared?

Solution

(a) $\bar{x} = 2$ and $\bar{y} = 11$

(b) $S_{xy} = ((1 - 2)(6 - 11) + (2 - 2)(15 - 11) + (3 - 2)(12 - 11)) = 6$

(c) $S_{xx} = ((1 - 2)^2 + (2 - 2)^2 + (3 - 2)^2) = 2$

(d) $m = \frac{6}{2} = 3$

(e) We are given $\bar{y} = m\bar{x} + b$, so $b = 5$. Then $\hat{y} = 3x + 5$.

(f) $\hat{y} = (3)(10) + 5 = 35$

(g) Sum of Residuals, $\sum(y - \hat{y}) = 0$. Sum Squares of Residuals, $\sum(y - \hat{y})^2 = 24$.

2 Day 2

1. Suppose that there is 35% that the wife but not the husband will win the raffle ticket. The chance that both of them will win is 20%. The chance that both of them lose is 40%. Let H=husband wins the raffle ticket and W=wife wins the raffle ticket.
 - (a) Calculate the probability that wife OR husband wins the raffle, $P(H \cup W)$.
 - (b) Calculate the probability that the husband wins the raffle, $P(H)$.
 - (c) What is the probability that the wife wins given that the husband wins, $P(W|H)$?
 - (d) What is the probability that the husband loses given that the wife has won, $P(H^c|W)$?

Solution

- (a) $P(W^c \cap H^c) = 1 - P(W \cup H)$, so $P(W \cup H) = 1 - .4 = .6$
- (b) $P(W \cup H) = P(H) + P(W) - P(H \cap W)$, so $P(H) = .6 - .35 = .25$
- (c) $P(W|H) = \frac{P(W \cap H)}{P(H)} = \frac{.2}{.25} = .8$
- (d) $P(H^c|W) = \frac{P(H^c \cap W)}{P(W)} = \frac{.35}{.55} = .64$

2. Suppose that amount of snow accumulation in Jasper is normally distributed. Suppose we take a random sample of 25 independent measurements and find the mean to be 32.5 and standard deviation 9. Let X represent the snow accumulation.
- (a) What is the probability that the amount of snow will be between 28.5 and 36.5?
 - (b) Find the 20th percentile for X .
 - (c) In order to have margin of error of 2 and be 95% confident, what is the smallest sample size we would need to have?

Solution

(a) $P(28.5 < X < 36.5) = P(-2.22 < Z < 2.22) = .9736$

Using Calculator: $normalcdf(28.5, 36.5, 32.5, \frac{9}{5})$

(b) $P(Z \leq \frac{X - 32.5}{9/5}) = .2 \rightarrow .84 = \frac{X - 32.5}{9/5} \rightarrow X = 30.99$

Using Calculator: $invNorm(0.2, 32, \frac{9}{5})$

(c) $n = \left(\frac{(Z_{\alpha/2})(s)}{MOE} \right)^2 = \left(\frac{(1.96)(9)}{2} \right)^2 = 77.79$, so $n = 78$.

3. A certain machine produces bottle caps for Coca-Cola. The machine produces 3% defectives. Total of 20,000 bottle caps are produced. Let X be the number of defective bottle caps out of 20,000.
- (a) What is the name of the exact distribution? (Specify the parameter values)
 - (b) What is the name of the approximate distribution? (Specify the parameter values) What are the conditions that allow us to use this approximation? (Name the Theorem as well)
 - (c) Find approximate probability that more than 550 bottle caps are defective.

Solution

- (a) $X \sim \text{Binomial}(p = .03, n = 20,000)$
- (b) $X \sim \text{Normal}(600, 582)$. Since n is large, by Central Limit Theorem we can approximate using Normal Distribution
- (c) Because we are approximating a discrete distribution with a continuous one, we have to use 0.5 continuity correction, so $P(X \geq 550.5) = P(Z \geq \frac{550.5 - 600}{\sqrt{582}}) = P(Z \geq -2.05) = P(Z < 2.05) = .9809$
Using Calculator: $\text{normalcdf}(550.5, 9^{10}, 600, \sqrt{582})$

4. A standardized test is given annually to all sixth grade students in the state of Washington. To determine the average score of students in a particular district, a school supervisor selects a random sample of 100 students. The sample mean of these students' scores is 320 and the sample standard deviation is 16.
- (a) Give a 99% confidence interval estimate of the average score of students in that supervisor's district.
 - (b) Interpret the confidence interval found above in words.
 - (c) Suppose that the district claimed that the average student score was 322. Should we believe the district at 95% confidence level?
 - (d) If we wanted to have $MOE = 5$, what is the smallest sample size we should have in order to be 90% confident?

Solution

(a) $(L, U) = \bar{X} \pm Z_{\alpha/2} \frac{s}{\sqrt{n}} = 320 \pm (2.576) \frac{16}{\sqrt{100}} = 320 \pm 4.12 \rightarrow (315.88, 324.12)$

- (b) With 99% confidence, we can say that the true population mean will fall in the interval found above.

