## Chapter 24 – Inference about Variables: Part IV Review

**24.1** (c) 111.2 to 118.6. The margin of error is $2.056(9.3)/\sqrt{27} = 3.7$.

**24.2** The sample has to be an SRS taken from the population. Also, the sample should be free of outliers. The SRS condition is important for the validity of the procedure ($n$ = 27 is large enough that only the most serious skew or outliers would cause a problem).

**24.3** (b) between 0.01 and 0.05. $t$ = 2.023 and df = 13.

**24.4** Each newt had both hind limbs cut and measured; we assume that the limb exposed to the electrical field change was randomly assigned. A 99% confidence interval is given by −2.79 to 14.21 micrometers.

**24.5** (a) $H_0\colon p_{\mathrm{M}} = p_{\mathrm{F}}$ versus $H_{\mathrm{a}}\colon p_{\mathrm{M}} \neq p_{\mathrm{F}}$.

**24.6** (a) 0.763. $\hat{p}_{\mathrm{F}} = \frac{709}{929} = 0.763$.

**24.7** (b) 0.793. $\hat{p} = \frac{636 + 709}{767 + 929} = 0.793$.

**24.8** (a) $P < 0.001$. $z$ = 3.34 and $P < 0.001$.

**24.9 (a)** Assuming the observations can be thought of as an SRS, yes, the conditions are met since there are more than 15 successes and failures. **(b)** $0.76 \pm 1.645\sqrt{\frac{0.76(1-0.76)}{418}}$ =0.726 to 0.794. **(c)** We are 90% confident that between 72.6% and 79.4% of extremely obese people who have gastric bypass surgery maintain at least a 20% weight loss six years after surgery.

**24.10** (b) 3.60 ± 0.39. $3.60 \pm 2.021\frac{1.26}{\sqrt{43}}$, using 40 df from Table C.

**24.11** (b) 2.33 ± 0.17. $2.33 \pm 1.984\frac{1.00}{\sqrt{135}}$, using 100 df from Table C.

**24.12** (d) 6.05. $t = \frac{3.60-2.33}{\sqrt{\frac{1.26^2}{43}+\frac{1.00^2}{135}}}$.

**24.13** (a) 42. $df = 43 - 1$ using the conservative method.
**24.14** (a) less than 0.01. $t$ = 10.417. Note that it is surprising that the two sample standard deviations are so small, suggesting that the wait times for the next mating times for butterflies are remarkably consistent, with very little variation.

**24.15** The standard deviations are larger than the means. Because PedMIDAS scores must be greater than or equal to zero, the distributions must be right-skewed. The sample sizes are fairly large ($n$ = 64 and 71), so the sample means should be approximately Normal by the central limit theorem.

**24.16** (c) 11.87 to 19.13. Using 60 df, $15.5 \pm 1.671\frac{17.4}{\sqrt{64}}$. Technology agrees to two decimal places.

**24.17** (b) 4.98 to 23.22. df $= 63$ using the conservative option.

**24.18** Note that the confidence interval found in Exercise 24.17 does not include zero, so there is a difference at the 0.10 level. Using the conservative option, we have $t = 2.582$, df = 63, and $0.01 < P < 0.02$. There is strong evidence of a difference in mean PedMIDAS scores for the two groups. It seems that the education group has a larger mean.

**24.19** (c) This test is reasonable because the counts of successes and failures are each five or more in both samples. We would have to view these children as random samples from the larger population of children who could be in her class.

**24.20** (b) 6765. Use $p^* = 0.5$ and $z^* = 1.645$.

**24.21** (b) 0.80 ± 0.0098. $SE = 1.645\sqrt{\frac{0.8(0.2)}{4500}} = 0.0098$.

**24.22** (a) would have a smaller margin of error than the 90% confidence interval. This is because $z^*$ is smaller for 80% confidence.

**24.23** (c) $n$ = 4148. $n = \frac{(2.576)^2(0.5)(0.5)}{(0.02)^2} = 4147.4$; take $n = 4148$.

**24.24** (a) an observational study. Subjects (babies) were not assigned to groups being compared.

**24.25** (c) $H_0$: $p_{\text{VLBW}} = p_{\text{control}}$ versus $H_a$: $p_{\text{VLBW}} < p_{\text{control}}$. It seems reasonable that the researchers suspect that VLBW babies are less likely to graduate from high school.

