6.4 Problems

NOTE: Before starting these problems, it might be useful to review pages 1.3-1 and 2.1-1.

1. Suppose that a random sample of $n = 102$ children is selected from the population of newborn infants in Mexico. The probability that a child in this population weighs at most 2500 grams is presumed to be $\pi = 0.15$. Calculate the probability that thirteen or fewer of the infants weigh at most 2500 grams, using…

(a) the exact binomial distribution (*Tip*: Use the function **`pbinom`** in R),

(b) the normal approximation to the binomial distribution (with *continuity correction*).

Suppose we wish to test the null hypothesis H_0: $\pi = 0.15$ versus the alternative H_A: $\pi \neq 0.15$, and that in this random sample of $n = 102$ children, we find thirteen whose weights are under 2500 grams. Use this information to decide whether or not to reject H_0 at the $\alpha = .05$ significance level, *and interpret your conclusion in context.*

(c) Calculate the ***p*-value**, using the "normal approximation to the binomial" *with continuity correction*. (*Hint*: See (b).) Also compute the **95% confidence interval**.

(d) Calculate the exact ***p*-value**, via the function **`binom.test`** in R.

2. A new "smart pill" is tested on $n = 36$ individuals randomly sampled from a certain population whose IQ scores are known to be normally distributed, with mean $\mu = 100$ and standard deviation $\sigma = 27$. After treatment, the sample mean IQ score is calculated to be $\bar{x} = 109.9$, and a two-sided test of the null hypothesis H_0: $\mu = 100$ versus the alternative hypothesis H_A: $\mu \neq 100$ is performed, to see if there is any statistically significant difference from the mean IQ score of the original population. Using this information, answer the following.

(a) Calculate the ***p*-value** of the sample.

(b) Fill in the following table, concluding with the decision either to reject or not reject the null hypothesis H_0 at the given significance level α.

Significance Level α	Confidence Level $1 - \alpha$	Confidence Interval	Decision about H_0
.10			
.05			
.01			

(c) Extend these observations to more general circumstances. Namely, as the significance level decreases, what happens to the ability to reject a null hypothesis? Explain why this is so, in terms of the *p*-value *and* generated confidence intervals.

3. Consider the distribution of serum cholesterol levels for all 20- to 74-year-old males living in the United States. The mean of this population is 211 mg/dL, and the standard deviation is 46.0 mg/dL. In a study of a subpopulation of such males who smoke and are hypertensive, it is assumed (not unreasonably) that the distribution of serum cholesterol levels is normally distributed, with unknown mean μ, but with the same standard deviation σ as the original population.

(a) Formulate the **null hypothesis** and complementary **alternative hypothesis**, for testing whether the unknown mean serum cholesterol level μ of the subpopulation of hypertensive male smokers is equal to the known mean serum cholesterol level of 211 mg/dL of the general population of 20- to 74-year-old males.

(b) In the study, a random sample of size $n = 12$ hypertensive smokers was selected, and found to have a sample mean cholesterol level of $\bar{x} = 217$ mg/dL. Construct a 95% **confidence interval** for the true mean cholesterol level of this subpopulation.

(c) Calculate the ***p*-value** of this sample, at the $\alpha = .05$ significance level.

(d) Based on your answers in parts (b) and (c), is the null hypothesis rejected in favor of the alternative hypothesis, at the $\alpha = .05$ significance level? ***Interpret your conclusion*:** What exactly has been demonstrated, based on the empirical evidence?

(e) Determine the 95% **acceptance region** and complementary **rejection region** for the null hypothesis. Is this consistent with your findings in part (d)? Why?

4. Consider a random sample of ten children selected from a population of infants receiving antacids that contain aluminum, in order to treat peptic or digestive disorders. The distribution of plasma aluminum levels is known to be approximately normal; however its mean μ and standard deviation σ are not known. The mean aluminum level for the sample of $n = 10$ infants is found to be $\bar{x} = 37.20$ μg/l and the sample standard deviation is $s = 7.13$ μg/l. Furthermore, the mean plasma aluminum level for the population of infants not receiving antacids is known to be only 4.13 μg/l.

(a) Formulate the **null hypothesis** and complementary **alternative hypothesis**, for a two-sided test of whether the mean plasma aluminum level of the population of infants receiving antacids is equal to the mean plasma aluminum level of the population of infants not receiving antacids.

(b) Construct a 95% **confidence interval** for the true mean plasma aluminum level of the population of infants receiving antacids.

(c) Calculate the ***p*-value** of this sample (as best as possible), at the $\alpha = .05$ significance level.

(d) Based on your answers in parts (b) and (c), is the null hypothesis rejected in favor of the alternative hypothesis, at the $\alpha = .05$ significance level? ***Interpret your conclusion*:** What exactly has been demonstrated, based on the empirical evidence?

(e) With the knowledge that significantly elevated plasma aluminum levels are toxic to human beings, reformulate the **null hypothesis** and complementary **alternative hypothesis**, for the appropriate *one-sided* test of the mean plasma aluminum levels. With the same sample data as above, how does the new ***p*-value** compare with that found in part (c), and what is the resulting conclusion and interpretation?

5. Refer to Problem **4.4**/2.

(a) Suppose we wish to formally test the null hypothesis H_0: $\mu = 25$ against the alternative H_A: $\mu \neq 25$, at the $\alpha = .05$ significance level, by using the random sample of $n = 80$ given.

- Calculate the *p*-value, and verify that in fact, this sample leads to an *incorrect* conclusion. [[*Hint:* Use the Central Limit Theorem to approximate the sampling distribution of $\bar{X}$ with the normal distribution $N(\mu, \sigma/\sqrt{n})$.]] Which type of error (Type I or Type II) is committed here, and why?

(b) Now suppose we wish to formally test the null hypothesis H_0: $\mu = 27$ against the specific alternative H_A: $\mu = 25$, at the $\alpha = .05$ significance level, using the same random sample of $n = 80$ trials.

