Exam Review 2016 same as Stats 2009.doc

[VALUE 7]

1. Indicate which of the following variables used as qualitative and which are quantitative. As indicate whether the type of data is nominal, ordinal , ratio or interval.

VARIABLE	Qualitative	Quantitative	Nominal	Ordinal	Interval	Ratio
University in which a student is enrolled	✓		✓			
Duration (in Minuets) of a caller to a customer support line		✓				✓
Length (in feet) of a roll in plastic wrap advertised to be 30 feet long		✓				✓
Annual wage in dollars		✓				✓
Years of education		✓				✓
Number of Wins in Hockey league		✓				✓
temperature of your coffee		✓			✓	
Ranking In Grad Class		✓		✓		

[VALUE 4]

2. Julie collects a **sample** of size 6 to determine the average weight in kilograms of turkeys bought at Value Foods on June 11th. Manually compute the following using her data: 6.3, 7.0, 7.2, 18.5, 9.2, 5.8. Show all steps
   a. Mean
   b. Median
   c. Standard deviation

a) $\bar{x} = 9$

b) median = 7.1

x	$x-\bar{x}$	$(x-\bar{x})^2$
5.8	−3.2	10.24
6.3	−2.7	7.29
7.0	−2	4
7.2	−1.8	3.24
9.2	0.2	0.04
18.5	9.5	90.25
		115.06

median

$$S = \sqrt{\frac{\Sigma (x-\bar{x})^2}{n-1}}$$

$$S = \sqrt{\frac{115.06}{6-1}}$$

$$\therefore \sqrt{23.012} = 4.8$$

x	x^2
5.8	33.64
6.3	39.69
7.0	49
7.2	51.84
9.2	84.64
18.5	342.25

$\Sigma x = 54$

$\Sigma x^2 = 601.06$

$$s = \sqrt{\frac{n(\Sigma x^2) - (\Sigma x)^2}{n(n-1)}}$$

$$= \sqrt{\frac{6(601.06) - 54^2}{6(5)}}$$

$$\sqrt{23.012}$$
$$= 4.8$$

[VALUE 5]

3. Given the following scores on a first year Mrs. Mean's first year Statistics exam:
   a. Determine the percentile rank for a student who scored 70
   b. Determine the test score for a student who ranked in the 80th percentile

11	19	23	27	32	45	49
50	53	55	57	60	62	66
67	69	70	75	77	79	81
85	88	90	90	94	96	99

Percentile score $= \dfrac{\text{number of scores below data value} + 0.5}{\text{total number of data values}} \cdot 100$

Finding a Data Value Corresponding to a given Percentile $c = \dfrac{np}{100}$

where c = data value position
n = total number of values
p = percentile

-if c is a whole number count go to the number between c and c+1

-if c is not a whole number round up (hayway)!

(a) Percentile Score $\dfrac{\#\text{ below} + 0.5}{\text{total}} \times 100$

$\dfrac{16+0.5}{28} \times 100 = 58.9$ 59th 59th percentile

(b) $c = \dfrac{28(80)}{100} = 22.4$ score ∴ 23rd value ∴ 88

locator

$c = \dfrac{np}{100}$

[VALUE 4]

4. A random sample of smoking individuals is classified in the following frequency Table. Determine the mean and standard deviation of the frequency distribution.

Age Limits	Frequency
10-19	8
20-29	16
30-39	22
40-49	27
50-59	33
60-69	15
70-79	3

Age Limits	Frequency	midpt x	$f \cdot x$	$f \cdot x^2$
10-19	8	14.5	116	1682
20-29	16	24.5	392	9604
30-39	22	34.5	759	26186
40-49	27	44.5	1201.5	53467
50-59	33	54.5	1798.5	98018
60-69	15	64.5	967.5	62404
70-79	3	74.5	223.5	16651
	$n = 124$		5458	268011

Mean of grouped data - $\dfrac{\Sigma f \cdot x}{n}$

$\text{mean} = \dfrac{5458}{124} = 44.02$

$$S = \sqrt{\frac{n[\Sigma(f \cdot x_m^2)] - [\Sigma(f \cdot x_m)]^2}{n(n-1)}}$$

$$S = \sqrt{\frac{124(268011) - (5458)^2}{124(123)}}$$

$S = \sqrt{225.78}$ 15.03

5. A study was carried out to look at the relationship between the smoking habits of high school students and the habits of their parents. The data was summarized in the following table. Find the probability :

a. that one of the students parents smokes
b. that the high school student does not smoke given that both their parents smoke
c. Smokes given that one parent smokes