**24.26** (b) $\hat{p}$ = 0.78. $\hat{p} = \frac{179+193}{242+233} = 0.78$.

**24.27** (b) $z = -2.34$. $z = \frac{0.7397-0.8283}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{242}+\frac{1}{233}\right)}} = -2.34$.

**24.28** (b) $t = -3.50$. $t = \frac{87.6-94.7}{\sqrt{\frac{15.1^2}{113}+\frac{14.9^2}{106}}} = -3.50$.

**24.29** (d) greater than 0.10. $t = \frac{86.2-89.8}{\sqrt{\frac{13.4^2}{38}+\frac{14^2}{54}}} = -1.25$, and the test is two-sided.

**24.30** (a) $H_0\colon p = 0.5$ versus $H_\text{a}\colon p \neq 0.5$.

**24.31** (c) between 0.05 and 0.10. $z = \frac{0.56 - 0.5}{\sqrt{\frac{0.5(0.5)}{250}}} = 1.897$, so $P = 0.058$ (using technology).

**24.32 (a)** $H_0\colon p = 0.5$ versus $H_\text{a}\colon p > 0.5$. **(b)** There are 22 successes and 10 failures. Assuming the sample can be thought of as an SRS, we can perform a large-sample test. The sample proportion is $\hat{p} = \frac{22}{32} = 0.6875$, and the test statistic is $z = \frac{0.6875 - 0.5}{\sqrt{\frac{0.5(0.5)}{32}}} = 2.12$. Using Table C, the $P$-value is between 0.01 and 0.02. There is moderately strong evidence that the candidate with the better face wins more than half the time.

**24.33** $\hat{p} = \frac{475}{625} = 0.76$. The interval is $0.76 \pm 1.645\sqrt{\frac{0.76(1-0.76)}{625}} = 0.732$ to $0.788$.

**24.34** $(0.76 - 0.41) \pm 1.96\sqrt{\frac{0.76(1-0.76)}{625} + \frac{0.41(1-0.41)}{1917}} = 0.31$ to $0.39$.

**24.35** Let $p_\text{H}$ be the proportion of human offers rejected and $p_\text{C}$ the proportion of computer offers rejected. There are at least five successes (offer rejected) and failures (offer accepted) in each group. Assuming these subjects can be thought of as an SRS, we can conduct a hypothesis test of the hypotheses $H_0\colon p_\text{H} = p_\text{C}$ versus $H_\text{a}\colon p_\text{H} > p_\text{C}$. $\hat{p}_\text{H} = \frac{18}{38} = 0.4737$ and $\hat{p}_\text{C} = \frac{6}{38} = 0.1579$. The pooled sample proportion is $\hat{p} = \frac{18 + 6}{38 + 38} = 0.3158$. The test statistic is $z = \frac{0.4737 - 0.1579}{\sqrt{0.3158(1 - 0.3158)\left(\frac{1}{38} + \frac{1}{38}\right)}} = 2.96$. Using Table C, the $P$-value is between 0.0010 and 0.0025. There is strong evidence that offers from another person are rejected more often than offers from a computer.

**24.36** (c) 0.614 to 0.646. $0.63 \pm 3.182\frac{0.01}{\sqrt{4}}$.

**24.37** (a) $t = 0.39$, $\text{df} = 3$.

**24.38** (d) −0.07 to 0.09. $0.64 - 0.63 \pm 3.182\sqrt{\frac{0.01^2}{4} + \frac{0.05^2}{4}}$.

**24.39** In all three cases, the observations must be able to be seen as random and representative samples of both types of diets. Also, the populations must be

Normally distributed. Because the sample sizes are very small, this is almost impossible to check with typical graphical methods.

**24.40** A two-sample $t$ test of the difference in average ratings for speeding drivers and noisy neighbors.

**24.41** A large-sample (or plus four) confidence interval would be used for estimating a population proportion.

**24.42** If the sample can be viewed as an SRS and the population is Normal, use a $t$ confidence interval for a population mean.