- How much power exists (i.e., what is the probability) of inferring the correct conclusion?
- Calculate the *p*-value, and verify that, once again, this sample in fact leads to an *incorrect* conclusion. [[Use the same hint as in part (a).]] Which type of error (Type I or Type II) is committed here, and why?

6. Two physicians are having a disagreement about the effectiveness of chicken soup in relieving common cold symptoms. While both agree that the number of symptomatic days generally follows a normal distribution, physician A claims that most colds last about a week; chicken soup makes no difference, whereas physician B argues that it does. They decide to settle the matter by performing a formal two-sided test of the null hypothesis H_0: $\mu = 7$ days, versus the alternative H_A: $\mu \neq 7$ days.

(a) After treating a random sample of $n = 16$ cold patients with chicken soup, they calculate a mean number of symptomatic days $\bar{x} = 5.5$, and standard deviation $s = 3.0$ days. Using either the 95% **confidence interval** or the ***p*-value** (or both), verify that the null hypothesis *cannot* be rejected at the $\alpha = .05$ significance level.

(b) Physician A is delighted, but can predict physician B's rebuttal: "The sample size was too small! There wasn't enough power to detect a statistically significant difference between $\mu = 7$ days, and say $\mu = 5$ days, even if there was one present!" Calculate the minimum **sample size** required in order to achieve at least 99% **power** of detecting such a genuine difference, if indeed one actually exists. (*Note*: Use s to estimate σ.)

(c) Suppose that, after treating a random sample of $n = 49$ patients, they calculate the mean number of symptomatic days $\bar{x} = 5.5$ (as before), and standard deviation $s = 2.8$ days. Using either the 95% **confidence interval** or the ***p*-value** (or both), verify that the null hypothesis *can* now be rejected at the $\alpha = .05$ significance level.

FYI: The long-claimed ability of chicken soup – sometimes referred to as "Jewish penicillin" – to combat colds has actually been the subject of several well-known published studies, starting with a 1978 seminal paper written by researchers at Mount Sinai Hospital in NYC. The heat does serve to break up chest congestion, but it turns out that there are many other surprising cold-fighting benefits, far beyond just that. "Who knew?" Evidently… Mama. See http://well.blogs.nytimes.com/2007/10/12/the-science-of-chicken-soup/.

7. Toxicity Testing. [*Tip*: See page 6.1-28] According to the EPA (Environmental Protection Agency), drinking water can contain no more than 10 ppb (parts per billion) of arsenic, in order to be considered safe for human consumption.* Suppose that the concentration X of arsenic in a typical water source is known to be normally distributed, with an unknown mean μ and standard deviation σ. A random sample of $n = 121$ independent measurements is to be taken, from which the sample mean $\bar{x}$ and sample standard deviation s are calculated, and used in formal hypothesis testing. The following sample data for four water sources are obtained:

- Source 1: $\bar{x} = 11.43$ ppb, $s = 5.5$ ppb
- Source 2: $\bar{x} = 8.57$ ppb, $s = 5.5$ ppb
- Source 3: $\bar{x} = 9.10$ ppb, $s = 5.5$ ppb
- Source 4: $\bar{x} = 10.90$ ppb, $s = 5.5$ ppb

(a) For ***each*** water source, answer the following questions to test the null hypothesis $H_0: \mu = 10$ ppb, vs. the *two*-sided alternative hypothesis $H_A: \mu \neq 10$ ppb, at the $\alpha = .05$ significance level.

(i) ***Just by intuitive inspection****, i.e., without first conducting any formal calculations,* does this sample mean suggest that the water might be safe, or unsafe, to drink? *Why??*

(ii) Calculate the ***p*-value** of this sample (to the closest entries of the appropriate table), and use it to draw a formal conclusion about whether or not the null hypothesis can be **rejected** in favor of the alternative, at the $\alpha = .05$ significance level.

(iii) ***Interpret:*** According to your findings, is the result **statistically significant**? That is… ***Is the water unsafe to drink?*** **Does this agree with your informal reasoning in *(i)*?**

(b) For the hypothesis test in (a), what is the *two-sided* 5% **rejection region** for this H_0? **Is it consistent with your findings?**

(c) *One-sided* hypothesis tests can be justifiably used in some contexts, such as situations where one direction (either $\leq$ or $\geq$) is impossible (for example, a human knee cannot flex backwards), or irrelevant, as in ***"toxicity testing"*** here. We are really not concerned if the mean is significantly *below* 10 ppb, only *above*. With this in mind, repeat the instructions in (a) above, to test the left-sided null hypothesis $H_0: \mu \leq 10$ ppb (i.e., safe) versus the right-sided alternative $H_A: \mu > 10$ ppb (i.e., unsafe) at the $\alpha = .05$ significance level.

(d) Suppose a fifth water source yields $\bar{x} = 10.6445$ ppb and $s = 5.5$ ppb. Repeat part (c).

(e) For the hypothesis test in (c), what is the *exact* cutoff ppb level for $\bar{x}$, above which we can conclude that the water is unsafe? (Compare Sources 4 and 5, for example.) That is, what is the *one-sided* 5% **rejection region** for this H_0? **Is it consistent with your findings?**

(f) Summarize these results, and make some general conclusions regarding advantages and disadvantages of using a *one*-sided test, versus a *two*-sided test, in this context.
[*Hint*: Compare the practical results in (a) and (c) for Source 4, for example.]

---

* This is known as the Maximum Contaminant Level (MCL).

8. Do the ***Exercise*** on page 6.1-20.

9.

(a) In R, type the following command to generate a data set called "**x**" of 1000 random values.

```
x = rf(1000, 5, 20)
```

Obtain a graph of its frequency histogram by typing **`hist(x)`**. *Include this graph as part of your submitted homework assignment. (Do not include the 1000 data values!)*

Next construct a "normal q-q plot" by typing **`qqnorm(x, pch = 19)`**. *Include this plot as part of your submitted homework assignment.*

(b) Now define a new data set called "**y**" by taking the (natural) *logarithm* of **x**.

```
y = log(x)
```

Obtain a graph of its frequency histogram by typing **`hist(y)`**. *Include this graph as part of your submitted homework assignment. (Do not include the 1000 data values!)*

Then construct a "normal q-q plot" by typing **`qqnorm(y, pch = 19)`**. *Include this plot as part of your submitted homework assignment.*

(c) Summarize the results in (a) and (b). In particular, from their respective histograms and q-q plots, what general observation can be made regarding the distributions of **x** and **`y = log(x)`**? (*Hint*: See pages 6.1-25 through 6.1-27.)

