Statistics 300
Exam #3 Name: Solution

Fall 2011
Tue/Thu 7:00 - 9:05 pm

(9 points; 10 minutes)

1. **Use the row percentages in the table to test the idea that the percentage of Phoenix Suns fans that live in California is the same as the percentage of Sacramento Kings fans that live in Arizona.**

**Use a 5% significance level for this test.**

**The data represent truly random samples of Suns, Kings, and Sonics fans.**

| Favorite Basketball Team | Home State AZ | CA | WA | Row Total |
|---|---|---|---|---|
| Phoenix Suns | 68% | 15% | 12% | 190 |
| Sacramento Kings | 21% | 68% | 7% | 191 |
| Seattle Sonics | 11% | 18% | 81% | 219 |

$\boxed{(p_K - p_S) = 0}$

H₀: $\cancel{p_{Kings|AZ} = p_{Suns|CA}}$

H₁: $(\cancel{p_{K|A} \neq p_{S|C}}) \neq 0$

$\alpha = 0.05$ in 2 tails

$\hat{p}_K = 0.21 \quad N_K = 191 \quad$ so $\quad X_K = 40$

$\hat{p}_S = 0.15 \quad N_S = 190 \quad X_S = 29$

Since "0" difference is hypothesized, do $\bar{p}$ and $\bar{q}$

$$\bar{p} = \frac{\text{all successes}}{\text{all trials}} = \frac{40 + 29}{190 + 191} = \frac{69}{381} = 0.181$$

$$\bar{q} = 0.819$$

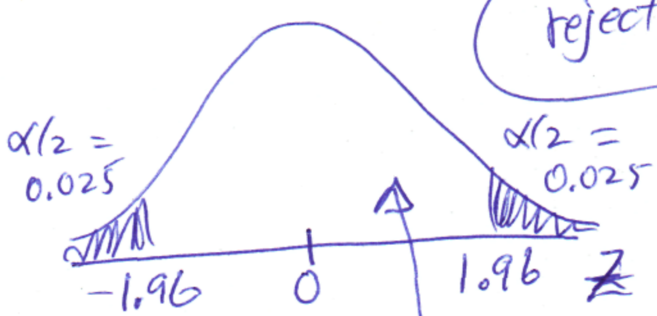
Test Statistic:

$$\frac{(\hat{p}_K - \hat{p}_S) - (p_K - p_S)_0}{\sqrt{\dfrac{\bar{p}\bar{q}}{N_K} + \dfrac{\bar{p}\bar{q}}{N_S}}}$$

$$= \frac{(0.21 - 0.15) - 0}{\sqrt{\dfrac{(0.181)(0.819)}{191} + \dfrac{(0.181)(0.819)}{190}}}$$

$$= \frac{0.06}{0.0395} = \boxed{1.52}$$

Do Not reject H₀



Statistics 300
Exam #3 Name: Solution

Fall 2011
Tue/Thu 7:00 - 9:05 pm

(8 points; 8 minutes)

**2. Is there a linear relationship between daily average temperature and daily average wind speed? Use the data in the table for a random sample of five daily values to test the claim that mean temperature and mean wind speed are negatively correlated.** (Let $\alpha = 0.10$ for this test.)

| Day | Mean Temp. °F | Mean Speed m/s |
|---|---|---|
| 1 | 91.9 | 15.7 |
| 2 | 81.4 | 13.8 |
| 3 | 93.2 | 21.5 |
| 4 | 70.8 | 33.6 |
| 5 | 100 | 2.1 |

Claim: $\rho < 0$ : daily avg. wind speed and daily average temp. are neg. correlated.

$H_0$: $\rho \geq 0$

$H_1$: $\rho < 0$

$\alpha = 0.10$ left tail

$N = 5$
$df = 3$

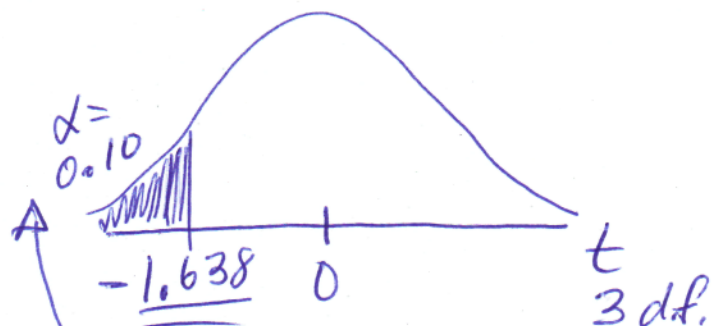
$\boxed{r = -0.805}$
from calculator

test statistic:

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

$$= \frac{-0.805}{\sqrt{\dfrac{1-(-0.805)^2}{5-2}}} = \frac{-0.805}{0.3425}$$

$$= \boxed{-2.350}$$




reject $H_0$:

**Statistics 300**
**Exam #3** **Name:** Solution

**Fall 2011**
**Tue/Thu 7:00 - 9:05 pm**

(9 points; 10 minutes)

**3. Use the summary statistics for a random selection of Fridays and Saturdays to test the claim that the average number of cars on a Sacramento freeway is at least 1000 more on Fridays than it is on saturdays.** (Use a 0.025 significance level for this test.) **Differences in average traffic on Fridays are known to be larger than they are on Saturdays.** $\sigma_F^2 > \sigma_S^2$ Do Not pool variances and do use smaller of $df_F$ and $df_S$

$$\mu_F \geq \mu_S + 1000$$

| Sample Statistic | Fridays | Saturdays |
|---|---|---|
| N = | 10 | 16 |
| Average = | 38,378 | 36,811 |
| Standard Deviation = | 838 | 901 |

$df_F = 9$ $\quad df_S = 15$

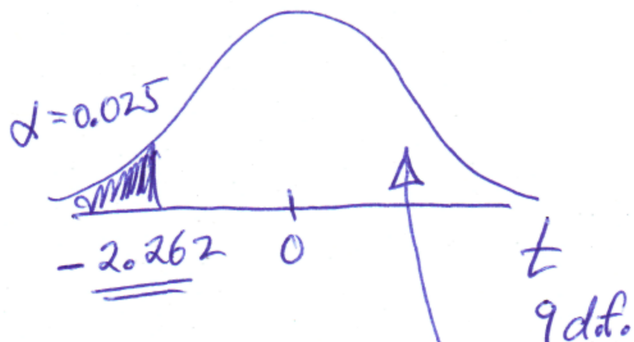
$H_0$: $(\mu_F - \mu_S) \geq 1000$

$H_1$: $(\mu_F - \mu_S) < 1000$

$\alpha = 0.025$ left tail

Test Statistic

$$\frac{(\bar{x}_F - \bar{x}_S) - (\mu_F - \mu_S)_0}{\sqrt{\dfrac{s_F^2}{n_F} + \dfrac{s_S^2}{n_S}}}$$




$$= \frac{(38378 - 36811) - 1000}{\sqrt{\dfrac{(838)^2}{10} + \dfrac{(901)^2}{16}}} = \frac{567}{348} = \boxed{1.629}$$

Do Not reject Ho:

Statistics 300 | Fall 2011
Exam #3 | Tue/Thu 7:00 - 9:05 pm

Name: Solution

(9 points; 10 minutes)

4. Use the survey results given in this problem to test the claim that the proportion of prison inmates who return to prison after being released is independent of the type of crime for which they were convicted. Use a Type I error rate of 0.05 for this test.

$\boxed{\text{Contingency Table}}$

| Type of Crime | Returned to Prison: Yes | Returned to Prison: No | Row total |
|---|---|---|---|
| Violent Felony | 35 / 31 | 65 / 69 | 100 |
| Non-violent Felony | 26 / 31 | 74 / 69 | 100 |
| Violent Misdemeanor | 31 / 31 | 69 / 69 | 100 |
| Non-violent Misdemeanor | 32 / 31 | 68 / 69 | 100 |
| Column Total | 124 | 276 | 400 |

$H_0$: (% return to prison) and (type of crime) are independent.

$H_1$: Not $H_0$!