**24.43** This is the entire population of Chicago Cubs players. Statistical inference is not appropriate.

**24.44** A matched pairs $t$ test or confidence interval would be used, because both partners in a couple were interviewed.

**24.45 (a)** A two-sample test or confidence interval for difference in proportions. **(b)** A two-sample test or confidence interval for difference in means. **(c)** A two-sample test or confidence interval for difference in proportions.

**24.46** The response rate for the survey was only about 20% (427/2100 = 0.203), which might make the conclusions unreliable.

**24.47 (a)** This is a matched pairs situation; the responses of each subject before and after treatment are not independent. **(b)** We need to know the standard deviation of the differences, not the two individual sample standard deviations. (Note that the mean difference is equal to the difference in the means, which is why we only need to know the standard deviation of the differences.)

**24.48** Each of a monkey's six trials are not independent. If a monkey prefers silence, it will almost certainly spend more time in the silent arm of the cage each time it is tested.

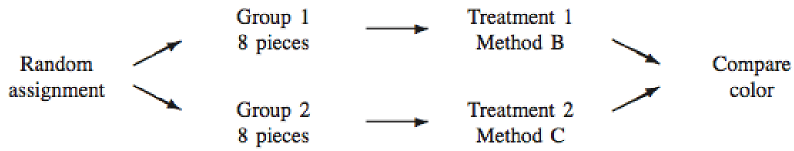
**24.49 (a)** $\hat{p}$ = 80/80 = 1, and the margin of error for 95% confidence (or any level of confidence) is 0 because $z^* = \sqrt{\frac{(1)(1-1)}{n}}$ = 0. Almost certainly, if more trials were performed, a rat would eventually make a mistake, so the actual success rate is less than 1. **(b)** The plus four estimate is $\tilde{p}$ = 82/84 = 0.9762, and the plus four 95% confidence interval is $\tilde{p} \pm z^*\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$ = 0.9762 $\pm$ 0.0326 = 0.9436 to 1.0088. Ignoring the upper limit, we are 95% confident that the actual success rate is 0.9436 or greater.

**24.50 (a)** PLAN: We will find a 99% confidence interval for $p_2 - p_1$, where $p_1$ is the proportion of subjects who have been vaccinated with Gardasil who develop cancer, and $p_2$ is the corresponding proportion for the control group. SOLVE: We assume that we have SRSs from each population. Because there were no cases of cervical cancer in the Gardasil group, we should use the plus four procedure. We have $\tilde{p}_1$= $\frac{0+1}{8487+2}$ = 0.000118 and $\tilde{p}_2$ = $\frac{32+1}{8460+2}$ = 0.0039. A 99% confidence interval for $p_2 - p_1$is then given by $\tilde{p}_2 - \tilde{p}_1 \pm 2.576\sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{8489} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{8462}}$ = 0.003782 $\pm$ 0.001772 = 0.002 to 0.0056. **(b)** PLAN: Now, let $p_1$ denote the proportion in the Gardasil group with genital warts, and let $p_2$ be the corresponding proportion for the control group. We find a 99% confidence interval for $p_2 - p_1$. SOLVE: Because we have fewer than 10 successes in the Gardasil group, conditions for using the large-sample interval are not met. However, we can use the plus four interval. We find that $\tilde{p}_1$ = 0.000253 and $\tilde{p}_2$ = 0.011644. A 99% confidence interval for $p_2 - p_1$ is then 0.0082 to 0.0145. **(c)** CONCLUDE: Gardasil is seen to be effective in reducing the risk of both cervical cancer (by between 0.002 and 0.0056, with 99% confidence) and genital warts (by between 0.0082 and 0.0145, with 99% confidence) These differences in proportions may not seem large, but if we consider the suffering that can be caused by both cervical cancer and genital warts, which are rather uncommon to begin with, the vaccine can be considered very effective.