10. In this problem, assume that population cholesterol level is normally distributed.

(a) Consider a small clinical trial, designed to measure the efficacy of a new cholesterol-lowering drug against a placebo. A group of six high-cholesterol patients is randomized to either a treatment arm or a control arm, resulting in two numerically balanced samples of $n_1 = n_2 = 3$ patients each, in order to test the null hypothesis H_0: $\mu_1 = \mu_2$ vs. the alternative H_A: $\mu_1 \neq \mu_2$. Suppose that the data below are obtained.

Placebo	Drug
220	180
240	200
290	220

Obtain the **95% confidence interval** for $\mu_1 - \mu_2$, and the ***p*-value** of the data, and use each to decide whether or not to reject H_0 at the $\alpha = .05$ significance level. Conclusion?

(b) Now imagine that the same drug is tested using another pilot study, with a different design. Serum cholesterol levels of $n = 3$ patients are measured at the beginning of the study, then re-measured after a six month treatment period on the drug, in order to test the null hypothesis H_0: $\mu_1 = \mu_2$ versus the alternative H_A: $\mu_1 \neq \mu_2$. Suppose that the data below are obtained.

Baseline	End of Study
220	180
240	200
290	220

Obtain the **95% confidence interval** for $\mu_1 - \mu_2$, and the ***p*-value** of the data, and use each to decide whether or not to reject H_0 at the $\alpha = .05$ significance level. Conclusion?

(c) Compare and contrast these two study designs and their results.

(d) Redo (a) and (b) using R (see hint). ***Show agreement between your answers and the output.***

11. In order to determine whether children with cystic fibrosis have a normal level of iron in their blood on average, a study is performed to detect any significant difference in mean serum iron levels between this population and the population of healthy children, both of which are approximately normally distributed with unknown standard deviations. A random sample of $n_1 = 9$ healthy children has mean serum iron level $\bar{x}_1 = 18.9$ μmol/l and standard deviation $s_1 = 5.9$ μmol/l; a sample of $n_2 = 13$ children with cystic fibrosis has mean serum iron level $\bar{x}_2 = 11.9$ μmol/l and standard deviation $s_2 = 6.3$ μmol/l.

(a) Formulate the **null hypothesis** and complementary **alternative hypothesis**, for testing whether the mean serum iron level μ_1 of the population of healthy children is equal to the mean serum iron level μ_2 of children with cystic fibrosis.

(b) Construct the **95% confidence interval** for the mean serum iron level difference $\mu_1 - \mu_2$.

(c) Calculate the ***p*-value** for this experiment, under the null hypothesis.

(d) Based on your answers in parts (b) and (c), is the null hypothesis rejected in favor of the alternative hypothesis, at the $\alpha = .05$ significance level? ***Interpret your conclusion*:** What exactly has been demonstrated, based on the sample evidence?

12. Methylphenidate is a drug that is widely used in the treatment of attention deficit disorder (ADD). As part of a **crossover study**, ten children between the ages of 7 and 12 who suffered from this disorder were assigned to receive the drug and ten were given a placebo. After a fixed period of time, treatment was withdrawn from all 20 children and, after a "washout period" of no treatment for either group, subsequently resumed after switching the treatments between the two groups. Measures of each child's attention and behavioral status, both on the drug and on the placebo, were obtained using an instrument called the Parent Rating Scale. Distributions of these scores are approximately normal with unknown means and standard deviations. In general, lower scores indicate an increase in attention. It is found that the random sample of $n = 20$ children enrolled in the study has a sample mean attention rating score of $\bar{x}_{\text{methyl}} = 10.8$ and standard deviation $s_{\text{methyl}} = 2.9$ when taking methylphenidate, and mean rating score $\bar{x}_{\text{placebo}} = 14.0$ and standard deviation $s_{\text{placebo}} = 4.8$ when taking the placebo.

(a) Calculate the **95% confidence interval** for μ_{placebo}, the mean attention rating score of the population of children taking the placebo.

(b) Calculate the **95% confidence interval** for μ_{methyl}, the mean attention rating score of the population of children taking the drug.

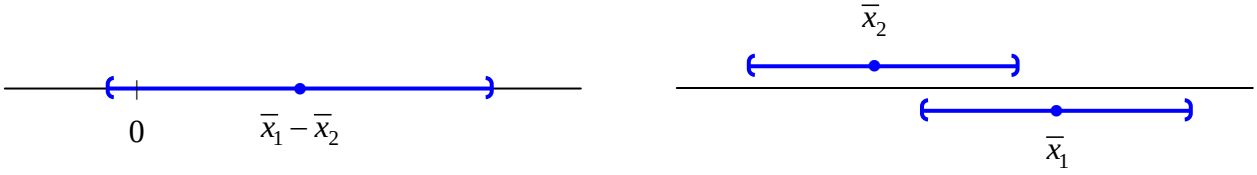
(c) Comparing these two confidence intervals side-by-side, develop an *informal* conclusion about the efficacy of methylphenidate, based on this experiment. Why can this not be used as a *formal* test of the hypothesis H_0: $\mu_{\text{placebo}} = \mu_{\text{methyl}}$, vs. the alternative H_A: $\mu_{\text{placebo}} \neq \mu_{\text{methyl}}$, at the $\alpha = .05$ significance level? (*Hint:* See next problem.)

13. A formal hypothesis test for two-sample means using the confidence interval for $\mu_1 - \mu_2$ is generally NOT equivalent to an informal side-by-side comparison of the *individual* confidence intervals for μ_1 and μ_2 for detecting overlap between them.

