华东理工大学

概率论与数理统计

作业簿（第十二册）

学　院 ____________专　业 ____________班　级 ____________
学　号 ____________姓　名 ____________任课教师____________

第23次作业

一．选择题

1. 若要通过抽样了解其某个服从正态分布的质量指标的方差是否在允许的范围内，宜采用的检验是 （ C ）

A. 双侧正态检验　　B. 双侧 t 检验

C. 单侧 χ^2 检验　　D. 单侧 F 检验

2. 设灯管寿命服从正态分布, 按规定寿命不低于 $\mu_0 = 2000$ 小时的灯管才算合格, 要检验某厂灯管是否合格,则原假设 H_0 应选为 （ C ）

A. $\mu = \mu_0$　　B. $\mu > \mu_0$

C. $\mu \geq \mu_0$　　D. $\mu \leq \mu_0$

3. 在显著性水平 α 下对原假设 $\boldsymbol{H_0}$ 的检验,本来应该做双侧检验的,某同学却错误地做了单侧检验,其检验的结果是拒绝 $\boldsymbol{H_0}$，则 (D)

A. 双侧检验的结果一定是拒绝 $\boldsymbol{H_0}$

B. 双侧检验的结果一定是接受 $\boldsymbol{H_0}$

C. 无论用单侧检验还是双侧检验, 犯第一类错误的概率都相同

D. 无论用单侧检验还是双侧检验, 犯第一类错误的概率都不超过 α

二．填空题：

1. 单侧检验与双侧检验的最大区别是__拒绝域在一侧还是两侧都有____
2. 如果要检验的是某品牌手机的辐射是否符合国家标准,应该选用_单侧检验
   如果要检验的是某厂生产螺栓的直径是否符合标准, 应该选用双侧检验
3. 设需要对某正态总体（方差 $\boldsymbol{\sigma^2}$ 已知）的均值进行假设检验 $H_0: \mu = 15$， H_1：$\mu < 15$，显著性水平为 $\boldsymbol{\alpha}$，采用的统计量为 $Z = \dfrac{\bar{X} - 15}{\sigma / \sqrt{n}}$， 拒绝域为 $\hat{Z} < -z_\alpha$

三. 计算题

1. 某种导线的电阻（单位：Ω）服从正态分布，按照规定，电阻的标准差不得超过 0.005。今在一批导线中任取 9 根，测得样本标准差 $\boldsymbol{S_{n-1}=0.007}$，这批导线的电阻的标准差比起规定的电阻的标准差来是否显著地偏大($\alpha=0.05$)？

解：检验 $H_0:\sigma^2\le 0.005^2, H_1:\sigma^2>0.005^2$，

考虑到均值 μ 未知，故采用单侧 χ^2 检验法。

取检验统计量 $\boldsymbol{\chi^2=\frac{(n-1)S_{n-1}^2}{\sigma_0^2}}$，在 $\boldsymbol{H_0}$ 真时服从 $\boldsymbol{\chi^2(n-1)}$。

计算统计量的值：$\hat{\chi}^2=\frac{(n-1)S_{n-1}^2}{\sigma_0^2}=\frac{8\times 0.007^2}{0.005^2}=15.68$

由 $\alpha=0.05$，查表得 $\chi_\alpha^2(n-1)=\chi_{0.05}^2(8)=15.507$，由于 15.68>$\chi_\alpha^2(n-1)$，

故拒绝 H_0，即认为电阻的标准差显著偏大.

2. 某冶金实验室对锰的熔化点作了四次试验，结果(单位: 度)分别为:

1269 ^{0}C　1271 ^{0}C　1263 ^{0}C　1265 ^{0}C

设数据服从正态分布 $N(\mu,\sigma^2)$，以显著性水平 $\alpha=5\ \%$ 作如下检验:

(1) 这些结果是否符合所公布的数字 1260 ^{0}C?

(2) 测定值的标准差是否不超过 2 ^{0}C?

解: 由样本得 $\bar{X}=1267$，$S=\sqrt{\frac{1}{3}\sum_{i=1}^{4}(X_i-\bar{X})^2}=\sqrt{40/3}=3.65$.

(1) 要检验的假设为 $H_0:\mu=1260$，$H_1:\mu\ne 1260$

检验用的统计量 $T=\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$，

拒绝域为 $|T|\ge t_{\frac{\alpha}{2}}(n-1)=t_{0.025}(3)=3.1824$.