**Note:** *This difference can be further explored using the idea of relative risk. Here, we estimate the relative risk of developing cancer in the group not vaccinated with Gardasil as 0.0039/0.000118 = 33.05. That is, those who are not vaccinated are about 33 times more likely to develop cervical cancer than those who are vaccinated. Similar calculations give relative risk = 46.02 for genital warts.*

**24.51** PLAN: We test $H_0 : \mu = 12$ versus $H_a : \mu > 12$, where $\mu$ denotes the mean age at first word (in months). SOLVE: We regard the sample as an SRS; a stemplot (not shown) shows that the data are right-skewed with a high outlier (26 months). If we proceed with the *t* procedures despite this, we find $\bar{x}$ = 13 and *s* = 4.9311 months. $t = \frac{13-12}{4.9311/\sqrt{20}}$ = 0.907, with df = 19, and *P* = 0.1879. (Note that if you delete the outlier mentioned above, $\bar{x}$ = 12.3158, *s* = 3.9729, and *t* = 0.346, yielding *P* = 0.3665; the conclusion will not change.) CONCLUDE: We cannot conclude that the mean age at first word is greater than one year.

**24.52** PLAN: We test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 < \mu_2$ versus, where $\mu_1$ is the mean number of new leaves on plants from the control population, and $\mu_2$ is the mean for the nitrogen population. SOLVE: With $\bar{x}_1$ = 13.2857, $\bar{x}_2$= 15.625, $s_1$ = 2.0587, $s_2$= 1.685, $n_1$ = 7, and $n_2$ = 8, we find $SE = \sqrt{\frac{2.0587^2}{7} + \frac{1.685^2}{8}} = 0.98$ and $t = \frac{13.3857-15.625}{SE} = -2.387$ With Option 2, df = 6 and $0.025 < P < 0.025$ (using technology, we find *P* = 0.0271). Or, using Option 1, df = 11.66 and *P* = 0.0174. CONCLUDE: We have strong evidence that nitrogen increases the mean number of new leaves formed.




**24.53** PLAN: We give a 90% confidence interval for $\mu$, the mean age at first word, measured in months. SOLVE: See results from Exercise 24.51. For df = 19, $t^*$ = 1.729, so the 90% confidence interval is $13 \pm 1.729\frac{4.9311}{\sqrt{20}}$ = 11.09 to 14.91 months. CONCLUDE: We are 90% confident that the mean age at first word for normal children is between 11 and 15 months.

**24.54** PLAN: Do a two-sided test, because we have no advance claim about the direction of the difference: $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$. SOLVE: We view the data as coming from two SRSs; the distributions show no strong departures from Normality. The means and standard deviations of the lightness scores are: $\bar{x}_1$= 48.9513 and $s_1$ = 0.2154 (cotton), and $\bar{x}_2$= 41.6488 and$s_2$ = 0.3922 (ramie). Ramie is darker, having a lower score for lightness. We find $SE$ = 0.1582 and $t$ = 46.16. With either df = 7 or df = 10.87 (using software), $P \approx 0$. CONCLUDE: There is overwhelming evidence that ramie is darker than cotton when dyed this way.

**24.55 (a)** The design is shown below. **(b)** PLAN: We test $H_0 : \mu_{\text{B}} = \mu_{\text{C}}$ versus $H_a : \mu_{\text{B}} \neq \mu_{\text{C}}$. SOLVE: We have $\bar{x}_{\text{B}}$ = 41.2825 and $s_{\text{B}}$ = 0.255, and $\bar{x}_{\text{C}}$= 42.4925 and $s_{\text{C}}$ = 0.2939; $n_{\text{B}} = n_{\text{C}}$ = 8. $SE$ = 0.1376 and $t = \frac{\bar{x}_{\text{B}}-\bar{x}_{\text{C}}}{SE}$ = –8.79. With df = 7 (or 13.73 from software), $P$ < 0.001. Using software, with df = 13.73, $P$ = 0.0000 to four places. There is overwhelming evidence that method B gives darker color on average. However, the magnitude of this difference may be too small to be important in practice.