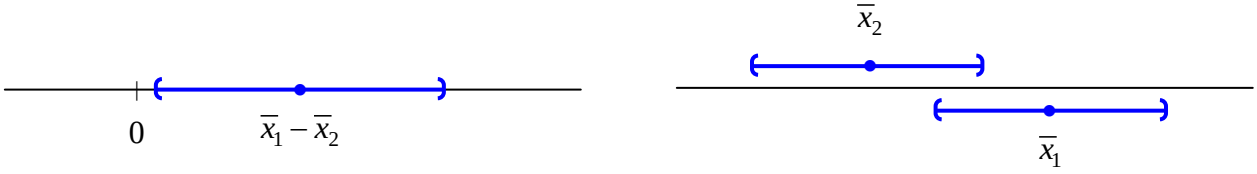
(a) Suppose that two population random variables X_1 and X_2 are normally distributed, each with standard deviation $\sigma = 50$. We wish to test the null hypothesis H_0: $\mu_1 = \mu_2$ versus the alternative H_0: $\mu_1 \neq \mu_2$, at the $\alpha = .05$ significance level. Two independent, random samples are selected, each of size $n = 100$, and it is found that the corresponding means are $\bar{x}_1 = 215$ and $\bar{x}_2 = 200$, respectively. Show that even though the two *individual* 95% confidence intervals for μ_1 and μ_2 overlap, the formal 95% confidence interval for the mean difference $\mu_1 - \mu_2$ does not contain the value 0, and hence the null hypothesis can be rejected. (See middle figure below.)

(b) In general, suppose that $X_1 \sim N(\mu_1, \sigma)$ and $X_2 \sim N(\mu_2, \sigma)$, with equal σ (for simplicity). In order to test the null hypothesis H_0: $\mu_1 = \mu_2$ versus the two-sided alternative H_0: $\mu_1 \neq \mu_2$ at the α significance level, two random samples are selected, each of the same size *n* (for simplicity), resulting in corresponding means $\bar{x}_1$ and $\bar{x}_2$, respectively. Let CI_{μ_1} and CI_{μ_2} be the respective $100(1-\alpha)\%$ confidence intervals, and let $\boxed{d = \dfrac{|\bar{x}_1 - \bar{x}_2|}{z_{\alpha/2}\left(\sigma/\sqrt{n}\right)}}$. (Note that the denominator is simply the margin of error for the confidence intervals.) Also let $\text{CI}_{\mu_1 - \mu_2}$ be the $100(1-\alpha)\%$ confidence interval for the true mean difference $\mu_1 - \mu_2$. Prove:

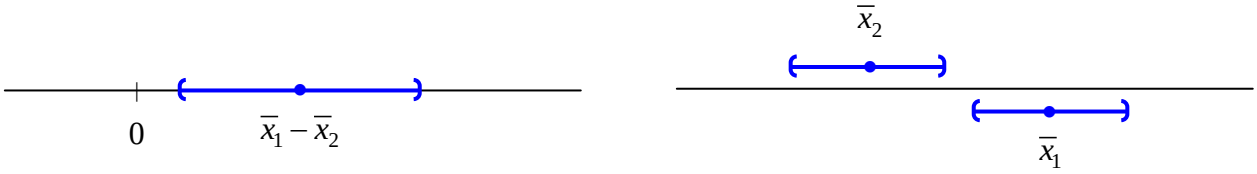
- If $d < \sqrt{2}$, then $0 \in \text{CI}_{\mu_1 - \mu_2}$ (i.e., "accept" H_0), and $\text{CI}_{\mu_1} \cap \text{CI}_{\mu_2} \neq \varnothing$ (i.e., overlap).



- If $\sqrt{2} < d < 2$, then $0 \notin \text{CI}_{\mu_1 - \mu_2}$ (i.e., reject H_0), but $\text{CI}_{\mu_1} \cap \text{CI}_{\mu_2} \neq \varnothing$ (i.e., overlap)!



- If $d > 2$, then $0 \notin \text{CI}_{\mu_1 - \mu_2}$ (i.e., reject H_0), and $\text{CI}_{\mu_1} \cap \text{CI}_{\mu_2} = \varnothing$ (i.e., no overlap).



14. *Z*-tests and Chi-squared Tests

(a) **Test of Independence (1 population, 2 random responses)**. Imagine that a marketing research study surveys a random sample of n = 2000 consumers about their responses regarding two brands (*A* and *B*) of a certain product, with the following observed results.

Do You Like Brand *B*?

Do You Like Brand *A*?	Yes	No	
Yes	335	915	1250
No	165	585	750
	500	1500	2000

First consider the null hypothesis H_0: $\pi_{A|B} = \pi_{A|B^c}$, that is, in this consumer population,

"The probability of liking *A*, given that *B* is liked, is equal to probability of liking *A*, given that *B* is not liked."

⇔ "There is no association between liking *A* and liking *B*."

⇔ "Liking *A* and liking *B* are **independent** of each other."
[Why? See Problem 3.5/22(a).]

Calculate the point estimate $\hat{\pi}_{A|B} - \hat{\pi}_{A|B^c}$. Determine the ***Z*-score** of this sample (and thus whether or not H_0 is rejected at $\alpha = .05$). Conclusion?

Now consider the null hypothesis H_0: $\pi_{B|A} = \pi_{B|A^c}$, that is, in this consumer population,

"The probability of liking *B*, given that *A* is liked, is equal to probability of liking *B*, given that *A* is not liked."

⇔ "There is no association between liking *B* and liking *A*."

⇔ "Liking *B* and liking *A* are **independent** of each other."

Calculate the point estimate $\hat{\pi}_{B|A} - \hat{\pi}_{B|A^c}$. Determine the ***Z*-score** of this sample (and thus whether or not H_0 is rejected at $\alpha = .05$). How does it compare with the previous *Z*-score? Conclusion?

Compute the **Chi-squared score**. How does it compare with the preceding *Z*-scores? Conclusion?

(b) **<u>Test of Homogeneity</u> (2 populations, 1 random response)**. Suppose that, for the sake of simplicity, the same data are obtained in a survey that compares the probability π of liking Brand *A* between two populations.