	Student Does Not Smoke	Student Smokes	Total
Both parents Smoke	1210	430	1640
Neither Parent Smokes	915	190	1105
One Parent Smokes	1350	310	1660
Total	3475	930	4405

[VALUE 5]

6. The probability density function for the number of blue m&m s in bags of m&m's is as follows
   a. Determine the mean and standard deviation of the probability distribution
   b. Is it unusual to get a bag of M&M's with 3 Blue M&M's. Explain

X	P(x)
4	0.09
5	0.12
6	0.18
7	0.20
8	0.19
9	0.11
10	?

$\mu = \Sigma[x \cdot P(x)]$

$\sigma = [\Sigma x \cdot P(x)] - \mu^2$

$\sigma = \sqrt{[\Sigma x^2 \cdot P(x)] - \mu^2}$

X	P(x)	$xP(x)$	$x^2P(x)$
4	0.09	0.36	1.44
5	0.12	0.60	3
6	0.18	1.08	6.48
7	0.20	1.40	9.8
8	0.19	1.52	12.16
9	0.11	0.99	8.91
10	? 0.11	1.1	11
		7.05	52.79

(a) $\mu = \Sigma x P(x)$
$= 7.05$

$\sigma = \sqrt{52.79 - 7.05^2}$

$\sigma \approx 1.76$

(b) $7.05 \pm 2\sigma$

$7.05 \pm 2(1.76)$

7.05 ± 3.52

3.53 to 10.57

$\therefore$ 3 is unusual

[VALUE 12]

7. The population of a large southern city is 63% African-American. A jury of 12 is selected at random from the citizens of the city and the courts are interested in the number of African-Americans on the jury.
   a. What is a success?
   b. What are the values of n and p?
   c. What is the probability that you get 3 African Americans on the jury?
   d. What is the mean number of African Americans on a jury of 12?
   e. What is the standard deviation of of African Americans on a jury of 12?
   f. What is the probability that all 12 jurors are African American?
   g. What is the probability that at least 10 jurors are not African American?

(a) success is selecting an african American

(b) $n = 12$, $p = 0.63$ $q = 0.37$

(c)

Binomial Probability Formula $P(X) = {}_nC_x \cdot p^x \cdot q^{n-x}$

(c) $P(X=3) = {}_{12}C_3 (0.63)^3 (0.37)^9$
$= 0.007$

(d) $\mu = np$
$= 12(0.63)$
$= 7.56$

(e) $\sigma = \sqrt{n \cdot p \cdot q}$
$= \sqrt{12(0.63)(0.37)}$
$= 1.67$

(f) $P(X=12) = {}_{12}C_{12}(0.63)^{12}(0.37)^0$
$= 0.0039$

(g) At least 10 jurors are not african American

not $p = 0.37$
$q = 0.63$

$P(\underset{\text{are not}}{X=10}) + P(\underset{\text{are not}}{X=11}) + P(\underset{\text{are not}}{X=12})$

${}_{12}C_{10}(0.37)^{10}(0.63)^2 + {}_{12}C_{11}(0.37)^{11}0.63^1 + {}_{12}C_{12}(0.37)^{12}(0.63)^0$

$= 0.00126 + 0.000135 + 0.00000658$
$= 0.0014$

(h) Is it unusual to get only 4 AA jurors

It is unusual

$7.56 \pm 2(1.67)$

4.22 to 10.9

The average caloric intake per day of males age 16-19 is 2850 cal. with a st. deviation of 280 cal.

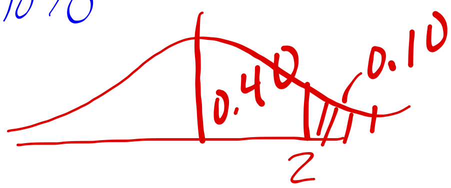
$z = \frac{x-\mu}{\sigma}$

(a) what percent of these males consume less than 2200 cal?

$P(x < 2200)$

$P(z <$

(b) How many calories do the top 10% of eaters consume



$X = z\sigma + \mu$

(c) If 100 males age 16-19 are examined, what is the prob. that their caloric intake is more than 2500 cal.