**24.56** PLAN: We test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$. SOLVE: We are told that the samples may be regarded as SRSs from their respective populations. Back-to-back stemplots show that $t$ procedures are reasonably safe because both distributions are only slightly skewed, with no outliers and with fairly large sample sizes. We have $\bar{x}_1$ = 4.1769, $s_1$ = 2.0261, $n_1$ = 65 (parent allows drinking), $\bar{x}_2$ = 4.5517, $s_2$ = 2.4251, and $n_2$ = 29 (parent does not allow drinking). $SE$ = 0.5157 and $t = \frac{4.1769-4.5517}{0.5157} =$ $-0.727$. This is close to zero, so we will certainly not reject the null hypothesis. Indeed, with df = 46.19 (using software), $P$ = 0.471. CONCLUDE: We do not have significant evidence that there is a difference in the mean number of drinks between females with a parent who allows drinking and those whose parents do not allow drinking.

| Parent allows drinking | | Parent does not allow drinking |
|---|---|---|
| 00000 | 1 | 000 |
| 55555550000 | 2 | 0 |
| 5500000000000 | 3 | 00000055 |
| 500000000000 | 4 | 000000 |
| 00000000 | 5 | 000 |
| 5000000 | 6 | 0 |
| 00000 | 7 | 000 |
| 00 | 8 | 0 |
| 0 | 9 | 00 |
| 0 | 10 | 0 |

**24.57** PLAN: We give a 95% confidence interval for $p$, the proportion of female students with at least one parent who allows drinking. SOLVE: We are told that the sample represents an SRS. Large-sample methods may be used, because the number of successes and the number of failures are both greater than 15. With $\hat{p}$ = 65/94 = 0.6915, we have $SE$ = 0.04764, so the margin of error is 1.96 $SE$ = 0.09337, and the interval is 0.5981 to 0.7849. CONCLUDE: With 95% confidence, the proportion of female students who have at least one parent who allows drinking is 0.598 to 0.785.

**24.58 (a)** Stemplots are provided. The diabetic potentials appear to be larger.

| Diabetic | | Normal |
|---:|---|---|
| 1 | 0 | |
| | 0 | |
| | 0 | 4 |
| 7 | 0 | 6777 |
| 988 | 0 | 8888999 |
| 1000000 | 1 | 00 |
| 3 | 1 | 233 |
| 5444 | 1 | 4 |
| 76 | 1 | |
| 9988 | 1 | |
| | 2 | |
| 2 | 2 | |

**(b)** PLAN: We test $H_0 : \mu_1 = \mu_2$ versus $H_a : \ \mu_1 \neq \mu_2$, where $\mu_1$ is the mean potential for the diabetic mice and $\mu_2$ is the mean for the normal mice. SOLVE: We assume we have two SRSs; the distributions appear to be safe for $t$ procedures. With $\overline{x}_1$ = 13.0896, $\overline{x}_2$= 9.5222, $s_1$ = 4.8391, $s_2$ = 2.5765, $n_1$ = 24, and $n_2$ = 18, we find $SE$ = 1.1595 and $t$ = 3.077. With Option 2, df = 17 and 0.005 < $P$ < 0.01. Or, using Option 1, df = 36.6 and $P$ = 0.0040. CONCLUDE: We have strong evidence that the electric potential in diabetic mice is different than the potential in normal mice. **(c)** If we remove the outlier, the diabetic mouse statistics change: $\overline{x}_1$ = 13.613, $s_1$ = 4.1959, $n_1$ = 23. Now, $SE$ = 1.065 and $t$ = 3.841. With df = 17, 0.001 < $P$ < 0.002. With df = 37.15, $P$ = 0.0005. CONCLUDE: With the outlier removed, the evidence that diabetic mice have higher mean electric potential is even stronger.