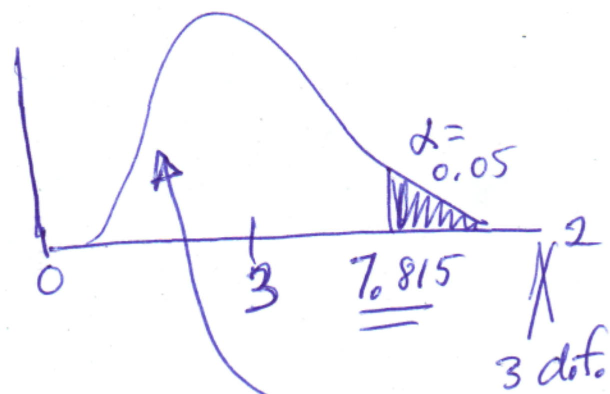
$\alpha = 0.05$ right tail

d.f. $= (r-1)(c-1) = (4-1)(2-1) = 3$

~~obser~~ Expected $= \dfrac{(\text{row total})(\text{col. total})}{\text{grand total}}$

$(O-E)^2/E$

| | | |
|---|---|---|
| 0.52 | 0.23 | .75 |
| 0.81 | 0.36 | +1.17 |
| 0.00 | 0.00 | +0 |
| 0.03 | 0.01 | +.04 |
| | | 1.96 |




$$\sum\left[\frac{(\text{Obs}-\text{Exp})^2}{\text{Exp}}\right] = 1.96$$

Do not reject $H_0$!

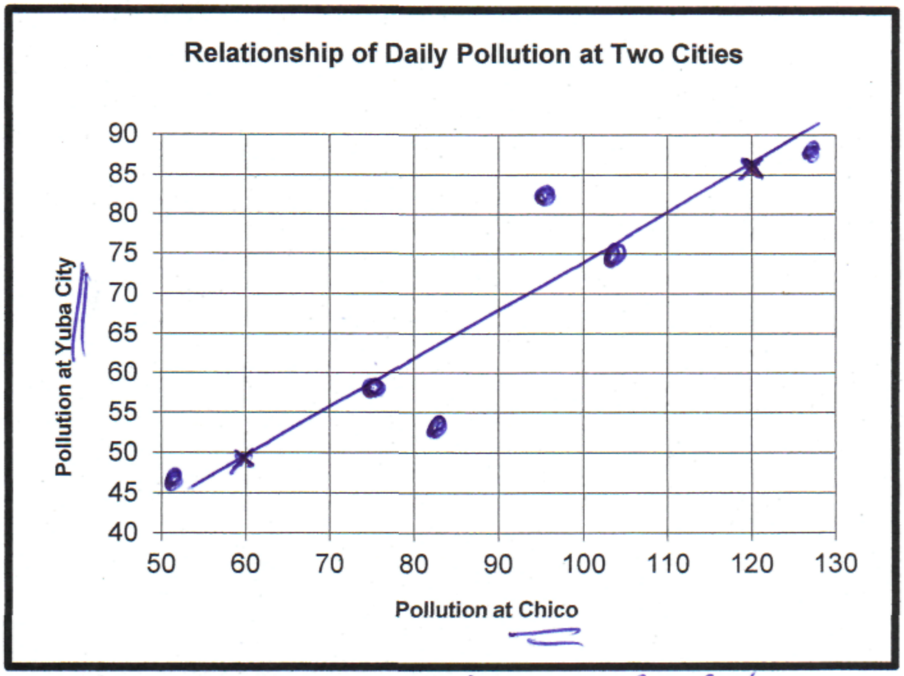
Statistics 300
Exam #3 Name: Solution

Fall 2011
Tue/Thu 7:00 - 9:05 pm

(13 points; 13 minutes)

**5. Daily air pollution measurements from communities that are near one another usually have a linear relationship to one another. Use the data for Chico and for Yuba City to answer the questions on this page and the next page.**

| Day | Yuba City (Y) | Chico (X) |
|---|---|---|
| 1 | 75 | 103 |
| 2 | 84 | 96 |
| 3 | 88 | 128 |
| 4 | 57 | 75 |
| 5 | 46 | 51 |
| 6 | 54 | 82 |




**(a) Plot the points on the graph.**

**(b) Use your calculator to determine the equation of the line that best predicts pollution at Yuba City based on pollution at Chico.**

from calculator

**equation =** $\hat{y} = 13.68 + (0.602)(X)$

**(c) Plot your line on the graph.**

**(d) What is the predicted pollution at Yuba City when pollution at Chico is 100?** $y'$ 73.9

$\hat{y} = 13.68 + (0.602)(100)$ OR from calculator

**(e) Estimate the correlation of pollution at Chico and Yuba City on all days?** 0.912 $= r$

**(f) What proportion of the variation in pollution levels at Yuba City for this set of six days is explained by the levels of pollution at Chico?** 0.8319 $= r^2$