$P(\bar{x} > 2500)$

$\bar{x}$

$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

The average calorie consumption of 17 yr olds is 1950 cal with st. dev of 260 cal.

(a) What is the prob. that a randomly chosen 17 year old consumes less than 1700 cal

$P(x < 1700)$

$P(z < -0.96)$

$z = \frac{x-\mu}{\sigma}$

$= \frac{1700-1950}{260}$

-0.96

$= 0.1685$

(b) 35 seventeen yr olds were studied what is the prob. that their average caloric consumption was less than 1700 cal

$P(\bar{x} < 1700$

$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

$= \frac{1700-1950}{260/\sqrt{35}}$

$= -5.69$

$= 0.0001$

[VALUE 8]

8. The length of elephant pregnancies from conception to birth varies according to an approximately normal distribution with mean 525 days and standard deviation 32 days.

a. What percent of pregnancies last more than 600 days (that's about 20 months)?

b. What percent of pregnancies last between 510 and 540 days (that's between 17 and 18 months)?

c. How short do the shortest 10% of all elephant pregnancies last?

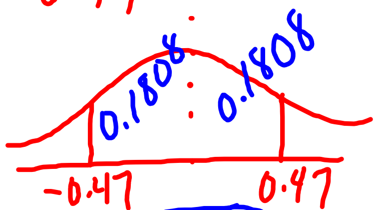
(a) $P(x>600)$

$$z = \frac{x-\mu}{\sigma} = \frac{600-525}{32} = 2.34$$

ans: $\boxed{0.0096}$

(b) $P(510 < x < 540)$

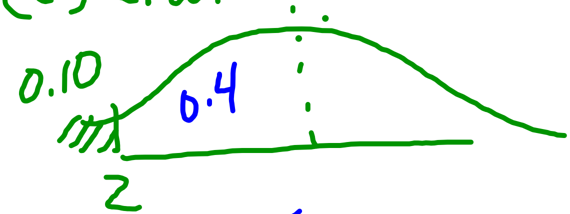
$$z = \frac{510-525}{32} = 0.47 \qquad z = \frac{540-525}{32} = -0.47$$



$\boxed{\therefore\ 0.3616}$

$$z = \frac{x - \mu}{\sigma}$$

(c) shortest 10%



$\therefore z = -1.28$

$$x = z\sigma + \mu = -1.28(32) + 525 = 484.04$$

[VALUE 4]

9. Assume that the heights of women are normally distributed with a mean given by $\mu = 63.6$ **inches** and a standard deviation given by $\sigma = 2.5$ **inches**. The Beanstalk Club, a social organization for tall people, has a requirement that women must be at least 70 in. tall. What percentage of women meet that requirement?

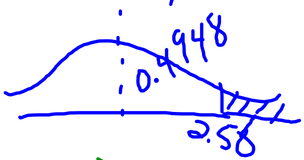
$P(x > 70)$

$P(z > 2.56)$

$z = \frac{x - \mu}{\sigma}$

$= \frac{70 - 63.6}{2.5}$

$= 2.56$



$P(x > 70) = 0.0052$

As reported by Runners World the average time to finish teh New York City 10km run is 61min with a standard deviation of 9 min.

a. What is the probability that a runner finishes in less than 56 min

b. If a sample of 35 runners were randomly selected what is the probability that they average less than 56 min

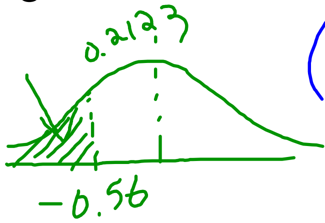
$P(x < 56)$

$z = \frac{x - \mu}{\sigma}$

$= \frac{56 - 61}{9}$

$= \frac{-5}{9}$

$z = -0.56$



$\boxed{0.2877}$

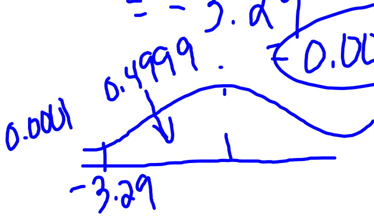
(b) $P(\bar{x} < 56)$

$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

$= \frac{56 - 61}{9/\sqrt{35}}$

$= -3.29$

0.0001



[VALUE 4]