Do You Like Brand *A*?	City 1	City 2	
Yes	335	915	1250
No	165	585	750
	500	1500	2000

Here, the null hypothesis is H_0: $\pi_{A|\text{City 1}} = \pi_{A|\text{City 2}}$, that is,

"The probability of liking *A* in the City 1 population is equal to probability of liking *A* in the City 2 population."

$\Leftrightarrow$ "City 1 and City 2 populations are **homogeneous** with respect to liking *A*."

$\Leftrightarrow$ "There is no association between city and liking *A*."

How do these corresponding *Z* and Chi-squared test statistics compare with those in (a)? Conclusion?

15. Consider the following 2×2 contingency table taken from a retrospective **case-control study** that investigates the proportion of diabetes sufferers among acute myocardial infarction (heart attack) victims in the Navajo population residing in the United States.

		MI		
		Yes	No	Total
Diabetes	Yes	46	25	71
	No	98	119	217
	Total	144	144	288

(a) Conduct a **Chi-squared Test** for the null hypothesis H_0: $\pi_{\text{Diabetes | MI}} = \pi_{\text{Diabetes | No MI}}$ versus the alternative H_A: $\pi_{\text{Diabetes | MI}} \neq \pi_{\text{Diabetes | No MI}}$. Determine whether or not we can reject the null hypothesis at the $\alpha = .01$ significance level. ***Interpret your conclusion:*** At the $\alpha = .01$ significance level, what exactly has been demonstrated about the proportion of diabetics among the two categories of heart disease in this population?

(b) In the study design above, the 144 victims of myocardial infarction (**cases**) and the 144 individuals free of heart disease (**controls**) were actually age- and gender-matched. The members of each case-control pair were then asked whether they had ever been diagnosed with diabetes. Of the 46 individuals who had experienced MI and who were diabetic, it turned out that 9 were paired with diabetics and 37 with non-diabetics. Of the 98 individuals who had experienced MI but who were not diabetic, it turned out that 16 were paired with diabetics and 82 with non-diabetics. Therefore, each cell in the resulting 2×2 contingency table below corresponds to the combination of responses for age- and gender- matched case-control pairs, rather than individuals.

		MI		
		Diabetes	No Diabetes	Totals
No MI	Diabetes	9	16	25
	No Diabetes	37	82	119
	Totals	46	98	144

Conduct a **McNemar Test** for the null hypothesis H_0: "The number of 'diabetic, MI case' - 'non-diabetic, non-MI control' pairs, is equal to the number of 'non-diabetic, MI case' - 'diabetic, non-MI control' pairs, who have been matched on age and gender," or more succinctly, H_0: "There is no association between diabetes and myocardial infarction in the Navajo population, adjusting for age and gender." Determine whether or not we can reject the null hypothesis at the $\alpha = .01$ significance level. ***Interpret your conclusion:*** At the $\alpha = .01$ significance level, what exactly has been demonstrated about the association between diabetes and myocardial infarction in this population?

(c) Why does the McNemar Test only consider **discordant** case-control pairs? *Hint*: What, if anything, would a **concordant** pair (i.e., either *both* individuals in a 'MI case - No MI control' pair are diabetic, or *both* are non-diabetic) reveal about a diabetes-MI association, and why?

(d) Redo this problem with R, using **`chisq.test`** and **`mcnemar.test`**.

16. The following data are taken from a study that attempts to determine whether the use of electronic fetal monitoring ("exposure") during labor affects the frequency of caesarian section deliveries ("disease"). Of the 5824 infants included in the study, 2850 were electronically monitored during labor and 2974 were not. Results are displayed in the 2×2 contingency table below.

Caesarian Delivery

EFM Exposure	Yes	No	Totals
Yes	358	2492	2850
No	229	2745	2974
Totals	587	5237	5824

(a) Calculate a **point estimate** for the population odds ratio *OR, and interpret*.

(b) Compute a **95% confidence interval** for the population odds ratio *OR*.

(c) Based on your answer in part (b), show that the null hypothesis H_0: $OR = 1$ can be rejected in favor of the alternative H_A: $OR \neq 1$, at the $\alpha = .05$ significance level. ***Interpret this conclusion:*** What exactly has been demonstrated about the association between electronic fetal monitoring and caesarian section delivery? Be precise.

(d) Does this imply that electronic monitoring somehow *causes* a caesarian delivery? Can the association possibly be explained any other way? If so, how?

17. The following data come from two separate studies, both conducted in San Francisco, that investigate various risk factors for epithelial ovarian cancer.

Study 1

Disease Status

Term Pregnancies	Cancer	No Cancer	Total
None	31	93	124
One or More	80	379	459
Total	111	472	583

Study 2

Disease Status

Term Pregnancies	Cancer	No Cancer	Total
None	39	74	113
One or More	149	465	614
Total	188	539	727

(a) Compute **point estimates** $\widehat{OR}_1$ and $\widehat{OR}_2$ of the respective odds ratios OR_1 and OR_2 of the two studies, *and* ***interpret***.

(b) In order to determine whether or not we may combine information from the two tables, it is first necessary to conduct a **Test of Homogeneity** on the null hypothesis H_0: $OR_1 = OR_2$, vs. the alternative H_A: $OR_1 \neq OR_2$, by performing the following steps.

Step 1: First, calculate $l_1 = \ln(\widehat{OR}_1)$ and $l_2 = \ln(\widehat{OR}_2)$, in the usual way.

Step 2: Next, using the definition of $\widehat{\text{s.e.}}$ given in the notes, calculate the *weights*

$$w_1 = \frac{1}{\widehat{\text{s.e.}}_1^{\,2}} \quad \text{and} \quad w_2 = \frac{1}{\widehat{\text{s.e.}}_2^{\,2}}.$$

Step 3: Compute the *weighted mean* of l_1 and l_2:

$$L = \frac{w_1 l_1 + w_2 l_2}{w_1 + w_2}.$$

Step 4: Finally, calculate the ***test statistic***

$$X^2 = w_1 (l_1 - L)^2 + w_2 (l_2 - L)^2,$$

which follows an approximate χ^2 distribution, with 1 degree of freedom.