(c) $(L, U) = \bar{X} \pm Z_{\alpha/2} \frac{s}{\sqrt{n}} = 320 \pm (1.96) \frac{16}{\sqrt{100}} = 320 \pm 3.14 \rightarrow (316.86, 323.14)$. Since 322 falls in the interval found at 95% level, we do not have sufficient evidence to reject the claim. It is plausible that the average student score was 322.

(d) $n = \left(\frac{(Z_{\alpha/2})(s)}{MOE} \right)^2 = \left(\frac{(1.645)(16)}{5} \right)^2 = 27.71 = 28$

5. A car is advertised as having a gas mileage rating of at least 30 mpg for highway driving. If the miles per gallon obtained in five independent experiments are $\{28, 30, 33, 25, 29\}$, is the advertisement correct?
- (a) State the Null and Alternative Hypothesis.
 - (b) What method would you use?
 - i. 1-sample T
 - ii. 1-sample Z
 - iii. 2-sample T
 - iv. 2-sample Z
 - (c) Calculate the test statistic.
 - (d) Find the Rejection Region using $\alpha = 0.05$.
 - (e) Should you believe the advertisement? Explain.
 - (f) What assumptions are you making?

Solution

- (a) $H_0 : \mu = 30$
 $H_a : \mu < 30$
- (b) 1-sample T because $n < 30$ and σ is unknown.
- (c) $t = \frac{\bar{X} - \mu}{s/\sqrt{5}} = \frac{29 - 30}{2.92/\sqrt{5}} = -.767$
Using Calculator: $T - Test(\mu_0 = 30, \bar{x} = 29, s = 2.92, n = 5, \mu < \mu_0)$
- (d) $RR : (t \leq t_{\alpha/2, df}) = (t \leq -2.132)$
Using Calculator: $invT(0.05, 4) = -2.132$
- (e) Do not Reject the null because $-.767 \not\leq -2.132$
- (f) We assume that the dataset is normally distributed and data sample is random.

6. Suppose we want to show that children have a different cholesterol level than the national average. It is known that the mean cholesterol level of Americans is 180 and $\alpha = .10$. We test 5 children and find the sample mean to be 200. We also know that the **population standard deviation** is 20.
- State the Null and Alternative Hypothesis.
 - State the parameters.
 - What method would you use?
 - 1-sample T
 - 1-sample Z
 - 2-sample T
 - 2-sample Z
 - Find the p-value
 - Do we reject or accept the null? Interpret your results.
 - Suppose that instead of 5 children, we tested 49 children and mean and standard deviation remained the same. Would we reject or accept the null? Interpret your results.
 - Given that the true value is 190, what sample size we would need to have to have a power of the test of 90%?

Solution

- $H_0 : \mu = 180$
 $H_a : \mu \neq 180$
- μ represents the average cholesterol level of children
- 1-sample Z test since σ is known
- $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{5}} = \frac{200 - 180}{20/\sqrt{5}} = 2.24$, from there we can look up our p -value = .0125 (For one sided test). Because it is a two sided test (\neq), the value should be double, so p -value = .0250.
 Using Calculator: $Z - Test(\mu_0 = 180, \bar{x} = 200, s = 20, n = 5, \mu \neq \mu_0) \rightarrow p = 0.0253$
- Since p -value $< \alpha$, we can reject the null at $\alpha = .10$ and state that children in fact do have a different cholesterol level than national average.
- Yes, you would still reject the null since your Z value would be even higher than at $n = 5$.

(g)

$$\text{Power} = 1 - P(\text{ Type II Error })$$

$$0.90 = 1 - P(\text{ Don't Reject } H_0 | \mu = 190)$$

$$0.10 = P(-1.645 < \frac{X - 180}{20/\sqrt{n}} < 1.645 | \mu = 190)$$

$$0.10 = P(\frac{(-1.645)(20)}{\sqrt{n}} + 180 < X < \frac{(1.645)(20)}{\sqrt{n}} + 180 | \mu = 190)$$

$$0.10 = P(\frac{(-1.645)(20)/\sqrt{n} + 180 - 190}{20/\sqrt{n}} < Z < \frac{(1.645)(20)/\sqrt{n} + 180 - 190}{20/\sqrt{n}})$$

$$0.10 = P(Z > \frac{(-1.645)(20)/\sqrt{n} + 180 - 190}{20/\sqrt{n}})$$

$$0.10 = P(Z > -1.645 + \frac{-10}{20/\sqrt{n}})$$

$$0.90 = P(Z \leq -1.645 + \frac{-10}{20/\sqrt{n}})$$

$$1.28 = -1.645 - \frac{10}{20/\sqrt{n}}$$

$$n = 34.22$$

$$n = 35$$

7. A famous reporter, Ian, wanted to figure out whether the students at IUPUI were more republican than at IU. From his two independent surveys, he found out that 60 students out of a 100 at IUPUI and 65 out of 120 at IU supported the Republican party. Assume $\alpha = .05$.
- State the Null and Alternative Hypothesis.
 - What do the parameter(s) represent?
 - What method would you use?
 - 1-sample T
 - 1-sample Z
 - 2-sample T
 - 2-sample Z
 - Calculate the test statistic.
 - State your conclusion and interpret the results in the context of the problem.