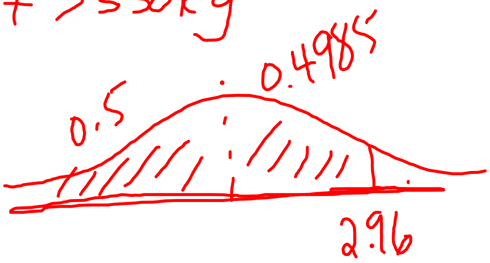
10. Government regulations indicate that the total weight of cargo in a certain kind of airplane cannot exceed 330 kg. On a particular day a plane is loaded with 100 boxes of goods. If the weight distribution for individual boxes is normal with mean 322 kg and standard deviation 27 kg, what is the probability that the regulations will NOT be met:

TOTAL weight cannot > 330kg

$P(\bar{x} > 330)$

$Z = \dfrac{330-322}{27/\sqrt{100}}$

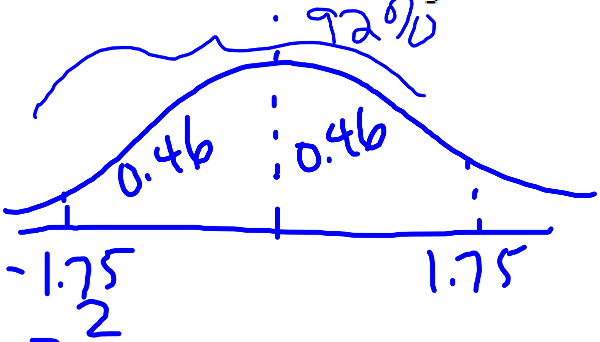
$Z = 2.96$



≈ 0.0015

[VALUE 3]

11. You wish to estimate the mean weight of machine components of a certain type and you require a 92% degree of confidence that the sample mean will be in error by no more than 0.008g. Find the sample size required. A pilot study showed that the population standard deviation is estimated to be 0.08g



Sample size

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 \quad \text{or} \quad n = \frac{(z_{\alpha/2})^2\, \hat{p}\,\hat{q}}{E^2}$$

$$n = \left[\frac{1.75(0.08)}{0.008}\right]$$

$n = 306.25$

[VALUE 3]

12. For a group of 14 men subjected to a stress test, the mean number of hearts beats per minute was $\overline{126}$ with a standard deviation of 9. Find the 90% confidence interval of the true mean. (assume normal)

n<30

Confidence Interval

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Confidence Interval for t-distribution

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$126 \pm 1.645\,\frac{9}{\sqrt{14}}$$

$$126 \pm 3.96$$

$$122.04 < \mu < 129.96$$

$$\bar{x} \pm t_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

90%

df n-1 = 13

$$126 \pm 1.771\left(\frac{9}{\sqrt{14}}\right)$$

$$126 \pm 4.26$$

$$\boxed{121.74 < \mu < 130.26}$$

[VALUE 6]

13. A wildlife biologist studying black bears measured the weights of 54 such animals. The data was entered into a computer and the results to the right were obtained:
   a. Find a 95% confidence interval for the mean weight of bears in the population. Interpret the interval.
   b. Would a 99% confidence interval be narrower or wider. Explain.

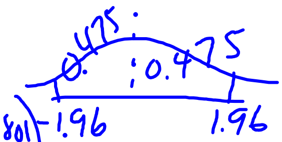
Summary of No Selector	Weight
Count	54
Mean	182.889
Median	150
MidRange	270
StdDev	121.801
Range	488
IntQRange	150

Confidence Interval

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

(a) $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

$182.889 \pm 1.96 \left(\frac{121.801}{\sqrt{54}} \right)$



182.889 ± 32.487

$150.402 < \mu < 215.376$

We are 95% conf. that the mean weight of a black bear is between 150.4 and 215.3

(b) $\bar{x} \pm 2.575 \frac{121.801}{\sqrt{54}}$

182.889 ± 42.681

$140.228 < \mu < 225.55$

wider

[VALUE 8]

14. The average cost to produce a major motion picture is \$62.9 million. In a recent sample of 20 movies the average production cost was \$67.2 million with a standard deviation of 8.8 million dollars. At the 0.05 level can it be concluded that is costs more thai \$62.9 million to make a movie?

   a. State the null and alternate hypothesis.

   b. Test the hypothesis at the 0.05 significance level.