$|\hat{T}|=\frac{1267-1260}{3.65/\sqrt{4}}=3.836>3.1824$，落在拒绝域内，

故拒绝原假设 H_0，即不能认为结果符合公布的数字 1260^{0}C.

(2) 要检验的假设为 $H_0:\sigma \le 2$，$H_1:\sigma > 2$

检验用的统计量 $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$，

拒绝域为 $\chi^2 > \chi_\alpha^2(n-1) = \chi_{0.05}^2(3) = 7.815$

$\hat{\chi}^2 = 40/4 = 10 > 7.815$,落在拒绝域内，

故拒绝原假设 H_0，即不能认为测定值的标准差不超过 2^0C.

3. 从某品牌的油漆中随机抽取 9 个样品，测得油漆的干燥时间（单位：h）为: 6.5， 5.8, 7.2, 6.6, 6.8, 6.3, 5.6, 6.1, 4.9

假设油漆的干燥时间 $\xi \sim N(\mu,\sigma^2)$，这里 μ,σ^2 未知。由样本观测数据算得样本均值为 6.2, 样本标准差为 0.6928, 厂家号称平均干燥时间 μ=5.5 小时, 且产品质量稳定, 干燥时间的误差 σ 不超过 0.5 小时. 试根据抽样数据, 在显著性水平 α=0.05 下分别对厂家的上述两个论断做假设检验.

解: (1) 要检验的假设为 $H_0:\mu = 5.5$，$H_1:\mu \ne 5.5$

检验用的统计量 $T = \dfrac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$，

拒绝域为 $|T| \ge t_{\frac{\alpha}{2}}(n-1) = t_{0.025}(8) = 2.306$.

$\left|\hat{T}\right| = \dfrac{6.2-5.5}{0.6928/\sqrt{9}} = 3.0312 > 2.306$，落在拒绝域内，

故拒绝原假设 H_0，即不能认为结果符合公布的数字 μ=5.5 小时. .

(2) 要检验的假设为 $H_0:\sigma \le 0.5$，$H_1:\sigma > 0.5$

检验用的统计量 $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$，

拒绝域为 $\chi^2 > \chi_\alpha^2(n-1) = \chi_{0.05}^2(8) = 15.507$

代入得 $\hat{\chi}^2$=15.3591<15.507, 未落在拒绝域内，

故不拒绝原假设 H_0，即可以认为测定值的标准差不超过 0.5 小时.

4. 从某锌矿的东、西两支矿脉中，各抽取样本容量分别为 9 与 8 的样本进行测试，样本含锌平均值及样本方差如下：

东支：$\overline{x}=0.230 \quad S_x^2=0.1337$；　西支：$\overline{y}=0.269 \quad S_y^2=0.1736$

若东、西两支矿脉的含锌量都服从正态分布且方差相同，

1) 显著性水平 $\alpha=0.05$ 下，是否可以认为东西两支矿脉含锌量的平均值一样？

2) 显著性水平 $\alpha=0.1$ 下，可否认为西矿含锌量比东矿含锌量高？

解：1) 检验 $H_0:\mu_1=\mu_2, H_1:\mu_1\neq\mu_2$，采用双侧 t 检验法

检验统计量为 $T=\dfrac{\overline{X}-\overline{Y}}{S_w\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}$，在 $\boldsymbol{H_0}$ 真时服从 $\boldsymbol{t(m+n-2)}$

计算 $S_w=\sqrt{\dfrac{8\times0.1337+7\times0.1736}{15}}=0.3903$，代入可得：

$$\hat{T}=\frac{\overline{X}-\overline{Y}}{S_w\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}=\text{ - }0.2056$$

由 $\alpha=0.05$ 查表得 $t_{\frac{\alpha}{2}}(m+n-2)=t_{0.025}(15)=2.1314$,

由于 $|\hat{T}|<t_{0.025}(15)$，故接受 H_0，即这两支矿脉含锌量的平均值看作一样.

2) 检验 $H_0:\mu_1\geq\mu_2$，$H_1:\mu_1<\mu_2$

$$\hat{T}=\frac{\overline{X}-\overline{Y}}{S_w\sqrt{\dfrac{1}{m}+\dfrac{1}{n}}}=\text{ - }0.2056$$

H_0 的拒绝域为: $T<-t_{0.1}(15)=-1.3406$，因 $\hat{T}>$-1.3406,

故接受 H_0(即拒绝 H_1)，即显著性水平 $\alpha=0.1$ 下，

不能认为西矿含锌量比东矿含锌量高.