Solution

- $H_0 : p_1 = p_2$
 $H_a : p_1 > p_2$
- p_1 = proportion of students at IUPUI that support republican party
 p_2 = proportion of students at IU that support republican party
- 2-sample Z proportion test since n is large
- Our sample proportions are $\hat{p}_1 = \frac{60}{100}$ and $\hat{p}_2 = \frac{65}{120}$. From the hypothesis, we conclude that $p_1 - p_2 = 0$, and then we can calculate our z statistic.

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} = \frac{.6 - .54}{\sqrt{.0024 + .00207}} = .87$$
 Using Calculator: $2 - PropZTest(x_1 = 60, n_1 = 100, x_2 = 65, n_2 = 120, p_1 > p_2) \rightarrow 0.87$
- We can find critical value for $\alpha = .05$ which is 1.645 (Using Calculator: $invNorm(1 - 0.05, 0, 1) = 1.645$).
 Since $.87 \not\geq 1.645$, we do not have sufficient evidence to reject the null hypothesis at 95% that there are more supporters at IUPUI than IU.

8. A researcher of the MAC wants to know whether visiting the MAC affects the grade the student receives in a class. He records the number of visits (x) and the score received in a class (y) on 30 randomly selected students. Summary statistics for the data are as follows: $\bar{x} = 8.5$, $s_x = 4$, $\bar{y} = 86.5$, $s_y = 12$, $n = 30$.
- In view of the question the researcher wants to answer, what is the independent variable, and the response variable?
 - Suppose that correlation between number of visits and exam score is $r = 0.95$. Find the least squares regression equation.
 - What is predicted score in a class if a student visits the MAC 5 times?
 - Is the model overestimating or underestimating for the student who visited the MAC 5 times and has a score of 94%?
 - Interpret the meaning of slope and the intercept.
 - Find the 95% confidence interval for the slope of the regression line. Does the 95% confidence interval support the claim that $\beta > 3$?

Solution

- Independent = number of visits. Dependent = Score received in class
- $\hat{\beta}_1 = r \frac{s_y}{s_x} = (0.95) \frac{12}{4} = 2.85$. Now, let's use the fact that the line has to go through the point (\bar{x}, \bar{y}) so, $\bar{y} = \hat{\beta}_1 \bar{x} + \hat{\beta}_0$. From this, we calculate $\hat{\beta}_0 = 62.275$. Lastly, $\hat{y} = 2.85x + 62.275$.
- $\hat{y} = (2.85)(5) + 62.275 = 76.53$
- Since $\hat{y} = 76.52\%$ and $y = 94\%$, the model is underestimating.
- Slope means that for every one visit in the MAC, the score of the student will increase by 2.85. Intercept tells us that the student who does not visit the MAC will receive a score of 62.275 on average.
- $\hat{\beta}_1 \pm t_{\alpha/2} se(\hat{\beta}_1) = \hat{\beta}_1 \pm t_{\alpha/2} \frac{\hat{\beta}_1}{r} \sqrt{\frac{1-r^2}{n-2}} = \hat{\beta}_1 \pm \frac{s}{\sqrt{S_{xx}}} = 2.85 \pm .347 \rightarrow (2.487, 3.213)$. Since the interval does cover some of the values for $\beta > 3$, we do not have sufficient evidence to reject the null. So, we can't be certain that $\beta > 3$ at 95% confidence level.

9. Suppose we have four different patient groups that we treat each with different medicine. Each of the groups has 7 people. The doctors want to know whether there is significant difference in the average recovery time between all of the groups, so they run an Anova test. The output is shown below with some missing entries.

Source	df	SS	MS	F	p-value
Regression	3	56310	18770	*	≈ 0
Residual	*	*	*		
Total	27	75998			

Answer the following questions.

- What are the residual degrees of freedom?
- What is the number of observations, N ?
- Find SSE.
- Find the standard deviation of the regression, σ .
- Calculate the F statistic and explain what it means (Include appropriate Hypothesis).
- Find R^2 .

Solution

Source	df	SS	MS	F	p-value
Regression	$k - 1$	SSR	$MSR = \frac{SSR}{k-1}$	$F = \frac{MSR}{MSE}$	$Fcdf()$
Residual	$k(n - 1)$	SSE	$MSE = \frac{SSE}{k(n-1)}$		
Total	$kn - 1$	$SST = SSR + SSE$			

- $df = 24$
- $n = 28$
- $SSE = SST - SSR = 75998 - 56310 = 19688$
- $\sigma = \sqrt{MSE} = \sqrt{\frac{19688}{24}} = 28.64$
- $F = \frac{MSR}{MSE} = \frac{56310}{820.33} = 22.88$. The hypothesis for an F-statistic is the following

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_a : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$

From the table we can see that the p-value is very small, thus, we would reject the null hypothesis for any reasonable α level. So, there is significant difference in the mean recovery times between the four groups.

- $R^2 = \frac{SSR}{SST} = \frac{56310}{75998} = 0.74$.