$n=20$

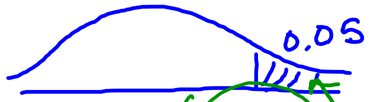
z-table one tail test

$\alpha=0.05$

$df=19$

$H_0: \mu \le 62.9$

$H_1: \mu > 62.9$ (claim)



t-table.. reject the null

test stat $= \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

$\dfrac{67.2-62.9}{8.8/\sqrt{20}}$

$= 2.19$

claim was supported movies cost more than 62.9 million

Reject H_0 if $P < \alpha$.

Nationwide, graduates entering the actuarial field earn \$40,000. A college placement officer feels that this number is too low. She surveys 36 graduates entering the actuarial field and finds the average salary to be \$41,000 with a standard deviation of \$3000. Can her claim be supported at $\alpha = 0.05$?

Source: BeAnActuary.org.

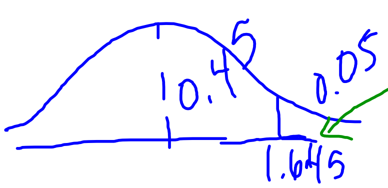
$H_0 \quad \mu \le 40000$

$H_a: \mu > 40000$ (claim)

test stat $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

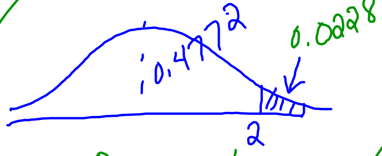
$= \dfrac{41000-40000}{3000/\sqrt{36}}$

$= 2$



reject null

p-value



$\underline{P}$ $\quad$ $\underline{\alpha}$

$0.0228 < 0.05$

reject null

[VALUE 6]

15. Suppose that in the last year all students at a certain university reported the number of hours spent on their studies during a certain week; the average was 40 hours. This year we want to determine whether the mean time spent on studies of all students at the university was not 40 hours per week. A random sample of 35 students at the university was selected and the mean number of hours was 43.1 with a standard deviation of 11.85 hours.

   a. State the null and alternate hypothesis.

   b. Test the hypothesis at the $\alpha = 0.10$ significance level.

$H_0: \mu = 40$

$H_a: \mu \neq 40$
(Two Tail Test)

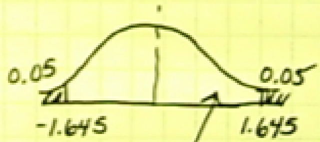
$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$= \frac{43.1 - 40}{11.85/\sqrt{35}}$

$= 1.55$

Trad. Method

$\alpha = 0.10$

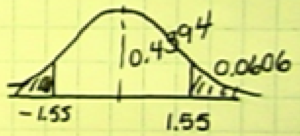


$z = 1.55$

∴ fail to reject the null

evidence says students do not spend 40 hrs studying

p-value method



$p = 2 \times 0.0606$
$= 0.1212$

p α

$0.1212 > 0.10$

16. 1500 randomly selected pine trees were tested for traces of the Bark Beetle infestation. It was found that 203 of the trees showed such traces. Test the hypothesis that more than 15% of the trees have been infested. (Use a $\alpha = 0.05$)

Hypothesis Testing: null always contains the equality
p-value method and critical value method

Test Statistic

$$z = \frac{x - \bar{\mu}}{\sigma/\sqrt{n}}$$

Test Statistic for a proportion

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

$H_0: p \le 0.15$

$H_a: p > 0.15$ (claim)

Rt. Tail Test

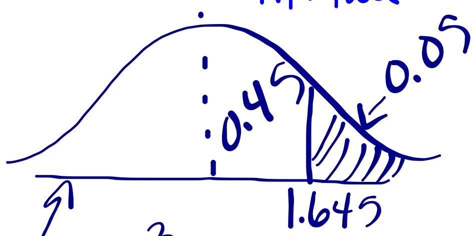
$\hat{p} = \frac{203}{1500}$

0.135

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

$$z = \frac{0.135 - 0.15}{\sqrt{\frac{0.15(0.85)}{1500}}}$$

$z = -1.63$



$z = -1.63$

$\therefore$ fail to reject the null do not support the claim

17. In the TV show "Sneak Preview", the late Gene Siskel and Roger Ebert reviewed the weeks new movies and rated with a Thumb Up, Mixed or Thumbs Down. Where the ratings given by Siskel and Ebert related? The answer to this question was the focus of the paper, " Evaluating Agreement and Disagreement Among Movie Reviewers", by Alan Agresti and Larry Winner. Following is a contingency table that summarizes the ratings of Siskel and Ebert for 160 movies