**(g) For the "total" variation in the Yuba City pollution data:**

**The expression is:** $\sum (y - \bar{y})^2$ **The value is:** 1503.3 $= s_y^2 (n-1)$

**(h) For the "explained" variation in the Yuba City pollution data:**

**The expression is:** $\sum (\hat{y} - \bar{y})^2$ **The value is:** 1250.6 $= (\text{total})(r^2)$

**(i) For the "unexplained" variation in the Yuba City pollution data:**

**The expression is:** $\sum (y - \hat{y})^2$ **The value is:** 252.7 $=$ total $-$ explained

**Statistics 300** | **Fall 2011**

**Exam #3** Name: Solution **Tue/Thu 7:00 - 9:05 pm**

(2 points; 2 minutes)

**6. Continue using the Chico and Yuba City pollution data to answer the questions below.**

**(a) For the "standard error of estimate" in relating Yuba City pollution to Chico pollution:**

**The expression is:** $\sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}}$ **The value is:** 7.95

$$\sqrt{\frac{\text{unexplained}}{n-2}} = \sqrt{\frac{252.7}{6-2}} = 7.95$$

Statistics 300
Exam #3

Name: Solution

Fall 2011
Tue/Thu 7:00 - 9:05 pm

$(\sigma_F^2 = \sigma_M^2)$ so, pool variances and add deg. of freedom

(10 points; 10 minutes)

7. Experts tell you that variability in annual income is the same for Engineering graduates, whether they are male or female. Use the data for two random samples in the table below to prepare a 90% confidence interval for the difference between the mean income of all male Engineering graduates and the mean income of all female engineering graduates.

| Incomes in 1000's of Dollars | | |
|---|---|---|
| Sample Statistic | Gender | |
| | Male | Female |
| N = | 12 | 18 |
| Average = | 44.8 | 47.8 |
| Std. Deviation = | 5.6 | 4.7 |

$\alpha = 1 - \text{confid.} = 1 - 0.90 = 0.10$ in 2 tails

$\boxed{t_{\alpha/2} = 1.701}$

$df = 11 + 17 = 28$ d.f.

$$\boxed{S_{pool}^2 = \frac{S_M^2(n_M - 1) + S_F^2(n_F - 1)}{(n_M - 1) + (n_F - 1)} = \frac{(5.6)^2(12-1) + (4.7)^2(18-1)}{(12-1) + (18-1)} = \frac{720.49}{28} = 25.73}$$

$$90\%\, CI(\mu_M - \mu_F) = (\bar{x}_M - \bar{x}_F) \pm t_{\alpha/2}\sqrt{\frac{S_{pool}^2}{N_M} + \frac{S_{pool}^2}{N_F}}$$

$$(44.8 - 47.8) \pm 1.701\sqrt{\frac{25.73}{12} + \frac{25.73}{18}}$$

$$= (-3) \pm (1.701)(1.890)$$

$$= (-3) \pm 3.21$$

$$\left[-6.21 < (\mu_M - \mu_F) < 0.21\right]$$

Based on your interval is it reasonable to claim that male Engineering graduates earn more on average than female Engineering graduates earn.

Yes    No    Why? Because if $\mu_M > \mu_F$, $(\mu_M - \mu_F) > 0$, and values $> 0$ are in the interval.

Based on your interval is it reasonable to claim that male Engineering graduates earn less on average than female Engineering graduates earn.

Yes    No    Why? Because if $(\mu_M < \mu_F)$, $(\mu_M - \mu_F) < 0$, and values $< 0$ are in the interval.