Step 5: Use this information to show that the null hypothesis *cannot* be rejected at the $\alpha = .05$ significance level, and that the information from the two tables may therefore be combined.

(c) Hence, calculate the **Mantel-Haenszel estimate** of the **summary odds ratio**:

$$\widehat{OR}_{\text{summary}} = \frac{(a_1 d_1 / n_1) + (a_2 d_2 / n_2)}{(b_1 c_1 / n_1) + (b_2 c_2 / n_2)}.$$

(d) To compute a **95% confidence interval** for the summary odds ratio OR_{summary}, we must first verify that the sample sizes in the two studies are large enough to ensure that the method used is valid.

Step 1: Verify that the *expected* number of observations of the $(i, j)^{\text{th}}$ cell in the first table, plus the *expected* number of observations of the corresponding $(i, j)^{\text{th}}$ cell in the second table, is greater than or equal to 5, for $i = 1, 2$ and $j = 1, 2$. Recall that the expected number of the $(i, j)^{\text{th}}$ cell is given by $E_{ij} = R_i C_j / n$.

Step 2: By its definition, the quantity L computed in part (b) is a weighted mean of log-odds ratios, and already represents a **point estimate** of $\ln(OR_{\text{summary}})$. The estimated **standard error** of L is given by

$$\widehat{\text{s.e.}}(L) = \frac{1}{\sqrt{w_1 + w_2}}.$$

Step 3: From these two values in Step 2, construct a 95% confidence interval for $\ln(OR_{\text{summary}})$, and exponentiate it to derive a 95% confidence interval for OR_{summary} itself.

(e) Also compute the value of the **Chi-squared test statistic** for OR_{summary} given at the end of § 6.2.3.

(f) Use the confidence interval in (d), and/or the χ_1^2 statistic in (e), to perform a **Test of Association** of the null hypothesis H_0: $OR_{\text{summary}} = 1$, versus the alternative H_A: $OR_{\text{summary}} \neq 1$, at the $\alpha = .05$ significance level. ***Interpret your conclusion:*** What exactly has been demonstrated about the association between the number of term pregnancies and the odds of developing epithelial ovarian cancer? Be precise.

(g) Redo this problem in R, using the code found in the link below, *and compare results*.

http://www.stat.wisc.edu/~ifischer/Intro_Stat/Lecture_Notes/Rcode/

18.

(a) Suppose a survey determines the political orientation of 60 men in a certain community:

	Left	Middle	Right	
Men	12	18	30	60

Among these men, calculate the *proportion* belonging to each political category. Then show that a **Chi-squared Test** of the null hypothesis of equal proportions

$$H_0: \pi_{\text{Left} \mid \text{Men}} = \pi_{\text{Mid} \mid \text{Men}} = \pi_{\text{Right} \mid \text{Men}}$$

leads to its rejection at the $\alpha = .05$ significance level. Conclusion?

(b) Suppose the survey also determines the political orientation of 540 women in the same community:

	Left	Middle	Right	
Women	108	162	270	540

Among these women, calculate the *proportion* belonging to each political category. How do these proportions compare with those in (a)? Show that a **Chi-squared Test** of the null hypothesis of equal proportions

$$H_0: \pi_{\text{Left} \mid \text{Women}} = \pi_{\text{Mid} \mid \text{Women}} = \pi_{\text{Right} \mid \text{Women}}$$

leads to its rejection at the $\alpha = .05$ significance level. Conclusion?

(c) Suppose the two survey results are combined:

	Left	Middle	Right	
Men	12	18	30	60
Women	108	162	270	540
	120	180	300	600

Among the individuals in each gender (i.e., row), the *proportion* belonging to each political category (i.e., column) of course match those found in (a) and (b), respectively. Therefore, show that a **Chi-squared Test** of the null hypothesis of equal proportions

$$H_0: \pi_{\text{Left} \mid \text{Men}} = \pi_{\text{Left} \mid \text{Women}} \text{ AND } \pi_{\text{Mid} \mid \text{Men}} = \pi_{\text{Mid} \mid \text{Women}} \text{ AND } \pi_{\text{Right} \mid \text{Men}} = \pi_{\text{Right} \mid \text{Women}}$$

leads to a 100% acceptance at the $\alpha = .05$ significance level. Conclusion?

NOTE: The closely-resembling null hypothesis

$$H_0: \pi_{\text{Men} \mid \text{Left}} = \pi_{\text{Women} \mid \text{Left}} \text{ AND } \pi_{\text{Men} \mid \text{Mid}} = \pi_{\text{Women} \mid \text{Mid}} \text{ AND } \pi_{\text{Men} \mid \text{Right}} = \pi_{\text{Women} \mid \text{Right}}$$

tests for *equal proportions of men and women within each political category*, which is very different from the above. Based on sample proportions (0.1 *vs.* 0.9), it is likely to be rejected, but each column would need to be formally tested by a separate Goodness-of-Fit.

(d) Among the individuals in each political category (i.e., column), calculate the *proportion* of men, and show that they are all equal to each other.

Among the individuals in each political category (i.e., column), calculate the *proportion* of women, and show that they are all equal to each other.

Therefore, show that a **Chi-squared Test** of the null hypothesis of equal proportions

$$H_0: \ \pi_{\text{Men} \mid \text{Left}} = \pi_{\text{Men} \mid \text{Mid}} = \pi_{\text{Men} \mid \text{Right}} \quad \text{AND} \quad \pi_{\text{Women} \mid \text{Left}} = \pi_{\text{Women} \mid \text{Mid}} = \pi_{\text{Women} \mid \text{Right}}$$

also leads to a 100% acceptance at the $\alpha = .05$ significance level. Conclusion?

MORAL: There is more than one type of null hypothesis on proportions to which the Chi-squared Test can be applied.

19. In a random sample of $n = 1200$ consumers who are surveyed about their ice cream flavor preferences, 416 indicate that they prefer vanilla, 419 prefer chocolate, and 365 prefer strawberry.