第 24 次作业

一. 选择题

1. 假设一个(一元或多元)线性回归问题的总离差平方和 SST=100, 残差平方和 SSE=19, 则错误的选项是 (C)

   A. 回归平方和 SSR=81　　　　B. 可决系数 R^2 为 0.81

   C. 样本相关系数 R 为 0.9　　　　D. 样本相关系数 R 为 0.9 或-0.9

2. 假设根据样本数据求得变元 X 与 Y 的样本相关系数 R = - 0.9, 则变元 X 与 Y 可能的回归方程是 (B)

   A. Y=1+2X　　　　B. Y=1-2X

   C. Y= -1+2X　　　　D. Y= -0.9+0.9X

3. 回归分析的前提是 (D)

   A. 正态性　　　　B. 独立性

   C. 方差齐性　　　　D. 以上选项全是

二.填空题

1. 一元线性回归的模型为 $\underline{Y=\beta_0+\beta_1 X+\varepsilon \quad \varepsilon \sim N(0,\sigma^2)}$;根据 n 组样本数据 (x_i, y_i)求回归模型中参数的极大似然估计(也是最小二乘估计)的公式为 $\underline{\hat{\beta}_1=\frac{L_{xy}}{L_{xx}},\ \hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}}$; 求变元 X 与 Y 的样本相关系数的公式为 $R=\underline{\frac{L_{xy}}{\sqrt{L_{xx}}\sqrt{L_{yy}}}}$, R^2 称为<u>可决系数;</u> 若残差平方和 SSE=0, 则可决系数 R^2=<u>1 ;</u>
2. 在 EXCEL 中对回归方程进行检验,当检验的 p 值 <u>小于</u> 给定的显著性水平时说明变元的线性相关关系显著(回归方程有意义).
3. 线性回归的离差分解公式为: <u>SST=SSE+SSR(或总离差平方和=残差平方和+回归平方和)</u>

三. 计算题:

1. 某公司营销人员对公司 6 个月来每月广告费用（记为 x, 单位:万元）和销售额（记为 y, 单位:万元）做了统计，得到如下数据：

广告费用（x)	1	2	3	4	5	6
销售额(y)	3	6	9	15	20	25

经计算得 $\sum x_i=21$，$\sum x_i^2=91$，$\sum y_i=78$，$\sum y_i^2=1376$，$\sum x_i y_i=352$：

1) 计算销售额（y）与广告费用(x)的相关系数

2) 求销售额对广告费用的直线回归方程

3) 若下月计划广告费支出 10 万元，试预测相应的销售额

解：1) 相关系数 $R=\dfrac{L_{xy}}{\sqrt{L_{xx}L_{yy}}}=\dfrac{n\sum xy-\sum x\sum y}{\sqrt{n\sum x^2-(\sum x)^2}\sqrt{n\sum y^2-(\sum y)^2}}$ =0.99

2) 由 $\hat{\beta}_1=\dfrac{L_{xy}}{L_{xx}}$，$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}$ 得：$\hat{\beta}_1=4.51$，$\hat{\beta}_0=-2.8$，

所以直线回归方程为 $\hat{y}=-2.8+4.51x$

3) $\hat{y}=-2.8+4.51\times 10=42.3$

2. 为了研究钢线含碳量（单位：%）x 对于电阻 Y 在 20^oC 下的影响，做了 7 次试验，得到数据如下：

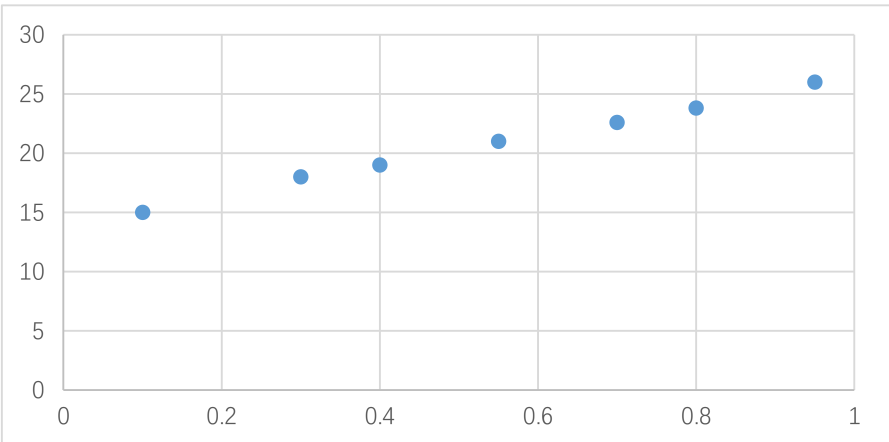
钢线含碳量 x	0.10	0.30	0.40	0.55	0.70	0.80	0.95
电阻 y	15	18	19	21	22.6	23.8	26