Statistics 300
Exam #3 Name: Solution

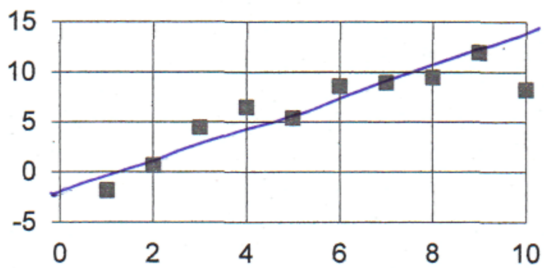
Fall 2011
Tue/Thu 7:00 - 9:05 pm

(6 points; 6 minutes)

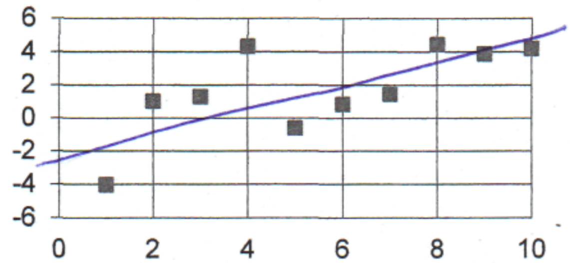
**8. Connect each picture with one of the candidate "r" values by writing the appropriate candidate "r" value in the space at the top of each graph.**

| Candidate values of "r", the sample correlation coefficient. | | | | | | |
|---|---|---|---|---|---|---|
| 0.00 | -0.70 | -0.90 | -1.00 | 0.70 | 0.90 | 1.00 |

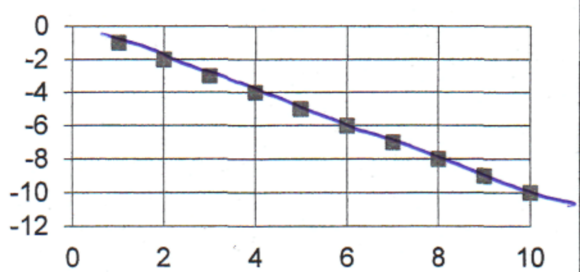
0.90



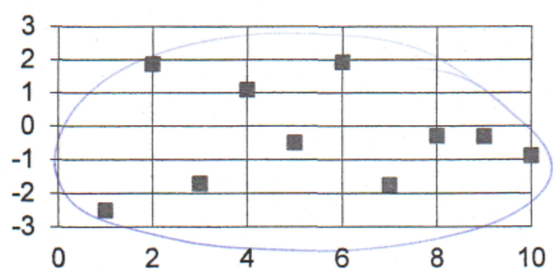

0.70



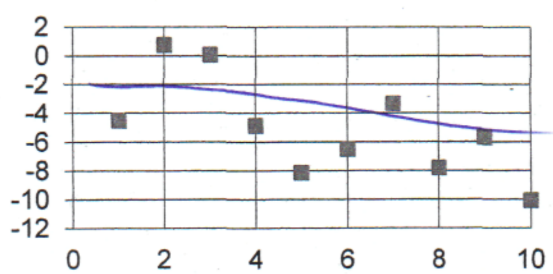

-1.00



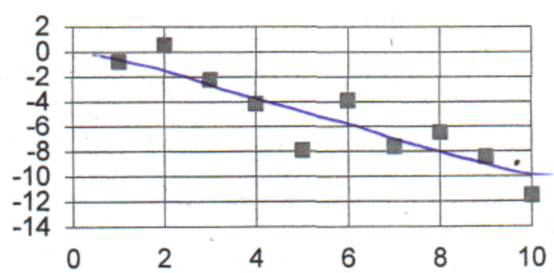

0.00




-0.70




-0.90



Statistics 300
Exam #3

Name: Solution

Fall 2011
Tue/Thu 7:00 - 9:05 pm

(9 points; 9 minutes)

**9. Based on the statistics shown below, construct an 84% confidence interval for the difference between the percentage of 15 year old girls that have a personal cell phone and the percentage of 15 year old boys that have a personal cell phone.** (For the test, let $\alpha = 0.05$.)