(a) Conduct a **Chi-squared "Goodness-of-Fit" Test** of the null hypothesis of equal proportions $H_0: \ \pi_{\text{Vanilla}} = \pi_{\text{Chocolate}} = \pi_{\text{Strawberry}}$ of flavor preferences, at the $\alpha = .05$ significance level.

Vanilla	Chocolate	Strawberry	
416	419	365	1200

(b) Suppose that the sample of $n = 1200$ consumers is equally divided between males and females, yielding the results shown below. Conduct a **Chi-squared Test** of the null hypothesis that flavor preference is not associated with gender, at the $\alpha = .05$ level.

	Vanilla	Chocolate	Strawberry	Totals
Males	200	190	210	600
Females	216	229	155	600
Totals	416	419	365	1200

(c) Redo (a) and (b) with R, using `chisq.test`. ***Show agreement with your calculations!***

20. In the late 1980s, the pharmaceutical company Upjohn received approval from the Food and Drug Administration to market Rogaine™, a 2% minoxidil solution, for the treatment of androgenetic alopecia (male pattern hair loss). Upjohn's advertising campaign for Rogaine included the results of a double-blind randomized clinical trial, conducted with 1431 patients in 27 centers across the United States. The results of this study at the end of four months are summarized in the 2×5 contingency table below, where the two row categories represent the treatment arm and control arm respectively, and each column represents a response category, the degree of hair growth reported. [Source: Ronald L. Iman, A Data-Based Approach to Statistics, Duxbury Press]

Degree of Hair Growth

	No Growth	New Vellus	Minimal Growth	Moderate Growth	Dense Growth	Total
Rogaine	301	172	178	58	5	714
Placebo	423	150	114	29	1	717
Total	724	322	292	87	6	1431

(a) Conduct a **Chi-squared Test** of the null hypothesis H_0: $\pi_{\text{Rogaine}} = \pi_{\text{Placebo}}$ versus the alternative hypothesis H_A: $\pi_{\text{Rogaine}} \neq \pi_{\text{Placebo}}$ across the five hair growth categories (That is, H_0: $\pi_{\text{No Growth} \mid \text{Rogaine}} = \pi_{\text{No Growth} \mid \text{Placebo}}$ *and* $\pi_{\text{New Vellus} \mid \text{Rogaine}} = \pi_{\text{New Vellus} \mid \text{Placebo}}$ *and* … *and* $\pi_{\text{Dense Growth} \mid \text{Rogaine}} = \pi_{\text{Dense Growth} \mid \text{Placebo}}$.) **Infer** whether or not we can reject the null hypothesis at the $\alpha = .01$ significance level. ***Interpret in context:*** At the $\alpha = .01$ significance level, what exactly has been demonstrated about the efficacy of Rogaine versus placebo?

(b) Form a 2×2 contingency table by combining the last four columns into a single column labeled **Growth**. Conduct a **Chi-squared Test** for the null hypothesis H_0: $\pi_{\text{Rogaine}} = \pi_{\text{Placebo}}$ versus the alternative H_A: $\pi_{\text{Rogaine}} \neq \pi_{\text{Placebo}}$ between the resulting **No Growth** versus **Growth** *binary* response categories. (That is, H_0: $\pi_{\text{Growth} \mid \text{Rogaine}} = \pi_{\text{Growth} \mid \text{Placebo}}$.) **Infer** whether or not we can reject the null hypothesis at the $\alpha = .01$ significance level. ***Interpret in context:*** At the $\alpha = .01$ significance level, what exactly has been demonstrated about the efficacy of Rogaine versus placebo?

(c) Calculate the *p*-value using a two-sample ***Z*-test** of the null hypothesis in part (b), and show that the *square* of the corresponding *z*-score is equal to the Chi-squared test statistic found in (b). Verify that the same conclusion about H_0 is reached, at the $\alpha = .01$ significance level.

(d) Redo this problem with R, using `chisq.test`. ***Show agreement with your calculations!***

21. Male patients with coronary artery disease were recruited from three different medical centers – the Johns Hopkins University School of Medicine, The Rancho Los Amigos Medical Center, and the St. Louis University School of Medicine – to investigate the effects of carbon monoxide exposure. One of the baseline characteristics considered in the study was pulmonary lung function, as measured by X = "Forced Expiratory Volume in one second," or FEV_1. The data are summarized below.

Johns Hopkins	Rancho Los Amigos	St. Louis
$n_1 = 21$	$n_2 = 16$	$n_2 = 23$
$\bar{x}_1 = 2.63$ liters	$\bar{x}_2 = 3.03$ liters	$\bar{x}_3 = 2.88$ liters
$s_1^2 = 0.246$ liters^2	$s_2^2 = 0.274$ liters^2	$s_3^2 = 0.248$ liters^2

Based on histograms of the raw data (not shown), it is reasonable to assume that the FEV_1 measurements of the three populations from which these samples were obtained are each approximately normally distributed, i.e., $X_1 \sim N(\mu_1, \sigma_1)$, $X_2 \sim N(\mu_2, \sigma_2)$, and $X_3 \sim N(\mu_3, \sigma_3)$. Furthermore, because the three sample variances are so close in value, it is reasonable to assume equivariance of the three populations, that is, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$. With these assumptions, answer the following.

(a) Compute the pooled estimate of the common variance σ^2 "within groups" via the formula

$$s_{\text{within}}^2 = \text{MS}_{\text{Error}} = \frac{\text{SS}_{\text{Error}}}{\text{df}_{\text{Error}}} = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2 + \ldots + (n_k - 1)\, s_k^2}{n - k}.$$

(b) Compute the grand mean of the $k = 3$ groups via the formula

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + \ldots + n_k \bar{x}_k}{n}, \qquad \text{where the combined sample size } n = n_1 + n_2 + \ldots + n_k.$$

From this, calculate the estimate of the variance "between groups" via the formula

$$s_{\text{between}}^2 = \text{MS}_{\text{Treatment}} = \frac{\text{SS}_{\text{Treatment}}}{\text{df}_{\text{Treatment}}} = \frac{n_1 (\bar{x}_1 - \bar{x})^2 + n_2 (\bar{x}_2 - \bar{x})^2 + \ldots + n_k (\bar{x}_k - \bar{x})^2}{k - 1}.$$

(c) Using this information, construct a complete ANOVA table, including the F-statistic, and corresponding p-value, relative to .05 (i.e., < .05, > .05, or = .05). **Infer** whether or not we can reject H_0: $\mu_1 = \mu_2 = \mu_3$, at the $\alpha = .05$ level of significance. ***Interpret in context:*** Exactly what has been demonstrated about the baseline FEV_1 levels of the three groups?