**24.59 (a)** PLAN: We want to compare the proportions $p_1$ (microwaved crackers that show checking) and $p_2$ (control crackers that show checking). We can do this either by testing hypotheses or with a confidence interval, but because the "microwave checked" count is only three, significance tests are not appropriate. We will use the plus four procedure and construct a confidence interval for $p_1 - p_2$. SOLVE: We find that $\tilde{p}_1 = \frac{3+1}{65+2}$ = 0.0597 and $\tilde{p}_2 = \frac{57+1}{65+2}$ = 0.8657. $SE = \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{67} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{67}}$ = 0.05073, and a 95% confidence interval is given by $\tilde{p}_1 - \tilde{p}_2 \pm 1.96(0.0507)$ = −0.9054 to −0.7066. CONCLUDE: We are 95% confident that microwaving reduces the percentage of checked crackers by between 70.7% and 90.5%. **(b)** PLAN: We want to compare $\mu_1$ and $\mu_2$, the mean breaking pressures of microwaved and control crackers. We test $H_0 : \mu_1 = \mu_2$ versus $H_a : \ \mu_1 \neq \mu_2$ and construct a 95% confidence interval for $\mu_1 - \mu_2$. SOLVE: We assume the data can be considered SRSs from the two populations and that the population distributions are not far from Normal. Now, $SE$ = 9.0546 and $t$ = 6.914, so the $P$-value is very small, regardless of whether we use df = 19 or df = 33.27. A 95% confidence interval for the difference in mean breaking pressures between these cracker types is 43.65 to 81.55 psi (using df = 19 and $t^*$ = 2.093), or 44.18 to 81.02 psi (using df = 33.27 and $t^*$ = 2.0339). CONCLUDE: There is very strong evidence that microwaving crackers changes their mean breaking strength. We are 95% confident that microwaving crackers increases their mean breaking strength by between 43.65 and 81.55 psi.

**24.60** PLAN: We give a 95% confidence interval for $\mu$, the mean date on which the tripod falls through the ice. SOLVE: We assume that the data can be viewed as an SRS of fall-through times and that the distribution is roughly Normal (the large sample size and the central limit theorem assure us that the mean is approximately Normal). We find $n$ = 97, $\overline{x}$ = 15.309, and $s$ = 6.172 days. Using df = 96, a 95% confidence interval is given by 14.065 to 16.665 days. CONCLUDE: We are 95% confident that the mean number of days for the tripod to fall through the ice is 14.065 days to 16.665 days from April 20, or between May 4 (close to midnight) and May 6 (about 4 p.m.).

**24.61** Two of the counts are too small to perform a significance test safely.

**24.62 (a)** "SEM" stands for "standard error of the mean"; SEM $=s/\sqrt{n}$. **(b)** Two-sample $t$ tests were done, because there are two independent groups of mice. **(c)** The observed difference between the two groups of mice was so large that it would be unlikely to occur by chance alone if the two groups were the same, on average. Specifically, if the two population means were the same, and if we repeated the experiment, an observed difference in sample means for insulin as large as this would occur by chance alone less than 0.5% of the time. The result for insulin has the stronger evidence of a difference, because its $P$-value is one-tenth that for glucose.

**24.63** The group means are $\overline{x}_1$ = 5.9 (insulin), $\overline{x}_2$ = 0.75 (glucose) ng/ml, and the standard deviations are $s_1 = 0.9\sqrt{10}$ = 2.85 and $s_2 = 0.2\sqrt{10}$ = 0.632 ng/ml. PLAN: We test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$. SOLVE: The estimated standard error of the difference in sample means is $SE = \sqrt{0.9^2 + 0.2^2}$ = 0.922, so $t = \frac{5.9-0.75}{SE} =$ 5.59. With either df = 9 or df = 9.89, $P$ < 0.001. CONCLUDE: The evidence is even stronger than the paper claimed.

**24.64 (a)** Large-sample methods require at least 15 successes and 15 failures in each group. Only 8 of 25 did not prefer Times New Roman for Web use, and only 5 said Gigi was not more attractive. However, both samples had 25 observations (more than 10), so the plus four interval conditions are met. **(b)** The 95% confidence interval for the proportion who would prefer Times New Roman for Web use is $0.655 \pm 1.96\sqrt{\frac{0.655(1-0.655)}{29}} = 0.482$ to 0.828. The 90% confidence interval for those who would think Gigi is more attractive is $0.759 \pm 1.645\sqrt{\frac{0.759(1-0.759)}{29}} = 0.628$ to 0.89.

**24.65** The sample proportion is $\hat{p} = 0.39$. A 95% confidence interval is $0.39 \pm 1.96\sqrt{\frac{0.39(1-0.39)}{808}} = 0.3564$ to $0.4236$. We are 95% confident the proportion of 18- to 34-year-olds who lived with their parents or had moved back in temporarily is between 0.3564 and 0.4236.