**Sample Statistics**

| Personal Cell Phone | 15 Year Old Girls | 15 Year Old Boys |
|---|---|---|
| Yes | 90 | 51 |
| No | 45 | 41 |
| | $n=135$ | $n=92$ |

$\hat{p}_g = \frac{90}{135} = 0.667$ $\quad \hat{p}_b = \frac{51}{92} = 0.554$

$\hat{q}_g = 0.333$ $\quad \hat{q}_b = 0.446$

$84\%\ CI(p_g - p_b) =$

$\cancel{(\hat{p}_g - \hat{p}_b) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_g \hat{q}_g}{n_g} + \frac{\hat{p}_b \hat{q}_b}{n_b}}}$

$(\hat{p}_g - \hat{p}_b) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_g \hat{q}_g}{n_g} + \frac{\hat{p}_b \hat{q}_b}{n_b}}$

$= (0.667 - 0.554) \pm 1.41\sqrt{\frac{(.667)(.333)}{135} + \frac{(.554)(.446)}{92}}$

$= 0.113 \pm (1.41)(0.0658)$

$= 0.113 \pm 0.093$

$= [0.020 < (p_g - p_b) < 0.206]$

confidence $= 0.84$

$\alpha = 1 - \text{confid} = 0.16$

$\alpha/2 = 0.08$

$z_{\alpha/2} = 1.41$

| | .01 |
|---|---|
| -1.4 | $\boxed{.0793}$ |

**Based on your interval is it reasonable to claim that the percentage of 15 year old boys that have a personal cell phone is greater than the percentage of 15 year old girls that have a personal cell phone?**

**Yes** **No**

**Why?** If $(p_b > p_g)$, then $(p_g - p_b)$ would be $< 0$, and No values less than zero are in the interval.

Statistics 300
Exam #3

Name: Solution

Fall 2011
Tue/Thu 7:00 - 9:05 pm

(9 points; 7 minutes)

10. Use the information on this page to complete the Analysis of Variance table and test the claim that milk from nine different producers has the same average "shelf life" (number of days before milk goes bad). Use a 10% significance level for the test.

Use p-value approach to hypothesis test.

AOV Table

$=(8)(9.818)$

$(9.818/4.820)$

| Source | SS | df | MS | F | p-value |
|---|---|---|---|---|---|
| Producer | 78.544 | 8 | 9.818 | 2.0369 | 0.0474 < 0.10 |
| Error | 592.866 | 123 | 4.820 | | |
| Total | 671.41 | 131 | | | |

$\alpha$

so reject $H_0$:

$N=132$

$(671.41-78.544)$

$\frac{592.866}{123}$

$H_0$: $\mu_1 = \mu_2 = \cdots = \mu_9$

$H_1$: Not $H_0$:

$\alpha = 0.10$

Shelf Lives ( in "days" ) of Milk samples from Nine Producers

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| | 15.6 | 14.7 | 17.2 | 16.1 | 14.5 | 16.2 | 14.1 | 13.2 | 17.6 |
| | 11.1 | 11.7 | 13.5 | 15.5 | 16.0 | 12.7 | 11.7 | 15.7 | 14.0 |
| | 14.7 | 17.9 | 14.1 | 10.4 | 14.9 | 14.4 | 16.6 | 16.2 | 12.5 |
| | 16.2 | 16.1 | 14.2 | 11.8 | 13.9 | 13.4 | 14.4 | 18.4 | 12.7 |
| | 18.2 | 16.2 | 13.9 | 12.6 | 17.1 | 13.7 | 10.4 | 13.8 | 15.8 |
| | 14.8 | 13.1 | 17.1 | 12.0 | 14.1 | 16.4 | 11.9 | 18.5 | 12.6 |
| | 15.3 | 12.9 | 14.0 | 13.7 | 13.8 | 15.8 | 10.9 | 16.1 | 12.0 |
| | 14.5 | 12.0 | 14.9 | 14.4 | 15.2 | 8.7 | 14.3 | 9.0 | 13.7 |
| | 14.3 | 13.1 | 17.2 | 17.2 | 17.8 | 14.6 | 10.3 | 13.6 | 14.8 |
| | 16.3 | 12.3 | 15.4 | 11.4 | 16.3 | 17.7 | 14.1 | 16.6 | 12.6 |
| | 13.8 | 13.0 | 13.2 | 15.9 | | 16.7 | 12.7 | 14.8 | 16.5 |
| | 18.8 | 13.9 | 18.2 | 16.6 | | 14.3 | 12.9 | 13.5 | 17.6 |
| | 11.6 | 18.1 | 13.9 | 10.9 | | | 12.4 | 13.1 | 10.9 |
| | 12.6 | 14.7 | 18.7 | 14.2 | | | 7.3 | 14.5 | 16.1 |
| | 12.5 | 13.4 | 14.0 | 13.0 | | | 12.2 | 14.3 | |
| | 10.5 | 10.7 | | | | | 13.3 | 17.6 | |
| | | 13.1 | | | | | 18.0 | | |
| n = | 16 | 17 | 15 | 15 | 10 | 12 | 17 | 16 | 14 |
| x = | 14.4 | 13.9 | 15.3 | 13.7 | 15.4 | 14.6 | 12.8 | 14.9 | 14.2 |
| s = | 2.4 | 2.1 | 1.9 | 2.2 | 1.4 | 2.4 | 2.5 | 2.4 | 2.2 |