22. ***Generalization of Problem 2.5/8***

(a) Suppose a random sample of size n_1 has a mean $\overline{x}_1$ and variance s_1^2, and a second random sample of size n_2 has a mean $\overline{x}_2$ and variance s_2^2. If the two samples are combined into a single sample, then algebraically express its mean $\overline{x}_{\text{Total}}$ and variance s_{Total}^2 in terms of the preceding variables. (*Hint*: If you think of this in the right way, it's easier than it looks.)

(b) In a study of the medical expenses at a particular hospital, it is determined from a sample of 4000 patients that a certain laboratory procedure incurs a mean cost of \$30, with a *standard deviation* of \$10. It is realized however, that these values inadvertently excluded 1000 patients for whom the cost was \$0. When these patients are included in the study, what is the adjusted cost of the mean and standard deviation?

23.

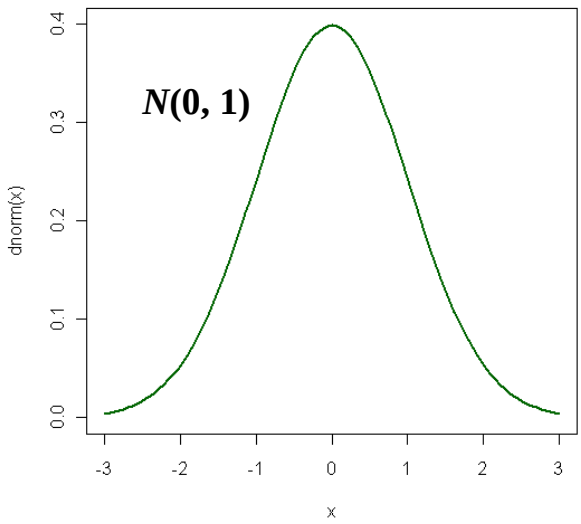
(a) For a generic 2×2 contingency table such as the one shown, prove that the Chi-squared test statistic reduces to $\chi_1^2 = \dfrac{n\,(ad - bc)^2}{R_1 R_2 C_1 C_2}$.

a	b	R_1
c	d	R_2
C_1	C_2	n

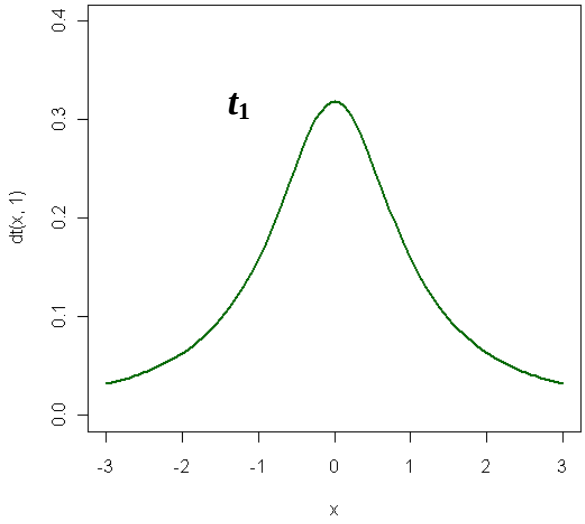
(b) Suppose that a z-test of two equal proportions results in the generic sample values shown in this table. Prove that the square of the z-score is equal to the Chi-squared score in (a).

24. Problem 5.3/1 illustrates one way that the normal and t distributions differ, as similar as their graphs may appear (drawn to scale, below). Essentially, any t-curve has heavier tails than the bell curve, indicating a higher density of outliers in the distribution. (So much higher in fact, that the mean does not exist!) Another way is to see this is to check the t-distribution for normality, via a Q-Q plot. The posted R code for this problem graphs such a plot for a standard normal distribution (with predictable results), and for a t-distribution with 1 degree of freedom (a.k.a. the Cauchy distribution). Run this code ***five*** times each, ***and comment on the results!***

```
curve(dnorm(x), -3, 3, lwd = 2, col = "darkgreen")
```




```
curve(dt(x, 1), -3, 3, ylim = range(0,.4), lwd = 2, col = "darkgreen")
```



25.

(a) In R, type the following command to generate a data set called "**x**" of 1000 random values.

```
x = rf(1000, 5, 20)
```

Obtain a graph of its frequency histogram by typing **hist(x)**. *Include this graph as part of your submitted homework assignment. (Do not include the 1000 data values!)*

(b) Next construct a "normal q-q plot" by typing the following.

```
qqnorm(x, pch = 19)
```

```
qqline(x)
```

Include this plot as part of your submitted homework assignment.

Now define a new data set called "**y**" by taking the (natural) *logarithm* of **x**.

```
y = log(x)
```

Obtain a graph of its frequency histogram by typing **hist(y)**. *Include this graph as part of your submitted homework assignment. (Do not include the 1000 data values!)*

Then construct a "normal q-q plot" by typing the following.

```
qqnorm(y, pch = 19)
```

```
qqline(y)
```

Include this plot as part of your submitted homework assignment.

(c) Summarize the results in (a) and (b). In particular, from their respective histograms and q-q plots, what general observation can be made regarding the distributions of **x** and **y = log(x)**? (*Hint*: See pages 6.1-25 through 6.1-27.)

26. Refer to the posted Rcode folder for this problem. ***Please answer all questions.***

27. Refer to the posted Rcode folder for this problem. ***Please answer all questions.***