	Siskel Thumbs Up	Siskel Mixed	Siskel Thumbs Down	
Ebert Thumbs Up	24 (12.58) 11.81	8 (0.023) 8.44	13 (5.58) 24.75	45
Ebert Mixed	8 (0.02) 8.4	13 (8.12) 6.0	11 (2.48) 17.6	32
Ebert Thumbs Down	10 (6.38) 21.79	9 (2.77) 15.56	64 (7.38) 45.65	83
	42	30	88	160

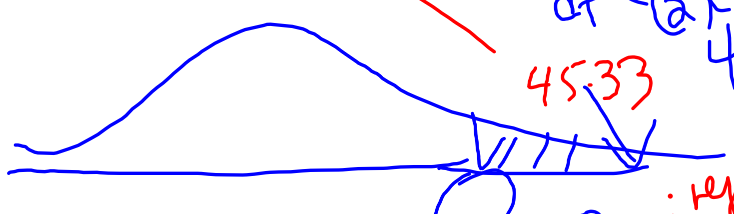
At $\alpha = 0.01$ does the data provide sufficient evidence to conclude that there is an association between Siskel and Eberts' ratings?

H_0: Siskel's review is independent of Eberts review

H_a: Siskels review is dependent on Eberts review

$\chi^2 = 12.58 + 0.023 + 5.58 + 0.02 + 8.12 + 2.48 + 6.38$
$2.77 + 7.38 = 45.33$

G Table df (3-1)(3-1)
(2)(2)
4



45.33

13.277

∴ reject null.

[VALUE 7]

18. A study was conducted to determine if employees perform better at work with music playing. The music was turned on during the working hours of a business with 45 employees. There productivity level averaged 5.2 with a standard deviation of 2.4. On a different day the music was turned off and there were 40 workers. The workers' productivity level averaged 4.8 with a standard deviation of 1.2. What can we conclude at the .01 level?

$H_0: \mu_1 \leq \mu_2$

$H_a: \mu_1 > \mu_2$ employees perform better with music (claim)

x_1	x_2
$\bar{x}_1 = 5.2$	$\bar{x}_2 = 4.8$
$S_1 = 2.4$	$S_2 = 1.2$
with music	without music
$n = 45$	$n = 40$

test stat $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$

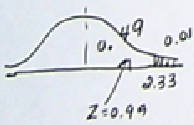
$z = \dfrac{(5.2 - 4.8) - (0)}{\sqrt{\dfrac{2.4^2}{45} + \dfrac{1.2^2}{40}}}$

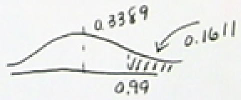
$= 0.988$

$= 0.99$

$\alpha = 0.01$

Trad. Method





p		α
0.1611	>	0.01

∴

∴ do not reject the null; do not support the claim

employees do not perform better with music

VALUE 7]

19. A study is designed to check the relationship between smoking and longevity (life expectancy). A sample of 15 men 50 years and older was taken and the average number of cigarettes smoked per day and the age at death was recorded, as summarized in the table. Test the significance of the correlation coefficient.

Cigarettes	5	23	25	48	17	8	4	26	11	19	14	35	29	4	23
Longevity	80	78	60	53	85	84	73	79	81	75	68	72	58	92	65

$r = -0.713$ this means a negative correlation (increase smoking decreasing longevity)

Is this due to chance or is there a significant linear relationship

$H_0: \rho = 0$ (no correlation & r value is due to change)

$H_a: \rho \neq 0$ (significant correlation)

t (test stat)

$t = r\sqrt{\frac{n-2}{1-r^2}}$

$= -0.713\sqrt{\frac{15-2}{1-(-0.713)^2}}$

$= -3.67$

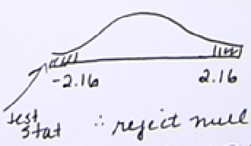
CRIT value

$n-2$

$15-2 = 13$ df

$\alpha = 0.05$ Two Tail

$\therefore \pm 2.16$



$\therefore$ reject null

there is a non-zero correlation between smoking and longevity

LinReg
y=ax+b
a=-.6282004052
b=85.72042119
r²=.5089826137
r=-.7134301744

Attachments

- Exam Review 2013 same as Stats 2009.doc
- Exam Review 2016 same as Stats 2009.doc
- Exam Review 2017 same as Stats 2009.doc