$N=132$

Statistics 300
Exam #3

Name: Solution

Fall 2011
Tue/Thu 7:00 - 9:05 pm

(9 points; 10 minutes)

**11. Two programs for encouraging school attendance were studied at five schools. At each shool, half of the students were randomly assigned to Method A and the other half were assigned to Method B. Use the data below to prepare a 98% confidence interval for the difference between the population means for the two methods.**

matched Pairs

| 1000's of Student-Days of Attendance | | | A-B diff |
|---|---|---|---|
| School | Method A | Method B | |
| 1 | 70.4 | 69.4 | 1.0 |
| 2 | 74.9 | 78.9 | -4.0 |
| 3 | 64.3 | 68.3 | -4.0 |
| 4 | 80.8 | 83.8 | -3.0 |
| 5 | 76.3 | 78.3 | -2.0 |
| mean = | 73.3 | 75.7 | $-2.4 = \bar{d}$ |
| st. dev. = | 6.27 | 6.65 | $2.07 = S_d$ |

$5 = n$

$4 = d.f.$

confidence = 0.98

$\alpha = 1 - \text{confid} = 0.02$ in 2 tails

$t_{\alpha/2} = 3.747$

$98\%\ CI(\mu_A - \mu_B) =$

$98\%\ CI(\mu_d) =$

$\bar{d} \pm t_{\alpha/2}\left(\frac{S_d}{\sqrt{n}}\right)$

$= (-2.4) \pm 3.747\left(\frac{2.07}{\sqrt{5}}\right)$

$= (-2.4) \pm (3.747)(0.9257)$

$= (-2.4) \pm 3.47$

$= \boxed{[-5.87 < (\mu_A - \mu_B) < 1.07]}$

**Statistics 300**
**Exam #3**

**Name:** Solution

**Fall 2011**
**Tue/Thu 7:00 - 9:05 pm**

(8 points; 8 minutes)

**12. Five schools competed for best daily attendance. The competition lasted for 180 days. Use the results below to test the claim that all of the schools were equally likely to win on each of the 180 days during the competition. (Let alpha be 0.025 for this test.)**

Goodness-of-Fit OR Multi-Nomial

| School | Count of Days school won (Obs.) | Exp. | $(O-E)^2/E$ |
|---|---|---|---|
| A | 36 | 36 | 0 |
| B | 29 | 36 | 1.36 |
| C | 37 | 36 | 0.03 |
| D | 47 | 36 | 3.36 |
| E | 31 | 36 | 0.69 |
| total = | 180 | | |

$$\sum\left[\frac{(O-E)^2}{E}\right] = 5.44$$

**H₀:** All schools were equally likely win on each day

**H₁:** Not Ho!

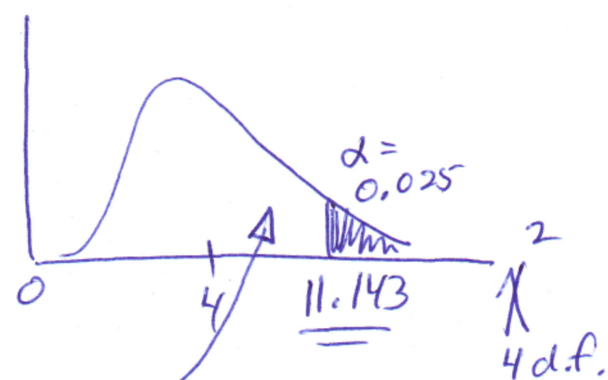
$\alpha = 0.025$ right tail

$k = 5$

$df = k - 1 = 4$

If all schools were equally likely to win on each day, then the expected count for each school is $(180)\left(\frac{1}{5}\right)$

$= 36$




Do not reject Ho.