1）画出散点图；

2）求电阻关于钢线含碳量的直线回归方程；

3）进行线性回归的显著性检验（$\alpha=0.01$）

解： 1)散点图:



2）利用 EXCEL 进行回归分析，可得如下结果:

SUMMARY OUTPUT							
回归统计							
Multiple R	0.998714						
R Square	0.99743						
Adjusted R	0.996916						
标准误差	0.207833						
观测值	7						
方差分析							
	df	SS	MS	F	Significance F		
回归分析	1	83.81831	83.81831	1940.48	1.13798E-07		
残差	5	0.215973	0.043195				
总计	6	84.03429					
	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 99.0%	上限 99.0%
Intercept	13.95839	0.173468	80.46646	5.62E-09	13.51247449	14.4043	13.25894	14.6578387
X Variable	12.55034	0.284905	44.05088	1.14E-07	11.81796278	13.28271	11.401556	13.6991151

所以直线回归方程为 $\hat{y}$ = 13.9584+12.5503x

3）由于 P-value 小于 α，故认为回归方程有显著意义

3. 试根据如下对变元 X 与 Y 样本数据的 EXCEL 回归分析的结果填空

SUMMARY OUTPUT						
回归统计						
Multiple R	0.9584					
R Square	0.9185					
Adjusted R	0.8913					
标准误差	1					
观测值	5					
方差分析						
	df	SS	MS	F	Significance F	
回归分析	1	33.8	33.8	33.8	0.010131424	
残差	3	3	1			
总计	4	36.8				
	Coefficien	标准误差	t Stat	P-value	Lower 95%	Upper 95%
Intercept	1.25	0.901388	1.38675	0.2596	-1.618618333	4.118618
X Variable	1.625	0.279508	5.81378	0.01013	0.735479216	2.514521

1) 这个回归分析的前提是_$Y=\beta_0+\beta_1X+\varepsilon \quad \varepsilon\sim N(0,\sigma^2)$,其中误差项满足__

正态性, 独立性, 方差齐性_;

2) 变元 X 与 Y 的样本相关系数是__0.9584___,可决系数是__0.9185___;
3) 变元 X 与 Y 的回归方程是_$\hat{Y}$=1.25+1.625X, 回归的残差平方和是___3__;
4) 在显著性水平 0.05 情况下, 变元 X 与 Y 的线性相关关系是否显著___是____
5) 根据分析结果,当 X=0 时, 预测变元 Y 的点估计为 __1.25___;
6) 回归方程中变元 X 系数的置信水平为95%的置信区间是___[0.7355, 2.5145]__

4. 对某种产品的表面腐蚀深度 y (微米)与腐蚀时间 t (秒)的 12 次调查, 根据测得结果计算:

$$\sum t_i = 660; \quad \sum y_i = 264; \quad \sum t_i^2 = 59250; \quad \sum y_i^2 = 8256; \quad \sum t_i y_i = 21935$$

(1) 求腐蚀深度与腐蚀时间的样本相关系数

(2) 求线性回归方程 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 t$

解: 1) 相关系数 $R = \dfrac{L_{xy}}{\sqrt{L_{xx}L_{yy}}} = \dfrac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$ =0.989268

2) 由 $\hat{\beta}_1 = \dfrac{L_{xy}}{L_{xx}}$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ 得: $\hat{\beta}_1 = 0.323094$, $\hat{\beta}_0 = 4.229847$,

所以直线回归方程为 $\hat{y} = 4.229847 + 0.323094t$

SUMMARY OUTPUT								
回归统计								
Multiple R	0.989268							
R Square	0.978652							
Adjusted R Square	0.976517							
标准误差	2.286052							
观测值	12							
方差分析								
	df	SS	MS	F	Significance F			
回归分析	1	2395.74	2395.74	458.424	1.1E-09			
残差	10	52.26035	5.226035					
总计	11	2448						
	Coefficients	标准误差	t Stat	P-value	Lower 95%	Upper 95%	下限 95.0%	上限 95.
Intercept	4.229847	1.060348	3.989113	0.002563	1.867245	6.59245	1.867245	6.59
X Variable 1	0.323094	0.01509	21.41084	1.1E-09	0.289471	0.356717	0.289471	0.356